

Probabilistic Design and Optimization

Class Notes, ECI 249 Winter, 2019

Jay R. Lund
Department of Civil and Environmental Engineering
University of California - Davis
4 January 2019

"Life is uncertain; eat dessert first." motto Pacific Desserts, Seattle (c.1990), attr. Ernestine Ulmer (1925)

"Holidays are unpredictable. Give blood." Bloodsource 2016

"Life is uncertain. Don't sip!" motto Lagunitas Brewing Company

"In these matters the only certainty is that nothing is certain." Pliny the Elder (23-79AD)

"Doubt is not a pleasant condition, but certainty is absurd." Voltaire (1694-1778)

"Not to be absolutely certain is, I think, one of the essential things in rationality." Bertrand Russell (1947)

"Because we cannot foresee, we must manage now." Leonardo da Vinci

"To know yet to think that one does not know is best; Not to know yet to think that one knows will lead to difficulty." Lao Tzu (c. 500 BC), *Tao Te Ching*, No. 71.

"And the trouble is, if you don't risk anything, you risk even more." Erica Jong

"Unless you expect the unexpected you will never find truth, for it is hard to discover and hard to attain."
Heraclitus (c. 500 BC)

"Plan, v.t. To bother about the best method of accomplishing an accidental result."
Ambrose Bierce's *Devil's Dictionary* (1906)

"Chance favors only the prepared mind." Pasteur, Louis (1854)

"Errors using inadequate data are much less than those using no data at all." C. Babbage (1792-1871)

"Some problems are so complex that you have to be highly intelligent and well informed just to be undecided about them." Laurence J. Peter

"... danger is never so near as when you are unprepared for it." Francis Parkman (1849), *The Oregon Trail*

"The policy of being too cautious is the greatest risk of all." Jawaharlal Nehru (1889 – 1964)

"Hope is not a strategy" – Anonymous

"There is no such thing as luck. There is only adequate or inadequate preparation to cope with a statistical universe." Robert Heinlein (1907 - 1988), *Time Enough for Love*

"Take calculated risks. That is quite different from being rash." George S. Patton (1885 – 1945)

"Chance is always powerful. Let your hook be always cast; in the pool where you least expect it, there will be a fish." Ovid (43 BC - 17 AD)

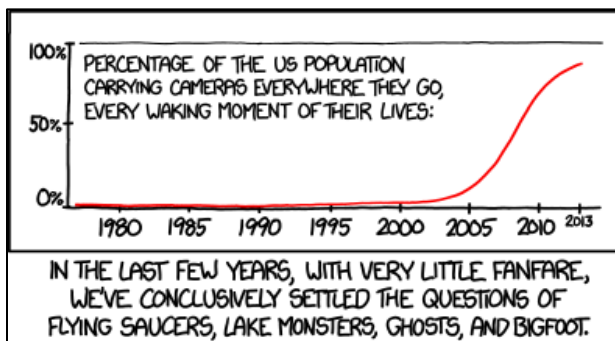
"2 is not equal to 3, not even for large values of 2." Grabel's Law

"Certainty is for losers." John Lund, age 15 (2006)



Table of Contents

1. Design Philosophy	3
2. Objectives which Consider Uncertainty	6
3. Expected Utility Theory	9
3b. Does Expected Utility Theory Work With Real People?	13
4. Dominance Decisions under Uncertainty	14
5. Structure of Decision Analysis	22
6. The Value of Imperfect Information	25
7. Example of Decision Analysis: Levee Design	29
8. A Probability Primer	33
9. "Maximum Entropy": Assigning Probabilities with Little Information	38
10. Simulation Modeling with Uncertainty	43
11. Analytical Propagation of Uncertainty	46
12. Approximating the Mean and Variance of Model Outputs	49
15. Stochastic Dynamic Programming	63
16. Probabilistic Linear Programming	67
17. Two-Stage Linear Programming	74
19. Markov Decision Theory	82
20. Probabilistic Optimal Search	85
21. Evolutionary Algorithms (Genetic Algorithms)	87
22. Some Thoughts on Probabilities in Engineering Calculations	88
Appendix - Worked examples	91
Appendix - Institutional Models for Integrating Knowledge and Management (adapted from NRC 2007)	92
Appendix - Buckingham π theorem (adapted from Wikipedia, the free encyclopedia)	100



xkcd.com/1235/

Each lecture is nominally 50 minutes, but sometimes longer and occasionally shorter. The course is about uncertainty after all.

Many thanks to the students who have helped de-bug and improve these notes over the years.

1. Design Philosophy

Most decisions have uncertain outcomes.

Some examples:

- The final cost of most construction projects are uncertain until after completion.
- The environmental impact of the Central Valley Project remains uncertain decades after its completion.
- If concrete is scheduled to be placed on a given day, will the temperature be sufficient for the concrete to cure properly?
- What is the public health effect of a particular water treatment process?
- What pesticides should be used (if any) and how should they be used, given their different risks and costs?
- Should major infrastructure be built in anticipation of:
 - Sea level rise and climate change?
 - Urban growth?
 - Possible changes in water treatment standards?
 - Possible increases in environmental requirements?
- If we build it, they will come. But how many will there be, how much service will they want, and will they be willing to pay enough for the service?
- Having decided which courses to take this quarter, what will your grades be?
- Having decided which courses to take this quarter, will you enjoy these classes or learn anything very interesting in them?
- Having selected a job offer, what will be your salary in 5 years? Will your boss be good?
- Having selected a graduate program, will they be able to fund you next year and will you be able to complete your degree within two years?

Here is a nice talk by David Rosenberg on optimality and near-optimality:

<https://www.youtube.com/watch?v=b96LsqcJyHM>

We will use probabilities to represent uncertainties.

Consider the following mutually exclusive decisions and their outcomes.

Table 1: Probability distributions of outcomes for three decision options

Decision	Outcome:	-\$1,000	-\$500	\$0	\$500	\$1,000	\$2,000
A	Probability:	0.05	0.3	0.3	0.2	0.1	0.05
B	Probability:	0	0.4	0.3	0.2	0.1	0
C	Probability:	0.1	0.1	0.2	0.3	0.3	0

Some properties of probabilities:

$$0 \leq p \leq 1$$

The sum of probabilities for all possible outcomes equals one.

What does probability represent? Two theories:

1. Relative frequency of outcomes. Example: Throwing a die.
(You can throw the die many times to determine the relative frequency of each side.)
2. "Subjective" certainty of outcome occurrence. Example: Grade in this class.
(You can only take this class once. What is your probability distribution for this class? How will you "update" this distribution as the quarter progresses? What decisions can you take if this probability distribution does not evolve as you would hope? [Drop the course.]

Are there other quantitative ways to represent uncertainty? Fuzzy sets? Why use probability?
For this course, we use only probabilities. Other ways are less rigorous.

How can we incorporate uncertainty in outcomes in selecting and formulating "optimal" engineering designs?

Let's return to the table of decisions, outcomes, and probabilities above.

Which alternative is best?

- The alternative with the highest expected value, or highest average value, is best. If you had to make this choice many times, which choice would give the highest yield? We calculate the expected value of alternative j (EV_j) as:

$$EV_j = \sum_{i=1}^n p_{ij} c_i,$$

where p_{ij} is the probability of outcome i given decision j and c_i is the value of the consequence of outcome i . We assume there are a total of n outcomes.

The expected values of the three alternatives are: $EV(A) = \$100$, $EV(B) = \$0$, and $EV(C) = \$300$. By this criterion, decision C is best.

- The alternative with the best worst outcome is best. Quite often, to be safe, people choose the alternative that can go the least wrong. Some parents encourage their children to become engineers because, even though the average salaries of doctors and lawyers might be higher, almost any reasonably good and diligent student can become an engineer and find a reasonable job someplace. Where the consequences are in terms of gains, this is called a MAXIMIN criteria. The MAXIMIN values of each decision are: $MAXIMIN(A) = -\$1,000$, $MAXIMIN(B) = -\$500$, and $MAXIMIN(C) = -\$1,000$. By this criterion, alternative B is the best. (This is a pessimistic choice or design criterion, assuming Murphy's Law will hold.)

- Select the alternative with the most optimistic possible outcome. Homer Simpson bets his entire small Christmas bonus on a long shot at the dog races, hoping he can make enough money to buy lavish presents for his family. This is a gambler's choice. This would be a MAXIMAX criterion. The MAXIMAX values for the decisions are: $MAXIMAX(A) = \$2,000$, $MAXIMAX(B) = \$1,000$, and $MAXIMAX(C) = \$1,000$. Decision A would be the best choice. (The mob will be after you if you don't pay your \$2,000 gambling debt by tomorrow.)

- How would you choose an alternative if only A and B were choices?

Some funny things can happen when you make decisions and designs in an uncertain world.

- You can look stupid in retrospect. If you chose alternative C and you were unlucky with an outcome of $-\$1,000$, you'd feel a bit embarrassed.
- You can get lucky, and look brilliant in retrospect. By choosing alternative A, a dumb choice on the average, you might ($p=0.05$) come out looking very smart.
- The designs you get by playing the averages can be very different than the designs selected by trying to minimize the chance of looking stupid or trying to maximize your chance of looking brilliant. (This poses a philosophical dilemma; you have to decide what you want.)
- *Recourse* or the flexibility of our designs to respond to events in the future becomes important for making initial decisions and designs. How does the design envision responding to uncertain events? The "optimal" design of a water supply system will vary depending on how well the system can respond to shortages via water conservation measures or water purchases. The design of how the system should respond to uncertain outcomes is integrally important in the design. How many spare tires do you keep in your car?

Some Roles for Probabilistic Approaches to Design

- 1) Formulating design problems as probabilistic analysis problems can help us think about the design problem, perhaps suggesting better approaches to the problem and perhaps formalizing the problem to make it more easily understood and worked on by a group.
- 2) Actual probabilistic analysis can sometimes help in evaluation of competing design alternatives.
- 3) The process of probabilistic analysis can make it easier to develop promising novel design approaches. Probabilistic optimization can be especially useful here.
- 4) Conducting probabilistic analysis can help give confidence to the engineer and ultimate decision-makers that a proposed design is desirable, when risks and uncertainties are considered.

Game theory example: The regulation game between Utilities and Regulators [Update from Madani 2009]

Cost outcomes for different combinations of decisions:

Table 2: Game-theoretic payoffs for regulator vs. water utility

(regulator value: utility value)		Water Utility: Minimize expected value cost; design for which standard?	
		Current Standard	New Standard
Regulator: minimize risk, but stay employed	Current Standard	0:0	2:-5
	New Standard	1: -10	2: -5

What should you do if you are a:

- o Water Utility? Which standard should you design your treatment plant for? Should you adapt a design of intermediate expense that has a contingency for each standard?
- o Regulator? Which standard should you adopt?

What criterion is important to you? EV, minimax, maximin?

Your optimal decision depends on your assessment of the probability of the other agency’s decision.

Let X_U be the utility’s decision (0 = build to current standard, 1 = build to new standard)

Let X_R be the regulator’s decision (0 = current standards, 1 = new standard)

Utility: $\text{Max } z_U = p_{cR}(-5 X_U) + p_{nR}(-10 + 5X_U)$

Regulator: $\text{Max } z_R = p_{cU}(1 X_R) + p_{nU}(2)$

What is the best strategy for each actor? Does this affect your estimate of their future decision?

What do you think will be the outcome of this “game”?

a) Would the regulator ever not adopt the new standard? $U_R(\text{new std}) \geq U_R(\text{current std})$ regardless of the utility’s decision

b) So, what standard should the utility design for, in this case? Does the decision change if construction to the current standards is “grandfathered” under new regulations (changing its impact from -10 to 0)?

The Course

- We will live and breathe probabilities.
- Manipulating probabilities requires lots of calculations, so there will be a lot of computer programming. I want this to be a very practical theory course.
- Look over syllabus.
- Readings will be from the lecture notes, text, and hand-outs.
- There will be 5 homework assignments, roughly at 2-week intervals. Project 2 is most challenging.
- There is a term project, but no final exam.
- I want this class to be a bit of ambitious, fast-paced fun. I think we can solve some problems in this class that some parts of the profession agonize over.

2. Objectives which Consider Uncertainty

Given the following predictions of alternative design performance, what criterion should we use to select the best decision or design?

Table 3: Probability distributions of outcomes for three decision options

Decision	Outcome:	-\$1,000	-\$500	\$0	\$500	\$1,000	\$2,000
A	Probability:	0.05	0.3	0.3	0.2	0.1	0.05
B	Probability:	0	0.4	0.3	0.2	0.1	0
C	Probability:	0.1	0.1	0.2	0.3	0.3	0

The evaluation of alternatives under uncertainty can consider both the consequences of different outcomes and their probabilities. Here are a few common and sometimes useful examples.

Probability of achieving a performance target

What is the probability that performance will be worse (or better) than a set performance target? Set targets can be somewhat arbitrary, and this approach neglects the range and likelihood of other performance levels.

Expected Value (Average Outcome)

Select the decision with the highest expected value of outcome. Expected value is defined as:

$$EV_j = \sum_{i=1}^n p_{ij} c_i,$$

where p_{ij} is the probability of outcome i given decision j and c_i is the value of the consequence of outcome i . We assume there are a total of n outcomes. If this criterion is used for all decisions under uncertainty, then the greatest average benefit will be gained over time. This works particularly well for making many, similar, small decisions (Arrow and Lind 1970).

MAXIMIN or MINIMAX

The MAXIMIN criterion selects the alternative with the best minimum outcome (the best worst-case). It is a very pessimistic criterion, but can be appropriate when the worst-case is particularly dire. Designs selected by performance under "worst-case scenarios" are MAXIMIN designs.

Where outcomes are considered benefits this approach is called MAXIMIN. Where outcomes are considered losses, this approach is called MINIMAX.

A common example of MAXIMIN design is the design of dam spillways, which use a concept called the "maximum probable flood," which is ideally the worst possible flood that could occur at the dam site. Another example that approaches MINIMAX design is traditional sizing of reservoirs, which has been to provide full water supply during the worst drought of record.

Maximize "Reliability"

Sometimes performance within a tolerable range is acceptable, but performance outside the range is not, without much other concern for performance. In this case, maximizing the probability of acceptable performance can be a suitable criterion for selecting a design. Minimizing the probability of a dam or structure failing or minimizing probability of a water shortage are commonly-stated objectives. However, these are not careful statements if not accompanied by other objectives or limitations. For example, the only ways to absolutely minimize the probability of a dam failing are to either a) not build a dam or b) build it so massively that it would be prohibitively expensive. Both these approaches would effectively minimize the probability of failure to zero. Maximizing reliability subject to a budget constraint and water supply performance would result in a more reasonable design.

Expected Utility

We often do not really value outcomes linearly, as assumed by expected value theory. I would prefer to have \$500,000 for sure than 50% chance at \$1 million. MINIMAX criteria are also a bit extreme, considering only the worst outcome and disregarding its probability.

Expected utility tries to avoid these problems. Expected utility evaluation takes the form:

$$EU_j = \sum_{i=1}^n p_{ij} U(c_i),$$

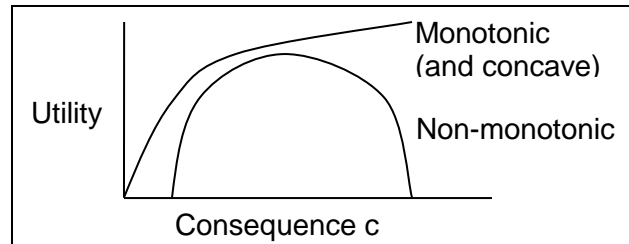
where EU_j is the expected utility of alternative j , n is the number of possible outcomes, p_{ij} is the probability of each outcome, and $U(c_i)$ is the utility of the consequence of outcome i . The utility function $U()$ is used to estimate the actual value to the decision-maker of a given outcome. This function can be difficult to obtain and we will discuss it a lot later.

(What should a utility function look like for beer, say, compared to the utility function for money or streamflow for fish?)

Important definitions and some implications

Monotonic vs. non-monotonic utility

Utility for water supply is monotonic; more water is better. Utility for water is non-monotonic; at first, more water is good, but too much water reduces benefits and eventually causes drowning. Utility for money is monotonic for most of us, but if we spoil our kids it can be non-monotonic.

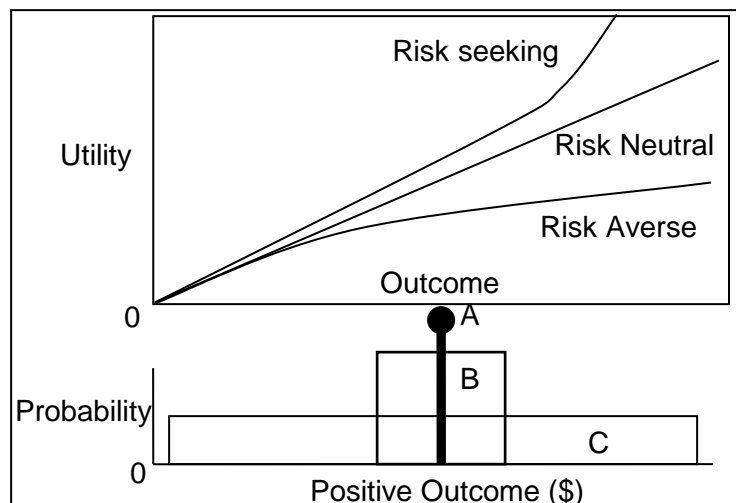


Risk Aversion versus Risk Seeking in Utility Functions

Consider three probability distributions of positive outcomes (A, B, and C) and three utility functions, one “risk seeking” another “risk averse”, and a third “risk neutral.” The expected value of each distribution is the same, with the distribution A being certainty at the expected value of the other distributions. For the risk averse, utility function, A (certainty) is always preferred. With the risk seeking utility function, the disproportionately high values away from the mean raise the expected utility more than the losses of utility at the low values, implying that the riskier distributions are preferred. The opposite is true for the risk-averse utility function, where the less certain distributions are preferred, all else being equal. For *risk neutral*, $dU/dx = \text{constant}$ and $d^2U/dx^2 = 0$; for “*risk averse*,” $d^2U/dx^2 < 0$; and for “*risk seeking*,” $d^2U/dx^2 > 0$.

Should *risk aversion* increase or decrease as you become wealthier? What would this say about the third derivative of the utility function? Should d^3U/dx^3 be negative (< 0) also for positive outcomes?

For *risk-neutral*, $EU(A) = EU(B) = EU(C)$ and $EV(A) = EV(B) = EV(C)$,
Risk-averse, $EU(A) > EU(B) > EU(C)$
Risk-seeking, $EU(A) < EU(B) < EU(C)$



Risk aversion vs. seeking and Expected values vs. Utilities: Let \tilde{c} be a set of uncertain consequences resulting from a decision. *Risk aversion* means people prefer the expected value of a set of consequences with certainty to the set of uncertain consequences, so $U(E[\tilde{c}]) > E[U(\tilde{c})]$. *Risk seeking* occurs then risky outcomes are preferred to certain expected value, so $U(E[\tilde{c}]) < E[U(\tilde{c})]$.

Certainty Equivalence: If someone is indifferent in choosing between a certain consequence c_c and a set of uncertain consequences $\{p_1 c_1, \dots, p_i c_i, \dots, p_n c_n\}$, then consequence c_k is the *certainty equivalent* to the set of uncertain outcomes (Keeney and Raiffa 1976). Non-monotonic utility functions can have more than one certainty equivalent. The idea of certainty equivalence is useful for estimating a utility function.

Risk premium: Risk premium (RP) is how much someone prefers the expected value of an uncertain consequence to its certainty equivalent, $RP = E[\tilde{c}] - c^c$. $c^c =$ certainty equivalent of \tilde{c} , where $U(c^c) = E[U(\tilde{c})]$. This is why we buy insurance. We would pay an *insurance premium* up to the certainty equivalent ($= RP + E[\tilde{c}]$). However, a large and perfectly-efficient insurance company sees only the expected value cost for the policy $E[\tilde{c}]$.

References

- Arrow, K.J. and R.C. Lind (1970), "[Uncertainty and the evaluation of public investment decisions](#)," *American Economic Review*, Vol. 60, No. 3, June, pp. 364-378.
- Keeney, R.L. and H. Raiffa (1976), *Decisions with Multiple Objectives*. New York: Wiley. Reprinted in 1993 by Cambridge University Press.

3. Expected Utility Theory

Two objectives commonly are given for utility theory:

1. Utility theory should guide rational decision-making.
2. Utility theory should mimic how real decisions are made.

Today's installment focuses on the use of utility theory to guide rational decision-making. In the next installment, we'll see how well it represents real decision-making.

Recall that in expected utility theory, the objective is to select alternative j that maximizes:

$$EU_j = \sum_{i=1}^n p_{ij} U(c_i).$$

What is rational decision-making? (You tell me.)

What characteristics should a measure of alternative performance have to be rational? Does expected utility have these characteristics?

An evaluation criterion function $F()$ allows for the ordering of alternatives.

Some notation: $A > B$, alternative A is preferred to B
 $A < B$, alternative B is preferred to A
 $A \sim B$, alternatives A and B are indifferently preferred.

Some commonly suggested characteristics (axioms) for rational utility measurement are:

1. *Existence of Relative Preference*

For any pair of alternatives, either $A > B$, $A < B$, or $A \sim B$. Thus, only $F(A) > F(B)$, $F(A) < F(B)$, or $F(A) = F(B)$ are possible.

2. *Transitivity*

For any three alternatives, if $A > B$ and $B > C$, then $A > C$. Thus, if $F(A) > F(B)$ and $F(B) > F(C)$, then $F(A) > F(C)$.

3. *Mixture Continuity*

More utility $F()$ is always better. If α is a constant, $0 \leq \alpha \leq 1$, and $F(A) > F(B) > F(C)$, then there is a value of α such that $F(B) = \alpha F(A) + (1-\alpha)F(C)$.

4. *Existence and Availability of Probabilities for Outcomes*

Probability values exist and can be quantified for each possible outcome of each alternative.

5. *Monotonicity of Probability Effect on Utility*

Higher probabilities make good outcomes better. For outcome X, where $F(X) > 0$, probability of outcome p, and $\Delta > 0$, $F(p+\Delta, X) > F(p, X)$.

6. *Substitution Axiom*

For any three outcomes, X1, X2, and X3, and probability p, if $X1 \sim X2$ ($F(X1) = F(X2)$), then the choice between bundles of outcomes (X1 or X3) and (X2 or X3) are indifferently preferred:

$$pF(X1) + (1-p)F(X3) \sim pF(X2) + (1-p)F(X3).$$

This allows substitution of X1 for X2, or vice versa, in decision problems.

7. *Independence of Irrelevant Alternatives*

The principle of the independence of irrelevant alternatives is that if $A > B$, then A cannot be made \sim or $<$ B by introducing some new alternative C. This axiom can be derived from a combination of axioms 3 and 6 (Luce and Raiffa 1957).

Expected utility theory satisfies these characteristics of rational evaluation.

Several "derivations" of expected utility exist from sets of axioms similar to those above (von Neumann and Morgenstern 1947; Luce and Raiffa 1957; Machina 1990). For us, it is sufficient that the expected utility form satisfies these characteristics (or definition) of "rationality." So expected utility is a common

definition of rational behavior (Fishburn 1988), $F() = EU_j = \sum_{i=1}^n p_{ij} U(c_i)$. [$F()$ is linear with probability.]

Some History of Utility Functions

Daniel Bernoulli (1738): $U = c \log(Y/Y_s)$, where U is utility, c is a constant, Y is income, and Y_s is subsistence income.

Weber-Fechner Law: $U = c \log(\Delta Y/Y_0)$, where U is utility, c is a constant, ΔY is income change, and Y_0 is initial income.

Von Neumann and Morgenstern's (1944) work revolutionized the basis and thinking about utility functions.

How can we estimate a utility function?

Now that we know that decision-making by expected utility is "rational," how can we estimate the necessary utility function $U()$ for a single type of consequence, such as monetary gains and losses?

Most modern approaches to estimating utility functions begin by establishing an arbitrary consequence as the origin and an arbitrary unit of utility. Some common origin and scale conventions are:

- Let $U(C_{\min}) = 0$, $U(C_{\max}) = 1$,
- Let $U(C_{\min}) = -1$, $U(C_{\max}) = 0$, or
- Let $U(C = 0) = 0$, $U(C_{\text{ref}}) = 100$,

where C_{\min} is the minimum possible consequence or outcome, C_{\max} is the greatest possible outcome, and C_{ref} is some arbitrary reference consequence (say, \$1,000). In theory, any other scale could be used.

Having established an origin and a scale, the certainty equivalent method seeks to define additional utility values for other consequence values by offering a choice between a certain outcome and a probabilistic outcome (von Neumann and Morgenstern 1944).

Let $U(\$0) = 0$, $U(\$1,000) = 100$. To find the utility of \$2,000, the question is asked, "At what probability p would the following two alternatives be indifferent to you?"

- A) \$1,000 for certain.
- B) probability p of \$2,000 and probability $1-p$ of \$0.

For indifference between these two alternatives, by expected utility theory, $EU_A = EU_B$, or

$$U(\$1,000) = pU(\$2,000) + (1-p)U(0), \text{ or}$$
$$100 = pU(\$2,000) + (1-p)0.$$

Therefore, $U(\$2,000) = 100/p$.

For the utilities of consequences less than the utility scale length, say \$500 in this case, similar choices of wagers can be made. At what p are you indifferent between the following choices:

- C) \$1,000 with a probability p and \$0 with a probability $1-p$.
- D) \$500 for certain.

Since at indifference, $pU(\$1,000) + 0 = U(\$500)$, $U(\$500) = 100p$.

Several variations exist on how to implement this general approach (de Neufville 1990). Most variations are attempts to overcome some problems of 1) propagating "errors" in interview responses and 2) responses which violate some of the axioms of utility theory (such as those we shall see when comparing certain, uncertain, and rare events).

Multi-objective Utility Functions

Many real problems have multiple dimensions to performance, where it is difficult or controversial to collapse all indicators of system performance into one overall objective. Truth, beauty, money, and viability of a population of endangered species cannot always be expressed and compared in the same units. Keeney and Raiffa (1976) and de Neufville (1990) propose the use of multi-objective utility models for such problems. [This work is interesting, and I recommend you read it. Someday, I'll add more about it here, when I figure out a way to distill it to a page or two. If you take this interesting approach too seriously, you never escape the multi-utility function estimation problem, especially with multiple decision-makers. The probabilistic dominance material in an earlier section has a multi-utility theory basis, I believe. It just hasn't been worked out. Stochastic dominance conditions also should be possible for multi-utility theory. Perhaps this is a good student exercise?]

Practical Use of Utility Theory

In the more than half-century since modern utility theory was first developed and despite its theoretical appeal and long-standing academic calls for its use, it is used very rarely in practice. Some reasons for this include:

- There are several psychometric problems in estimating utility functions from individual survey responses. We shall see some of these later. Also, it is possible to make utility estimates for only about 30-50 values of a single type of consequence (de Neufville 1990).
- You can imagine how utility responses might vary between individuals and even vary for the same individual on different days. Furthermore, many planning problems evolve over decade-long periods, during which utility functions are likely to change considerably. People's valuations are likely to change as they learn and think more about a problem.
- It is often difficult to decide who in a large organization or group should be interviewed for estimating a utility function. For a private company, should the president, middle-manager, board members, or stockholders be interviewed? For public decisions, should the utility function be that of the agency head, private citizens, or elected officials?
- High-level decision-makers are busy people; it is often hard to get their time for interviews.
- Decision-makers are often unsure what their preference structure/utility function is. For newly controversial issues, these preferences may be highly unstable as new information develops.
- Many decision-makers have political reasons to avoid revealing their true preference or utility functions.
- Utility theory and processes for estimating utility are somewhat abstract and subjective. The process of estimating utility functions is somewhat artificial. These aspects do not help potential users "feel good" about the process.
- Utility function risk aversion is likely to vary between interests. For dam operations for flood control, a large federal interest should probably have a linear utility function (expected cost), but downstream homeowners or elected officials are likely to be risk-averse, and some short-term land developers might even be risk-seeking (given their economies of scale and short time-horizon).

Utility estimation is by no means an easy or exact exercise.

When can I avoid utility functions?

Several problems exist with this approach or virtually any approach to estimating utility functions. This is why the ideas of stochastic dominance and the Arrow-Lind theorem can be so useful. (See references in notes for Lecture 2.)

- Arrow-Lind (1970) prove that expected value is as good as expected utility for decisions with relatively "small" ranges of consequences, especially where the consequences are "small" relative to the total budget or economic scale of the enterprise. For such small changes in performance relative to overall performance of the system, the shape of the utility function should be rather linear.

- Stochastic dominance can help weed out many inferior alternatives under a wider range of conditions (Levy 1992). With some luck, stochastic dominance can eliminate any need for utility function estimation.

- Where the magnitude and range of consequences is a little larger than that discussed by Arrow and Lindh (1970), but still small compared to the economic scale of the enterprise, a mean-variance approximation of the utility function should be possible. But this requires a separate (utility-based) weighting of performance variance, relative to mean.

- Bouchart and Goulter (1998) suggest using an artificial neural network (ANN) to represent water users' preferences for different probability distributions of water availability. The ANN would be calibrated based on user preferences among many hypothetical shortage probability distributions. This approach avoids need for establishing a utility function, but rests more on empirical preferences based on water users' ability to compare probability distributions.

- Change the design to affect the probability distribution of outcomes. Structure recourses in the design so low-probability big costs are transformed into higher probabilities for smaller costs, in essence, insurance. Ideally, this allows you to come closer to using expected value as a decision criterion.

Mean-variance approximation of a utility function for small consequences

If consequences are all very small compared to the scale of the overall enterprise, a common condition for public policy, then utility functions within this small range should be linear, and expected value is an equivalent decision-making criterion to expected utility (Arrow and Lindh 1970). But, if the consequences are merely small compared to the overall enterprise, then a mean-variance approach will better approximate a utility function, at the cost of estimating one parameter (Keeney and Raiffa 1976, p. 161). For this case,

$$\boxed{EU_j = U(E[c_j]) - k \sigma_{c_j}^2},$$

where $E[c_j]$ is the average consequence for decision j , $\sigma_{c_j}^2$ is the variance of consequences with decision j , and k is the ration of the second to first derivatives of the utility function at the magnitude of the average consequence.

This result can be derived from the definition of certainty-equivalence, in the context of the average consequence μ_c , so $U(c) = U(\mu_c + \tilde{c}')$, where $E[\tilde{c}'] = 0$. Begin with the definition of certainty-equivalence: $E[U(\tilde{c})] = U(c^c)$ and risk premium $r_p = E[\tilde{c}] - c^c$, where $c^c =$ certainty equivalent of \tilde{c} . Expand about the mean:

$$E[U(\tilde{c})] = U(\mu_c - r_p) = E[U((\mu_c + \tilde{c}') - r_p)], \text{ where } \tilde{c}' = \tilde{c} - E[\tilde{c}], E[\tilde{c}] = \mu_c.$$

Taking a first-order Taylor series expansion for $U(\mu_c - r_p)$ gives:

$$U(\mu_c - r_p) \approx U(\mu_c) - r_p dU(\mu_c)/dc$$

and a second-order Taylor series expansion for $E[U((\mu_c + \tilde{c}'))]$ gives:

$$E[U((\mu_c + \tilde{c}'))] \approx E[U(\mu_c) - \tilde{c}' dU(\mu_c)/dc + 0.5 \tilde{c}'^2 d^2U(\mu_c)/dc^2]$$

Setting equivalence, simplifying (note $E[\tilde{c}'] = 0$), and solving for r_p :

$$r_p \approx E[-0.5 \tilde{c}'^2 (d^2U(\mu_c)/dc^2)] / (dU(\mu_c)/dc) = k(\mu_c) \sigma_c^2.$$

For these conditions, the risk premium is proportional to the variance of the consequence σ_c^2 . Returning this result to develop an estimate of utility based on mean and variance:

$$\boxed{E[U(\tilde{c})] \approx U(\mu_c) - k(\mu_c) \sigma_c^2}.$$

So where the mean consequence is the same, and the range of consequences is small enough to neglect higher order terms, it is sufficient to make decisions based on minimizing the variance of the consequences.

Where the range of consequences around the mean consequence for each decision is small (and the mean consequence differs), then the full utility analysis is reduced to consideration of only the mean and variance of consequences for each decision alternative.

References

- Arrow, K.J. and R.C. Lind (1970), "[Uncertainty and the evaluation of public investment decisions](#)," *American Economic Review*, Vol. 60, No. 3, June, pp. 364-378.
- Bouchart, François J.-C. and Ian C. Goulter (1998), "Is Rational Decision-Making Appropriate for Management of Irrigation Reservoirs?," *Journal of Water Resources Planning and Management*, ASCE, Vol. 124., No. 6, November-December, pp. 301-309.
- de Neufville, Richard (1990), *Applied Systems Analysis*, McGraw-Hill Book Co., N.Y.
- Fishburn, Peter C. (1988), "Expected Utility: An Anniversary and a New Era," *Journal of Risk and Uncertainty*, Vol. 1, No. 3, September, pp. 267-283.
- Keeney, R.L., and H. Raiffa (1976), *Decisions with Multiple Objectives*. New York: Wiley. Reprinted in 1993 by Cambridge University Press
- Luce, R. Duncan and Howard Raiffa (1957), *Games and Decisions*, John Wiley and Sons, N.Y.
- Machina, Mark J. (1990), "Expected Utility Hypothesis," in J. Eatwell, et al. (eds), *Utility and Probability*, W.W. Norton & Co., N.Y., pp. 79-85.
- Savage L.J. (1954), *The Foundations of Statistics*, John Wiley and Sons, N.Y.
- von Neumann, John and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.

3b. Does Expected Utility Theory Work With Real People?

The last lecture was meant to establish expected utility as a rational way to evaluate decision options. How closely does expected utility mimic real decision-making?

Allais Paradox

We begin with the famous Allais Paradox (Allais 1953).

Consider 2 sets of choices:

Choice Set 1:

A) $p=0.8$ of \$4,000 and $p=0.2$ of zero

B) $p=1$ of \$3,000

Choice Set 2:

C) $p=0.2$ of \$4,000 and $p=0.8$ of zero

D) $p=0.25$ of \$3,000 and $p=0.75$ of zero

For each choice set, select a choice, A or B and C or D.

Let $U(0)=0$. Therefore, $EU(A)=0.8 U(\$4,000)$
 $EU(B)=U(\$3,000)$
 $EU(C)=0.2 U(\$4,000)$
 $EU(D)=0.25 U(\$3,000)$.

Most people prefer B and C.

If $EU(B) > EU(A)$, then $0.8 U(\$4,000) < U(\$3,000)$.

If $EU(C) > EU(D)$, then $0.2 U(\$4,000) > 0.25 U(\$3,000)$,
or $0.8 U(\$4,000) > U(\$3,000)$.

These two statements seem to conflict and illustrate some problems with getting actual human choices to mimic the axioms of utility theory. Allais holds, and the example illustrates, that people have a preference for a sure thing (high reliability) greater than that predicted by expected utility.

Other Work

Work by several people has illustrated various interesting deviations of real decision-making from expected utility decision-making (Kahneman and Tversky 1979; Kahneman, et al. 1982). Simon (1976) is a leader of another interesting approach, "bounded rationality".

There is sometimes professional dilemma between making rational designs and giving people what they say they want. Ideally, good management tries to make these converge.

References

- Allais, M. (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine," *Econometrica*, Vol. 21, No. 4, October, pp. 503-546.
- Allais, M. (1990), "Allais Paradox," in J. Eatwell, et al. (eds), *Utility and Probability*, W.W. Norton & Co., N.Y., pp. 3-9.
- Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, Vol. 47, No. 2, March, pp. 263-291.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky (eds.) (1982), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, UK.
- Simon, Herbert A. (1976), *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, Free Press, N.Y.

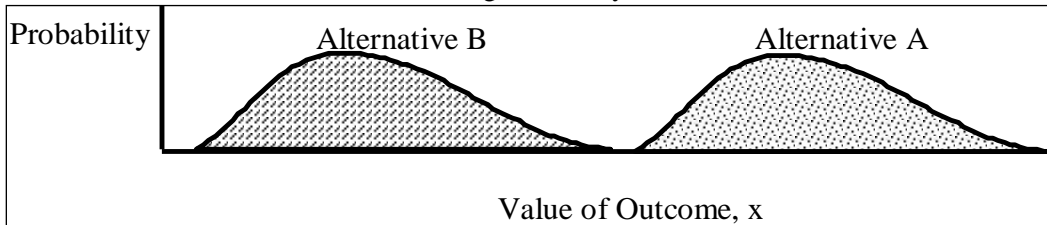
4. Stochastic Dominance Decisions under Uncertainty

Stochastic Dominance

It is often possible to select an alternative using the expected utility criterion without exact knowledge of the utility function $U(\cdot)$. This approach is known as *stochastic dominance*.

Zero-Degree Stochastic Dominance

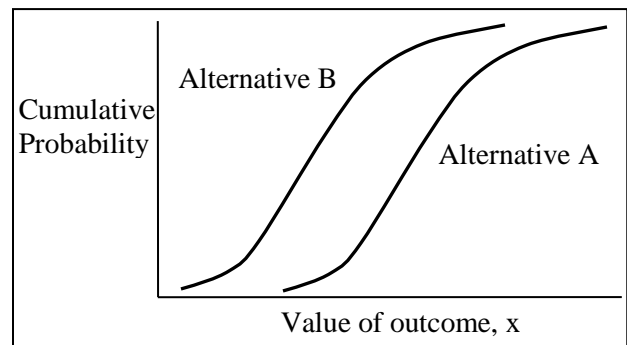
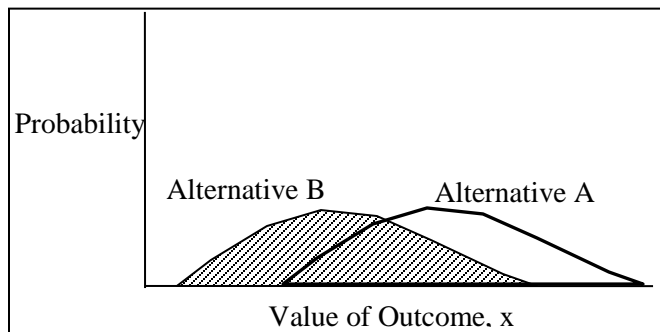
Which alternative below is better, if x is good? Why?



This can be used to eliminate bad alternatives (B) for any utility function that is non-decreasing with x . (Larger values of x are never worse, always better or equal to, smaller values of x .)

First-Degree Stochastic Dominance

Which alternative is better? Why?



Alternative A will usually be better. This too can be used to eliminate lesser alternatives for any monotonically increasing utility function. The mathematical definition of first-degree stochastic dominance for alternative A over B requires,

$$\int_{-\infty}^x p_A(s) ds \leq \int_{-\infty}^x p_B(s) ds, \text{ for all values of } x,$$

and strict inequality holds for at least one value of x , where $p_A(s)$ and $p_B(s)$ are the probability density functions of the performance of alternatives A and B. These conditions are often written in terms of the cumulative distributions of A and B's performance $F_A(s)$ and $F_B(s)$: $F_A(x) \leq F_B(x)$, for all values of x , and strict inequality holds for at least one value of x . For first-degree stochastic dominance to apply, the utility function may be any non-decreasing function with finite values for finite values of x (risk averse, risk seeking, or risk neutral).

Second-Degree Stochastic Dominance

The mathematical definition for second-degree stochastic dominance for A dominates B is:

$$\int_{-\infty}^x \int_{-\infty}^t p_A(s) ds dt \leq \int_{-\infty}^x \int_{-\infty}^t p_B(s) ds dt, \text{ for all values of } x,$$

and strict inequality holds for at least one value of x , where $p_A(s)$ and $p_B(s)$ are the probability density functions of the performance of alternatives A and B.

These conditions are often written in terms of the cumulative distributions of A and B's performance $F_A(s)$ and $F_B(s)$:

$$\int_{-\infty}^x F_A(s) ds \leq \int_{-\infty}^x F_B(s) ds, \text{ for all values of } x,$$

and strict inequality holds for at least one value of x . For second degree dominance to apply, the utility function must be non-decreasing and concave (any form of risk averse).

Third-Degree Stochastic Dominance

If the utility function is non-decreasing, concave, and has a non-negative third derivative, third degree stochastic dominance can be applied (Whitman 1970; Bawa 1975). A non-negative third derivative for the utility function implies decreasing risk aversion as absolute positive outcomes increase – Richer people should be less risk averse. These conditions for A dominates B are:

$$\int_{-\infty}^x \int_{-\infty}^t F_A(s) ds dt \leq \int_{-\infty}^x \int_{-\infty}^t F_B(s) ds dt, \text{ for all values of } x,$$

and strict inequality holds for at least one value of x . Bawa (1975) finds some simplifications of 3rd-degree stochastic dominance for several common classes of probability distributions.

If an appropriate stochastic dominance condition is met, it is a sufficient condition for proving that Alternative A is better than Alternative B. However, if a stochastic dominance condition is not met, Alternative A might still be better than Alternative B. Stochastic dominance is not a necessary condition.

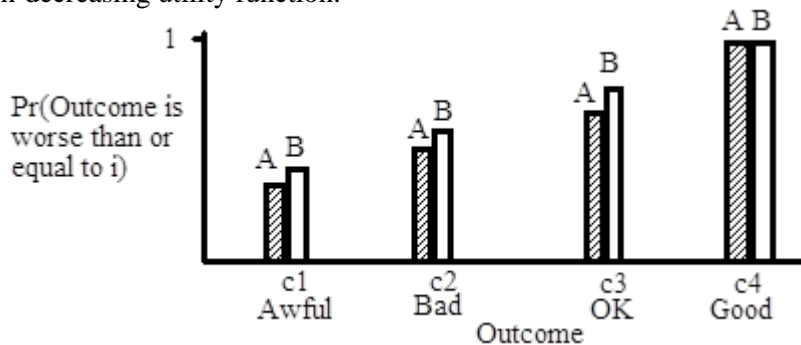
While stochastic dominance cannot always yield selection of the best alternative, these criteria are easy to apply and can sometimes yield a single optimal solution or a smaller sub-set of solutions without the need to specify a complete utility function. As we shall see, this can be quite an advantage. Good general references for stochastic dominance are Chapter 11 of Park and Sharp-Bette (1990), Bawa (1975), and Levy (1992). A nice application is Su, et al. (1991).

Ordinal Stochastic Dominance

Frequently, the potential outcomes of alternatives are not interval-scales (\$0-\$1,000), but fall on an ordinal scale (awful, bad, good). Spector, et al. (1996) develop first and second degree stochastic dominance methods for this ordinal performance case. Given our frequent inability to assess outcomes on a numerical interval scale, ordinal stochastic dominance can be useful.

Ordinal First Degree Stochastic Dominance (OFSD)

Where performance outcomes are ordinal, Alternative A dominates Alternative B if $F_A(i) \leq F_B(i)$, for all outcome levels i , ordered from worst to best. This case is depicted below. Like FSD, it can be applied to any case with a non-decreasing utility function.



Ordinal Second Degree Stochastic Dominance (OSSD)

Ordinal Second Degree Stochastic Dominance applies where the relative ordering of the differences in outcomes is decreasing. This would be the case if the difference between "Awful" and "Bad" were considered to be greater (worse) than the difference between Bad and OK. By OSSD, Alternative A dominates Alternative B if

$$\sum_{j=1}^k \sum_{i=1}^j P_{Ai} \leq \sum_{j=1}^k \sum_{i=1}^j P_{Bi}, \text{ for all performance levels } k,$$

where p_{Ai} and p_{Bi} are the probabilities for each outcome i .

OSSD applies only if the utility function is non-decreasing (positive first derivative) and has a negative second derivative, as with ordinary Second Degree Stochastic Dominance.

Anyone want to derive or research ordinal third degree stochastic dominance?

Example of Ordinal Stochastic Dominance

Consider the case of two alternative designs (of equal cost) for a habitat restoration project intended to protect a species. Three potential outcomes are considered, as shown in the table below.

Table 4: Comparison of two alternatives with ordinal stochastic dominance

Outcome	<u>Alternative outcome probabilities</u>		<u>OFSD (cumulative probabilities)</u>		<u>OSSD (twice cumulative probabilities)</u>	
	A	B	$F_A(i)$	$F_B(i)$	$\sum \sum P_{Ai}$	$\sum \sum P_{Bi}$
Extinction	0.20	0.30	0.20	0.30	0.20	0.30
Species OK	0.20	0.05	0.40	0.35	0.60	0.65
<u>Recovery</u>	<u>0.60</u>	<u>0.65</u>	1.00	1.00	1.60	1.65
Total	1.00	1.00				

By OFSD it is unclear if Alternative A is better than Alternative B, since B has a higher chance of Recovery, but also a higher chance of Extinction. This is shown in the OFSD columns of the table.

If the difference between extinction and OK is seen as greater (worse loss) than the difference between OK and recovery, then OSSD can be applied, as in the last pair of columns. Since each sum is greater for Alternative B than Alternative A, then if it is worse to go from OK to extinction than from recovered to OK, then Alternative A dominates Alternative B.

Multi-objective stochastic dominance

There is some literature on multi-objective stochastic dominance for continuous (Denuit et al. 2013) and ordinal performance (Yalonetzky 2013). Stochastic dominance might be useful for bringing uncertainty better into Pareto-optimality analyses.

Multi-objective Probabilistic Dominance

Pareto-optimality or deterministic dominance is used to eliminate poorer-performing solutions for multi-objective problems. How can such ideas of dominance be extended to multi-objective problems with probabilistic performance of alternatives? If the boxed areas in Figure 1 (Table 5) are uniformly and independently distributed probabilities of cost and environmental performance for each alternative, then we can calculate the probability of each alternative having superior performance for each objective, separately and together, Table 6 (Lund, et al. 2008). This indicates the probability of multi-objective Pareto-optimality for each alternative (Cohon 1978). This is called "probabilistic dominance" because the dominance between alternatives is not certain (Teich 2001; Fieldsend and Everson 2005, Hudgins 1994).

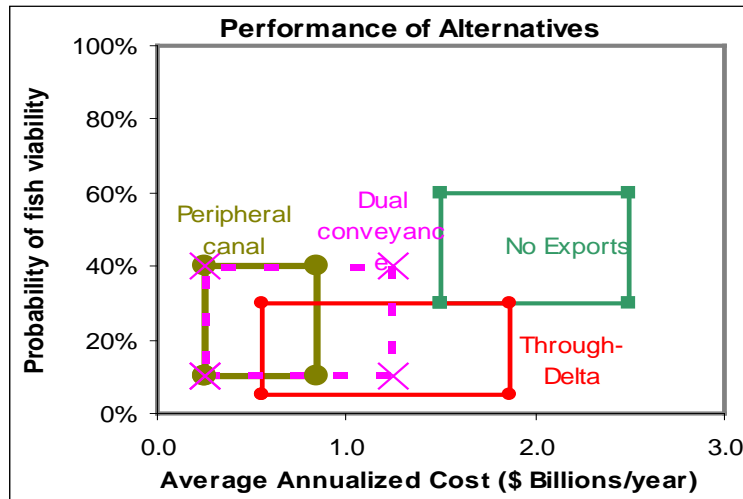


Figure 1 - Performance and Superiority

Table 5 - Costs & Fish Population Viability for Delta Alternatives (Lund, et al. 2008)

Alternative	Average Cost (\$ billion/year)	Likelihood of Viable Fish Population (%)
Continued through-Delta exports	0.55-1.86	5-30
Peripheral canal	0.25-0.85	10-40
Dual conveyance	0.25-1.25	10-40
No exports	1.50-2.50	30-60

How can we evaluate the overlap of performance in multi-objective space? What is the probability that each alternative is superior to others? (Table 6) How can we calculate such results?

Table 6 - Probabilities That Each Export Alternative is Superior to Others, With Fish Viability as the Environmental Objective (%)

Alternatives	Fish Viability	Statewide Cost	Both Objectives
Through-Delta better than peripheral canal	27	6	1
Through-Delta better than dual conveyance	27	19	5
Through-Delta better than no exports	0	95	0
Peripheral canal better than no exports	6	100	6
Peripheral canal better than through-Delta	73	94	69
Dual conveyance better than no exports	6	100	6
Dual conveyance better than through-Delta	73	81	60
No Exports better than peripheral canal	94	0	0
No Exports better than through-Delta	100	5	5

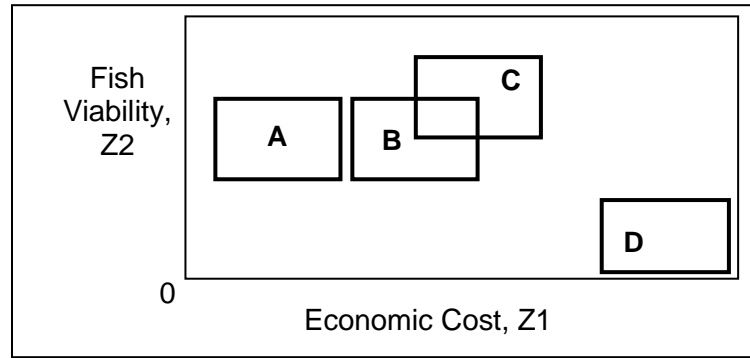


Figure 2 – Ranges of estimated performance for four alternatives

Figure 2 helps illustrate this concept. Here, alternative D is clearly inferior to alternatives A, B, and C, having no possibility of performing better than A, B, or C on either the fish or economic objective. Let $P(A \leq B)$ be the probability that alternative A is superior to alternative B. Then, from Figure 2, $P(A \leq D) = P(B \leq D) = P(C \leq D) = 1$ for either objective, and for both objectives taken together. Alternative A is always superior to B in terms of cost. Half the time A will be superior to B in terms of fish viability, with an equal probability that B will be superior to A for fish. Overall, there is a 50% chance that A is superior to B for both fish and cost, and zero probability that B will be superior to A on both objectives.

If C overlaps area B by 25%, as shown with the southwest corner of C at the centroid of B, there is a 87.5% chance that C is better for fish than B (or A) ($=0.5 + 0.5(0.5+0.25)$), with a 12.5% chance that B (or A) is better for fish ($=0.5*0.5*0.5$). For cost, there is a 87.5% chance that B is superior to C, and 12.5% chance that C is better than B. Overall, there is a 10.9% chance that B is superior to C for both objectives, and a 10.9% chance that C is superior to B for both objectives. C will never be superior to A for both objectives, but there is a 12.5% chance that A is superior to C for both objectives. For the remaining probability, each alternative would be better on a different objective, indicating a performance trade-off requiring a value judgment. These calculations are explained more in the next section.

For many cases, the performance will not be uniformly distributed, but have some central tendency, often near the centroid. In these non-uniformly distributed cases with a centroidal central tendency, the above calculations are somewhat skewed, especially for overlaps of corners far from the centroid of an alternative's performance region.

A More Formal Derivation

Given two alternatives, each with uncertain and probabilistic performance on two objectives, what is the probability that the actual (point) performance of each alternative will be superior (non-dominated or Pareto-optimal) relative to the other? Consider the problem as in Figure 3, with the assumption that more of Z1 is inferior and more of Z2 is superior (monotonic single-objective utility functions), so ideal performance is in the Northwest corner of the diagram.

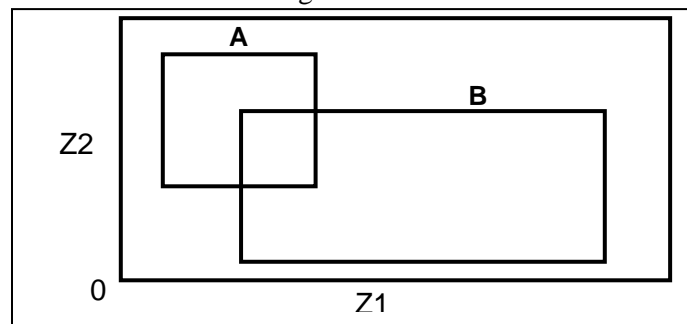


Figure 3 - Performance and Superiority for Two Overlapping Alternatives

Let actual performance of alternatives A and B be points in this multi-objective space, but our knowledge of their future performance is only given by probability distributions $P_A(Z_1, Z_2)$ and $P_B(Z_1, Z_2)$. What is the probability that alternative A will be superior to B in terms of each objective and what is the probability that the performance of A will be superior to B for both objectives (i.e., B is a dominated solution)?

Let $P(AsB)$ be the probability that A is superior to B for both objectives, and let $PZi(AsB)$ be the probability that A is superior to B with respect to objective i.

To define overall probabilistic dominance or superiority of A over B,

$$P(AsB) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(P_A(Z_1, Z_2) \int_{Z_1}^{\infty} \int_{-\infty}^{Z_2} P_B(z_1, z_2) dz_2 dz_1 \right) dZ_1 dZ_2,$$

where the probabilistic performance of each alternative is independent. If not independent, the probabilistic performance of alternative B is a joint probability with alternative A, then

$$P(AsB) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(P_A(Z_1, Z_2) \int_{Z_1}^{\infty} \int_{-\infty}^{Z_2} P_B(z_1, z_2 | A(Z_1, Z_2)) dz_2 dz_1 \right) dZ_1 dZ_2,$$

where $P_B(z_1, z_2 | A(Z_1, Z_2))$ is the probability distribution of B's performance given A's performance at Z_1, Z_2 .

For a single objective, the probability that A is superior to B is

$$PZ1(AsB) = \int_{-\infty}^{\infty} \left(P_A(Z_1) \int_{Z_1}^{\infty} P_B(z_1) dz_1 \right) dZ_1$$

for independent probabilistic performance, where less of objective Z1 is better. Where more of objective

Z2 is better,

$$PZ2(AsB) = \int_{-\infty}^{\infty} \left(P_A(Z_2) \int_{-\infty}^{Z_2} P_B(z_2) dz_2 \right) dZ_2$$

This should apply to any two alternatives in a two-dimensional performance space. Higher dimensional objective spaces can be incorporated with suitable additional integrals.

Consider special conditions where the point performance of A and B are independent probability distributions, each of which is uniform between different ranges of performance on each objective, with some overlap of these distributions as in Figure 3. Overlapping and partially overlapping sub-areas of each distribution can be defined as in Figure 4, with more Z2 and less Z1 being better.

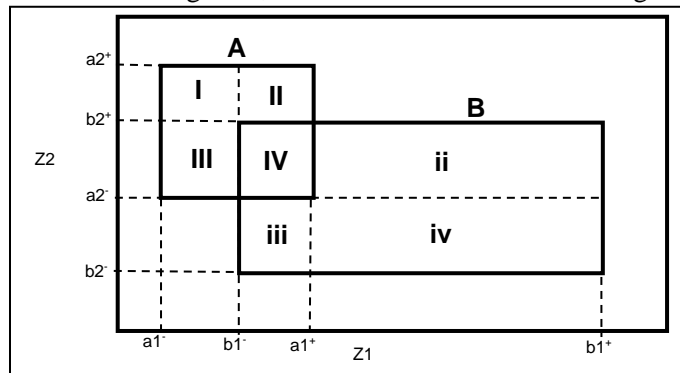


Figure 4 - Performance and Superiority for Two Overlapping Objectives - Detailed

$$P(AsB) = \frac{I}{A} + \frac{II}{A} \left(\frac{ii+iv}{B} + 0.5 \frac{IV+iii}{B} \right) + \frac{III}{A} \left(\frac{iii+iv}{B} + 0.5 \frac{IV+ii}{B} \right) + \frac{IV}{A} \left(\frac{iv}{B} + 0.5 \frac{ii+iii}{B} + 0.25 \frac{IV}{B} \right)$$

where A is the total area of A, B is the total area of B, and each Roman number represents a different quadrant area, as shown in Figure 4. If A falls in area I, there is 100% chance that A is superior to B for both objectives. If A falls in area II, there is 100% chance of A being superior for both objectives if B falls in areas ii or iv, and a 50% chance of A being superior if B falls in areas IV or iii. If both A and B fall in area IV, there is a 25% chance of A being superior to B and an identical chance of B being superior to A for both objectives. With complete overlap for both objectives, there is a 25% change of dominance. With an overlap for one objective, there is a 50% chance of dominance for that objective. When there is no overlap for any individual objective dimension, one solution will completely dominate the other.

The area-based equation above can be expanded to the coordinate-based equation:

$$P(AsB) = \frac{(a2^+ - b2^+)(b1^- - a1^-)}{A} + 0.25 \frac{\left((b2^+ - a2^-)(a1^+ - b1^-) \right)^2}{AB} +$$

$$\frac{(a2^+ - b2^+)(a1^+ - b1^-)}{AB} \left((b2^+ - a2^-)(b1^+ - a1^+) + (a2^- - b2^-)(b1^+ - a1^+) + \frac{(b2^+ - a2^-)(a1^+ - b1^-) + (a2^- - b2^-)(a1^+ - b1^-)}{2} \right)$$

$$+ \frac{(b2^+ - a2^-)(b1^- - a1^-)}{AB} \left((a2^- - b2^-)(a1^+ - b1^-) + (a2^- - b2^-)(b1^+ - a1^+) + \frac{(b2^+ - a2^-)(a1^+ - b1^-) + (b2^+ - a2^-)(b1^+ - a1^+)}{2} \right)$$

$$+ \frac{(b2^+ - a2^-)(a1^+ - b1^-)}{AB} \left((a2^- - b2^-)(b1^+ - a1^+) + \frac{(b2^+ - a2^-)(b1^+ - a1^+) + (a2^- - b2^-)(a1^+ - b1^-)}{2} \right)$$

$$+ \frac{(b2^+ - a2^-)(a1^+ - b1^-)}{AB} \left((a2^- - b2^-)(b1^+ - a1^+) + \frac{(a2^- - b2^-)(b1^+ - a1^+) + (a2^- - b2^-)(a1^+ - b1^-)}{2} \right)$$

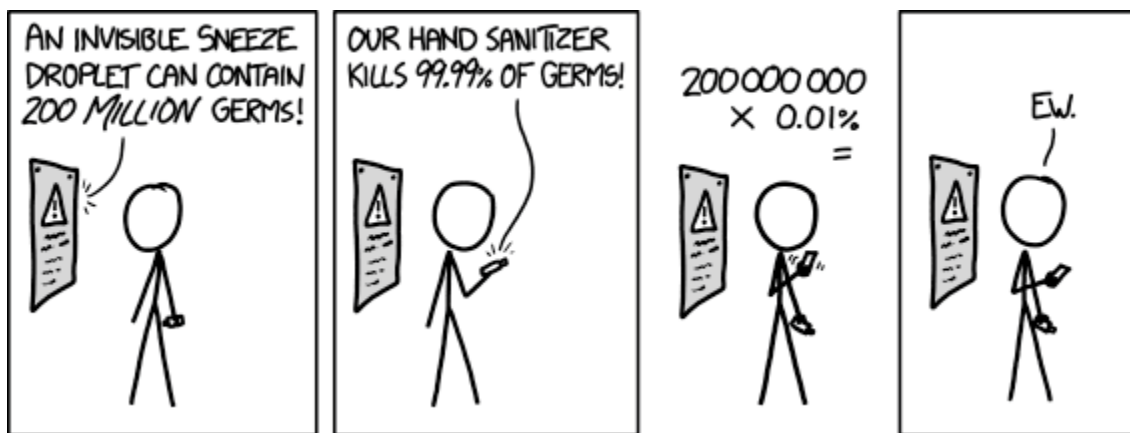
If terms are given values of zero when they become negative, this equation should be generalizable without foreknowledge of the geometric arrangement of alternatives in objective space. The calculations are much simpler when the alternatives overlap on only one objective, or if they do not overlap at all. Where probabilities of performance are independent, the product of the probabilities of dominance for each single objective is the probability that an alternative is superior for all objectives.

This is the basis for calculations in Table 6.

References

- Bawa, Vijay S., "Optimal Rules for Ordering Uncertain Prospects," *Journal of Financial Economics* 2, 1975, 95-121.
- Cohon, J.L. (1978), *Multiobjective Programming and Planning*, Academic Press, NY.
- Fieldsend, J.E. and R.M. Everson (2005), "Multiobjective optimisation in the presence of uncertainty," *2005 IEEE Congress on Evolutionary Computing*, proceedings.
- Hudgins, W.R. (1994), "Patients' Attitude about Outcomes and the Role of Gamma Knife Radiosurgery in the Treatment of Vestibular Schwannomas," *Neurosurgery*, Vol. 34(3), March, p 459-465.
- Levy, H. (1992), "Stochastic Dominance and Expected Utility: Survey and Analysis," *Management Science*, Vol. 38, No. 4, April, pp. 555-593.
- Lund, J.R., E. Hanak, W.E. Fleenor, W. Bennett, R.E. Howitt, J.F. Mount, and P.B. Moyle (2008), *Appendix J: Decision Analysis of Delta Strategies*, Public Policy Institute of California, San Francisco, CA. (www.ppic.org)
- Park, C.S. and G.P. Sharp-Bette (1990), *Advanced Engineering Economics*, John Wiley & Sons, N.Y., 740 pp.
- Spector, Y., M. Leshno, and M. Ben Horin (1996), "Stochastic dominance in an ordinal world," *European Journal of Operational Research*, Vol. 93, pp. 620-627.
- Su, M-D., R. Narayanan, T.C. Hughes, and A.B. Bishop (1991), "Evaluation of Municipal Water Supply

- Operating Rules using Stochastic Dominance," *Water Resources Research*, Vol. 27, No. 7, July, pp. 1519-1527.
- Teich, J. (2001), "Pareto-front exploration with uncertain objectives," in E. Zitzler, et al, (eds.), *Evolutionary Multi-Criterion Optimization, First International Conference, EMO 2001 Zurich*, Switzerland, March 7-9, 2001 Proceedings, Springer-Verlag, Berlin, pp. 314-328.
- Whitmore, G.A. (1970), "Third-Degree Stochastic Dominance," *American Economic Review*, Vol. 60, No. 3, pp. 457-459.
- Yalonetzky, G. (2013) "[Stochastic Dominance with Ordinal Variables: Conditions and a Test](#)," *Econometric Reviews*, 32:1, 126-163, DOI: 10.1080/07474938.2012.690653 – He also has a nice web lecture on the subject: <https://ophi.org.uk/multidimensional-dominance-2/>
- Denuit, M., L. Eeckhoudt, I. Tsetlin, and R.L. Winkler (2013), "[Multivariate Concave and Convex Stochastic Dominance](#)," in F. Biagini, A. Richter, and H. Schlesinger (eds), *Risk Measures and Attitudes*, Springer,



Xkcd.com

5. Structure of Decision Analysis

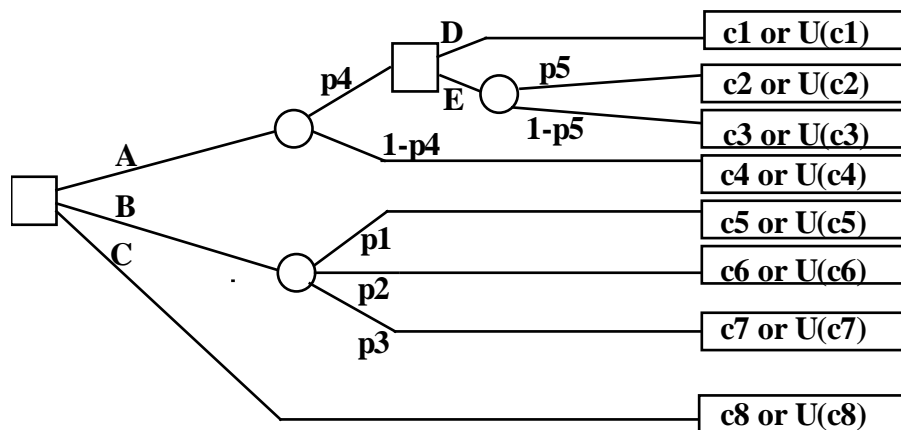
"The test of usefulness of decision theory lies mainly in the future, and it will be made by those who learn about decision theory early enough in their lives that practical experience will not yet have made them feel that orderly, careful thinking about human decisions is futile." – H.V. Roberts, *Journal of Business*, Vol. 43, No. 1 (Jan., 1970), pp. 66-68.

One of the hardest parts of examining choices in a probabilistic framework is depicting the logic and structure of the choices to be made and the precise structure of uncertainties. This becomes particularly difficult and important when choices can be made contingently, sequentially, or in stages (allowing recourse) in a context with multiple uncertainties. Creating designs which are flexible in responding to uncertainties usually requires the representation of sequential and contingent decisions within a structured context of uncertainties.

Decision Trees

Decision trees are commonly used to depict the structure of decision-making problems. Each branch of the tree represents a different choice or outcome. Where the branches connect, represent opportunities for choices or chance.

Here's a decision tree:



Boxes denote *choice nodes*, where a decision must be made.

Circles denote *chance nodes*, where outcomes are uncertain.

Rectangles denote the *values of outcomes* in raw form or in utility form.

Each branch emanating from a choice node is an alternative; a name is attached.

Each branch emanating from a chance node is a possible outcome, with a probability attached.

While many decision analysis problems are too complex to be represented by a full tree (forest?) diagram, the technique is useful for small problems and can help one think about or talk about how a large problem is structured.

An Example with a Twist

Here's an example where a decision tree can illustrate the difference between two seemingly identical choices (Lund 1991). Consider an engineer deciding between two water supply sources, surface water or ground water. Both are uncertain. Streamflow is certainly uncertain. The yield of the aquifer is also uncertain in this case.

Imagine that by an astounding coincidence the initial costs of developing each source is the same (K) and the probability distribution of the annual supply from each source is exactly the same $P_S(Q) = P_G(Q)$. In the event of a shortfall, the engineer has two alternatives, either constructing additional capacity (at cost K_a) or buying additional water (at a present value cost of C/r). Each outcome cost and the probability distributions are identical for the two alternatives. Are the values of the two initial choices (surface versus ground water) identical?

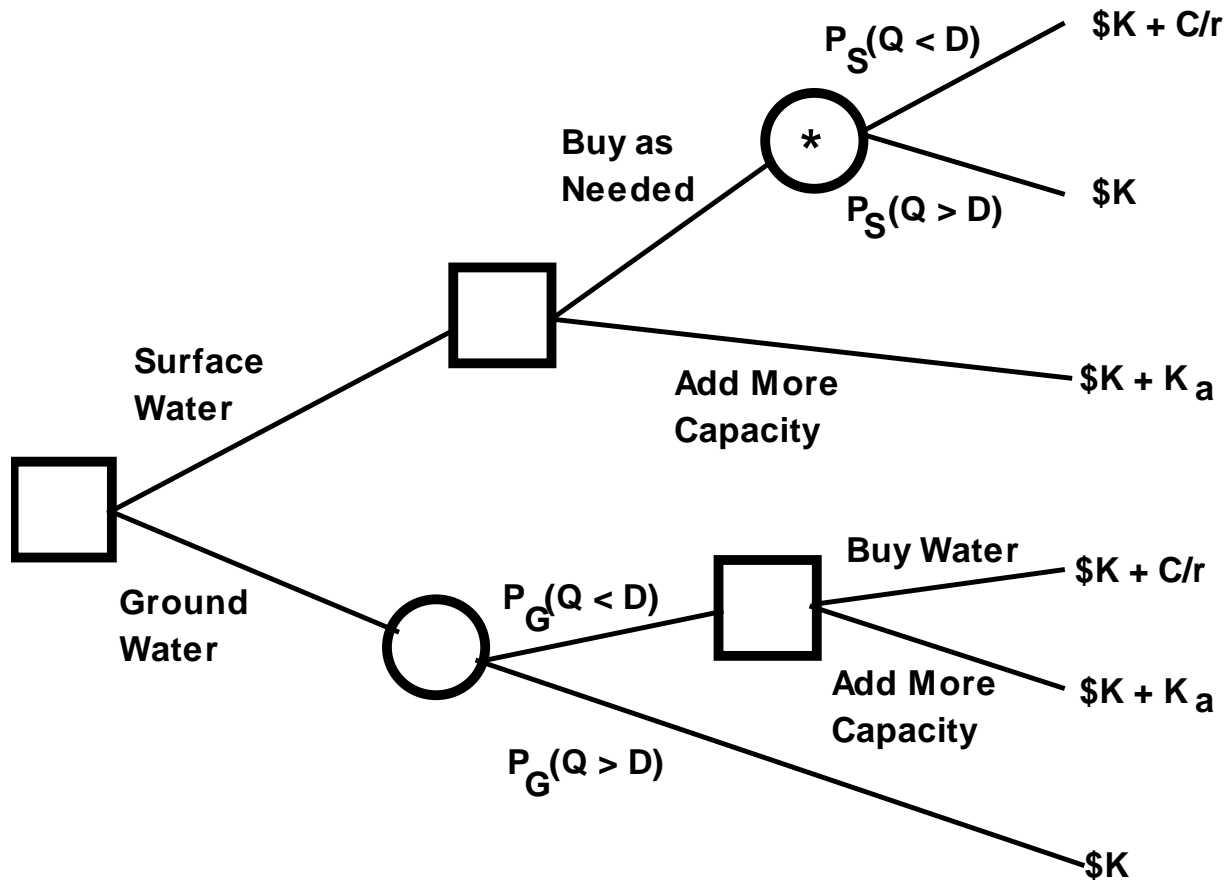
Let's see, using an expected value criterion.

For the surface water source: $EVC_S = K + \text{Min}[P_S(Q < D) C/r, K_a]$,

For the ground water source: $EVC_G = K + P_S(Q < D) \text{Min}[C/r, K_a]$,

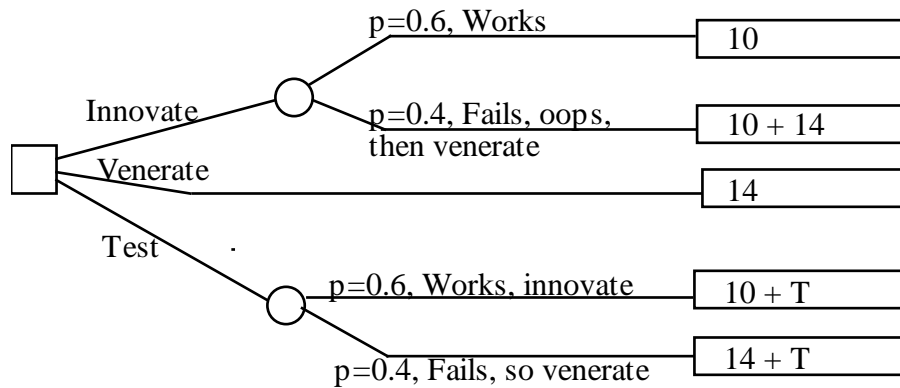
Wait, these are different. Why? Look at the decision tree below. If surface water shortages occur their magnitude will be the same in every year, very much like the random variables of probability theory. If groundwater yield is insufficient, it is unlikely that the permeability or recharge into the aquifer will change enough over time to vary the aquifer's yield. P_S happens every year, with Q being different every year. P_G happens mostly now, but once the aquifer has been in operation for some time P_G should be a much narrower distribution. The decision tree shows the logic and structure of the decision-making problem, including the recourses available after the initial design decisions have been made.

Figure 1: Decision Tree for Water Supply Example



The Value of Information: Perfect Tests

A common use of decision analysis is to estimate the value of a test that would help us make a decision. Consider the example below adapted from Stokey and Zeckhauser (1978), where we want to know if we should adopt an innovative design or a time-tested venerable design. We can also test the innovative design at cost T ; this test is perfect. (If the innovation passes the test, there will be no problems with the design.) How much should we be willing to pay for the test. First, some decision forestry:



Let's resolve this. Each decision has an expected value:

- Innovate: 15.6
- Venerate: 14
- Perfect Test: $T + 11.6$.

What is the value of the perfect test? 4 or 2.4?

It is too late to estimate the value of an imperfect test. We'll do that next time.

References

Corner, J.L. and C.W. Kirkwood (1991), "Decision Analysis Applications in the Operations Research Literature, 1970-1989," *Operations Research*, Vol. 39, No. 2, March-April, pp. 206-219.
 de Neufville, Richard (1990), *Applied Systems Analysis*, McGraw-Hill, N.Y.
 Lund, Jay R. (1991), "Random Variables versus Uncertain Values: Stochastic Modeling and Design," *Journal of Water Resources Planning and Management*, ASCE, Vol. 117, No. 2, March, pp. 179-194.
 Stokey, Edith and Richard Zeckhauser (1978), *A Primer for Policy Analysis*, W.W. Norton & Co., N.Y.

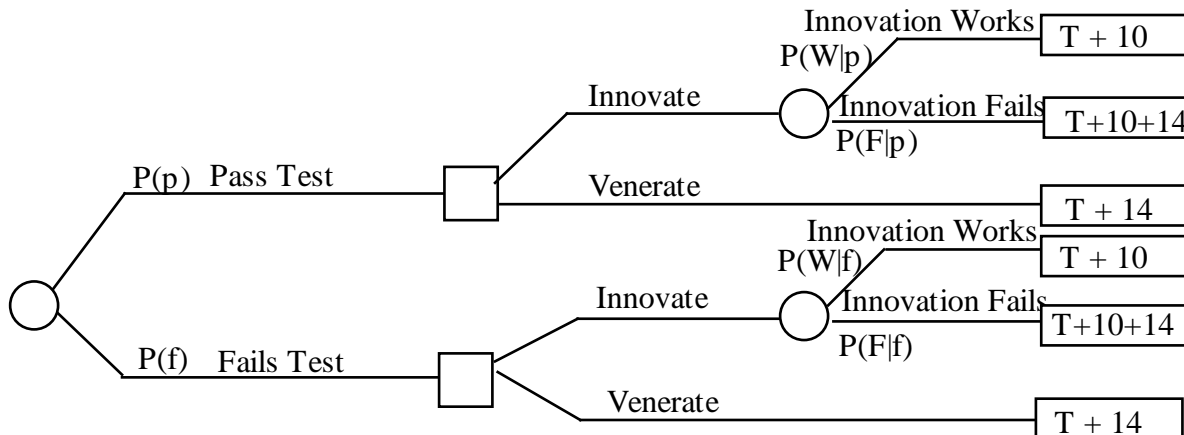
6. The Value of Imperfect Information

Let's modify the example from Lecture 5 on the value of perfect information to see the value of imperfect information from an imperfect test. This involves adding another test type to the decision tree.

With imperfect information, there is a chance that if the innovation passes the test, the prototype will fail and a chance that if the prototype fails the test, that the prototype would succeed. This can be expressed as the set of *conditional probabilities*:

Test Result	P(result)	Prototype Result	
		Works	Fails
Pass	$P(p) = 4/7$	$P(W p) = 0.9$	$P(F p) = 0.1$
Fail	$P(f) = 3/7$	$P(W f) = 0.2$	$P(F f) = 0.8$

This complicates the tree structure of the decisions and outcomes of the test option in the Lecture 5 problem. The following set of branches represents the outcomes and probabilities for the new, imperfect test. You could replace the perfect test branch in the Lecture 5 problem with this, if you like.



Note that $P(W) = P(p) P(W|p) + P(f) P(W|f)$ and $P(F) = P(p) P(F|p) + P(f) P(F|f)$. $P(W)$ and $P(F)$ are called the *prior* probabilities (prior to the test) and the conditional probabilities $P(W|p)$, $P(W|f)$, $P(F|p)$, and $P(F|f)$ are the post-test *posterior* probabilities. For this case the priors are $P(W) = 0.6$ and $P(F) = 0.4$, the probabilities without testing.

If the testing decision is chosen and the imperfect test is passed, then the innovation should be chosen if $10 P(W|p) + (10 + 14) P(F|p) < 14$. This reduces to $10 + 14 P(F|p) < 14$ or $P(F|p) < 4/14$. Since $P(F|p) = 0.1$, innovate if the test is passed.

If testing is chosen and the test is failed, then innovate if $10 P(W|f) + (10 + 14) P(F|f) < 14$, or $P(F|f) < 4/14$. Fortunately, this is not the case here, and the test does help us make decisions. To be useful, a test must yield a different decision when the test fails than when the test is passed.

If a passed test encourages innovation and a failed test encourages veneration, then the expected cost of this total test option will be $T + 10P(p) + (P(f) + P(F|p)P(p))14 = T + (4/7)(10) + [(3/7) + (0.1)(4/7)]14 = T + 12.514$. What is the value of the imperfect test, compared to the perfect test and other options? How much should we be willing to pay for the imperfect test? How much cheaper must the imperfect test be before it is preferred to the perfect test?

Updating, Combining, and Reversing Conditional Probabilities

Let $P(p)$ and $P(f)$ be our estimates of whether the innovation will work or fail without the test. Let us assume that experiments with the test were made on both working and failed innovations; thus, we have $P(p|W)$, $P(p|F)$, $P(f|W)$, and $P(f|F)$. *Bayes' Theorem* lets us *reverse* these conditional probabilities to those needed to solve the problem above.

$$P(W|p) = \frac{P(W) P(p|W)}{P(W) P(p|W) + P(F) P(p|F)} = \frac{P(W) P(p|W)}{P(p)}$$

$$P(F|p) = 1 - P(W|p),$$

$$P(F|f) = \frac{P(F) P(f|F)}{P(F) P(f|F) + P(W) P(f|W)} = \frac{P(F) P(f|F)}{P(f)}, \text{ and}$$

$$P(W|f) = 1 - P(F|f).$$

Derivation of Bayes' Theorem

Jaynes (1996) has a nice derivation of Bayes' Theorem using the product rule from probability theory. The probability *product rule* is: $P(A \text{ and } B) = P(A) P(B|A) = P(B) P(A|B)$. Bayes' Theorem is found by merely dividing the product rule by $P(B)$ to yield:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

By using the *sum rule* of probability, $P(B) = P(B|A) P(A) + P(B|\text{not } A) P(\text{not } A)$, the more common form of Bayes' Theorem can be produced:

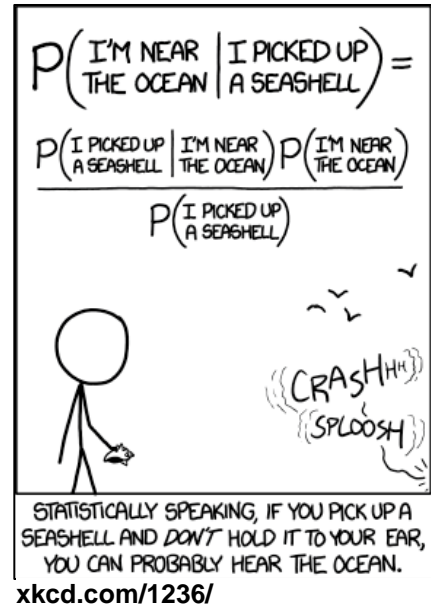
$$P(A|B) = \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|\text{not}A)P(\text{not}A)}$$

Example: Bayes' Theorem and Water Meter Readings

This is a minor application of Bayes' Theorem (Lund 1988), but it illustrates how Bayes' Theorem can get you out of a bind. We want to estimate the probability a meter has failed by inspecting the meter reading Q and knowing the meter's age, $P(f|Q,t)$. Meter maintenance records indicate the proportion (probability) of failed meters for every meter age, $P(f|t)$. From distributions of household consumption and assuming that meters fail randomly between meter readings $P(Q|f,t)$ can also be found. $P(Q|\text{not } f,t)$ is merely the unmodified probability distribution of household consumption (i.e., the meter works), assuming only a very few failed meters exist in the population. The failure probability is updated to incorporate the additional information of the meter reading using Bayes' Theorem:

$$P(f|Q,t) = \frac{P(f|t)P(Q|f,t)}{P(f|t)P(Q|f,t) + P(\text{not}f|t)P(Q|\text{not}f,t)}$$

Thus, we combine a number of probabilities that we know ($P(f|t)$, $P(Q|f,t)$, $P(\text{not } f|t)$, and $P(Q|\text{not } f,t)$), using Bayes' Theorem, to find a probability of failure based on things we know (like flow measurement Q and meter age t) to help identify failed meters for maintenance.



Example: Life, Death, Treatment, and Testing

This is a grim example. Only 5% of a region’s population has Sleeping Sickness (SS), a disease spread by the bite of the tse fly. If untreated, an infected person has a 20% chance of dying. However, treatment is hazardous as well; 5% of treated patients die of a reaction to the treatment (whether they are infected or not). The treatment cures SS in an infected person 90% of the time, if they survive treatment. A screening test is available to test a patient for SS infection. The test has the following reliabilities, estimated from clinical trials.

- 1) When should a patient be treated?
- 2) Is this test worthwhile?

	SS	No SS
P()	0.05	0.95
P(+ _)	0.90	0.20
P(- _)	0.10	0.80

P(+|_) is the probability the test result is positive given condition _ . Note that for each sickness event, the sum of P(+|_) and P(-|_) must equal one.

Use Bayes Theorem for reversing the probabilities,

$$P(SS|+) = \frac{P(SS)P(+|SS)}{P(+|SS)P(SS) + P(+|NSS)P(NSS)} = \frac{P(SS)P(+|SS)}{P(+)}$$

and similarly for P(SS|-), P(NSS|+), and P(NSS|-). [Have class derive these.]

This gives the following table of probabilities: (students can fill these in, completed in appendix to notes)

	P(disease test result)		Test Result
	SS	NSS	Probability
+	0.1915		
-			0.765

For decision-making, the following table of results can be used: (again to be filled in by students)

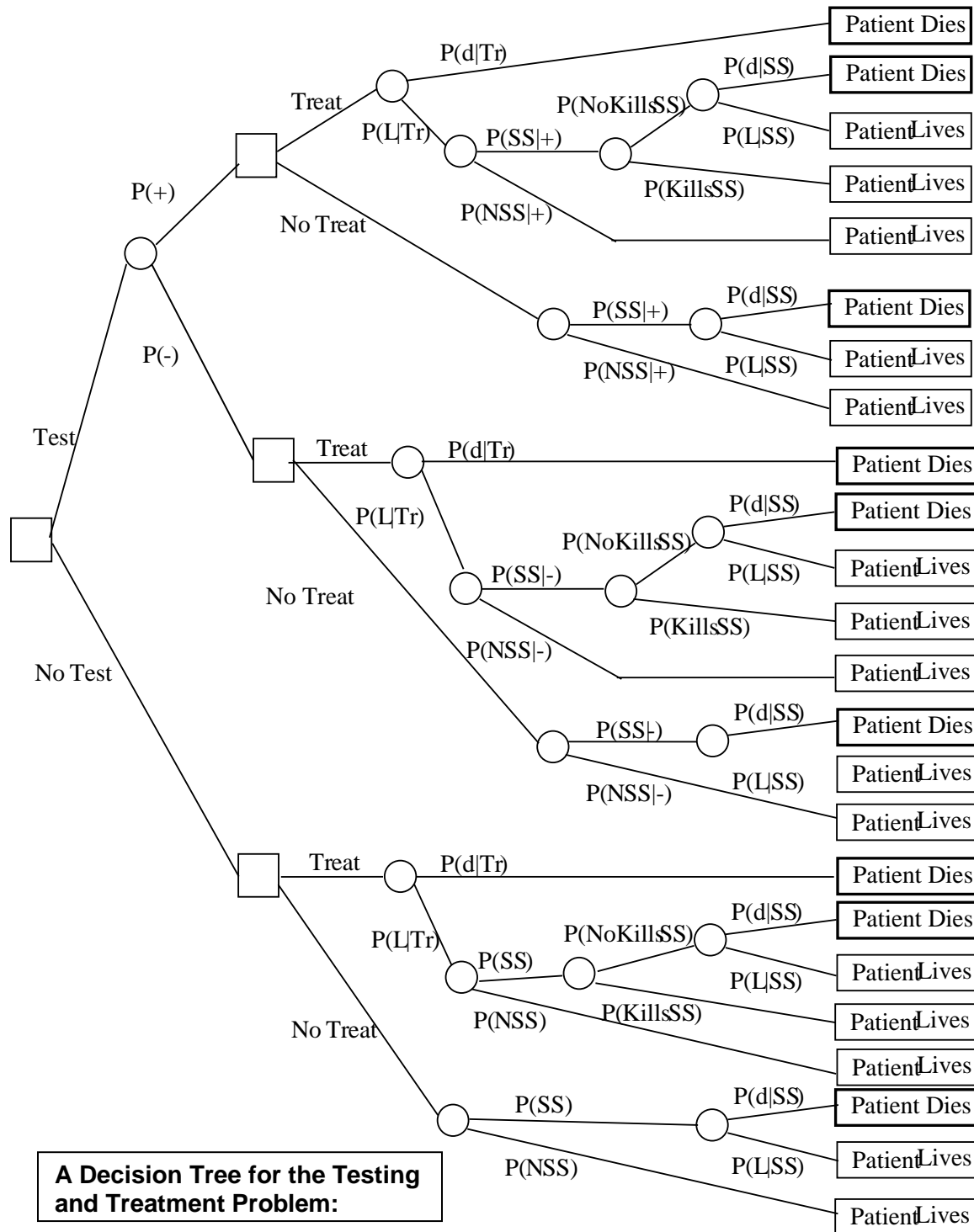
Test	Overall Probability of Death	
	Treated	Untreated
No Test	0.05095	
+		0.0383
-	0.050123	

P(D|no test, Treat) = 0.05 + 0.95*0.05*0.10*0.20 [Have class derive other equations.]

	Probability of Death
Test	
No Test	0.01

A decision tree for this situation appears on the next page.

[Add a new homework question: If a new treatment is 99% effective with only a 3% fatality from treatment, is the test now worthwhile?]



References

Jaynes, E.T. (1996), *Probability Theory - The Logic of Science*, unpublished manuscript.
<http://omega.albany.edu:8008/JaynesBook.html> or <http://bayes.wustl.edu>
 Lund, J.R. (1988), "Metering Utility Services: Evaluation and Maintenance," *Water Resources Research*,
 Vol. 24, No. 6, June, pp. 802-817.
 Stokey, E. and R. Zeckhauser (1978), *A Primer for Policy Analysis*, W.W. Norton & Co., N.Y.
 Nice web site on Bayes' Theorem: <http://yudkowsky.net/bayes/bayes.html>

7. Example of Decision Analysis: Levee Design

The following method for designing a levee illustrates many of the good and bad features of probabilistic design (Davis, et al. 1972). The earliest probabilistic levee design is van Dantzig (1956) which formed the basis of modern Dutch flood management practice (Voortman 2003).

The Problem

- Find the design levee height which minimizes net economic costs (construction and damage costs).
- Flood damages are related to peak annual flood flows, Q
- The design parameter for the levees is the design flow, q
- Higher levees (higher q) cost more and lower q s result in greater and more frequent flood damages.
- Flood flows are probabilistic and are imperfectly described by a probability distribution with a vector of uncertain parameters, $\underline{\theta}$. For most flood problems, $\underline{\theta}$ would include the mean annual peak flood Q_{av} , the standard deviation of annual flood peaks s_Q , and the skew of annual flood peaks γ_Q .
- $P(Q, \underline{\theta})$ is the probability density function of flood peak Q , given parameter values $\underline{\theta}$.
- Let $f(\underline{\theta})$ be the probability density function of the parameter values $\underline{\theta}$.
- $D(Q)$ is the damage wrought by a flood of Q
- $C(q)$ is the annualized cost of building a levee for design flow q

Problem 1: certain parameter values

Find the best levee design q^* for a given certainly known set of flood frequency parameter values $\underline{\theta}$. This expected value cost is given by:

$$(1) \quad \underset{q}{\text{Min EVC}}(q | \underline{\theta}) = \int_q^{\infty} D(Q) P(Q | \underline{\theta}) dQ + C(q)$$

Solve for q^* by enumeration, iterative search, or whatever.

This is how this problem is often addressed in advanced states of practice. But it neglects our uncertainty in the parameter values $\underline{\theta}$.

Problem 2: uncertain parameter values

Here, the parameters in the flood frequency distribution $\underline{\theta}$ are also uncertain. The mathematical form of the flood frequency distribution (lognormal, Gumbel, etc.) is the only thing assumed to be known with certainty. Now the objective function to minimize for the least-cost levee design is:

$$(2) \quad \underset{q}{\text{Min EVC}}(q | f(\underline{\theta})) = \int_q^{\infty} \int_{\theta_1} \int_{\theta_2} \int_{\theta_3} D(Q) P(Q | \underline{\theta}) f(\underline{\theta}) d\theta_3 d\theta_2 d\theta_1 dQ + C(q),$$

where θ_1 , θ_2 , and θ_3 are three parameters of the flood frequency distribution (mean, standard deviation, and skew, say). Again, we solve by enumeration or iterative search.

This gives a more realistic optimal design, incorporating both the probabilistic variation in flood flows and uncertainty in flood frequency parameters. (We might also include uncertainty in the flood frequency distribution's mathematical form (Wood and Rodríguez-Iturbe 1975b) or nonstationarity in the flood frequency distribution (Zhu et al. 2007).)

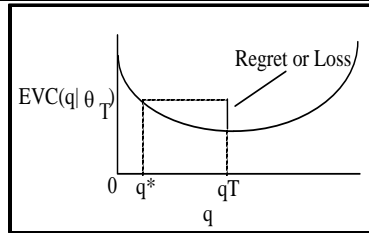
Solving this problem gives an optimal design q^* if we had to build the levee today. Would it be worthwhile waiting a year or several years to collect more flood experience (data) to improve our estimates of the parameters $\underline{\theta}$, probably narrowing the distribution $f(\underline{\theta})$.

Expected Loss from imperfect information: a digression

What is the expected loss in our design for not having perfect information regarding the values of $\underline{\theta}$? Surely, if we knew $\underline{\theta}$ with certainty, we would feel more comfortable about the optimality of our design q^* .

If we did know the true parameter values $\underline{\theta}_T$, we would probably choose a better (more omniscient) levee design flow q_T . The difference between the expected economic costs of the designs q^* and q_T is the "loss" L . You can think of it as our "regret" associated with our design q^* on finding out that the true parameter values were $\underline{\theta}_T$ with the associated best design was q_T . Here's the loss function:

$$(3) \quad L(q^*, q_T) = EVC(q^* | \underline{\theta}_T) - EVC(q_T | \underline{\theta}_T)$$



But $\underline{\theta}_T$ and q_T are unknown. However, we have or can develop probability distributions for q_T and $\underline{\theta}_T$. This distribution of $\underline{\theta}_T$ is merely $f(\underline{\theta})$, and we have already used it. Using this insight, the *expected loss* resulting from design q^* is:

$$(4) \quad XL(q^* | f(\underline{\theta})) = \int_{\theta_1} \int_{\theta_2} \int_{\theta_3} \{EVC(q^* | \underline{\theta}) - EVC(q_T(\underline{\theta}) | \underline{\theta})\} f(\underline{\theta}) d\theta_3 d\theta_2 d\theta_1$$

Here, q^* is a constant, the one we found in problem 2. And $q_T(\underline{\theta})$ varies for each individual set of parameter values $\underline{\theta}$. $q_T(\underline{\theta})$ is the value of q which solves the problem given perfectly known parameter values (problem 1):

$$(5) \quad \text{Min}_q EVC(q, \underline{\theta}) = \int_q^\infty D(Q)P(Q, \underline{\theta}) dQ + C(q).$$

XL , the expected loss, represents the expected value of having perfect estimates of the parameters $\underline{\theta}$.

The Benefits of Waiting a Year

If we wait one year to make a decision we will have more data to estimate $f(\underline{\theta})$. This new data will be in the form of next year's flood peak Q_N , whose flow we will record.

If next year's flood were known today, we could update our current $f(\underline{\theta})$ using Bayes theorem:

$$(6) \quad f(\underline{\theta} | Q_N) = \frac{f(\underline{\theta})P(Q_N | \underline{\theta})}{\int_{\theta} f(\underline{\theta})P(Q_N | \underline{\theta})d\theta}$$

So next year we can use $f(\underline{\theta} | Q_N)$ to update $XL(q^*, f(\underline{\theta}))$ to see if waiting was worthwhile.

Unfortunately, we must decide today if the additional flood observation is worth waiting for. While we do not know Q_N , we do know its probability distribution. The distribution of Q_N is merely the current flood frequency distribution with the current uncertain parameter distribution:

$$(7) \quad P(Q_N) = \int_{\underline{\theta}} P(Q_N | \underline{\theta})f(\underline{\theta})d\underline{\theta}$$

When we put all this together, the expected value of the benefits (reduced expected loss) of waiting for one year of additional flood flow record is:

$$(8) \quad \text{EVB}_{Q_N} = \text{XL}(q^* | f(\theta)) - \int_0^\infty \text{XL}[q^*(f(\theta | Q_N)) | f(\theta | Q_N)] P(Q_N) dQ_N$$

Whew!

The Costs of Waiting One Year

The expected value of costs of waiting one year is the damage to be expected during the next year, given that there is no levee, plus the annualized cost avoided by deferring construction by one year. Thus,

$$(9) \quad \text{EVC}_{Q_N} = \int_0^\infty D(Q_N) P(Q_N) dQ_N - C(q^*)$$

The Big Decision

If $\text{EVB}_{Q_N} \geq \text{EVC}_{Q_N}$ then we should wait a year for more data.

Some Thoughts on Practice

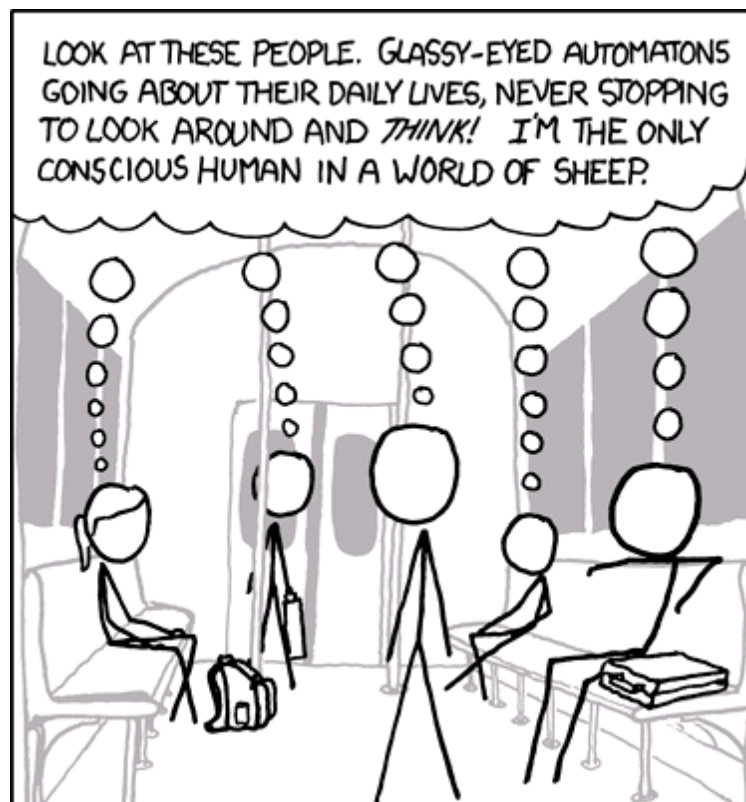
The foregoing method requires little that is not available in practice. The major theoretical problem is estimating $f(\theta)$, but this can be done without too much difficulty. For instance, the Central Limit Theorem would hold that the distribution for Q_{av} is Gaussian. Distributions for s_Q and γ_Q might be gotten from regional data, as the USGS does (Wood and Rodríguez-Iturbe 1975a). Everything else, even $D()$ and $C()$, is fairly common. Probably the biggest practical problems with this method are explaining it to regular engineers and doing the greater amount of computations required. (But with computers, this latter problem should be minimal.)

Some Extensions

1. Flood flows and the parameters of the flood frequency distribution are not the only uncertain aspects of this problem. The cost and damage functions are also somewhat uncertain, as is the mathematical form of the flood frequency distribution. Of likely particular importance is uncertainty in the stage-discharge relationship which, in this formulation, is embedded in the functions $D()$ and $C()$. Stage-discharge relationships can be particularly uncertain in the region of high flows, where simultaneous measurements of stage and discharge are rare and fairly crude approximations are often unavoidable. Uncertainties in these aspects of the problem could be incorporated into more elaborate formulations, adding probability distributions for parameters in the functions $D()$ and $C()$, for example. Data gathering and probability distribution estimation for this problem are problematic in practice. Adding additional uncertainties also adds computational burden, since in the formulation we add an additional integration for each new uncertain parameter. [Cite USACE studies of these things.]
2. Levees commonly fail by other mechanisms than by overtopping (e.g., undermining, wave action, piping, slope failure). A probabilistic model of levee failure by other mechanisms can be developed and incorporated into this scheme (Wood 1977; Tung and Mays 1981a; Voortman 2003).
3. Many levee projects take place in multi-reach systems. Tung and Mays (1981b) propose solution of this problem by dynamic programming.
4. The problem so far has assumed that once the levee is built, it cannot be altered if in the future we find that it was too low or too high. In practice, it is hardly economical to reduce levee heights, but it is common to find that levees should be increased in height. What happens if the problem is modified further to allow for increasing levee heights in the future? This becomes a multi-stage decision-making problem. It is solvable by stochastic dynamic programming, a later topic in the course. (Zhu, et al. 2007)

References

- Davis, D.R., C.C. Kisiel, and L. Duckstein (1972), "Bayesian Decision Theory Applied to Design in Hydrology," *Water Resources Research*, Vol. 8, No. 1, February, pp. 33-41.
- Duckstein, D., I. Bogardi, F. Szidarovszky, and D.R. Davis (1975), "Sample Uncertainty in Flood Levee Design: Bayesian versus Non-Bayesian Methods," *Water Resources Bulletin*, Vol. 11, No. 3, June, pp. 425-440.
- Tung, Y-K and L.W. Mays (1981a), "Risk Models for Flood Levee Design," *Water Resources Research*, Vol. 17, No. 4, August, pp. 833-841.
- Tung, Y-K and L.W. Mays (1981b), "Optimal Risk-Based design of Flood Levee Systems," *Water Resources Research*, Vol. 17, No. 4, August, pp. 843-852.
- Van Dantzig, D. (1956), "Economic decision problems for flood prevention," *Econometrica*, Vol. 24, July, pp. 276-287.
- Wood, E.F. (1977), "An Analysis of Flood Levee Reliability," *Water Resources Research*, Vol. 13, No. 3, June, pp. 665-671.
- Wood, E.F. and I. Rodríguez-Iturbe (1975a), "Bayesian Inference and Decision Making for Extreme Hydrologic Events," *Water Resources Research*, Vol. 11, No. 4, August, pp. 533-542.
- Wood, E.F. and I. Rodríguez-Iturbe (1975b), "A Bayesian Approach to Analyzing Uncertainty Among Flood Frequency Models," *Water Resources Research*, Vol. 11, No. 6, December, pp. 839-843.
- Voortman, H. (2003), Risk-based design of large-scale flood defence systems," PhD dissertation, Technical University of Delft, Netherlands.
<http://hydr.ct.tudelft.nl/wbk/public/gelder/thesisvoortman.pdf>
- Zhu, T., J.R. Lund, M.W. Jenkins, and G. Marques (2007), "Climate Change, Urbanization, and Optimal Long-term Floodplain Protection," *Water Resources Research*, Vol. 43, No. 6, June.



8. A Probability Primer

“When one admits that nothing is certain one must, I think, also admit that some things are much more nearly certain than others.” Bertrand Russell (1947)

"Probability theory is nothing but common sense reduced to calculation." Laplace 1819

Probability and Plausible Reasoning

The subjective or Bayesian school of probability theory holds that probabilities should represent the plausibility of uncertain events (Raiffa 1968). More plausible events have higher probability. Events occurring with certainty have probabilities of one. Impossible events have probabilities of zero.

The calculus of probability is based fundamentally on the analogy between mathematics and deductive logic. So, much of probability comes from the use of symbolic logic, or Boolean algebra. In addition to Boolean algebra and set theory, probability theory rests on three fundamental desiderata that correspond to desirable goals of plausible reasoning (Jaynes 2003).

Desiderata 1: Degrees of plausibility are represented by real numbers. As a convention, a larger number means greater plausibility. The plausibility number we assign also should depend on evidence; therefore the plausibility of an event A should be affected by what we already know B. This conditional nature of probabilities creates the need for a notation $P(A|B)$, the probability of A given B.

Desiderata 2: Qualitative correspondence to common sense. If B is evidence that A will occur, and B gets updated to B' so as to increase the plausibility of A, then $P(A|B') > P(A|B)$. More supportive evidence of A should make A more plausible.

Desiderata 3: Consistency. The desirability of consistency has three more detailed meanings:

- If a conclusion can be reasoned out in more than one way, then every possible way of reasoning through the problem must lead to the same result.
- All evidence relevant to evaluating plausibility must be taken into account.
- Equivalent levels of plausibility must be represented by equal probabilities.

Rules have been developed for probability calculations using these desirable features for a calculus of plausible reasoning (Jaynes 2003).

Fundamental Rules of Probability

Scale: The scale of a probability theory is somewhat arbitrary. The traditional scale of [0,1] does not violate the desiderata and is used. If $P(A) = 0$, then A is impossible; if $P(A) = 1$, then A is certain to occur.

Product Rule: The product rule allows us to evaluate the plausibility of jointly-occurring conditional events.

$$P(A \cap B|C) = P(A|B \cap C) P(B|C) = P(B|A \cap C) P(A|C)$$

The product rule is that the probability of two events A and B both happening is $P(A \cap B) = P(A|B) P(B)$. If events A and B are *independent*, $P(A|B) = P(A)$ and the probability of these two independent events both happening is $P(A \cap B) = P(A) P(B)$.

Sum Rule: The sum rule allows us to evaluate the plausibility of aggregated events such as $(A \cup B|C)$. The fundamental sum rule is: $P(A|B) + P(\text{not } A|B) = 1$. A more general form of the sum rule is:

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$$

Mutual Exclusivity: Two events are mutually exclusive if $P(A \cap B) = 0$. Mutually exclusive events by definition cannot both occur. For mutually exclusive events A and B, $P(A \cup B|C) = P(A|C) + P(B|C)$.

Approaches to Assigning Probability Values

Principle of Insufficient Reason

Laplace (1819?) found that for n mutually exclusive and collectively exhaustive events, and in the absence of any other plausibility information, the probability of each event should be $1/n$.

Proportion of Total Outcomes

By combining the sum rule with the principle of insufficient reason, we can find James Bernoulli's original (1713) definition of probabilities. If there are n equally likely outcomes, of which m would be classified as event A , then $P(A|m,n) = m/n$.

Observed Relative Frequency

True subjective probabilists poo-poo using relative frequency ideas in determining probabilities. Still, if I flip a coin 100 times and get 90 heads, I'm tempted to believe the probability of heads for this almost certainly weighted coin is roughly 0.90. Relative frequency may be theoretically impure, but empiricism is often useful for engineers.

Subjective Probability

Sometimes an experienced person's opinions regarding the likelihood of events is a useful guide for assessing probability values for events. There are some well-known biases in such assessments, however (Morgan and Henrion 1990).

Bayes' Theorem

Bayes' theorem reverses conditional probabilities and is used commonly to update probability distributions to incorporate new information. It is derived quickly from the product rule. The product rule states: $P(A|B \cap C) P(B|C) = P(B|A \cap C) P(A|C)$. Dividing both sides by $P(B|C)$ yields:

$$P(A|B \cap C) = \frac{P(B|A \cap C)P(A|C)}{P(B|C)}.$$

Using the sum rule gives Bayes' Theorem in its more familiar form:

$$P(A|B \cap C) = \frac{P(B|A \cap C)P(A|C)}{P(B|A \cap C)P(A|C) + P(B|notA \cap C)P(notA|C)}.$$

A graphical derivation follows this section.

In the Bayesian view, we are always updating our distributions of plausibility/probability as new observations become available. Logically, we should do this using Bayes Theorem. The distribution $P(A|C)$ is the *prior probability distribution* (or just *prior*), before a new observation. Event B is the new observation, a particular event. $P(B|A \cap C)$ is the likelihood of the new observation B given A .

The resulting $P(A|B \cap C)$ is the *posterior probability distribution* (or just *posterior*), accounting for the new observation B . In essence, we are multiplying the prior $P(A|C)$ by the ratio of the probability of observation B given A and the probability of B : $\frac{P(B|A \cap C)}{P(B|C)}$. If seeing A enhances the probability

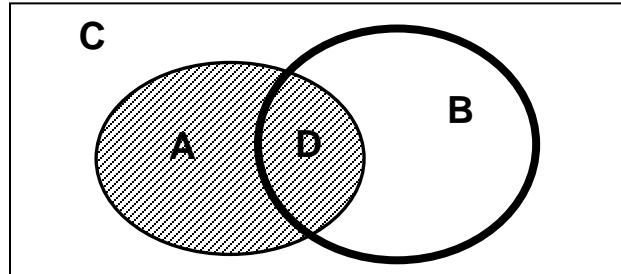
of seeing event B , then this ratio > 1 .

If our prior distribution is very broad and with very "fat" tails, we obviously have little initial idea what will happen. Such a prior is called an *uninformative prior*.

Using Bayes' Theorem to provide probability distributions can require a lot of computation. There are some particular forms of prior distribution for which Bayes' Theorem can be applied analytically to give an analytical probability distribution, so-called *conjugate priors*. Typically, we must be prepared for manipulation of approximate discrete, numerically derived distributions when employing Bayes' Theorem. (Agarwal and Daumé 2010) [add references for conjugate priors]

Graphical derivation of Bayes' Theorem

Given the situation represented below by a Venn diagram, let's look at the probability of different non-mutually-exclusive events, A, B, C, and D each represented by the areas a, b, c, and d.



Let $P(A|C) = a/c$, $P(B|C)=b/c$, $P(C)=c$, and $P(D|C)=d/c$. These are the *priors*. It helps to assume $c = 1$.

From the diagram, $P(D|C) = P(D|B \cap C) P(B|C)$ or $d/c = P(D|B \cap C) b/c$ or $P(D|B \cap C) = d/b$.

Since the only way A can occur, given B is through D, then $P(A|B \cap C) = P(D|B \cap C) = d/b$

Similarly, $P(B|A \cap C) = d/a$.

Therefore, $P(A|B \cap C) = \frac{P(A|C)P(B|A \cap C)}{P(B|C)} = \frac{(a/c)(d/a)}{(b/c)} = d/b$, Bayes' Theorem.

This provides a conceptual and mathematical way to update our prior probability of A, given that B has occurred. To do this we need the priors for A and B ($P(A|C)$ and $P(B|C)$) as well as the probability of B, given A and C, $P(B|A \cap C) = d/a$ here. The result $P(A|B \cap C)$ is called the *posterior* probability of A.

Probability Distributions

Probability distributions can be either discrete or continuous. In engineering, we often are forced to use more numerically-friendly discrete distributions. In fact, this may be a blessing, since it is often easier for engineers to solve problems numerically than analytically. If we want to be in this business, we have a choice of being good with computers or very good with math.

Probability Density Functions (pdfs) are equations or numerical results that represent the probability that an observation falls within a small unit range of possible outcomes. The dimensions of a pdf for a flood frequency distribution are the probability that the greatest flood occurring in a given year will be within 1 cfs of some Q. Therefore a pdf for a variable with a very narrow possible range, such as Manning's n, can have a probability density greater than one.

Cumulative Density Functions (CDFs) are integrals, or cumulative summations, of pdfs or discrete probability distributions.

$$CDF(Q) = \int_{-\infty}^Q pdf(q)dq$$

Probability Distributions from Random Processes

If a particular organization of events yields a particular and consistent probabilistic result, that process is called a random process. Rolling a die is a random process that yields a discrete uniform distribution of outcomes in the range of integers [1,6]. A flipped coin does the same, but in the range [heads, tails]. Many common probability distributions result from particular random processes.

Binomial Distribution: If only two events can occur, A or C, and $P(A)$ and $P(C)=1-P(A)$ are constant, then the probability of seeing A m times out of n independent observations is:

$$B(m|n, P(A)) = \frac{n!}{m!(n-m)!} P(A)^m P(C)^{n-m}.$$

Geometric Distribution: If the probability of failure in a given year is constant at $P(f)$, then the probability distribution of the year of the first failure t is:

$$G(t|P_f) = P_f (1-P_f)^{t-1}.$$

Poisson Distribution: If N is the average number of independent discrete events that occur in a unit area or time period, then the probability that M such events will occur within that region or time period is:

$$P(M|N) = \frac{e^{-N} N^M}{M!}.$$

Gamma Distribution: For events that happen according to a Poisson process, with an arrival rate of N per time-step, the probability distribution of having t time-steps elapse before the r -th event is a Gamma distribution. (Sometimes called an Erlang distribution for integer values of r .)

$$G(t) = \frac{N}{(r-1)!} (Nt)^{r-1} e^{-Nt}, \text{ for } t \geq 0 \text{ and } = 0 \text{ for } t < 0.$$

Exponential Distribution: An exponential distribution describes the time until the first arrival of a Poisson process. The exponential distribution is a Gamma distribution for $r = 1$.

$$f_E(t) = Ne^{-Nt}, \text{ for } t \geq 0 \text{ and } = 0 \text{ for } t < 0.$$

Gaussian (Normal) Distribution: If multiple continuous random errors (with any distributional form) occur in an additive way, the probability distribution of the measurement will be Gaussian.

$$f(x) = [1/(\sigma_x \text{SQRT}(2\pi))] \exp(-0.5((x-\mu_x)/\sigma_x)^2), \text{ where } \sigma_x \text{ is the standard deviation of } x \text{ and } \mu_x \text{ is its mean.}$$

Log-Normal Distribution: The product of multiple random variables is a log-normal distribution – very useful. (Proof: $X = aX_1X_2X_3$. $\ln(X) = a + \ln X_1 + \ln X_2 + \ln X_3$. The natural log of X is normally distributed by the central limit theorem.)

$$f(x) = \frac{1}{(\sqrt{2\pi})x\sigma_{\ln x}} \exp\left(-0.5\left(\frac{\ln x - \mu_{\ln x}}{\sigma_{\ln x}}\right)^2\right).$$

To transform mean and variance values to the ln-scale for the log-normal distribution, use:

$$\sigma_{\ln x}^2 = \ln\left(\frac{\sigma_x^2}{\mu_x^2} + 1\right) \quad \text{and} \quad \mu_{\ln x} = \ln(\mu_x) - (\sigma_{\ln x}^2)/2,$$

The coefficients of skew and kurtosis are given by:

$$\gamma = [\exp(\sigma_{\ln x}^2) + 2]\sqrt{\exp(\sigma_{\ln x}^2) - 1}, \text{ and } \kappa = \exp(4\sigma_{\ln x}^2) + 2\exp(3\sigma_{\ln x}^2) + 3\exp(2\sigma_{\ln x}^2) - 3$$

A “shifted lognormal” has $y = \ln(x + c)$ being normally distributed, with a lower bound of c .

Rosenberg (2007) has a nice application of log-normal distributions for water conservation problems.

An odd implication of lognormal distributions is something called “Benford’s Law”, that the first digit in a list of lognormally-distributed random numbers will be a one, about 30% of the time. This can be extended to predict the probability of later digits and is a technique can be used for fraud detection. (Hill 1995; http://en.wikipedia.org/wiki/Benford%27s_law) (I wonder if this idea could be used to detect stream gage errors.)

Gumbel Distribution: If independently distributed floods occur each week, and the probability distribution of the peak flow of each weekly flood is exponential, then the probability distribution of the biggest flood peak flow in each year, approaches a Gumbel (Extreme Value 1) distribution (Gumbel 1958). The cumulative Gumbel distribution is:

$$F(Q) = e^{-e^{-a(Q-u)}}, \text{ where } a \text{ is a scale parameter and } u \text{ is the modal value of the distribution.}$$

These parameters have values $a = \pi/(\text{SQRT}(6)\sigma)$ and $u = \mu - 0.5772/a$, where μ is the mean annual flood

peak and σ is the standard deviation of flood peaks.

Other distributions: A great number of probability distributions have been derived for particular random processes. Still, it is often difficult to find the right distribution for a particular engineering problem. There are many approximate and numerical approaches for these cases, such as first-order second-moment approximations and Monte Carlo methods.

Fitting Parameters to Probability Distributions

Once a probability distribution is selected, establishing parameter values for that distribution is the next problem, which we largely avoid here (Stedinger 1980).

Uncertain Probabilities in Decision Analysis

The probabilities we use in decision analysis are usually uncertain. What should be done? Must we develop and carry entire probability distributions representing the uncertainty of probability estimates? Fortunately, no. For expected utility (or expected value) decision criteria, only the mean or expected value of an uncertain probability is needed (Howard 1988). This works for any evaluation measure that is linear with probability (as expected utility and expected value are). (Mosleh and Bier 1996 provide some counter-examples.)

Given the expected utility of an alternative A, where each outcome i has an uncertain probability P_i distributed $f_i(P_i)$ for each outcome utility U_i :

$$EU_A = \sum_i \int_0^1 f_i(P_i) P_i U_i dP_i,$$

Since the definition of the expected value of event probability is, $EV(P_i) = \int_0^1 f_i(P_i) P_i dP_i$, this becomes:

$$EU_A = \sum_i EV(P_i) U_i.$$

This is simply the expected utility, calculated using the expected value of the probability of each outcome. (Whew! This saves us a world of worry and estimation effort.)

References

- Gumbel, E.J. (1958), *Statistics of Extremes*, Columbia University Press, New York.
- Hill, T.P. (1995), "A statistical derivation of the significant-digit law," *Statistical Science*, Vol. 10, No. 4, pp. 354-363.
- Howard, Ronald A. (1988), "Uncertainty about Probability: A Decision Analysis Perspective," *Risk Analysis*, Vol. 8, No. 1, pp. 91-98.
- Morgan, M.G. and M. Henrion (1990), *Uncertainty*, Cambridge University Press, N.Y.
- Mosleh, A. and V.M. Bier (1996), "Uncertainty about probability," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 26, No. 3, May, pp. 303-310.
- Raiffa, Howard A. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, MA.
- Rosenberg, D. (2007), "Probabilistic Estimation of Water Conservation Effectiveness." *Journal of Water Resources Planning and Management*. Vol. 133 (1), pp. 39-49.
- Stedinger, J.R. (1980), "Fitting Log Normal Distributions to Hydrologic Data," *Water Resources Research*, Vol. 16, No. 3, pp. 481-490, June.
- Walpole, R.E. and R.H. Meyers (1975), *Probability and Statistics for Engineers and Scientists*, Macmillan, N.Y.
- Agarwal, A. and H. Daumé (2010), "A geometric view of conjugate priors," *Mach Learn*, 81, pp. 99-113. <https://www.vox.com/future-perfect/2018/11/30/18096751/bayes-theorem-rule-rational-thinking-reason>
- Wikipedia (<http://en.wikipedia.org>) has lots of information on a wide variety of probability distributions.

9. "Maximum Entropy": Assigning Probabilities with Little Information

We often must assign a probability distribution to outcomes when we have little information regarding the relative frequency or likelihood of each event. Laplace's "principle of insufficient reason" is an extreme illustration of this case. In the late 1940s, Shannon (1948) developed an indicator of the information content of a probability distribution. In his approach, one would assign probability values to outcomes such that the probability distribution has the minimum amount of "information" while still representing aspects of the outcomes and their frequency that are felt to be known.

What is the information content of a probability distribution? A probability distribution consisting of one outcome with probability of one and other outcomes with zero probability obviously is more informative than a uniform distribution of probability among outcomes. How can the information content of a probability distribution be estimated?

Shannon (1948) developed a consistent measure of the "amount of uncertainty" represented by a probability distribution. He suggests several principles for developing such a quantitative index of a probability distribution's information content:

- 1) Assume some numerical measure $H_n(p_1, \dots, p_n)$ exists describing the amount of uncertainty contained in the probability distribution p_i .
- 2) Assume H_n is a continuous function of p_i . If H_n were not continuous, then a small change in p_i would lead to a great change in information or uncertainty content. This would be unreasonable.
- 3) Having more possibilities (larger n) should increase uncertainty. Where all p_i are equal, $H(n) = H_n(1/n, 1/n, \dots, 1/n)$. So $H(n)$ should always increase with n .
- 4) If there is more than one way to calculate the value of H_n , then all ways must give the same value.

Shannon proposes the following measure of the amount of uncertainty in a probability distribution, often called its "information entropy" as:

$$(1) \quad H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \ln(p_i).$$

Using Shannon's approach, we would like to assign probabilities to events so that we maximize this information entropy, subject to constraints representing the things we feel we know. Such a distribution contains the minimum amount of information consistent with what we know.

What might we know as constraints in assigning probabilities? We might feel we know the mean of the distribution, its variance, skew, or upper or lower limits of the distribution. Such an approach can be used to develop probability distributions for treatment effectiveness, Manning's n , or various environmental outcomes. The solution for the probabilities is by mathematical optimization, maximizing information entropy (1), subject to constraints.

Solution Approach for Maximum Entropy Probability Distributions

Several kinds of knowledge constrain the assignment of probabilities to events. First, for all cases the sum of probabilities over collectively exhaustive events must equal one:

$$(2) \quad \sum_{i=1}^n p_i = 1.$$

If we solve the maximum entropy problem, subject to this constraint only, we find a uniform probability distribution; Laplace's Principle of Insufficient Reason rules.

We often also feel we know the mean, variance, or other statistical moments of a probability distribution. For each statistical moment we have a constraint

$$(3) \quad \sum_{i=1}^n g_r(x_i) p_i = E(g_r), \quad \text{for all functions } r,$$

where $E(g_r)$ is the expected value of function g_r . For example, $g_r(x_i) = x_i$ for the mean, $g_r(x_i) = x_i^2$ for the variance, etc.

The mathematical programming formulation is then to Maximize Equation 1, subject to Equation 2 and Equations 3, with the p_i as decision variables. To solve, the Lagrangian can be formed and differentiated:

$$(4) \quad \frac{\partial L}{\partial p_i} = 0 = -1 - \ln p_i + \lambda_0 + \lambda_1 g_1(x_i) + \lambda_2 g_2(x_i) + \lambda_3 g_3(x_i) + \dots, \text{ for all } i.$$

Then, including the constant -1 in λ_0 , we find;

$$(5) \quad p_i = \exp\left(\lambda_0 + \sum_r \lambda_r g_r(x_i)\right) \quad \text{for all } p_i.$$

Noting that the p_i sum to one, leading to the following expression for λ_0 :

$$(6) \quad \lambda_0 = -\ln \sum_i \exp\left(\sum_r \lambda_r g_r(x_i)\right).$$

Then, following Tribus (1969):

$$(7) \quad \exp(-\lambda_0) = \sum_i \exp\left(\sum_r \lambda_r g_r(x_i)\right)$$

and differentiating with respect to λ_k ;

$$(8) \quad -\exp(-\lambda_0) \frac{\partial \lambda_0}{\partial \lambda_k} = \sum_i \left(\exp\left(\sum_r \lambda_r g_r(x_i)\right) g_k(x_i) \right);$$

or;

$$(9) \quad -\frac{\partial \lambda_0}{\partial \lambda_k} = \sum_i \exp\left(\lambda_0 + \sum_r \lambda_r g_r(x_i)\right) g_k(x_i)$$

Putting equation 5 into 9 gives the following result;

$$(10) \quad \frac{\partial \lambda_0}{\partial \lambda_k} = -E(g_k).$$

Equation 6 can be differentiated, then, to give a useful version of (10):

$$(11) \quad \frac{\sum_i g_r(X_i) \exp\left(\sum_r \lambda_r g_r(X_i)\right)}{\sum_i \exp\left(\sum_r \lambda_r g_r(X_i)\right)} = E(g_k), \text{ for all } k.$$

Hence the Lagrange multipliers can be determined by the mean values of arbitrary functions $g_r(X)$ using equations 6 and 11.

Where only the mean of X is known, $g_r(x_i) = g_1(x_i) = x_i$, $\lambda_r = \lambda_1$, and equations 5, 6, and 11 reduce to;

$$(12) \quad p_i = \exp(\lambda_0 + \lambda_1 x_i), \text{ for all } p_i;$$

$$(13) \quad \lambda_0 = -\ln \sum_i \exp(\lambda_1 x_i), \text{ and};$$

$$(14) \quad \frac{\sum_i x_i \exp(\lambda_1 x_i)}{\sum_i \exp(\lambda_1 x_i)} = E(X).$$

Equations 5, 6, and 11 can be used to find discrete distributions for uncertain variables in practical problems, where the range of X and/or the mean of any functions of X can be specified, by discretizing over the range of possible X.

While this looks intimidating, the final results are fairly simple. Where no statistical moments are known, a uniform distribution results. With a mean constraint only, an exponential distribution results, and with mean and variance constraints a truncated Gaussian distribution results. The problem is analogous to that of finding the probability of each face of a possibly unfair die turning up, given the average of many rolls (Jaynes 1988). For a fair die the average would be 3.5 and the probabilities would each be 1/6, representing a uniform distribution. What if the average of many rolls is not 3.5?

Maximum Entropy Example: Annual Sediment Deposition in a Reservoir

Tribus (1969), Englehardt and Lund (1992), and others use this approach for a variety of problems. For a simple example, let's look at a maximum-entropy estimate of the annual sediment deposition rate in a reservoir (or detention pond). For this problem, all we know is that over the last 20 years, the reservoir has lost 40,000 acre-ft of storage capacity. What is the annual probability distribution of sedimentation volume?

The mean is a relatively robust statistic. An accumulation over 20 years is a fairly averaged process. Therefore, we will assume we know the mean of the annual sediment volume distribution is 2,000 acre-ft/year = s. We can also feel fairly confident that negative sedimentation rates do not occur, so the annual sedimentation rate, $S \geq 0$. What is the maximum sedimentation rate? This is hard to fix, so let's start with no upper bound for S. The mathematical program for assigning probabilities to annual sedimentation rates becomes:

$$\text{Max } z = - \int_0^{\infty} P(S) \ln(P(S)) dS$$

Subject to:

$$\int_0^{\infty} P(S) dS = 1$$

$$\int_0^{\infty} P(S) S dS = s = 2,000$$

Solving this problem by Lagrange multipliers gives the Lagrangian,

$$L = - \int_0^{\infty} P(S) \ln(P(S)) dS + \lambda_0 \left(\int_0^{\infty} P(S) dS - 1 \right) + \lambda_1 \left(\int_0^{\infty} P(S) S dS - s \right)$$

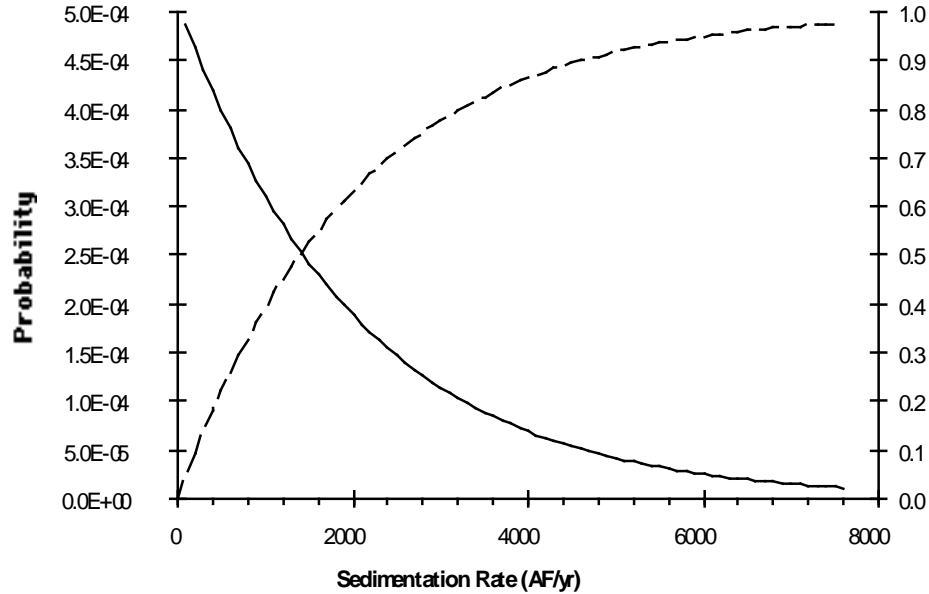
The first-order conditions are,

$$\partial L / \partial P(S) = 0 = -1 - \ln(P(S)) + \lambda_0 + \lambda_1 S.$$

This becomes,

$$P(S) = \exp(\lambda_0 - 1 + \lambda_1 S),$$

an exponential probability distribution form. This exponential distribution must have a mean of $s = 2,000$ AF/year and an area of one, shown below.



Updating Prior Distributions with Data 1

A new water treatment device is being considered. The device has failed in 7 out of 28 tests. What is the probability that it will work in a given application?

Here is a Bayesian perspective on this question. For prior information, before the tests, we know not what the device's failure probability is. We might judge, as a prior, that the probability distribution of the failure probability is uniformly distributed between zero and one, a uniform probability density of $P(p_f) = 1$. The probability of 7 failures of 28 tests is a binomial probability process, so the likelihood function is binomial:

$$P(x|p_f, n) = \frac{n!}{x!(n-x)!} p_f^x (1-p_f)^{n-x},$$

where x is the number of failures in n trials, and p_f is the probability that a single trial will end in failure. Bayes Theorem is:

$$P(p_f|x, n) = \frac{P(p_f)P(x|p_f, n)}{P(x|n)}, \text{ where}$$

$$P(x|n) = \int_0^1 P(x|p_f, n)P(p_f)dp_f$$

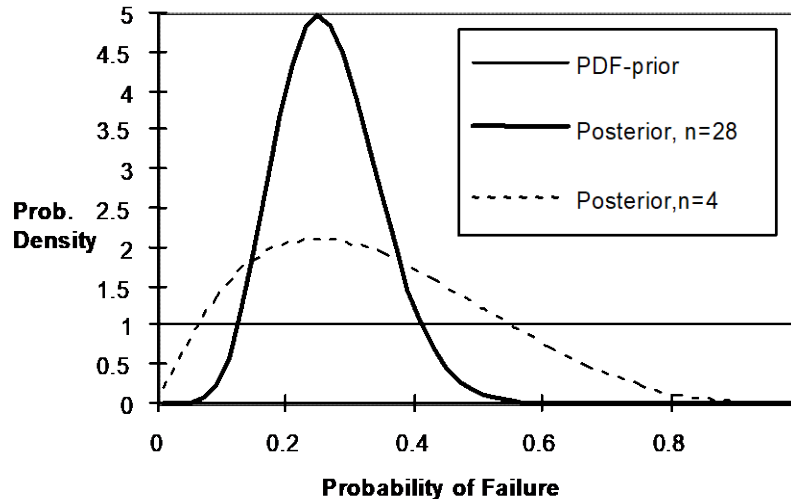
For this case, Bayes Theorem becomes:

$$P(p_f|x, n) = \frac{1 \frac{n!}{x!(n-x)!} p_f^x (1-p_f)^{n-x}}{\int_0^1 P(x|p_f, n)P(p_f)dp_f},$$

which is a Beta Distribution:

$$f_B(p_f|a, b) = \frac{1}{B(a, b)} p_f^{a-1} (1-p_f)^{b-1},$$

where $a = x+1$ and $b = n-x+1$, for an initial uniform prior. For this case, the prior and posterior distributions can be compared in the figure below. Note how the posterior distribution tightens much less if only 1 test fails out of four (the same proportion of failures, but much less information).



The uniform distribution is a special case of Beta Distribution, a very general form of distribution when the independent variable varies only between 0 and 1. Where a Beta Distribution is a prior, and the likelihood function Binomial, the posterior distribution is also a Beta Distribution, with somewhat different parameter values updated to reflect the new data (Pratt, et al. 1996).

Updating Prior Distributions with Data 2

Given some prior probability distribution, how can we update this information with new observations or data? Prior probability distributions can have several origins. Jaynes (1968) suggests the use of such "maximum entropy" distributions as priors for updating with data using Bayes Theorem.

$$P(A|data) = \frac{P(A)P(data | A)}{P(data)}$$

References

- Benjamin, Jack R. and C. Allin Cornell (1970), *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill Book. Co., N.Y.
- Englehardt, J.D. and J.R. Lund (1992), "Information Theory in Risk Analysis," *Journal of Environmental Engineering*, ASCE, Vol. 118, No. 6, pp. 890-904, Nov.-Dec.
- Jaynes, E.T. (2003), *Probability Theory - The Logic of Science*, Cambridge University Press, Cambridge, UK. unpublished manuscript available at. <http://omega.albany.edu:8008/JaynesBook.html> or <http://bayes.wustl.edu>
- Jaynes, E. T. (1988), "How does the brain do plausible reasoning?" In Erickson, G. J. and C. R. Smith (eds.), *Maximum-entropy and Bayesian Methods in science and engineering; volume 1: foundations*. Kluwer Academic Publishers.
- Jaynes, E.T. (1968), "Prior Probabilities," *IEEE Transactions Systems, Science, and Cybernetics*, Vol. SSC-4, pp. 227-241
- McBride, G.B. and J.C. Ellis (2001), "Confidence of Compliance: a Bayesian approach for percentile standards," *Water Research*, Vol. 35, No. 5 (April), pp. 1117-1124.
- Pratt, John W., Howard Raiffa, and Robert Schlaifer (1996), *Introduction to Statistical Decision Theory*, MIT Press, Cambridge, MA.
- Shannon, C.E. (1948), "A Mathematical Theory of Communication," *Bell Systems Technical Journal*, Vol. 27, pp. 379-623. Reprinted as C.E. Shannon and W. Weaver (1949), *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.
- Tribus, M. (1969), *Rational Descriptions, Decisions, and Designs*, Pergamon Press, New York.
- Nice web site on Bayes' Theorem: <http://yudkowsky.net/bayes/bayes.html>
- Eames, Ray & Charles (1953), *A Communications Primer*, IBM, 21 minute video.

10. Simulation Modeling with Uncertainty

Why do simulation models work?

Models are a quantitative representation of our understanding of a real or proposed system. They represent an integrated assembly of our knowledge of a system, developed for some scientific or engineering purpose. This knowledge can come from derived application of fundamental theoretical principles (conservation of mass, energy, momentum, geometry, mathematics, etc.) and empirical relationships (such as regressions) and measurements. Klemes (1982) and Beven (1993) have some interesting thoughts on this, including the many ways that models can work when their fundamental representation of the system is physically wrong.

Gass (1983) and Beck (2002), and USEPA (2003) review the literature on the development and testing of models for policy and operational applications. An appendix to these notes presents more (NRC 2007).

Sources of Model Error and Uncertainty

"All models are wrong, but some are useful." G.E.P. Box (1979)

Output of all models contains conceptual and factual errors. Four general sources of model error are:

1. *Error in Model Equation Form.* Errors in the form of model equations arise from model mathematics not fully representing the actual phenomenon under study. Such errors are common, even unavoidable, and arise from simplifications required in model development. All empirical equations (Manning's formula, etc.) necessarily include some error in model form. While common, these errors are not always significant for engineering uses of models.

2. *Error in Parameter Values.* Most model equations include some parameter values that are largely held constant throughout each simulation run. Examples include Manning's "n" values, kinetic rate constants, unit cost values, etc. These values are often empirical in nature, having no "real" existence - as with Manning's "n", or rely on being set by empirical observations, such as the use of recent bid information to estimate unit costs. In all these cases, there is likely to be some error in the value of each parameter value in a model. Sometimes these errors are important.

3. *Error in Input Data Values.* Input streamflow, precipitation, water demand, or other data are typically somewhat in err, although these errors are not always important for engineering. Even field data have errors.

4. *Numerical Solution Error.* Numerical error is error introduced in the solution of model equations by numerical or approximate analytical methods. Numerical integration is a common requirement for solving the equations of a model. Numerical solution schemes are varied in form and implementation (convergence tolerances and discretization coarseness) and commonly introduce some error. Much work has been done in estimating and controlling these errors.

Levels of Simulation Uncertainty Analysis

A wide variety of approaches have been developed to examine and evaluate model errors or uncertainties. These range from simple identification of variable of concern to various forms of analysis of the effects of uncertainty on model results, as categorized below:

1. Identification of outputs of concern
2. Identification of inputs of concern
3. Identify, discuss, and document ranges of inputs of concern
4. Sensitivity analysis of such inputs (independent)
5. Error analysis of such inputs (independent) (analytical or numerical)
6. Probabilistic error analysis, with pdfs of independent inputs (analytical or numerical)
7. Cross-correlation of inputs of concern
8. Scenario-based sensitivity analysis
9. Monte Carlo probabilistic error analysis (for correlated inputs)

Under each category, there are many specific techniques. Frey and Patil (2002) discuss some of these.

Sensitivity Analysis

Sensitivity analysis is a simple approach to examining model uncertainty. In general, sensitivity analysis perturbs one or more assumptions of a model and compares the resulting model outputs with those of some "base" case. While conceptually simple, sensitivity analysis can be a computationally burdensome and pose difficult interpretation problems. The major benefits of sensitivity analysis is the lack of need to estimate probability distributions of for each source of model uncertainty/error, its relatively low computational requirements, and usual lack of prolonged analysis/formulation/interpretation time. Three types of sensitivity analysis are available to the engineer. Each of these approaches can be implemented in many ways.

1. Sensitivity Analysis Based on Approximated First Derivatives. This type of sensitivity analysis examines the effects of "small" changes in model assumptions, approximating the first derivative of model outputs with respect to model parameter or input values. Such sensitivity results can be estimated for selected parameters or input variables, for all parameter and input variables, or for many combinations of parameter or input variables that may co-vary. These forms of sensitivity analysis respectively incur increasing computational requirements and interpretation difficulties. A major problem with sensitivity analysis based on approximated first derivatives is that it does not include any notions of the magnitude of error plausible for each parameter or input value. Thus a parameter with a relatively small first derivative might actually be quite important if its variance is large.

Given some model output variable y which is a function of a vector of input variables \underline{x} and parameter values \underline{a} , $y = f(\underline{x}, \underline{a})$. Some first-derivative-based measures of sensitivity analysis could include:

- a. Scalar values of $\partial y / \partial a_i$ for selected of single parameters.
- b. A vector of values for $\partial y / \partial a_i$ for all parameters.
- c. For bi-variate changes in parameter values, a matrix of sensitivities can be found for small changes in combinations of parameter values. For any two parameters in the matrix, a_i and a_j :

$$\frac{\Delta y}{\Delta} = \frac{f(x, a, a_i + \Delta, a_j + \Delta) - f(x, a)}{\Delta}$$

This provides a matrix of sensitivity information to identify particularly pathological or promising co-variations in parameter values. The same approach can be extended to model input variables.

A common variation of sensitivity analysis using first derivatives is to range of some set of parameter values by $\pm 10\%$ from a base case.

2. Error Analysis, Range-based Sensitivity Analysis. Responding to this deficiency, range-based sensitivity analysis estimates the change in model output values over some range of plausible values for selected parameter or input values. As with earlier forms of sensitivity analysis, this can be performed for single or multiply-varied parameter and input values. However, this approach requires some quantifications of a range of "plausible" or "representative" errors in each parameter or input variable of concern.

Range-based sensitivity analysis requires specification of some upper and lower bounds for selected or all parameters \underline{a} or input variables \underline{x} in the model, $y = f(\underline{x}, \underline{a})$. These upper and lower bounds can be the most extreme values possible, less extreme percentiles (at the 99%, 90%, or 75% levels for example), or, more qualitatively, at the extremes of a "reasonable" range of values expected by the engineer. For mathematical illustration, let these bounds be defined as $a_{i\min}$ and $a_{i\max}$. Using these values, changes in y can be found for different parameter or input variables. These changes in y can be found and arranged in ways similar to the first derivatives found above, replacing the first derivatives with $\Delta y_i = f(\underline{x}, \underline{a}, a_{i\max}) - f(\underline{x}, \underline{a}, a_{i\min})$. This approach creates an indicator of sensitivity which is more

standardized in terms of the likely uncertainty in output y arising from some similar amounts of uncertainty in each parameter a_i .

3. *Selected Scenarios.* This form of sensitivity analysis judiciously selects particularly worrisome or hoped-for cases of parameter and input values and examines their relative effects on model results. "Worst-case" and "best-case" scenario analyses are so common they have made it into the lay vernacular.

4. *Sensitivity output from optimization algorithms.* Some optimization methods, particularly linear programming, can produce a great deal of sensitivity analysis output as a by-product of their solution algorithm. [Add David Rheinheimer's CALSIM sensitivity screening work]

5. *Working backwards to find critical levels of inputs which cause failures,* and then back-calculate their probabilities.

In all cases, the purposes of sensitivity analysis are to a) identify the likely uncertainty in model outputs, b) assess which parameter values deserve the greatest attention for reducing the uncertainty in model outputs, and c) assessing if the model results are stable and insensitive enough for use and application. [Expand this section, adding work from the Klamath NRC report.]

References

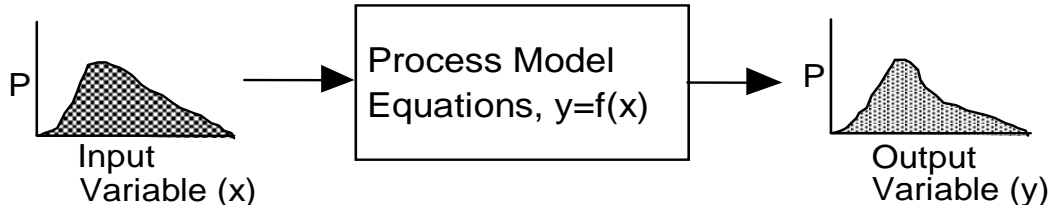
- Beck, M.B. (2002), "Model evaluation and performance," in A.H. El-Shaarawi and W.W. Piegorsch (eds.), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd, Chichester, Volume 3, pp 1275–1279. http://www.cgrer.uiowa.edu/cleaner/italy/Beck_Venicem.pdf
- Beven, K. (1993), "Prophecy, reality and uncertainty in distributed hydrological modelling," *Advances in Water Resources*, Vol. 16, pp. 41 – 51.
- Frey, H.C. and S.R. Patil (2002), "Identification and review of sensitivity analysis methods," *Risk Analysis*, Vol. 22, No. 3, pp. 553 – 578.
- Gass, S.I. (1983), Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis, *Operations Research*, Vol. 31, No. 4 (Jul. - Aug., 1983), pp. 603-631.
- Klemes, V. (1982), "Empirical and Causal Models in Hydrology," in National Research Council, Geophysics Study Committee, *Scientific Basis of Water-Resource Management*, National Academy Press, Washington, DC., pp. 95-104.
- National Research Council Committee on Hydrology, Ecology, and Fishes of the Klamath River (2007), *Hydrology, Ecology, and Fishes of the Klamath River Basin*, National Academy Press, Washington, DC, November.
- USEPA (2003), *Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models*, USEPA Office of Science Policy, Office of Research and Development, Washington, DC.



<http://xkcd.com/1445/>

11. Analytical Propagation of Uncertainty

Analytical propagation of uncertainty combines the uncertainties in input and parameter estimates and the model equations to produce analytically derived probability distributions for model output variables, as depicted below.



For probabilistic design, we are interested in using probability distributions for model inputs to find probability distributions for design performance measures. An example might be incorporating a probability distribution for annual flood flows through a flood damage model to obtain a probability distribution for flood damage. It is more common to do these things numerically, but it is probably best to with some analytical cases.

Case of One Input Variable and One Output Variable with 1-to-1 Mapping

For this case, there is a single output value that comes from each uncertain input variable and for each output value there is only one input value. Let's look at probabilistic input variable x and output variable y . We know $P(x)$ and we want to know $P(y)$.

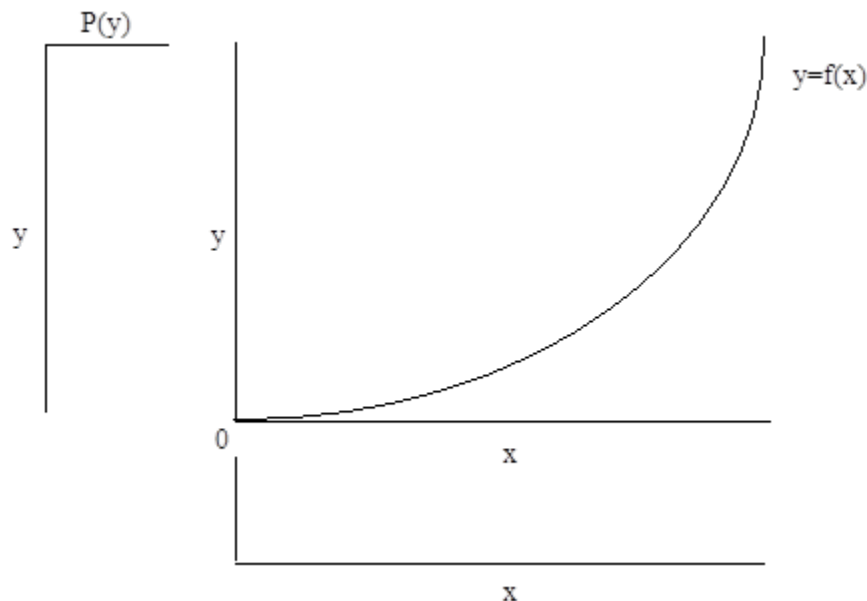
Discrete Case: If y and x are discrete variables with a one-to-one mapping between them, then $P(y) = P(f(x)) = P(x|y=f(x))$.

Continuous Case: If x and y are continuous and $f(x)$ is a one-to-one mapping, $f(x)$ is monotonic. This implies that:

$$(1) \quad \int_{-\infty}^{y=f(x)} P(y)dy = \int_{-\infty}^x P(x)dx .$$

Taking the derivative of this gives the probability density function, $P(y)dy = P(x)dx$, or

$$(2) \quad P(y) = \text{ABS} \left(\frac{dx}{dy} \right) P(x).$$



Case Where Each Output Value Can be Caused by Several Input Values

Where both x and y are continuous:

$$(3) \quad P(y)dy = \int_{-\infty}^{\infty} P(x | y = f(x))dx.$$

Where both x and y are discrete:

$$(4) \quad P(y) = \sum_x P(x | y = f(x)).$$

Example: Probabilistic Present Value Cost from Probabilistic Time-to-Failure

Consider the probabilistic landfill failure in the first homework. If t was the time of the first landfill failure and P_f is the probability of landfill failure in a given year if it has not failed previously, $P(t) = P_f(1-P_f)^{t-1}$ (a geometric distribution). This is the input probability function for the present value cost model. It is a discrete probability distribution for positive integer values of t .

The present value cost equation is $PVC_f(t) = C_f e^{-rt}$, where r is the real continuous discount rate, C_f is the cost of failure in the year of failure, and t is the time of failure. The function is monotonic.

$$(5) \quad P(PVC_f(t)) = P(t) \left| \frac{dt}{dPVC_f(t)} \right|, \text{ for continuous distributions of } PVC_f \text{ and } t.$$

Since $\left| \frac{dt}{dPVC_f(t)} \right| = C_f r e^{-rt}$, $P(PVC_f(t)) = \frac{P(t)}{rC_f e^{-rt}}$, or

$$(6) \quad P(PVC_f(t)) = P(t), \text{ if given a discrete distribution of } t.$$

Example: Probabilistic Present Value Cost from Probabilistic Discount Rates

The real discount rate r is a continuous parameter in the present value cost equation, $PVC_f = C_f e^{-rt}$. It is always hard to know exactly what value should be used for r , so let r have a uniform probability distribution between 0.01 and 0.05. If the failure occurs in year 10, what is the probability distribution of the present value cost of this failure?

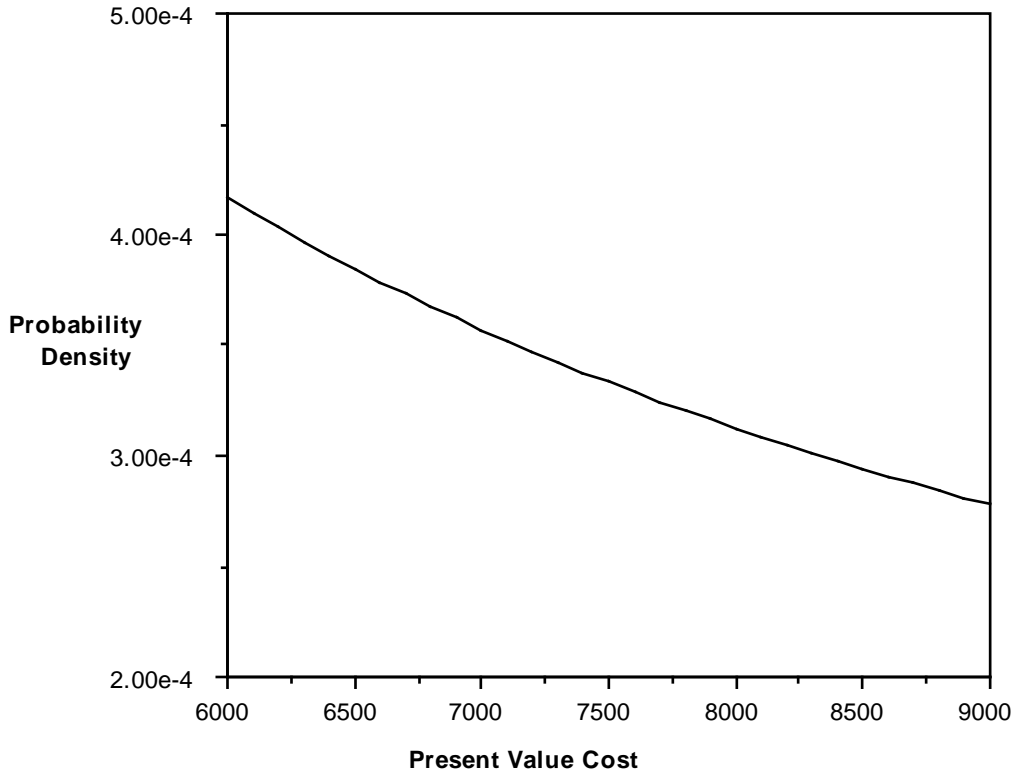
Using Equation 2, and noting that $dPVC_f/dr = -t C_f e^{-rt}$,

$$(7) \quad P(PVC_f(r)) = \text{ABS} \left(\frac{e^{-rt}}{-tC_f} \right) P(r) = \left(\frac{e^{-rt}}{tC_f} \right) P(r).$$

For a uniformly distributed r between 0.01 and 0.05, $P(r) = 25$. Solving the present value cost equation for r yields $r = \ln(C_f/PVC_f)/t$. Substituting these into Equation 7 gives:

$$(8) \quad P(PVC_f) = \left(\frac{C_f / PVC_f}{tC_f} \right) P(r) = \left(\frac{25}{tPVC_f} \right),$$

for the range of $C_f e^{-0.01t} \leq PVC_f \leq C_f e^{-0.05t}$, or $\$6065 \leq PVC_f \leq \9048 , for $t = 10$ and $C_f = \$10,000$. Outside this range, $P(PVC_f) = 0$. This result is plotted roughly below.



Note:

$$EV(PVC_t) = \int_{6065}^{9049} PVC_f P(PVC_f) dPVC_f = \frac{25}{t} PVC_f \Big|_{6065}^{9049} = \$7,460$$

$$PVC_t(r\text{-average}) = PVC_t(0.03) = 10,000 e^{-0.03(10)} = \$7,408.182$$

So for this case, there is not such a big difference.

Water Conservation Effectiveness

Rosenberg (2007), in an expansion of a term paper for this class, estimates the probability density of water conservation effectiveness where water use is the product of several factors. If each factor is normally distributed, then water use will therefore be log-normally distributed. [add more from this example]

References

Rosenberg, D.E. (2007), "Probabilistic Estimation of Water Conservation Effectiveness," *Journal of Water Resources Planning and Management*, Vol. 133, No. 1, January.

12. Approximating the Mean and Variance of Model Outputs

Sometimes the mean and variance of model outputs or design performance are of special interest, and the model of design performance and the model uncertainties are too complex for analytical propagation of uncertainty. One approach is to approximate the mean and variance of the model outputs by a truncated Taylor series, from the mathematical definitions of mean, standard deviation, skew, and covariance.

Definitions of Mean, Variance, Skew, and Covariance

The moment definitions of mean, standard deviation, and skew are:

$$\mu_y = \int_{-\infty}^{\infty} yp(y)dy ,$$

where μ_y is the mean of y and $p(y)$ is its probability density function.

$$\mu_y^2 + \sigma_y^2 = \int_{-\infty}^{\infty} y^2 p(y)dy , \text{ or } \sigma_y^2 = \int_{-\infty}^{\infty} (y - \mu_y)^2 p(y) dy ,$$

where σ_y is the standard deviation of y .

$$\gamma_y \sigma_y^3 + 3 \mu_y \sigma_y^2 + \mu_y^3 = \int_{-\infty}^{\infty} y^3 p(y)dy , \text{ or } \gamma_y \sigma_y^3 = \int_{-\infty}^{\infty} (y - \mu_y)^3 p(y) dy ,$$

where γ_y is the coefficient of skew. And the covariance of x and y is:

$$\text{cov}_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_y)(x - \mu_x)p(x, y) dx dy ,$$

Model $z = f(x,y)$ and its Taylor Series

Let $z = f(x,y)$ be the functional representation of the model of our design performance, with z being some measure of performance and x and y being some uncertain parameters or input variables described by their means, variances, and skews. The complete 3rd-order Taylor series expansion of $f(x,y)$ in two dimensions about the mean values of the parameters follows. This is as accurate in two dimensions as the more typical 3rd-order expansion is in one dimension.

$$\begin{aligned} \hat{f}(x,y) \cong & f(\mu_x, \mu_y) + \frac{\partial f(\mu_x, \mu_y)}{\partial y} (y - \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{(y - \mu_y)^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial y^3} \frac{(y - \mu_y)^3}{3!} \\ & + \frac{\partial f(\mu_x, \mu_y)}{\partial x} (x - \mu_x) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} (x - \mu_x)(y - \mu_y) + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x \partial y^2} \frac{(x - \mu_x)(y - \mu_y)^2}{2!} \\ & + \frac{\partial^4 f(\mu_x, \mu_y)}{\partial x \partial y^3} \frac{(x - \mu_x)(y - \mu_y)^3}{3!} \\ & + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{(x - \mu_x)^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x^2 \partial y} \frac{(x - \mu_x)^2 (y - \mu_y)}{2!} \\ & + \frac{\partial^4 f(\mu_x, \mu_y)}{\partial x^2 \partial y^2} \frac{(x - \mu_x)^2 (y - \mu_y)^2}{2! 2!} + \frac{\partial^5 f(\mu_x, \mu_y)}{\partial x^2 \partial y^3} \frac{(x - \mu_x)^2 (y - \mu_y)^3}{2! 3!} \\ & + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x^3} \frac{(x - \mu_x)^3}{3!} + \frac{\partial^4 f(\mu_x, \mu_y)}{\partial x^3 \partial y} \frac{(x - \mu_x)^3 (y - \mu_y)}{3!} \\ & + \frac{\partial^5 f(\mu_x, \mu_y)}{\partial x^3 \partial y^2} \frac{(x - \mu_x)^3 (y - \mu_y)^2}{3! 2!} + \frac{\partial^6 f(\mu_x, \mu_y)}{\partial x^3 \partial y^3} \frac{(x - \mu_x)^3 (y - \mu_y)^3}{3! 3!} . \end{aligned}$$

To estimate the mean and variance of $z = f(x,y)$ we integrate the parts of the Taylor series over the range from $-\infty$ to $+\infty$ with $\hat{f}(x,y)$ within the integral sign for the mean and $\hat{f}(x,y)^2$ for the variance or standard deviation.

Approximate Mean of f(x,y)

We can estimate the mean of the output/performance values as:

$$\begin{aligned}
 \mu_z &\approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x,y) p(x,y) dx dy . \\
 &\stackrel{\approx}{=} f(\mu_x, \mu_y) + 0 + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{\sigma_y^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial y^3} \frac{\gamma_y \sigma_y^3}{3!} \\
 &\quad + 0 + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} \text{cov}_{xy} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x \partial y^2} \frac{1}{2!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x - \mu_x)(y - \mu_y)^2) p(x,y) dx dy \\
 &\quad + \frac{\partial^4 f(\mu_x, \mu_y)}{\partial x \partial y^3} \frac{1}{3!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y)^3 p(x,y) dx dy \\
 &\quad + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{\sigma_x^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{2! \partial x^2 \partial y} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 (y - \mu_y) p(x,y) dx dy \\
 &\quad + \frac{\partial^4 f(\mu_x, \mu_y)}{2! 2! \partial x^2 \partial y^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 (y - \mu_y)^2 p(x,y) dx dy \\
 &\quad + \frac{\partial^5 f(\mu_x, \mu_y)}{\partial x^2 \partial y^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 (y - \mu_y)^3 p(x,y) dx dy \\
 &\quad + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x^3} \frac{\gamma_x \sigma_x^3}{3!} + \frac{\partial^4 f(\mu_x, \mu_y)}{3! \partial x^3 \partial y} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^3 (y - \mu_y) p(x,y) dx dy \\
 &\quad + \frac{\partial^5 f(\mu_x, \mu_y)}{3! 2! \partial x^3 \partial y^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^3 (y - \mu_y)^2 p(x,y) dx dy + \frac{\partial^6 f(\mu_x, \mu_y)}{3! 3! \partial x^3 \partial y^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^3 (y - \mu_y)^3 p(x,y) dx dy .
 \end{aligned}$$

Fortunately, this simplifies for some common special cases where:

1. Input variables x and y are independent.

$$\mu_z \stackrel{\approx}{=} f(\mu_x, \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{\sigma_x^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x^3} \frac{\gamma_x \sigma_x^3}{3!} + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{\sigma_y^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial y^3} \frac{\gamma_y \sigma_y^3}{3!}$$

2. There is only one uncertain input variable.

$$\mu_z \stackrel{\approx}{=} f(\mu_x) + \frac{\partial^2 f(\mu_x)}{\partial x^2} \frac{\sigma_x^2}{2!} + \frac{\partial^3 f(\mu_x)}{\partial x^3} \frac{\gamma_x \sigma_x^3}{3!}$$

3. Only first order estimates are desired.

$$\mu_z \stackrel{\approx}{=} f(\mu_x, \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} \text{cov}_{xy}$$

Approximate Variance of f(x,y)

We can estimate the variance of the output/performance values from the definition of variance as:

$$\sigma_z^2 \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{f}(x, y) - f(\mu_x, \mu_y)]^2 p(x, y) dx dy.$$

The same approach can be taken to derive the third-order approximation as was taken above, except it is just a little messier. The proof for the following results is "left to the interested reader."

For cases where input variables are not independent, only first-order approximation seems reasonable,

$$\sigma_z^2 \stackrel{\approx}{=} \frac{\partial f(\mu_x, \mu_y)}{\partial x} \sigma_x^2 + \frac{\partial f(\mu_x, \mu_y)}{\partial y} \sigma_y^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial f(\mu_x, \mu_y)}{\partial y} \text{cov}_{xy}.$$

For independent input variables, a second-order approximation is available:

$$\sigma_z^2 \stackrel{\approx}{=} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \sigma_y^2 + \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \gamma_y \sigma_y^3 + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \sigma_x^2 + \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \gamma_x \sigma_x^3$$

Mays and Tung (1992) review a bunch of problems that this approach has been applied to.

Application to Manning's Equation

Consider the error in the flow estimate from application of Manning's equation, $Q = (A/n) (A/P)^{2/3} S^{1/2}$, in metric, where there is a mean and variance for the Manning's roughness n, the slope s, the wetted perimeter P, and the cross-sectional area A.

If there is no covariance among the independent variables, and since $\frac{\partial^2 Q}{\partial n^2} = \frac{2A^{5/3} S^{1/2}}{P^{2/3} n^3}$,

$$\frac{\partial^2 Q}{\partial A^2} = \frac{10S^{1/2}}{9A^{1/3} P^{2/3} n}, \quad \frac{\partial^2 Q}{\partial P^2} = \frac{10A^{5/3} S^{1/2}}{9P^{8/3} n}, \quad \text{and} \quad \frac{\partial^2 Q}{\partial S^2} = -\frac{A^{5/3}}{4P^{2/3} S^{3/2} n},$$

$$\mu_Q \stackrel{\approx}{=} (A/n) (A/P)^{2/3} S^{1/2} + \frac{2A^{5/3} S^{1/2}}{P^{2/3} n^3} \frac{\sigma_n^2}{2!} + \frac{10S^{1/2}}{9A^{1/3} P^{2/3} n} \frac{\sigma_A^2}{2!} + \frac{10A^{5/3} S^{1/2}}{9P^{8/3} n} \frac{\sigma_P^2}{2!} - \frac{A^{5/3}}{4P^{2/3} S^{3/2} n} \frac{\sigma_S^2}{2!}$$

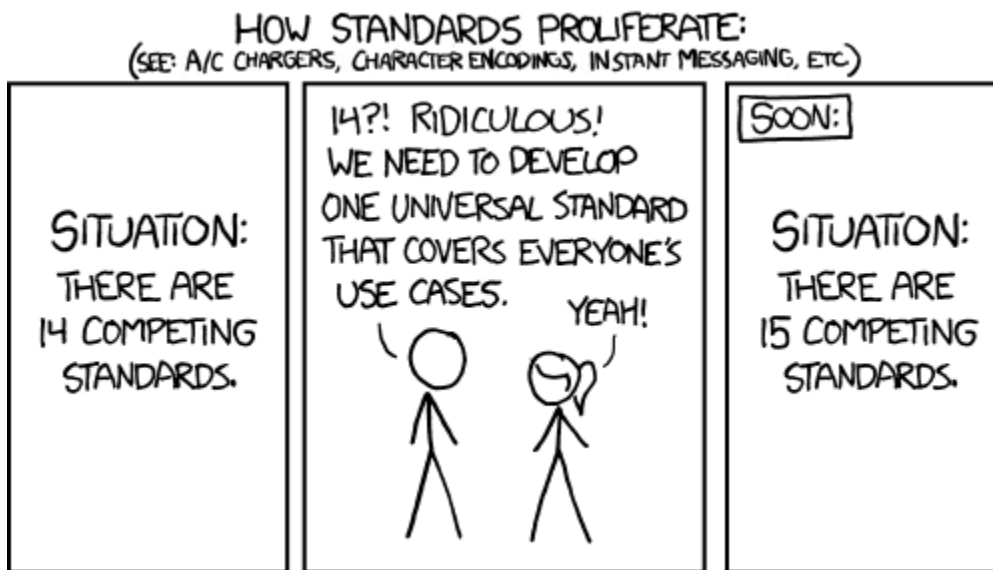
If the skews of the independent variables were known then a further refinement could be made.

For an estimate of the variance, without knowing the skew of each independent variable,

$$\sigma_Q^2 \approx \left(\frac{A^{5/3} S^{1/2}}{P^{2/3} n^2} \right)^2 \sigma_n^2 + \left(\frac{5S^{1/2} A^{2/3}}{3P^{2/3} n} \right)^2 \sigma_A^2 + \left(\frac{2A^{5/3} S^{1/2}}{3P^{5/3} n} \right)^2 \sigma_P^2 + \left(\frac{A^{5/3}}{2P^{2/3} S^{1/2} n} \right)^2 \sigma_S^2$$

References

- Benjamin, Jack R. and C. Allin Cornell (1970), *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill Book Co., N.Y.
- Burges, S. J., and D. P. Lettenmaier (1975), "Probability methods in stream quality management," *Water Resources Bulletin*, Vol. 11, pp. 115–130.
- Mays, Larry W. and Teou-Koung Tung (1992), *Hydrosystems Engineering and Management*, McGraw-Hill, N.Y.
- Tyagi, A. and C. T. Haan (2001), "Uncertainty analysis using corrected first-order approximation method," *Water Resources Research*, Vol. 37, No. 6, June, pp. 1847–1858.



Appendix: Some derivations regarding approximations to the variance:

$$\begin{aligned}
 \sigma_Z^2 &\approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{f}(x,y) - f(\mu_x, \mu_y)]^2 p(x,y) dx dy \\
 \sigma_Z^2 &\stackrel{\text{3}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{f}(x,y) + \hat{f}(x,y)^2] p(x,y) dx dy \\
 &= f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x,y)p(x,y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x,y)^2 p(x,y) dx dy \\
 &= f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_Z + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x,y)^2 p(x,y) dx dy
 \end{aligned}$$

The last of the three terms above becomes ugly. There are 16 terms in the 3rd-order Taylor-series expansion of $f(x,y)$; thus, there are $16^2 = 256$ terms in the 3rd-order Taylor series expansion $\hat{f}(x,y)^2$. This would be too much. Instead, let's look at a 1st-order expansion for the case where the input variables are not independent.

$$\begin{aligned}
 \sigma_Z^2 &\stackrel{\text{1}}{=} f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_Z \\
 &+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(f(\mu_x, \mu_y) + \frac{\partial f(\mu_x, \mu_y)}{\partial y}(y - \mu_y) + \frac{\partial f(\mu_x, \mu_y)}{\partial x}(x - \mu_x) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y}(x - \mu_x)(y - \mu_y) \right)^2 p(x,y) dx dy \\
 \sigma_Z^2 &\stackrel{\text{1}}{=} f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_Z + f(\mu_x, \mu_y)^2 \\
 &+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ 2f(\mu_x, \mu_y) \frac{\partial f(\mu_x, \mu_y)}{\partial y}(y - \mu_y) + 2f(\mu_x, \mu_y) \frac{\partial f(\mu_x, \mu_y)}{\partial x}(x - \mu_x) \right. \\
 &+ 2f(\mu_x, \mu_y) \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y}(x - \mu_x)(y - \mu_y) + \frac{\partial f(\mu_x, \mu_y)^2}{\partial y}(y - \mu_y)^2 + \frac{\partial f(\mu_x, \mu_y)^2}{\partial x}(x - \mu_x)^2 \\
 &+ 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial f(\mu_x, \mu_y)}{\partial x}(x - \mu_x)(y - \mu_y) + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y}(x - \mu_x)(y - \mu_y)^2 \\
 &\left. + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y}(x - \mu_x)^2(y - \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial x \partial y}(x - \mu_x)^2(y - \mu_y)^2 \right\} p(x,y) dx dy
 \end{aligned}$$

$$\begin{aligned} \sigma_z^2 \stackrel{1}{=} & \frac{\partial f(\mu_x, \mu_y)^2}{\partial y} \sigma_y^2 + \frac{\partial f(\mu_x, \mu_y)^2}{\partial x} \sigma_x^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial f(\mu_x, \mu_y)}{\partial x} (\text{cov}_{xy}) \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} (x - \mu_x)(y - \mu_y)^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} (x - \mu_x)^2 (y - \mu_y) \right. \\ & \left. + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial x \partial y} (x - \mu_x)^2 (y - \mu_y)^2 \right\} p(x, y) dx dy \end{aligned}$$

The first line of this is usually used as the first-order approximation, with other terms assumed to be insignificant. To add a bit of challenge, here's the 2nd-order expansion for the case where the input variables are independent.

$$\begin{aligned} \sigma_z^2 \stackrel{2}{=} & f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_z + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(x, y)^2 p(x, y) dx dy \\ = & f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_z \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ f(\mu_x, \mu_y) + \frac{\partial f(\mu_x, \mu_y)}{\partial y} (y - \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{(y - \mu_y)^2}{2!} \right. \\ & + \frac{\partial f(\mu_x, \mu_y)}{\partial x} (x - \mu_x) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x \partial y} (x - \mu_x)(y - \mu_y) + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x \partial y^2} \frac{(x - \mu_x)(y - \mu_y)^2}{2!} \\ & + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{(x - \mu_x)^2}{2!} + \frac{\partial^3 f(\mu_x, \mu_y)}{\partial x^2 \partial y} \frac{(x - \mu_x)^2 (y - \mu_y)}{2!} \\ & \left. + \frac{\partial^4 f(\mu_x, \mu_y)}{\partial x^2 \partial y^2} \frac{(x - \mu_x)^2 (y - \mu_y)^2}{2! 2!} \right\}^2 p(x, y) dx dy. \end{aligned}$$

Assuming the input variables are independent implies that all the co-varying terms have a value of zero. Thus,

$$\begin{aligned} \sigma_z^2 \stackrel{2}{=} & f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_z \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ f(\mu_x, \mu_y) + \frac{\partial f(\mu_x, \mu_y)}{\partial y} (y - \mu_y) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{(y - \mu_y)^2}{2!} \right. \\ & \left. + \frac{\partial f(\mu_x, \mu_y)}{\partial x} (x - \mu_x) + \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{(x - \mu_x)^2}{2!} \right\}^2 p(x, y) dx dy. \end{aligned}$$

The elimination of the co-varying terms also simplifies the elimination of the square.

$$\sigma_z^2 \stackrel{2}{=} f(\mu_x, \mu_y)^2 - 2f(\mu_x, \mu_y)\hat{\mu}_z$$

$$\begin{aligned}
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ f(\mu_x, \mu_y)^2 + 2f(\mu_x, \mu_y) \frac{\partial f(\mu_x, \mu_y)}{\partial y} (y - \mu_y) + 2f(\mu_x, \mu_y) \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{(y - \mu_y)^2}{2!} \right. \\
 & + \frac{\partial f(\mu_x, \mu_y)^2}{\partial y} (y - \mu_y)^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{(y - \mu_y)^3}{2!} + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial y^2} \frac{(y - \mu_y)^4}{2!} \\
 & + 2f(\mu_x, \mu_y) \frac{\partial f(\mu_x, \mu_y)}{\partial x} (x - \mu_x) + 2f(\mu_x, \mu_y) \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{(x - \mu_x)^2}{2!} \\
 & \left. + \frac{\partial f(\mu_x, \mu_y)^2}{\partial x} (x - \mu_x)^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{(x - \mu_x)^3}{2!} + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial x^2} \frac{(x - \mu_x)^4}{2!} \right\} p(x,y) dx dy.
 \end{aligned}$$

$$\begin{aligned}
 \sigma_z^2 & \stackrel{\cong}{=} -2f(\mu_x, \mu_y) \hat{\mu}_z \\
 & + 2f(\mu_x, \mu_y)^2 + 0 + 2f(\mu_x, \mu_y) \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{\sigma_y^2}{2!} \\
 & + \frac{\partial f(\mu_x, \mu_y)^2}{\partial y} \sigma_y^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{\gamma_y \sigma_y^3}{2!} \\
 & + 0 + 2f(\mu_x, \mu_y) \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{\sigma_x^2}{2!} + \frac{\partial f(\mu_x, \mu_y)^2}{\partial x} \sigma_x^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{\gamma_x \sigma_x^3}{2!} \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial y^2} \frac{(y - \mu_y)^4}{2!} + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial x^2} \frac{(x - \mu_x)^4}{2!} \right\} p(x,y) dx dy.
 \end{aligned}$$

$$\begin{aligned}
 \sigma_z^2 & \stackrel{\cong}{=} \frac{\partial f(\mu_x, \mu_y)^2}{\partial y} \sigma_y^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial y} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial y^2} \frac{\gamma_y \sigma_y^3}{2!} \\
 & + \frac{\partial f(\mu_x, \mu_y)^2}{\partial x} \sigma_x^2 + 2 \frac{\partial f(\mu_x, \mu_y)}{\partial x} \frac{\partial^2 f(\mu_x, \mu_y)}{\partial x^2} \frac{\gamma_x \sigma_x^3}{2!} \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial y^2} \frac{(y - \mu_y)^4}{2!} + \frac{\partial^2 f(\mu_x, \mu_y)^2}{\partial x^2} \frac{(x - \mu_x)^4}{2!} \right\} p(x,y) dx dy.
 \end{aligned}$$

We neglect the higher ordered terms remaining within the integral to get the earlier result.

13. Monte Carlo Modeling

Monte Carlo modeling is a numerical approach to estimating the probability distribution of model outputs. It consists of three steps:

1. Create a large number of statistically representative input data sets. Usually between 100 and several thousand separate data sets are needed, with each data set considered equally probable.
2. Run each input data set through the simulation model, independently.
3. From the resulting large number of statistically representative output data sets, infer a probability distribution for the output variables.

Random Number Generation

To implement the Monte Carlo approach, there must be an ability to generate random numbers on a computer that represent our understanding of the uncertainty in the problem. The handout reviews some methods for generating random numbers for specific distributions. Numerical methods for generating random numbers for many common named distributions can be found at <http://en.wikipedia.org>, along with other useful information.

Example: Synthetic Streamflow Modeling

The study of reservoir operations is usually hampered by short streamflow records. A 40-year streamflow record is unlikely to contain a 100-year flood or a 100-year drought. This makes it difficult to characterize the probabilistic performance of proposed reservoir construction and operation decisions, particularly where performance under extreme conditions is important. The generation of synthetic streamflows, based on Monte Carlo analysis, is sometimes proposed to overcome this problem.

A typical problem consists of a simple reservoir model (conservation of mass), with a release rule, and a fixed rate of water demand. The streamflow input is generated by a Monte Carlo method. Since streamflows are correlated in time (temporal autocorrelation), a time series model is used. The Thomas-Fiering or Lag-1 Markov model is one such model.

$$Q_t = \mu + \rho(Q_{t-1} - \mu) + z_t \sigma(1-\rho^2)^{0.5},$$

where Q_t is the streamflow in time t , μ is the mean streamflow, ρ is the lag-1 correlation between streamflows in time, σ is the variance of streamflows, and z_t is a "standard normal deviate" generated randomly for each time period. A standard normal deviate is a random number, normally distributed with a mean of zero and a standard deviation of one. The parameters μ , ρ , and σ can be constants whose values are selected before the numerical study, can themselves be part of a larger Monte Carlo study (where the parameter values are also considered to be uncertain).

This model can then be used to generate an arbitrary number of arbitrarily long "synthetic" streamflow records. These synthetic records are then run through the reservoir model to generate a distribution of estimated reservoir performances (costs, benefits, failures to supply fixed levels of water demand, etc.).

Example: Numerical Integration Using Monte Carlo

Another common use of Monte Carlo analysis is to integrate difficult functions. The Monte Carlo method of integrating a function is to evaluate the value of the function at a large number of randomly-selected points in the range of its independent variables. The integral of the function is then approximately the average value of the function found for the randomly-selected points, multiplied by the range or area of the independent variables of the integration.

Tail Behavior and Sample Size

With a small sample size, Monte Carlo experiments may not "see" important extreme events. Thus, if performance under extreme events is important, a large sample size is often needed.

Reference Law, A.M. and W.D. Kelton (1991), *Simulation Modeling and Analysis*, McGraw Hill, N.Y

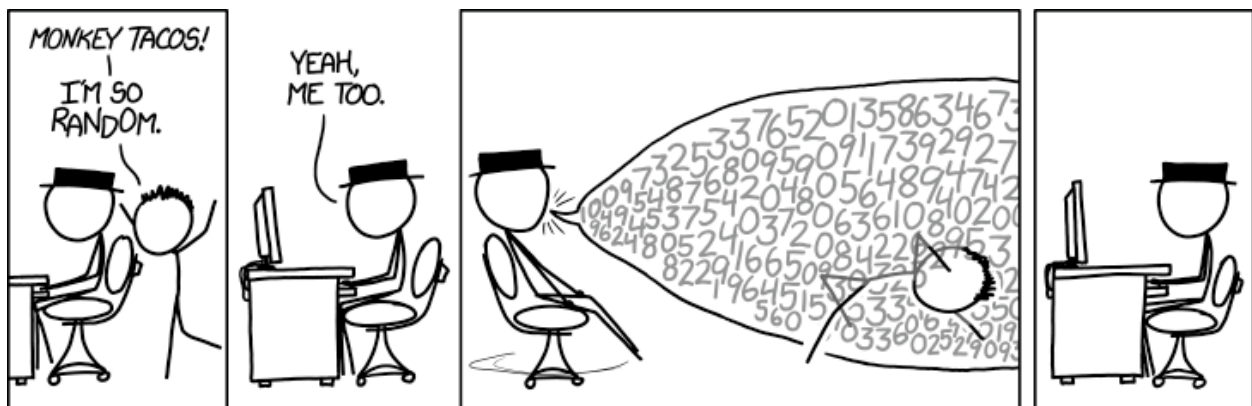
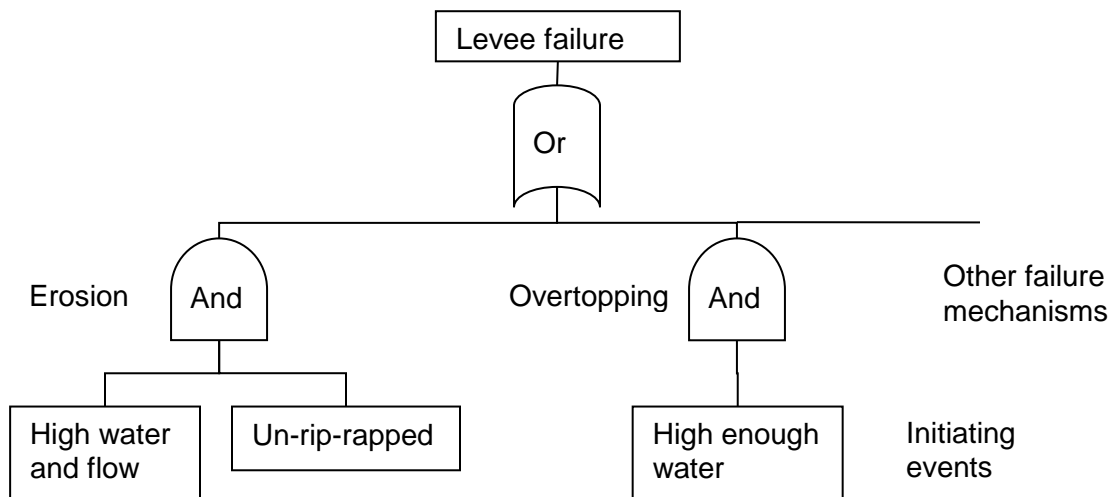
Rugen, Pamela and Callahan, Barbara(1996), “An overview of Monte Carlo, a fifty year perspective”, *Human and Ecological Risk Assessment: An International Journal*, 2: 4, 671 — 680.

Fault Tree Analysis

Fault tree analysis is used to depict the range of ways that a failure or other desirable outcome can occur and calculate the probabilities of failure and particular failure mechanisms. For example, a levee can fail by several mechanisms, such as overtopping, slump failure, piping, wave erosion, toe scour, etc. Such analyses are commonly applied in aerospace, nuclear, military, and industrial systems.

Typically, a fault tree diagram has a particular “top event” failure, and “bottom” initiating events, with a logic of failure mechanisms in between, with a particular nomenclature. By assigning probabilities to the initiating events, the probability of overall failure can be calculated, and means can be identified to reduce the overall probability of failure.

There are many web tutorials on this topic.



xkcd.com/1210/

14. Markov Chains

We are often interested in how the probability distribution of the condition or state of a system will evolve over time. Markov chains can help us do this in some cases.

Markov processes have the probability of a state occurring in the next time period depend only on the present state of the system. The particular path by which the current state came about is unimportant in determining the probability of what happens next. Stated more mathematically,

$$(1) \quad P(X_{t+1} = s_j | X_t, X_{t-1}, X_{t-2}, \dots) = P(X_{t+1} = s_j | X_t), \text{ for all states } s_j \text{ and periods } t.$$

If the *transition probabilities* do not change over time, then the process is called *stationary*,

$$(2) \quad P(X_{t+1} = s_j | X_t = s_i) = P(X_{t=1} = s_j | X_{t=0} = s_i), \quad \text{for all states } i \text{ and } j \text{ and periods } t.$$

An Example: Sedimentation in a Ship Channel

Given a probability distribution for the annual amount of sedimentation in a shipping channel, what is the probability distribution of accumulated sediment in each future year? Also, how many years are likely to elapse until accumulated deposition encroaches to a depth that would hinder navigation?

Let $S(t)$ be the amount of sediment that has accumulated in the channel over t years and s_i be the amount of sediment deposited in a particular year i ,

$$(3) \quad S(t) = s_1 + s_2 + s_3 + \dots + s_t$$

Let $P_1(s)$ be the probability that amount s is deposited in a particular year. This probability distribution is assumed not to vary in time. The probability distribution of the accumulation of sediment after one year would therefore be $P_1(S_1 = s) = P_1(s)$.

For the second year, $t=2$, the probability distribution for the 2-year accumulation becomes more complex. There are a number of ways for each amount of accumulation to occur,

$$(4) \quad P_2(S) = \int_0^S P_1(s) P_1(S - s) ds .$$

This represents the accumulation of the probabilities for all the ways that amount S can accumulate over two years.

The nice trait of Markov chains is that, once the probability distribution for states at one time period are calculated, it is simple to calculate the probability distribution for the next time period. For the third year, $t=3$, we can adapt the previous equation to find $P_3(S)$,

$$(5) \quad P_3(S) = \int_0^S P_1(s) P_2(S - s) ds .$$

We can then step blithely on for as many time periods as we like with relative computational efficiency using the general formula:

$$(6) \quad P_t(S) = \int_0^S P_1(s) P_{t-1}(S - s) ds .$$

Generalization

This last equation is a specific case of the more general Chapman-Kolmogorov equations (Hillier and Lieberman 1995). Let $p_{ij}^{(n)}$ be the probability of the process going from state i to state j in n steps. Then,

$$(7) \quad p_{ij}^{(n)} = \sum_{k=0}^m p_{ik}^{(v)} p_{kj}^{(n-v)} ,$$

for all i, j , and n where $0 \leq v \leq n$, m is the number of possible states, and v is some arbitrary intermediate number of steps. In words, the probability of going from state i to state j in n steps is the sum of the joint probabilities of going from i to k in v steps and then from k to j in the remaining $n-v$ steps. For $v = 1$, the equation reduces to Equation 8 below, which is just a discrete version of Equation 6:

$$(8) \quad p_{ij}^{(n)} = \sum_{k=0}^m p_{ik} p_{kj}^{(n-1)} .$$

Example Continued: Time to First Passage

For the sedimentation problem above, it is useful to have estimates of the probability distribution of the time when sediment accumulation will first exceed some volume V , $P(t|V)$. Lund (1990) represents the probability distribution of time of channel encroachment, given that volume V had just been dredged beyond the navigation depth for the channel. Mathematically $P(t|V)$ is,

$$(9) \quad P(t|V) = P(S(t) > V | S(t-1) < V).$$

This is equivalent to:

$$(10) \quad P(t|V) = \int_V^\infty P_t(S) dS - \int_V^\infty P_{t-1}(S) dS , \text{ or}$$

$$(11) \quad P(t|V) = \left(\begin{matrix} V \\ 1 - \int_0^V P_t(S) dS \\ 0 \end{matrix} \right) - \left(\begin{matrix} V \\ 1 - \int_0^V P_{t-1}(S) dS \\ 0 \end{matrix} \right) ,$$

the probability that V will be exceeded by year t , minus the probability that V would be exceeded before t . Combining terms and simplifying gives,

$$(12) \quad P(t|V) = \int_0^V [P_{t-1}(S) - P_t(S)] dS ,$$

which can be resolved using Equation 6.

Other Applications

Similar formulations can be used for beach erosion, sedimentation in a reservoir or detention basin, or growth in population, water demand, or waste loads.

We have a host of problems where we are interested in the probability distributions of demands or loads on our systems and the probability distributions of when a demand or load will first exceed some system capacity.

Markov chains work well where there is little or no long-term persistence or memory in the transition probabilities. (This makes it difficult to use Markov chains for droughts.) In the example case above, sedimentation was assumed to be independently distributed between years and unaffected by the amount

of sediment that had already accumulated. These assumptions limit the range of cases that the method would apply to.

Much else can be said of Markov chains. There is a huge literature on the subject. It is probably more important to get a feel for their application and fundamentals, and leave you to read the literature as you see problems arise.

A Discrete Example: Valve Failure and Replacement Rates

A city has many thousands of valves in its water distribution system. Generally, these valves are "exercised" and inspected each year. Upon inspection, valves are classified as OK, Stiff (hard to turn), or Failed (cannot be turned). From years of experience, the state transition probability matrix \underline{T} is as follows:

State at t	State at t+1		
	OK	Stiff	Failed
OK	0.90	0.08	0.02
Stiff	0.20	0.60	0.20
Failed	0	0	1

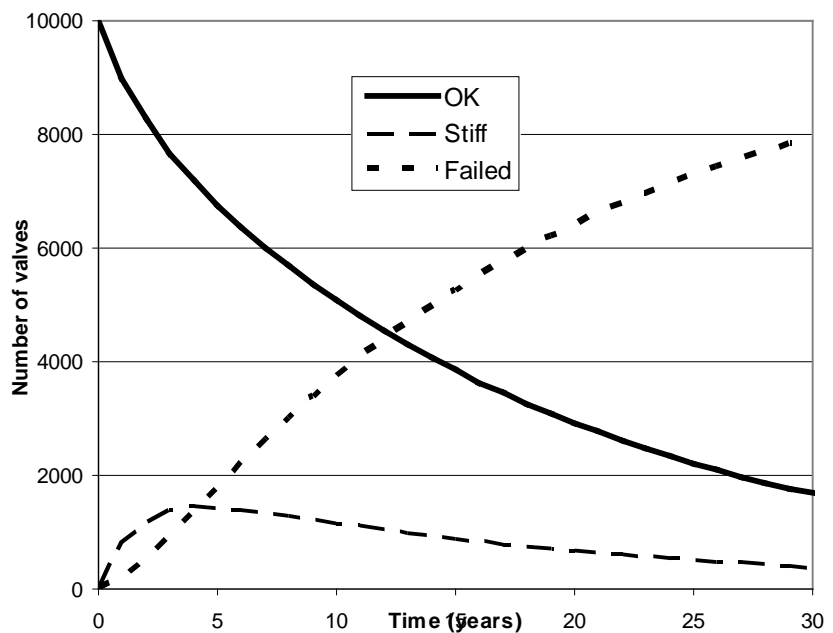
If only failed valves are replaced, what is the expected proportion of OK, Stiff, and Failed values each year? Assume an initial installation of all new valves (in OK condition). What is the probability distribution of valve lifetime? For inventory and budget purposes, what is the proportion of valves needing replacement in each year?

Probabilities for each year, without replacement of failed valves:

t=	0	1	2	t+1
OK	1	0.90	$0.90(0.9) + 0.08(0.2)$	$OK_t(0.9) + Stiff_t(0.2)$
Stiff	0	0.08	$0.90(0.08) + 0.08(0.6)$	$OK_t(0.08) + Stiff_t(0.6)$
Failed	0	0.02	$0.90(0.02) + 0.08(0.2) + 0.02$	$OK_t(0.02) + Stiff_t(0.2) + Fail_t$

These are plotted in the adjacent figure.

Without replacing any valves, the number of OK valves in the system gradually decays, until by about the year 180, all valves have failed.



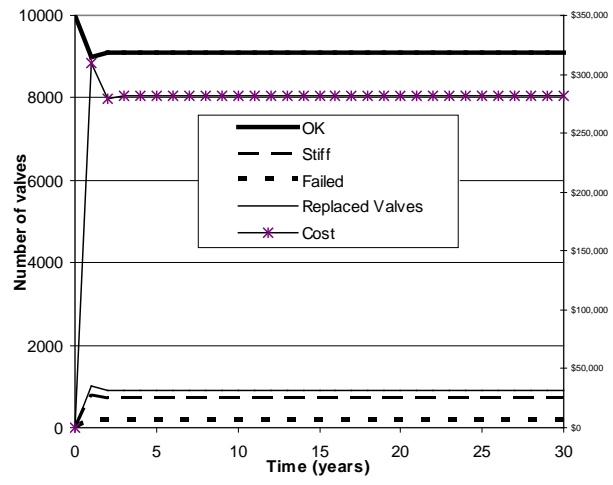
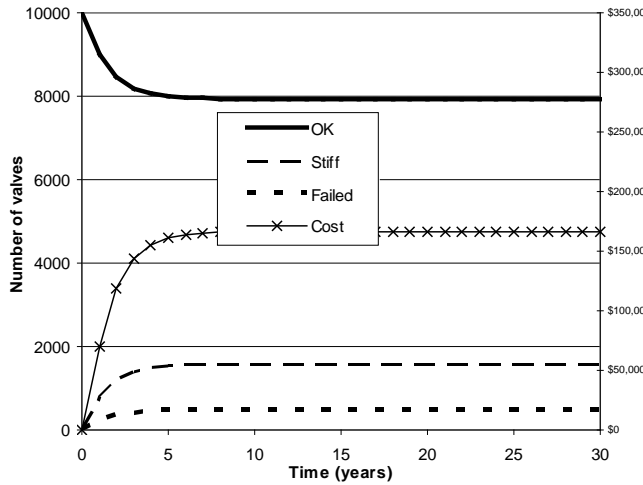
Examining Replacement Policies

If failed valves are replaced with new valves in OK condition, the “sticky” failure state is eliminated. The probabilities for each year, then change:

t=	0	1	2	t+1
OK	1	0.90	$(0.90+0.02)(0.9) + 0.08(0.2)$	$(OK_t+Fail_t)(0.9) + Stiff_t(0.2)$
Stiff	0	0.08	$(0.90+0.02)(0.08) + 0.08(0.6)$	$(OK_t+Fail_t)(0.08) + Stiff_t(0.6)$
Failed	0	0.02	$(0.90+0.02)(0.02) + 0.08(0.2)$	$(OK_t+Fail_t)(0.02) + Stiff_t(0.2)$

Replace only Failed Valves

Replace both Failed and Stiff Valves



Transition probability matrix with Failed Valves Replaced

	State at t+1		
State at t	OK	Stiff	Failed
OK	0.9	0.08	0.02
Stiff	0.2	0.6	0.2
Failed	0.9	0.08	0.02

Transition probability matrix with Failed and Stiff Valves Replaced

	State at t+1		
State at t	OK	Stiff	Failed
OK	0.9	0.08	0.02
Stiff	0.9	0.08	0.02
Failed	0.9	0.08	0.02

[Perhaps replace failed and stiff valves with {0.9, 0.08, 0.02} states, the characteristics of OK valves.]

With failed valves being replaced, we have essentially changed the state transition probability matrix so that Failed states transition to OK states in the next time-step. Similarly for a strategy of replacing both failed and stiff valves. The plots above both include annual costs, where the cost of a failed valve is \$50/year and the cost of replacing a valve is \$300.

To find the equilibrium condition (for large t), we set $OK_t = OK_{t+1}$, $Stiff_t = Stiff_{t+1}$, and $Fail_t = Fail_{t+1}$. (This is, after all, the definition of an equilibrium condition.) These would be known as the *stationary probabilities* for this process. Where all failed valves are replaced by OK valves, this provides a solution by three simultaneous linear equations:

$$\begin{aligned}
 OK_t &= (OK_t+Fail_t)(0.9) + Stiff_t(0.2) \\
 Stiff_t &= (OK_t+Fail_t)(0.08) + Stiff_t(0.6) \\
 Fail_t &= (OK_t+Fail_t)(0.02) + Stiff_t(0.2)
 \end{aligned}$$

This is the same as using the state transition probability matrix as: $\underline{X} = \underline{X}^T \underline{T}$

For this particular example, noting that the state transition probability matrix has been changed to reflect failed states transitioning to OK states, $P(OK) = 0.794$, $P(Stiff) = 0.159$, $P(Fail) = 0.047$.

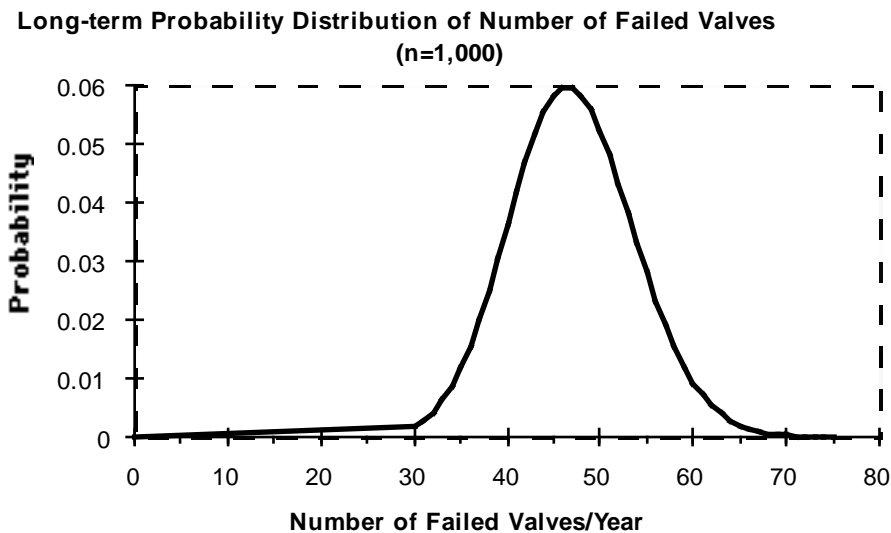
To find the *optimal equilibrium replacement policy*, $\underline{X}^T \underline{T} - \underline{X} = \underline{R}$, where \underline{R} is a vector of replacements, R_{OK} = the additional OK valves installed, and R_{stiff} and R_{failed} are the number of failed and stiff valves replaced. The sum of \underline{R} elements must equal zero. The sum of \underline{X} must equal the number of valves in the system. And of course R_{stiff} and R_{failed} are limited by the number of failed and stiff valves (\underline{X}). \underline{X} and \underline{R} are decision variables in a linear program with an objective function of minimizing the costs from failed valves plus the cost of valve replacements.

How many replacement valves should be ordered?

How would you estimate the probability distribution of the number of valves that need replacement each year? This would be a Binomial distribution with n as the number of valves in the system, and $P_{fail} = P(\text{Fail})$ above.

$$B(x|n, P_{fail}) = \frac{n!}{x!(n-x)!} P_{fail}^x (1-P_{fail})^{n-x}.$$

Where n = 1,000 this distribution is plotted below, for the stationary long-term probabilities.



Example: Markov Chains Applied to Inventory Stocks

[Gosh! This pdf of the number of replacement valves needed each year, is a nice start for an example of an inventory problem. Given this pdf for removal of valves from inventory, how many years will a given stock of, say 150 valves last? Given an ordering cost and unit cost/valve, what is the optimal inventory of valves to order? Have to make assumptions about how fast valves can be delivered.]

References

Benjamin, Jack R. and C. Allin Cornell (1970), *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill Book. Co., N.Y.
 Hillier, F.S. and G.J. Lieberman (1995), *Introduction to Operations Research*, McGraw-Hill, N.Y.
 Lund, J.R. (1990), "Scheduling Maintenance Dredging on a Single Reach with Uncertainty," *Journal of Waterway, Port, Coastal, and Ocean Engineering*, ASCE, Vol. 116, No. 2, pp. 211-231.
 Parzen, Emanuel (1960), *Modern Probability Theory and Its Applications*, John Wiley & Sons, N.Y.

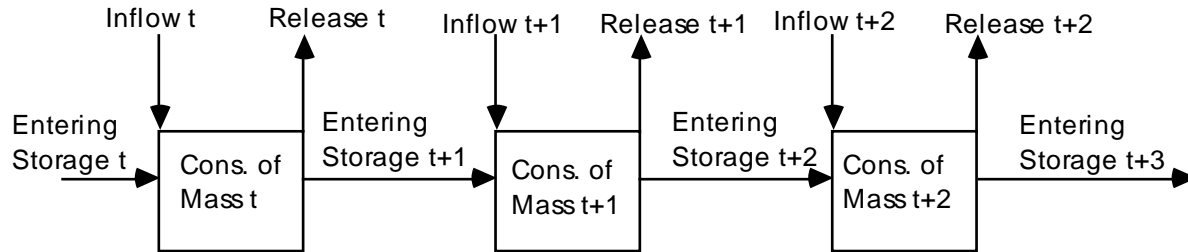
15. Stochastic Dynamic Programming

(2 lectures)

Bellman, the major originator of dynamic programming, is said to have referred to dynamic programming as a process for multi-stage decision-making. This is the approach taken here.

Review of Deterministic Dynamic Programming

Recall that dynamic programming can be used where the problem can be structured to consist of a series of staged decisions, where the decisions at each stage of the problem are contingent on the state of the system. Thus, for reservoir operations problems, the problem is divided into release decisions at each time period, with the state of the system being the amount of reservoir storage at each time period. For a single-reservoir system with three time periods, the problem is schematically,



Each time period is a *stage*. At each stage, a *decision* must be made, in this case the volume of water to be released. The decision at each stage will have some effect on the condition of the system, described by a *state variable*, in this case a reservoir storage volume. In this case the state of the system is governed by conservation of mass.

The decisions and states also have economic consequences, described by an *objective function*. In this case, let $B_t(Q_t, S_t)$ be the economic value of releasing Q_t from the reservoir and having storage S_t at time period t . The overall objective is to:

$$\text{Maximize } \sum_{t=0}^T B_t(Q_t, S_t),$$

over a operating horizon of T .

Decisions are linked between stages using a recursive formula, really an accumulated objective function. This accumulation can proceed either forward or backward. Oddly, backward is often easier. For this case the *backward recursion* is:

$$f_t(Q_t, S_t) = B_t(Q_t, S_t) + f_{t+1}^*(S_{t+1} = S_t + I_t - Q_t).$$

This recursion consists of two parts. The first term $B_t(Q_t, S_t)$ contains the direct benefits of the decision in stage t , Q_t , given the state, S_t . The second term is the accumulation of the best decisions from all later stages (remember we're going backwards).

$$f_t^*(S_t) = \text{MAX}_{Q_t} f_t(Q_t, S_t).$$

This $f_t^*(S_t)$ is stored, along with the accompanying decision, to facilitate calculations for other stages. Only the best decision for each state at each stage is stored.

The *forward recursion* would be:

$$f_t(Q_t, S_t) = B_t(Q_t, S_t) + f_{t-1}^*(S_{t-1} = S_t - I_t + Q_t).$$

For the forward formulation, the state variable is the outgoing state from stage t , $S_t = S_{t-1} + I_t - Q_t$. The $f_t^*(S_t)$ is found in the same way as in the backward formulation.

The solution of the problem typically involves setting up a computational table for each stage. In the backward formulation, the solution starts with stage $t = T$ and moves backwards until $t = 1$. Both storage and release decisions must be discretized. Here Release has m discrete values and Storage has p discrete values. The tables below are for the solution by a backwards formulation.

Stage T

$n = T$	$f_T(S_T, Q_T) = B(S_T, Q_T)$			
State S_T	$Q_T: 0$	-----	Q_m	$f^*(S_T)$
0		-----		

Reservoir Capacity = s_p		-----		

...

Stage t

$n = t$	$f_t(S_t, Q_t) = B(S_t, Q_t) + f_{t+1}^*(S_t - Q_t + I_t)$			
State S_t	$Q_t: 0$	-----	Q_m	$f^*(S_t)$
0		-----		

Reservoir Capacity = s_p		-----		

...

Stage t = 1

$n = 1$	$f_1(S_1, Q_1) = B(S_1, Q_1) + f_2^*(S_1 - Q_1 + I_1)$			
S_1	$Q_1: 0$	-----	Q_m	$f^*(S_1)$
$S_1 = \text{initial}$		-----		
storage value		-----		

There is only one reservoir state in the present, S_1 .

Once the tables are filled up to the present state (stage $t=1$), the optimal objective function value ($f^*(S_1)$) and optimal initial decision (Q_1^*) can be found. The initial decision and state are then used to find the optimal state in the next stage, with the table identifying the optimal decision for that state. By chaining forward in this way, the optimal solutions (and states) can be found for all stages.

While this method seems complex, it compares very favorably with enumeration, and is a form of implicit enumeration. The number of cost calculations grow with the number and discretization of state variables.

Stochastic Dynamic Programming without Temporal Autocorrelation

A dynamic program can incorporate several forms of uncertainty without temporal autocorrelation.

Uncertainty in the Direct Benefit Function

The easiest form of uncertainty to incorporate into dynamic programming is uncertainty in the benefit function. In the reservoir case above, the benefits from operation in a particular time step may be uncertain. $B_t(S_t, Q_t)$ may not be a single number, but may really be a probability distribution, $P(B_t(S_t, Q_t))$. In this case, the recursive equation would be modified to:

$$f_t(Q_t, S_t) = \sum_{k=0}^n [P(B_{tk}(Q_t, S_t)) B_k] + f_{t+1}^*(S_{t+1}=S_t+I_t-Q_t),$$

where the range of direct benefits has been divided into n increments with values B_k and a probability $P(B_{tk}(Q_t, S_t))$, conditioned on the state of the system (S_t), the stage (t), and the decision (Q_t). For this case, this is the only change in the dynamic program and its solution from the deterministic case.

Uncertainty in the State Transition

For reservoirs, there is usually uncertainty in future inflows (I_t) and relative certainty in the benefits of operations $B_t(Q_t, S_t)$. Uncertainty in the inflows affects only the *state transition* part of the recursive equation. Thus, we are not certain of the future state of the system resulting from a given decision made at an incoming state. To accommodate this uncertainty, the recursion can be modified to:

$$f_t(Q_t, S_t) = B_t(Q_t, S_t) + \sum_{k=0}^n [P(I_{tk}) f_{t+1}^*(S_{t+1}=S_t+I_{tk}-Q_t)] , \text{ for a backward formulation}$$

or

$$f_t(Q_t, S_t) = B_t(Q_t, S_t) + \sum_{k=0}^n [P(I_{tk}) f_{t-1}^*(S_{t-1}=S_t-I_{tk}+Q_t)] , \text{ for a forward formulation.}$$

Again, the change in the recursive equation is the only change in the dynamic program and solution method from the deterministic case.

The nice thing about the incorporation of these types of uncertainty into dynamic programming is that the computational size of the problem does not change greatly from the deterministic case.

Stochastic Dynamic Programming with Markov Processes

In many cases the state transition probabilities are conditioned on the previous period's events. For example, the probability of a streamflow in one period is assumed to be affected by the streamflow in the previous period. This makes the state transition a Markov process, and requires a little different change in the dynamic programming recursive equation.

Uncertainty in the State Transition

For the reservoir operation problem, where the probabilities of reservoir inflows I_t are conditioned on the previous periods flow, the dynamic programming recursion becomes, for a backward formulation:

$$f_t(Q_t, S_t, I_{t-1}) = B_t(Q_t, S_t) + \sum_{k=0}^n [P(I_{tk}|I_{t-1}) f_{t+1}^*(S_{t+1}=S_t+I_{tk}-Q_t, I_{tk})] ,$$

This addition of Markov uncertainty adds another state variable to the formulation (I_t), and can greatly increase the computational effort needed to solve the problem. Because the present inflows are conditioned on those of the immediate past, each possible past flow must be represented as a possible state in the recursive equation,

$$f_{t+1}^*(S_{t+1}, I_t) = \text{MAX}_{Q_{t+1}} f_{t+1}(Q_{t+1}, S_{t+1}, I_t).$$

The optimal decision here is conditioned on both states: reservoir storage and recent inflow.

A *forward* recursion for this problem would be:

$$f_t(Q_t, S_t, I_t) = B_t(Q_t, S_t) + \sum_{k=0}^n [P(I_{t-1,k}|I_t) f_{t-1}^*(S_{t-1}=S_t-I_t+Q_t, I_{t-1,k})] ,$$

where S_{t-1} is the outgoing storage from stage $t-1$. Note the reversal of the conditional probability for this formulation. (Bayes' theorem?) (There might be a better forward recursion.)

Uncertainty in the Direct Benefit Function

Sometimes the probability distribution for the direct benefits portion of the recursive equation are conditioned on events in the previous period/stage. For the case of the reservoir, if growth in recreation/fishing potential changes in a Markovian way with reservoir storage in the previous period, there would be consequent changes in the probabilities of how storage and releases would be valued. The following *backward* recursion reflects this:

$$f_t(Q_t, S_t, S_{t-1}) = \sum_{k=0}^n P(R_{kt}|S_{t-1}) [B_t(Q_t, S_t, R_{kt})] + f_{t+1}^*(S_{t+1}=S_t+I_t-Q_t, S_t).$$

Again, this adds another state variable to the formulation. Also, both parts of the recursion become probabilistic.

Another potential application of time-correlated uncertainty in the direct benefit function might be where hydropower value (price) is correlated in time, due to weather (e.g., air conditioning peaking) or hydrologic conditions. This might be formulated as in the following backward recursion:

$$f_t(Q_t, S_t, H_{t-1}) = \sum_{k=0}^n P(H_{kt}|H_{t-1}) [B_t(Q_t, S_t, H_{kt}) + f_{t+1}^*(S_{t+1}=S_t+I_t-Q_t, H_{kt})].$$

Applications

Stochastic dynamic programming (SDP) has been applied to a variety of problems. Some major examples include: reservoir problems, capacity expansion problems, inventory problems, and optimal maintenance and replacement problems. Kelman, et al. (1990), Faber and Stedinger (2001), and Stedinger et al (2013) have another variant of SDP, Sampling SDP.

[Would be good to add Sampling SDP and SDP with climatic change uncertainty – expanding on Vicuna 2007 dissertation. Ideally, expand Zhu et al 2007 DP to include uncertainty in future change in flood frequency.]

References

- Faber, B.A. and J.R. Stedinger (2001), "Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts," *Journal of Hydrology* 249 (2001) 113-133.
- Hillier, F.S. and G.J. Lieberman (1990), *Introduction to Operations Research*, McGraw-Hill, N.Y., Chapters 11 and 20.
- Kelman, J., J.R. Stedinger, L.A. Cooper, E. Hsu, and S-Q Yuan (1990), "Sampling stochastic dynamic programming applied to reservoir operation," *Water Resources Research*, Vol. 26, No. 3, March, pp. 447-454
- Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ, chapter 7 in particular.
- Stedinger, J.R., B.A. Faber, and J.R. Lamontagne (2013), "Developments in Stochastic Dynamic Programming for Reservoir Operation Optimization," *World Environmental and Water Resources Congress 2013: Showcasing the Future*, ASCE, pp. 1266-1278.
- Tejada-Guibert, J.A., Johnson, S.A., Stedinger, J.R. (1995), "The value of hydrologic information in stochastic dynamic programming models of a multireservoir system," *Water Resources Research* 31(10), 2571-2579.
- Yakowitz, S.J., (1982), "Dynamic programming applications in water resources," *Water Resources Research*, 18(4), 673 – 696.
- Wagner, H.M. (1975), *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs, NJ, chapters 17 and 18.

16. Probabilistic Linear Programming

While many probabilistic optimization problems can be formulated as stochastic dynamic programs, several special classes of probabilistic optimization problems can be formulated and solved more quickly using linear programming methods. We examine two major special cases: chance-constrained linear programming and two-stage linear programming.

Brief Review of Linear Programming

A linear program is any optimization problem that is formulated as:

$$\text{Max or Min} \quad z = \underline{c}^T \underline{X}$$

$$\text{Subject to:} \quad \mathbf{A} \underline{X} \leq \underline{b},$$

where the vector \underline{c} represents the unit cost or benefit of each unit of each decision variable in \underline{X} . The linear equations in the matrix equation $\mathbf{A} \underline{X} \leq \underline{b}$ each represent a constraint on the solution. Together, the constraints define the feasible region of solutions.

The defining aspect of linear programming is the objective function and all the constraints are linear.

Linear Programming with Probabilistic Objective Function Coefficients

Let the elements of the objective function coefficient vector c_i be probabilistic and independent, drawn from a known probability distribution, $P_i(c_i)$, and the true value of c_i is not known until after the decisions X_i are made. If expected value decision-making is to be used, the objective function becomes,

$$\begin{aligned} \text{Max or Min} \quad z &= \sum_{i=1}^n \int_{-\infty}^{\infty} P_i(c_i) c_i X_i dc_i \\ &= \sum_{i=1}^n X_i \int_{-\infty}^{\infty} P_i(c_i) c_i dc_i = \sum_{i=1}^n EV(c_i) X_i = \sum_{i=1}^n \bar{c}_i X_i, \end{aligned}$$

where \bar{c}_i is the average value of c_i .

The result is simple. If the coefficients in the objective function of a linear program are probabilistic (and not correlated) until after the decisions are made, and expected value decision-making is to be used, replace the coefficients with their expected values. (Wagner, 1975, p. 668)

Chance-Constraints

Chance-constrained linear programming considers problems where a specific reliability is desired for satisfying individual constraints in a linear program (Charnes and Cooper 1963; nicely developed by Taha 1992). Two types of uncertainty can be expressed this way, uncertainty in the coefficients in the \mathbf{A} matrix and uncertainty in the vector \underline{b} .

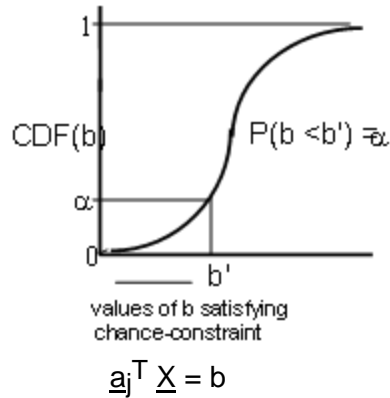
The basic form of a chance constraint is:

$$P(\underline{a}_j^T \underline{X} \geq b_j) \leq \alpha_j,$$

the probability that constraint j is violated must be less than or equal to α_j . Chance-constraint linear programming typically relies on being able to formulate a *deterministic equivalent* to this type of constraint.

Uncertainty in the RHS Coefficient b_j

Let b_j have a probability distribution, but other constants in the constraint (\underline{a}_j) are known exactly. The chance constraint $P(\underline{a}_j^T \underline{X} \geq b_j) \leq \alpha_j$, is represented in the figure below. From this figure, the deterministic equivalent of the chance constraint would be $\underline{a}_j^T \underline{X} \leq b'_j$, where b'_j is set at the α_j probability level.



Often, this is expressed as: $b'_j = EV(b_j) + K\alpha_j \sqrt{var(b_j)}$. The derivation of this form is simple for the normal distribution. The chance constraint is $P(\underline{a}_j^T \underline{X} \geq b_j) \leq \alpha_j$. Let b_j have a mean $EV(b_j)$ and a variance $var(b_j)$.

If we "normalize" the chance-constraint (subtracting the mean and dividing by the standard deviation), the chance constraint becomes:

$$P\left(\frac{\underline{a}_j^T \underline{X} - EV(b_j)}{\sqrt{var(b_j)}} = \frac{b_j - EV(b_j)}{\sqrt{var(b_j)}}\right) \leq \alpha_j.$$

This rephrases the chance-constraint in terms of the standard "z-scores" of the probability distribution. The deterministic equivalent would then be:

$$\frac{\underline{a}_j^T \underline{X} - EV(b_j)}{\sqrt{var(b_j)}} \leq K\alpha_j,$$

where $K\alpha_j$ is the z-score of the normal distribution for a probability of α_j of being exceeded. A bit of algebra transforms this into the final equivalent deterministic constraint:

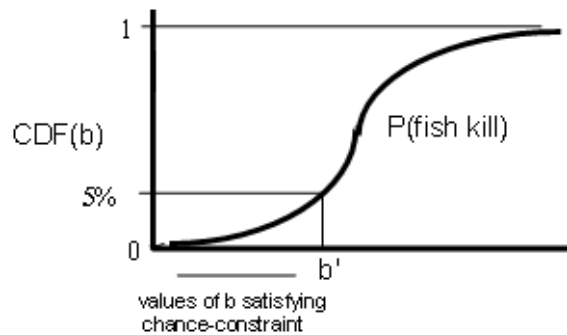
$$\underline{a}_j^T \underline{X} \leq EV(b_j) + K\alpha_j \sqrt{var(b_j)}.$$

Note that for high values of α_j , $K\alpha_j$ is negative. This deterministic equivalent can be incorporated readily into the linear program.

Chance Constraint Example: Fish kills from effluent loads

An urban stormwater and wastewater utility hopes to minimize the cost of a variety of control activities, with the objective function $\min z = \underline{c}^T \underline{X}$, where \underline{X} is the vector of control activities and \underline{c} is the vector of unit costs. There are constraints for effluent loads and flows.

The effluent load for copper is a linear function of the control measures with the form $\underline{a}_1^T \underline{X}$, where \underline{a}_1 is the vector of load reduction effectiveness. Fish in the stream are affected by copper concentration, but given uncertainties in the toxicity of copper concentrations and dilution flows entering the stream, there is a probability distribution of the effluent load which triggers a fish kill. The utility can tolerate a 5% chance of a fish kill. What should the chance-constraint look like? Consider the fish kill probability distribution below.



The fish kill chance constraint is $P(\underline{a}_1^T \underline{X} \geq b) \leq 5\%$ or $P(\underline{a}_1^T \underline{X} \leq b) \leq 95\%$. The deterministic-equivalent chance constraint would then be:

$$\underline{a}_1^T \underline{X} \leq b',$$

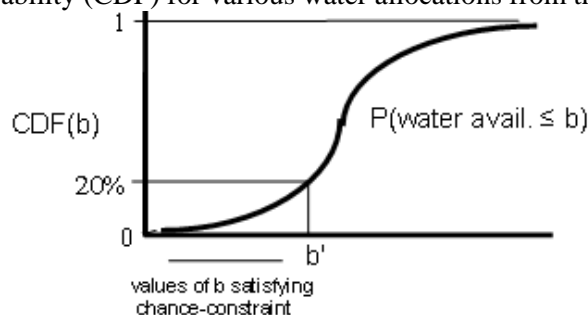
where b' is the effluent load with a 95% chance of not causing a fish kill.

Chance Constraint Example: Water availability for a cropping pattern

Consider the problem of a farmer selecting a mix of crops for her irrigated farmland. She hopes to maximize her profit from this year, with the objective function $\max z = \underline{c}^T \underline{X}$, where \underline{X} is the vector of crop acreages for various crops and \underline{c} is the vector of unit profits for each crop. Like all of us, she also has constraints, in this case limited land and water.

The farmer receives water from the Central Valley Project (CVP), which provides a variable amount of water, depending on how wet the year has been. Farm water consumption is given by $\underline{a}_1^T \underline{X}$, where \underline{a}_1 is the vector of water requirements per acre of each crop. The farmer wants to solve this optimization problem using a linear program, with only a 20% chance of the crops planted requiring more water than is available. (The cropping season requires that the planting be done before water availability is known.) What would this constraint look like?

The non-exceedence probability (CDF) for various water allocations from the CVP is given below.



The farmer's risk constraint is $P(\underline{a}_1^T \underline{X} \geq b) \leq 20\%$ or $P(\underline{a}_1^T \underline{X} \leq b) \leq 80\%$. The deterministic-equivalent chance constraint would then be:

$$\underline{a}_1^T \underline{X} \leq b'$$

where b' is the amount of water from the CVP with an 80% chance of being exceeded, or a 20% chance of not being met.

Uncertainty in the Elements of Matrix A

What if the elements in the **A** matrix have probabilistic values? Let the elements of **A**, a_{ij} , have mean values $E(a_{ij})$, variance $\text{var}(a_{ij})$, and covariance $\text{cov}(a_{ij}, a_{kl})$. Let the right-hand-side parameters b_j be known exactly. For the j -th constraint, the chance-constraint is:

$$P\left(\sum_{i=1}^n a_{ij} X_i \geq b_j\right) \leq \alpha_j.$$

Define a variable, $h_j = \sum_{i=1}^n a_{ij} X_i$. If the **A** matrix values are probabilistic, h_j should be approximately

normally distributed (by the Central Limit Theorem). The mean of h_j will be $EV(h_j) = \sum_{i=1}^n EV(a_{ij}) X_i$,

and its variance will be $\text{var}(h_j) = \underline{X}^T \mathbf{D}_j \underline{X}$, where \mathbf{D}_j is the j -th covariance matrix:

$$\mathbf{D}_j = \begin{pmatrix} \text{var}(a_{1j}) & \dots & \text{cov}(a_{1j}, a_{nj}) \\ \dots & \dots & \dots \\ \text{cov}(a_{nj}, a_{1j}) & \dots & \text{var}(a_{nj}) \end{pmatrix}$$

Given the definition of h_j , the chance-constraint becomes:

$$P(h_j \geq b_j) = P\left(\sum_{i=1}^n a_{ij} X_i \geq b_j\right) \leq \alpha_j.$$

Normalizing the chance-constraint, yields:

$$P\left(\frac{h_j - EV(h_j)}{\sqrt{\text{var}(h_j)}} \geq \frac{b_j - EV(h_j)}{\sqrt{\text{var}(h_j)}}\right) \leq \alpha_j.$$

Using the same argument used above, the deterministic equivalent to this chance-constraint is:

$$\frac{b_j - EV(h_j)}{\sqrt{\text{var}(h_j)}} \geq K\alpha_j,$$

where $K\alpha_j$ is the standard normal deviate with a probability α_j of being exceeded. Expanding this equation to eliminate reference to h_j yields:

$$b_j - \sum_{i=1}^n E(a_{ij}) X_i \geq K\alpha_j \sqrt{\underline{X}^T \mathbf{D}_j \underline{X}},$$

or

$$\sum_{i=1}^n E(a_{ij}) X_i + K\alpha_j \sqrt{\underline{X}^T \mathbf{D}_j \underline{X}} \leq b_j.$$

Unfortunately, this is a non-linear deterministic equivalent constraint and would have to be solved using a non-linear programming method, that can incorporate non-linear constraints (yuk!).

For the special case without covariance between elements of \mathbf{A} ,

$$\sum_{i=1}^n E(a_{ij})X_i + K\alpha_j \sqrt{\sum_{i=1}^n \text{var}(a_{ij}) X_i^2} \leq b_j.$$

If only one a_{ij} is uncertain, then $\text{Var}(a_{ki}) = 0, k \neq i$, and this case yields a linear constraint.

Uncertainty in Both the \mathbf{A} Matrix and \mathbf{b} Vector

If both the elements of the \mathbf{A} matrix and vector \mathbf{b} are uncertain and normally distributed, a similar approach can be taken to develop an equivalent deterministic set of constraints. The chance-constraint is:

$$P\left(\sum_{i=1}^n a_{ij}X_i - b_j \geq 0\right) \leq \alpha_j.$$

Define a variable, $h_j = \sum_{i=1}^n a_{ij}X_i - b_j$, which is approximately normally distributed. The mean of h_j will

be $EV(h_j) = \sum_{i=1}^n EV(a_{ij}) X_i - EV(b_j)$, and its variance will be $\text{var}(h_j) = \mathbf{X}^T \mathbf{D}_j \mathbf{X} + \text{var}(b_j)$, where \mathbf{D}_j is the j -th covariance matrix defined above.

$$P(h_j \geq 0) = P\left(\sum_{i=1}^n a_{ij} X_i - b_j \geq 0\right) \leq \alpha_j.$$

Normalizing the chance-constraint yields:

$$P\left(\frac{h_j - EV(h_j)}{\sqrt{\text{var}(h_j)}} \geq 0\right) \leq \alpha_j.$$

This leads to the deterministic equivalent constraint:

$$\frac{-EV(h_j)}{\sqrt{\text{var}(h_j)}} \leq K\alpha_j.$$

Substituting to remove h_j yields,

$$\sum_{i=1}^n EV(a_{ij}) X_i + K\alpha_j \sqrt{\mathbf{X}^T \mathbf{D}_j \mathbf{X} + \text{var}(b_j)} \leq EV(b_j).$$

Unfortunately, this too is a non-linear constraint.

Chance-Constraints Requiring Satisfaction of Several Constraints

Sometimes several probabilistic constraints must be satisfied for a particular system reliability to be achieved. This is the case in pipelines, where each pipe segment in a series must function for the entire pipeline to be functional. It is also the case in ship channels, where all reaches must be navigable for the channel to be useful (Ratick, et al. 1992).

This represents a limitation for conventional chance-constraint programming. These are sometimes called *joint chance-constraints* (Wagner 1975, p. 676). They must typically be solved by non-linear programming.

Example: Multi-Season Dredging of a Single Reach (Ratick, et al. 1992)

Consider the problem of dredging a single reach over several periods where sedimentation in the reach is probabilistic. If $C_t(V_t)$ is the cost of dredging in each period, the objective function becomes:

$$\text{MIN } z = \sum_{t=0}^n C_t(V_t).$$

The constraints would be:

$$P\left(\sum_{\tau=0}^t V_{\tau} \geq S_t\right) \geq \alpha, \text{ for all } t,$$

where α is the reliability of the channel for each time period and S_t is the probabilistic amount of sediment that accumulates in the reach from time $\tau=1$ until time t .

If the amount of sedimentation in each period s_1 is log-normally distributed and independent in time, how would chance constraints be formulated for each period? Let $EV(s_1) = 100,000 \text{ yd}^3$ and $SD(s_1) = 120,000 \text{ yd}^3$. Here's the lognormal distribution:

$$f(x) = \frac{1}{(\sqrt{2\pi})x\sigma_{\ln x}} \exp\left(-0.5 \left(\frac{\ln x - \mu_{\ln x}}{\sigma_{\ln x}}\right)^2\right).$$

To transform these parameter values to the ln-scale for the log-normal distribution, we use:

$$\sigma_{\ln x}^2 = \ln\left(\frac{\sigma_x^2}{\mu_x^2} + 1\right) \text{ and } \mu_{\ln x} = \ln(\mu_x) - (\sigma_{\ln x}^2)/2, \text{ so } \sigma_{\ln x}^2 = 0.892 \text{ and } \mu_{\ln x} = 11.067.$$

For the first time period, $t=1$, the deterministic equivalent to the chance-constraint is easy. If we want only a 5% chance of sedimentation exceeding dredging (95% chance of $V_0 + V_1 \geq S_1$), the z-score from the normal distribution is +1.645. Thus the deterministic equivalent sediment deposition would be:

$$\frac{\ln(S_{D1}) - \mu_{\ln x}}{\sigma_{\ln x}} = +1.645 \text{ or } \ln(S_{D1}) = \mu_{\ln x} + 1.645 \sigma_{\ln x} = 12.621, \text{ or } S_{D1} = 302,740.$$

The deterministic equivalent constraint for the first period then becomes:

$$V_0 + V_1 \geq 302,740.$$

For later periods, the probability distributions of accumulated deposition will not be lognormal, but can be found using a Markov process. The resulting probability distribution can then be integrated to find the amount of accumulated sediment that represents the 95% extreme. This amount would then appear as the right-hand-side constant in the equivalent deterministic constraint for each subsequent period.

Joint Chance-Constraints Example

Consider the dredging problem above, but with two reaches. Assume the unlikely case where sedimentation rates in each reach are independent events. The objective function becomes:

$$\text{MIN } z = \sum_{t=0}^n \sum_{i=1}^2 C_{ti}(V_{ti}),$$

and the constraints become:

$$P\left(\sum_{\tau=0}^t V_{\tau 1} \geq S_{t1} \text{ and } \sum_{\tau=0}^t V_{\tau 2} \geq S_{t2}\right) \geq \alpha, \text{ for each } t,$$

where α is the probability that both reaches are navigable for each time period (i.e., the channel reliability). Both reaches must be navigable for the navigation channel to work.

Assuming that S_{t1} and S_{t2} are independent events, let's break this constraint into pieces closer to a linear

program.

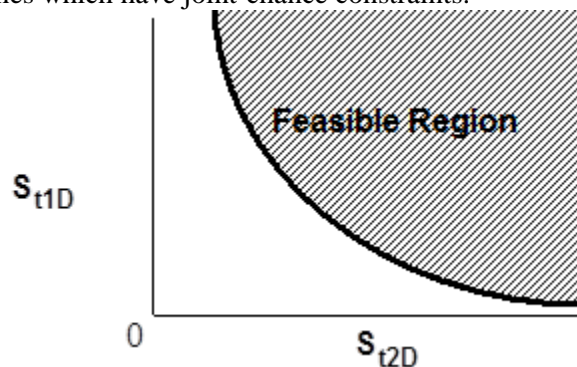
$$(1) \quad \sum_{\tau=0}^t V_{\tau 1} \geq S_{t1D}, \quad (2) \quad \sum_{\tau=0}^t V_{\tau 2} \geq S_{t2D},$$

and

$$(3) \quad P(S_{t1} \leq S_{t1D}) P(S_{t2} \leq S_{t2D}) \geq \alpha.$$

Unfortunately, this third constraint is not linear. Here's one way to get around this, but it requires solving several linear programs.

To satisfy constraint 3, only some values of S_{t1D} and S_{t2D} are feasible, as depicted below. By solving several linear programs for the combinations of values that just satisfy the edge of this feasible region, and selecting the best of these solutions, the least-cost chance-constrained dredging schedule can be found. Unfortunately, the number of linear programs which needs to be run increases greatly with the number of periods and reaches which have joint-chance constraints.



References

Charnes, A. and W.W. Cooper (1963), "Deterministic Equivalents for Optimizing and Satisfying Chance-Constraints," *Operations Research*, Vol. 11, pp. 18-39.

Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ, chapter 7 in particular.

Mays, L.W. and Y-K Tung (1992), *Hydrosystems Engineering and Management*, McGraw-Hill, N.Y.

Ratick, S.J., W. Du, and D.A. Moser (1992), "Development of a Reliability Based Dynamic Dredging Decision Model," *European Journal of Operational Research*, Vol. 58, pp. 318-334.

Taha, Hamdy A. (1992), *Operations Research*, Macmillan Publishing Co., N.Y., pp. 770-773.

Wagner, H.M. (1975), *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs, NJ, chapters 16, 17, and 18.

Look up:

Watanabe, T., and Ellis, H., A joint chance-constrained programming model with row dependence, *European Journal of Operational Research*, 77, pp. 325-343, 1994.

Morgan, D.R., J.W. Eheart, and A.J. Valocchi (1993). "Aquifer Remediation Design Under Uncertainty Using a New Chance Constrained Programming Technique," *Water Resources Research*, 29, (3), pp. 551-561.

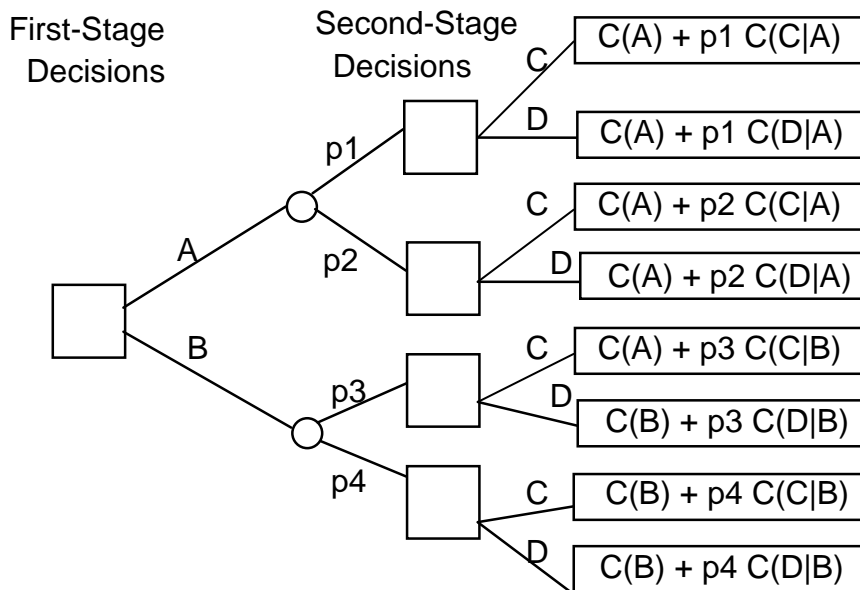
Kataoka, Shinji, "A Stochastic Programming Model," *Econometrica*, 31, pp 181-196, 1963.

17. Two-Stage Linear Programming

Two-stage and multi-stage linear programming is the second class of probabilistic linear programming methods that can represent a wide range of problems and often can be solved with relative ease (Birge and Louveaux 1997). Multi-stage formulations of linear programming problems often allow for better and more decision-oriented analysis of uncertainties (Higle and Wallace 2003).

Two-Stage Problems

Many decision-making problems occur in two stages. Some decision must be made, then something happens involving chance, and a follow-up decision/response is made which depends on the chance event and the first-stage decision. A decision-tree for this situation appears below.



The trick is to make the best first-stage decision, considering the best responses and recourses for a variety of later chance events. These problems are similar, if not identical, to the problems we have solved using stochastic dynamic programming.

Two-stage linear programming is especially attractive when the decisions have continuous values, as opposed to the discrete values shown in the decision-tree. Where the problem can be formulated as a linear program, the linear programming solution sometimes has great computational advantages. Linear programming solution also provides much more sensitivity analysis information and is less subject to discretization problems than dynamic programming solutions.

Two-Stage Linear Programming Example: Floodplain Planning Decisions and Economic Values

Two-stage optimization problems often can be formulated as a linear program, particularly where expected value is to be used to select an optimum. Consider the following two-stage problem.

For floodplain management, the overall objective is to minimize the sum of expected annual damages and annualized expected flood response costs (Lund 2002). Increases in permanent and emergency flood responses will reduce flood damages, but there will come a point where additional flood mitigation actions are no longer economical. The expected annualized cost equation for a combination of implemented floodplain options and resulting flood event damages is:

$$(1) \quad \text{Min } z_2 = \sum_{i=1}^m c_{P_i} X_{P_i} + \sum_{s=1}^q p_s \left(\sum_{j=1}^n c_{E_{js}} X_{E_{js}} + D_s \right)$$

where c_{P_i} and X_{P_i} are the unit cost and level of implementation of permanent response i (first-stage decisions), $c_{E_{js}}$ and $X_{E_{js}}$ are the unit cost and level of implementation of emergency response j in response to flood event s (second-stage decisions), D_s is the damage resulting from a managed event size (flow or stage) s . Given a set of structural flood control options (implemented outside of the floodplain damage area) that provide a probability distribution of event sizes, p_s , the floodplain management objective should be to minimize expected annual damages plus expected annual flood control costs. The minimization of total expected cost is limited by several types of constraints.

Flood damage calculation. Overall damage reduction is assumed to be the sum of incremental damage reductions arising from implementation of management options, as in Equation 2. Floodplain damage for each flood event (D_s) is calculated to prevent the impossible situation of reductions in damages exceeding base-case damages, d_s . The constraints distinguish calculated damages for each event (CD_s) from corrected damages (D_s) that cannot be negative.

$$(2) \quad d_s - \sum_{i=1}^m b_{P_{is}} X_{P_i} - \sum_{j=1}^n b_{E_{js}} X_{E_{js}} = CD_s, \forall s$$

$$(3) \quad D_s \geq CD_s, \forall s$$

$$(4) \quad D_s \geq 0, \forall s$$

Limits on option implementation. Limits exist for each decision variable (the X 's), representing limits of implementation for each option. Usually,

$$(5) \quad X_{P_i} \leq 1, \forall i, \quad X_{E_{js}} \leq 1, \forall j, s.$$

Some options can only be implemented either completely or not at all, or in integer values. For example, floodplain evacuation typically is implemented in discrete stages. This can be accommodated using a limited number of integer variables in the formulation, Integer X_{P_i} .

Interaction of implemented options. To make the formulation more realistic, the interaction of floodplain management options can be specified in constraints, allowing representation of some forms of non-additive costs and damage reduction. This has been done for some similar water-supply applications (Lund 1995). Cases where implementing a permanent option i precludes implementing an emergency option j during event s can be represented by the following constraint,

$$(6) \quad X_{P_i} + X_{E_{js}} \leq 1, \forall j, s \text{ precluded by implementing } i.$$

Such a case might be where a particular land use control precludes the relevance or benefits of evacuation measures. Conversely, where a particular emergency option j for an event s requires prior establishment of a particular permanent option i , the following constraint can be used,

$$(7) \quad X_{P_i} - X_{E_{js}} \geq 0, \forall j, s \text{ requiring implementation of } i.$$

This might be the case where the closure of a floodwall gate (an emergency option) would require prior construction of a floodwall (a permanent option). In such a case, the permanent option might confer no flood damage reduction benefits without implementation of the emergency option during flooding episodes.

Multi-Stage Linear Programming

The proposed approach has several limitations. First, the implementation of emergency options presumes very good (ideally perfect) forecasting of flood events. Emergency flood-response options are not implemented futilely or without necessity because the stage forecast for each flooding episode is assumed to be perfect. Significantly imperfect flood forecasts might be represented by conditional probabilities and a third stage might be added to the linear program. Adding a third stage allows contingencies for

outcomes in the second state, such as imperfect forecasts. The resulting linearized objective function is

$$(8) \quad z_3 = \sum_{i=1}^m c_{P_i} X_{P_i} + \sum_{s=1}^q p_s \left[\sum_{j=1}^n c_{E_{js}} X_{E_{js}} + \sum_{s_2=1}^q p(s_2 | s) D_{s_2|s} \right],$$

and primary constraints

$$(9) \quad D_{s_2|s} = d_{s_2} - \bar{b}_{P_{s_2}}^T \bar{X}_P - \bar{b}_{E_{s_2}}^T \bar{X}_{E_s}, \forall s, s_2$$

where s is the forecast flood stage, s_2 is the actual flood stage, and $p(s_2|s)$ is the probability of flood stage s_2 occurring if flood stage s is forecast. Damage reduction parameters are also modified for the occurrence of the actual flood stage s_2 , rather than the forecast flood stage s .

Expansion to a three-stage linear program greatly expands the computational and calibration effort needed for the model, requiring solution for $m + qn + q^2$ decision variables and at least q^2 constraints and calibration for at least $2m + q(1+2n) + q^2$ parameters. Calibration of the additional parameters for a three-stage model would require estimates of conditional flow probabilities for flood forecasts, perhaps not an easy estimation.

Second, reductions in damage arising from implementation of permanent and emergency options are assumed to be additive. This will often not be strictly the case. Constraint Equations 2, 3, and 4 provide this additive model, limited to positive damages. Some interactions of options, restricting additivity, can be handled as described above in Equations 6 and 7.

Multi-stage optimization quickly becomes ugly and computationally burdensome for most problems. But it might be useful for some problems. Hollinshead found 3-stage linear programming useful for identifying optimal water purchase strategies for California's Environmental Water Account (Hollinshead 2005, 2006). PG&E use a three-stage LP for long-term hydropower scheduling.

Representing Multiple Uncertainties in the Second Stage with Scenarios

At first glance, it appears that multi-stage linear programming can work with only one uncertain variable. (In the above case, only uncertain flood flow was considered, along with perfectly correlated response costs.) However, if we consider each event-probability combination s to represent a scenario (consisting of a joint occurrence of many parameter values), then discrete representations of joint events can be incorporated easily into a two-stage formulation. In this example, emergency response costs, flood levels, and damage potential also can be considered as uncertain. Often these parameters are not only uncertain, but also correlated; for example severe flood events might correlate with higher response costs. By defining a set of scenarios s over which all these parameters vary in a statistically representative way, the two-stage LP can be expanded to include all these uncertainties, and their correlations. (Of course, this means there are more parameters and joint probabilities to calibrate.)

Where the different variables are imperfectly correlated, many more scenarios must be generated to cover the range of likely combinations (m becomes larger) and the joint-probabilities must be estimated. So while multiple uncertainties can often be represented as probability weighted discrete joint-event scenarios, where correlations are imperfect, the problem can become too large to solve. Conceivably, scenarios could be generated by a Monte Carlo method to create a wide range of equally-probable scenarios, but this can easily create too many scenarios to be solved (m too large).

Independent Continuous Uncertainties in Multiple Parameters

Where each parameter in the formulation is uncertain and independent of other parameter values, it is often possible to separately solve for each second-stage decision with a resulting convex penalty function which can be incorporated into solution for the first-stage decisions (Birge & Louveaux 1997, pp.11-15). This separation of the problem into sub-problems can reduce the overall computation burden.

Decomposition of Multi-Stage Problems

More generally, depending on the structure of the multiple-stage decision problem, the second stage decisions are often only loosely related to the first-stage decisions, say by only a few decision variables. In such cases, the solution of larger multiple-stage problems can often be accomplished using decomposition techniques. Benders decomposition, the next topic, is one such technique.

Examples

Multi-stage linear programming and related techniques have been applied to a wide variety of problems. Birge and Louveaux (1997) offer a wide range of examples and a much more in-depth treatment of solution methods and mathematical interpretations. Wilchfort (1997), Garcia (2002, 2006), and Rosenberg (2007, 2008, 2009) apply this technique to water demand management and Lund and Israel (1995) apply it to developing water market portfolios. Sahinidis (2004) nicely reviews optimization with uncertainty.

References

- Birge, John R. and François Louveaux (1997), *Introduction to Stochastic Programming*, Springer, N.Y., NY, 421 pp.
- Garcia, Roberto (2002), "Derived Willingness-to-Pay for Water: Effects of Probabilistic Rationing and Price," MS Report, Department of Civil and Environmental Engineering, University of California, Davis. Available at: <http://cee.engr.ucdavis.edu/faculty/lund/>
- Garcia Alcubilla, R. and J.R. Lund (2006), "Derived Willingness-To-Pay for Household Water Use with Price and Probabilistic Supply," *Journal of Water Resources Planning and Management*, Vol. 132, No. 6, pp. 424-433, November/December.
- Higle, J.L. and S.W. Wallace (2003), "Sensitivity analysis and uncertainty in linear programming," *Interfaces*, Vol. 33, No. 4, pp. 53-60.
- Hollinshead, S.P. (2005), "Optimization of Environmental Water Account Purchases with Uncertainty," MS Thesis, Department of Civil and Environmental Engineering, University of California, Davis. Available at: <http://cee.engr.ucdavis.edu/faculty/lund/>
- Hollinshead, S.P. and J.R. Lund (2006), "Optimization of Environmental Water Account Purchases with Uncertainty," *Water Resources Research*, Vol. 42, No. 8, W08403, August.
- Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ.
- Lund, J.R. (2002), "Floodplain Planning with Risk-based Optimization," *Journal of Water Resources Planning and Management*, Vol. 127, No. 3, May.
- Lund, J.R. and M. Israel (1995), "Optimization of Water Transfers for Urban Water Supply," *Journal of Water Resources Planning and Management*, ASCE, Vol. 121, No. 1, January/February.
- Rosenberg, D.E. and J.R. Lund (2009), "Modeling Integrated Decisions for a Municipal Water System with Recourse and Uncertainties: Amman, Jordan," *Water Resources Management*, Volume 23, Number 1, pp. 85-115, January.
- David Rosenberg, Richard Howitt, and Jay Lund (2008), "Water Management with Water Conservation, Infrastructure Expansions, and Source Variability in Jordan," *Water Resources Research*, Vol. 44, doi:10.1029/2007WR006519.
- Rosenberg, D.E., T. Tarawneh, R. Abdel-Khaleq, and J.R. Lund (2007), "Modeling Integrated Water-User Decisions in Intermittent Supply Systems," *Water Resources Research*, Vol. 43, No. 7, July.
- Sahinidis, N.V. (2004), "Optimization under uncertainty: state-of-the-art and opportunities," *Computers and Chemical Engineering*, Vol. 28, pp. 971-983.
- Wagner, H.M. (1975), *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs, NJ.
- Wilchfort, O. and J.R. Lund (1997), "Shortage Management Modeling for Urban Water Supply Systems," *Journal of Water Resources Planning and Management*, ASCE, Vol. 123, No. 4, July/August.
- Look up:** Alborno, V. M., Benario, Pablo & Rojas, Manuel E. (2004) "A two-stage stochastic integer programming model for a thermal power system expansion." *International Transactions in Operational Research* **11** (3), 243-257.

18. Benders Decomposition (skip this year)

Benders Decomposition is an optimization method that solves a large, difficult problem by breaking it down into smaller, simpler problems, and solving and modifying these smaller problems iteratively until an optimal solution is found. Was first developed by Benders (1962) and generalized by Geoffrion (1972). Consider the general and difficult Problem 1 with two vectors of decision variables \underline{x} and \underline{y} . For this general version, both the objective function $f()$ and $\underline{G}()$ can be non-linear.

Problem 1:

Maximize $f(\underline{x}, \underline{y})$ subject to $\underline{G}(\underline{x}, \underline{y}) \geq 0$ $\underline{x}, \underline{y}$
--

This problem can often be simplified by solving only for the x vector of decision variables, for a fixed set of y values. This is Problem (1- y).

Problem (1- y):

$v(\underline{y}) =$ Maximize $f(\underline{x}, \underline{y})$ subject to $G(\underline{x}, \underline{y}) \geq 0$ \underline{x}
--

Benders and Geoffrion found that Problem 1 could be restated into the following equivalent "master problem."

Benders Equivalent Master Problem 1:

Maximize [infimum[\underline{y} maximum[$\underline{u} \geq 0$ maximum[\underline{x} $f(\underline{x}, \underline{y}) + \underline{u}^t G(\underline{x}, \underline{y})$]]]] Subject to: Maximum[\underline{x} $\underline{\lambda}^t G(\underline{x}, \underline{y}) \geq 0$ for all $\underline{\lambda} \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$] [infimum = greatest lower bound]

This "master problem" can also be re-defined as:

Benders Equivalent Master Problem 2:

Maximize y_0 \underline{y} Subject to: $y_0 \leq$ Maximum[\underline{x} $f(\underline{x}, \underline{y}) + \underline{u}^t G(\underline{x}, \underline{y})$ for all $\underline{u} \geq 0$] \underline{x} Maximum [$\underline{\lambda}^t G(\underline{x}, \underline{y}) \geq 0$ for all $\underline{\lambda} \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$] \underline{x}
--

Solution Method: (Geoffrion 1972)

Step 1: Fix a value of $\underline{y} = \underline{y}'$. Solve the subproblem (1- \underline{y}') and use this solution $(\underline{x}^*, \underline{y}')$ and the multiplier vector \underline{u}' ($\underline{u}' \geq 0$) to obtain the function $L^*(\underline{y}'; \underline{u}')$. Put $p=1$, $q=0$, $\underline{u}^1 = \underline{u}'$, and $LBD = v(\underline{y}')$. LBD is the *lower bound* of the optimal value of Problem A. Select a convergence tolerance parameter $\epsilon > 0$.

$$L^*(\underline{y}'; \underline{u}') = \underset{\underline{x}}{\text{maximum}} [f(\underline{x}, \underline{y}') + \underline{u}'^t G(\underline{x}, \underline{y}')] = f(\underline{x}^*, \underline{y}') + \underline{u}'^t G(\underline{x}^*, \underline{y}')$$

Step 2: Solve the relaxed master problem, relaxed in the sense that not all constraints are included.

$$\underset{\underline{y}, z}{\text{maximize}} \quad z \quad \text{subject to:} \quad z \leq L^*(\underline{y}; \underline{u}^j), j=1, \dots, p,$$

$$L^*(\underline{y}; \underline{\lambda}^j) \geq 0, j=1, \dots, q.$$

Let the optimal solution be (\underline{y}^*, z^*) . The resulting value of z^* is an *upper bound* on the optimal solution of Problem A. If $LBD \geq z^* - \epsilon$, STOP.

Step 3: Solve the revised problem (A- \underline{y}). One of the following cases must occur:

Step 3A: *The quantity $v(\underline{y})$ is finite.* If $v(\underline{y}) \geq z - \varepsilon$, STOP. Else, find the optimal multiplier vector \underline{u} and the function $L^*(\underline{y}; \underline{u})$. Increase p by 1 and put $\underline{u}^p = \underline{u}$. If $v(\underline{y}) \geq \text{LBD}$, put $\text{LBD} = v(\underline{y})$. GO TO Step 2.

Step 3B: *Problem (1- \underline{y}) is infeasible.* Find $\underline{\lambda}$ satisfying $\text{Maximum}_{\underline{x}} [\underline{\lambda}^t G(\underline{x}, \underline{y})] < 0$ and the function

$L^*(\underline{y}; \underline{\lambda})$. Increase q by 1 and put $\underline{\lambda}^q = \underline{\lambda}$. GO TO Step 2. [What is Lamda here, really? These Lamda represent which constraints force the problem to be infeasible, and, for linear programs, could be found from identifying constraints associated with "artificial variables" remaining in the basis.]

$$L^*(\underline{y}; \underline{\lambda}) = \text{Maximum}_{\underline{x}} [\underline{\lambda}^t G(\underline{x}, \underline{y})], \underline{\lambda} \geq 0.$$

Steps 1 and 3 provide lower bound estimates of the optimal value of the objective function by solving the Problem (1- \underline{y}). The lower bound increases with refined values of \underline{y} . Solutions to Problem (1- \underline{y}) include all constraints, thereby testing the feasibility of proposed values of \underline{y} , and defining the optimal solution of the \underline{x} decision variables, given \underline{y} .

Step 2 solves a "relaxed" version of the problem for \underline{y} , assuming away some constraints and decisions regarding the \underline{x} decisions. As constraints are found to be important, via infeasibilities in Step 3, they are included in this problem. The solution to this problem refines the value of the \underline{y} decision variables and provides an upper bound to the solution.

When the upper and lower bounds get close, within ε , STOP.

References:

- Basson, M.S., R.B. Allen, G.G.S. Pegram, and J.A. van Rooyen (1994), *Probabilistic Management of Water Resource and Hydropower Systems*, Water Resources Publications, Highlands Ranch, CO, pp. 242-248.
- Benders, J.F. (1962), "Partitioning procedures for solving mixed-variables program problems," *Numerische Mathematik*, Vol. 4, pp. 238-252.
- Geoffrion, A.M. (1972), "Generalized Benders Decomposition," *Journal of Optimization Theory and Applications*, Vol. 10, No. 4, pp. 237-260.
- McCusker, S.A., B.F. Hobbs, and Y. Ji (2002), "Distributed Utility Planning Using Probabilistic Production Costing and Generalized Benders Decomposition," *IEEE Transactions On Power Systems*, Vol. 17, No. 2, May pp. 497-505.
- Susan A., Benjamin F. Hobbs, Senior Member, IEEE, and Yuandong Ji, Jacobs, J., G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus, and J. Stedinger (1995), "A Stochastic Multi-Reservoir Hydroelectric Scheduling Model," in M. Domenica, *Integrated Water Resources Planning for the 21st Century*, ASCE, N.Y., pp. 828-831.
- Jacobs, J., G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus, and J. Stedinger (1995), "SOCRAATES: A system for scheduling hydroelectric generation under uncertainty," *Annals of Operations Research*, Vol 59, pp. 99-133.
- Periera and Pinto (1985), ... *Water Resources Research*, p. 790 has a nice appendix on Benders decomposition and a nice application to hydropower application.
- Ziari, H.A. and B.A. McCarl (1995), "A nonlinear mixed integer program model for evaluating runoff impoundments for supplemental irrigation," *Water Resources Research*, Vol. 31, No. 6, June, pp. 1585-1594. [nice references to Benders' codes]

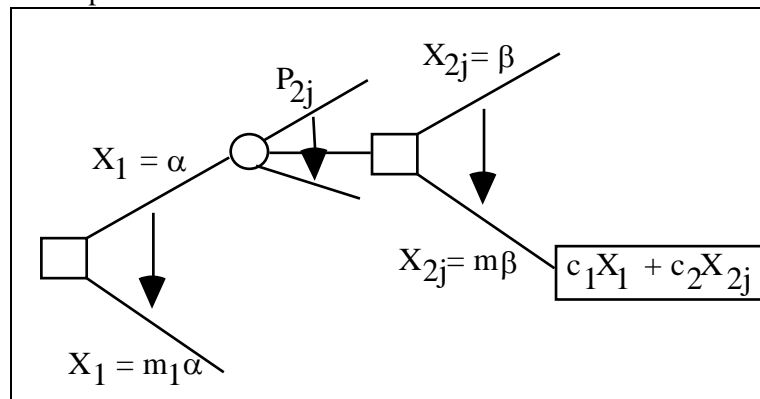
Example: Benders' Decomposition Applied to a Two-Stage Decision Problem

Consider the two-stage linear programming problem below. (Underlined terms are vectors; bold terms are matrices.)

Problem 1:

$$\begin{aligned} & \text{Maximize } z = c_1 X_1 + \sum_{j=1}^m p_{2j}(c_2 X_{2j}) \\ & \underline{X}_1 \quad \underline{X}_2 \\ & \text{Subject to:} \\ & \underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2 - \underline{b} \geq 0 \\ & X_1 \geq 0 \\ & \underline{X}_2 \geq 0 \end{aligned}$$

This problem can be represented in the decision tree below:



Problem 1 can be simplified for a given value of X_1 to:

Problem (1- X_1):

$$\begin{aligned} & \text{Maximize } z_1 = c_1 X_1 + \sum_{j=1}^m p_{2j}(c_2 X_{2j}) \\ & \underline{X}_2 \\ & \text{Subject to:} \\ & X_1 = X_{1i} \\ & \underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2 - \underline{b} \geq 0 \\ & X_1 \geq 0 \\ & \underline{X}_2 \geq 0 \end{aligned}$$

where X_{1i} is some feasible value for X_1 .

Solution by Benders Decomposition:

Step 1: Solve Problem (1- X_1) by linear programming for some value of $X_1 = X_{1i}$. For the first run, let $i = 1$. This provides both the optimal solution $\underline{X}_2 = \underline{X}_2^*$, given X_{1i} , and the Lagrange multipliers for the constraints, \underline{u}^i . From this,

$$L^*(X_{1i}; \underline{u}^i) = c_1 X_1 + \sum_{j=1}^m p_{2j}(c_2 X_{2j}^*) + \underline{u}^i (\underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2^* - \underline{b}).$$

Set $p = 1$, $q = 0$, and $\underline{u}^1 = \underline{u}^i$. $LBD = z_1(X_{1i})$. Set error tolerance $\epsilon > 0$.

Step 2: Solve the relaxed master problem for X_1 :

Maximize z_2
 X_1, z_2

Subject to:

$$z_2 \leq L^*(X_1; \underline{u}^i) = \underset{X_1}{\text{maximum}}(c_1 X_1 + \sum_{j=1}^m p_{2j}(c_2 X_{2j}^*) + \underline{u}^i (\underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2^* - \underline{b})), j = 1, \dots, p,$$

$$L^*(X_1; \underline{\lambda}^j) \geq 0, j=1, \dots, q.$$

Let $X_{1,i+1}^* = X_{1,i+1}$. If $LBD \geq z_2 - \epsilon$, STOP.

Step 3: Solve the revised sub-problem (1- X_1) for $X_1 = X_{1,i+1}$.

Step 3A: The revised sub-problem (1- X_1) is feasible. If $z_1(X_{1,i+1}) \geq z_2 - \epsilon$, STOP.

Otherwise, let $i = i+1$. This provides both the optimal solution $\underline{X}_2 = X_{2,i}^*$, given $X_{1,i}$, and the Lagrange multipliers for the constraints, \underline{u}^i . From this,

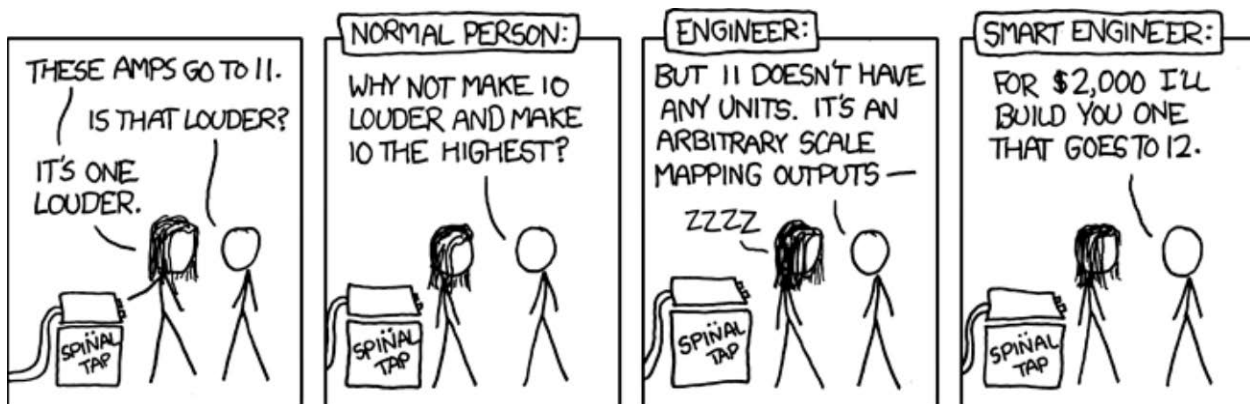
$$L^*(X_{1,i}; \underline{u}^i) = c_1 X_1 + \sum_{j=1}^m p_{2j}(c_2 X_{2j}^*) + \underline{u}^i (\underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2 - \underline{b}).$$

Set $p = p+1$, $q = 0$, and $\underline{u}^p = \underline{u}^i$. If $z_1(X_{1,i}) \geq z_2 - \epsilon$, then $LBD = z_1(X_{1,i})$. GP TO Step 2.

Step 3B: The revised sub-problem (1- X_1) is infeasible. Find $\underline{\lambda}^q$ satisfying $\text{Maximum}_{\underline{X}} [\underline{\lambda}^t (\underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2 - \underline{b})] < 0$ and the function $L^*(X_1; \underline{\lambda}^q)$. Increase q by 1 and put $\underline{\lambda}^q = \underline{\lambda}^q$. GO TO Step 2. [What is Lamda here, really?]

$$L^*(X_1; \underline{\lambda}^j) = \text{Maximum}_{\underline{X}} [\underline{\lambda}^t (\underline{A}_1 X_1 + \underline{A}_2 \underline{X}_2 - \underline{b})], \underline{\lambda} \geq 0.$$

Note: As commercial linear programming solvers have become faster, several applications of Benders' decomposition for multi-stage problems have found that direct linear programming solution has been easier and sufficiently rapid. The PG&E application for hydropower operations, SOCRATES (Jacobs, et al. 1995) is one example.



19. Markov Decision Theory

[This section needs improvement.]

Markov Decision Process – Steady-State Linear Program

Consider the problem of levee planning for a floodplain. The problem also can be optimized as a Markov Decision Process (Olsen, et al. 2000). This formulation is solved by linear programming (or can be solved by SDP). The Markov decision process optimization minimizes the expected cost of the system being in various conditions (flooded, not flooded, flooded with a high levee, etc.). For a steady-state condition of flood probabilities and net benefits (i.e., no change in climate or economic conditions) the formulation below applies.

$$\text{Max} \sum_i \sum_j \sum_k b_{ijk} X_{ijk} \quad (1)$$

Subject to:

$$\sum_i \sum_j \sum_k X_{ijk} = 1 \quad (2)$$

$$X_{ijk} \geq 0, \forall i, j, k \quad (3)$$

$$\left(\sum_j \sum_k X_{ijk} \right) P(j|i) = \sum_k X_{ijk}, \forall i, j \quad (\text{Prob. of flood state } j \text{ given initial state } i) \quad (4)$$

(Equation 4 reduces to $P_i P(j|i) = P_{ij}$)

$$\sum_j \sum_k X_{ijk} = \sum_k \sum_j X_{kji}, \forall i \quad (\text{steady-state condition; reduces to } P_i = P_i) \quad (5)$$

In this formulation, the decision variables X_{ijk} are the probability of the system being in a particular state ijk , where i is the levee state, j is the flooding state given levee state i , and k is the levee state in the next decision period (differences in i and k implying a decision). The levee state index can include levee location, levee height, flood storage, and other options. The constants b_{ijk} are the net benefits of being in state ijk (including the costs of actions taking the state from i to k , maintenance costs and land-use benefits from being in state i , and the costs of flooding that occurs in state i). $P(j|i)$ is the probability of flooding state j given previous decision i ; this is derived from the flood frequency. If climate or economic conditions are non-stationary, then an additional state variable (t for time) is added to the decision variable X , and probabilities $P(j|i)$ and net benefits b have an added time subscript, respectively (Olsen et al. 2000).

The objective function (Equation 1) is to maximize the expected value of net benefits from the system (averaging over all conditions ijk). This objective is limited by several constraints. The first constraint, Equation 2, insists that the probabilities over all states sum to one, a standard requirement of probability. The second constraint, Equation 3, insists that there be no negative probabilities, also a standard requirement of probabilities. Equation 4 represents the transition probabilities of the systems, essentially the probability of flooding given various previous states of the system. Equation 4 defines the flood state i , given the initial state j . Verbally, Equation 4

states that the probability of being in state i times the probability of flooding state j in state i must equal the probability of flooding state j in state i , no matter what state k you go to from there. Equation 5 links the levee state and levee state decision, ensuring that for steady-state conditions the probability of beginning in state i is the same as the probability of ending in state i .

The number of decision variables in this problem is given by the (number of levee locations and heights i)*(number of levee locations and heights in the next time period k)*(number of flood states j). Since there are generally the same number of levee locations and heights in each time period, say n and generally two flooding states (yes and no), the number of decision variables is $2n^2$. The number of constraints in this problem would be $1 + 3n + 2n^2$. For application with 10 levee locations and 10 levee heights ($n=100$), this would imply 20,000 decision variables and 20,301 constraints, a large but not unsolvable linear program.

Markov Decision Process – LP for Changing Climate and Economic Conditions

With climate change and/or changes in floodplain net benefits, the above linear programming formulation can be expanded to become dynamic.

$$Max \sum_t \sum_i \sum_j \sum_k b_{ijkt} e^{-rt} X_{ijkt} \quad (6)$$

Subject to:

$$\sum_i \sum_j \sum_k X_{ijkt} = 1, \forall t \quad (7)$$

$$X_{ijkt} \geq 0, \forall i, j, k, t \quad (8)$$

$$\left(\sum_j \sum_k X_{ijkt} \right) P_t(j|i) = \sum_k X_{ijkt}, \forall i, j, t \quad (9)$$

$$\sum_j \sum_i X_{ijkt} = \sum_j \sum_i X_{kjit+1}, \forall k, t \quad (10)$$

The formulation above is similar to the previous steady-state linear program (Equations 1-5), with the addition of time subscripts on parameters and decision variables. Replacing the steady-state constraint 5 in the previous formulation, constraint 10 ensures that the probability of ending state k in time t is the same as the probability of state k at the beginning of time $t+1$, for each state k .

A similar formulation is:

$$Max \sum_t \sum_i \sum_k B_{ikt} e^{-rt} X_{ikt} \quad (11)$$

Subject to:

$$\sum_i \sum_k X_{ikt} = 1, \forall t \quad (12)$$

$$X_{ikt} \geq 0, \forall i, k, t \quad (13)$$

$$\sum_i X_{ikt} = \sum_i X_{kit+1}, \forall k, t \quad (14)$$

where $B_{ikt} = (LB_{it} - BC_{ikt}) - FD_{ikt}(P_f|i,k,t)$, and LB_{it} = land benefits, BC_{ikt} = levee building costs, and FD_{ikt} = flood damage. In this formulation, X_{ikt} is the decision to move from state i to k at time t .

This problem becomes somewhat more computationally demanding to solve with $2n^2t$ decision variables and $t + 3nt + 2n^2t$ constraints. For 3 time periods and 100 levee states ($n=100$) and 2 flooding states, this problem has 60,000 decision variables and 60,903 constraints, rather on the largish size given that linear program solution times increase roughly with the square of the number of constraints and decision variables. These numbers can probably be reduced considerably if many state transitions are precluded, such as reducing levee heights.

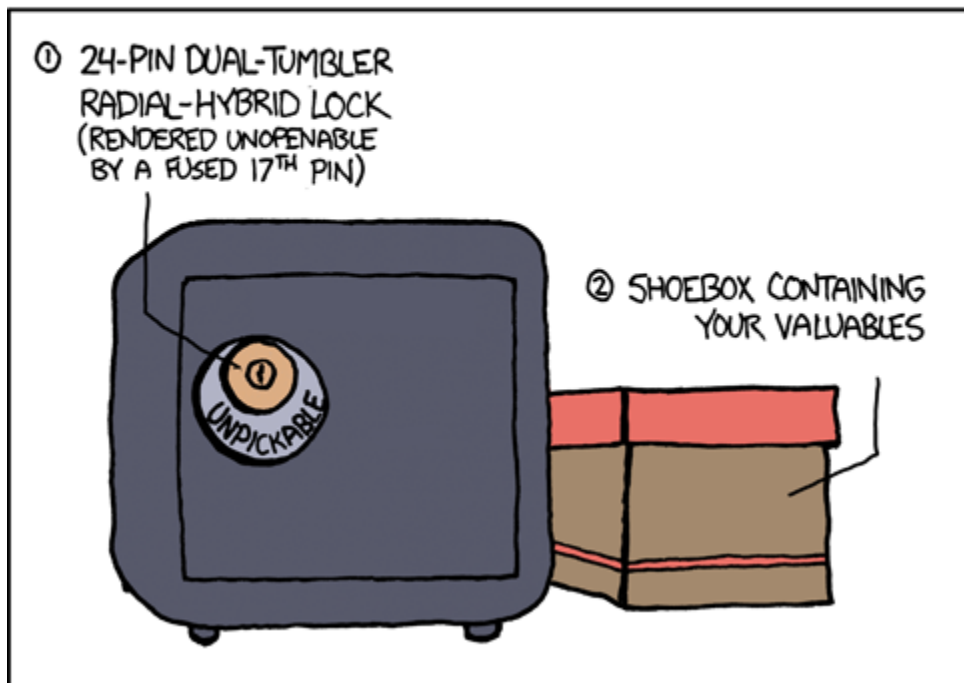
Both solution methods (stochastic dynamic programming and linear programming) allow the probability of flooding and floodplain benefits to change with time, allowing the examination of optimal flood control decisions with various forms of climate and economic change. The two solution methods are not difficult. From the standpoint of computational requirements, it appears that the stochastic dynamic programming approach might allow a larger number of time periods, providing a smoother representation of climate and economic change.

Hillier, F.S. and G.J. Lieberman (1995), *Introduction to Operations Research*, McGraw-Hill, N.Y., Chapter 19.

Olsen, J.R., Beling, P.A. and, Lambert, J.H. (2000), "Dynamic Models for Floodplain Management," *Journal of Water Resources Planning and Management*, Volume 126, No. 3, May/June, pp. 167-175.

Berenguer, C., Chu, C., and Grall, A. Inspection and Maintenance Planning: An Application of Semi-Markov Decision Processes. *Journal of Intelligent Manufacturing*, 1997, Vol. 8, No. 5, pp.467-476.

HACKER SHIELD GEEK-PROOF SAFE SYSTEM:



20. Probabilistic Optimal Search

In deterministic optimization, a host of methods exist that attempt to solve optimization problems with non-linear objective functions and non-linear constraints. The most general approaches are search methods, which find the gradient of the objective function at a starting point and then find a new trial solution in the direction of the steepest gradient. Eventually, this approach will find a local optimum, which may or may not be bound by a constraint.

This same approach can be applied to problems where the non-linear objective function and/or constraints are probabilistic. Unfortunately, like most probabilistic problem solving, probabilistic non-linear programming is usually much more computationally intensive than deterministic non-linear programming. Still, it is sometimes worthwhile.

Adapting Deterministic Search Methods to Probabilistic Problems

Several changes are frequently needed to adapt deterministic non-linear programming methodologies to probabilistic problems. These adaptations can include:

- evaluation of probabilistic objective functions,
- handling probabilistic constraints, and
- incorporating penalties associated with violation of probabilistic constraints.

Like deterministic non-linear search techniques, probabilistic non-linear search techniques guarantee only a local optimum solution. There may be better solutions outside of the region searched.

Probabilistic Objective Functions

Non-linear objective functions with probabilistic outcomes can be difficult to handle. Even expected value estimation requires some form of numerical integration of the objective function over the range of values with estimation of the probability of each value.

One approach to abbreviating this calculation is through Monte Carlo evaluation of the objective function for each trial solution to the problem. A few outcomes are chosen in proportion to their likelihood, and these results are used to estimate an expected value. This approach is especially attractive where the evaluation of an outcome requires a great deal of calculation, such as the solution of a large groundwater model (Gates, et al. 1989).

Where probabilistic constraints exist, penalties for violating these constraints should be included in the objective function evaluation.

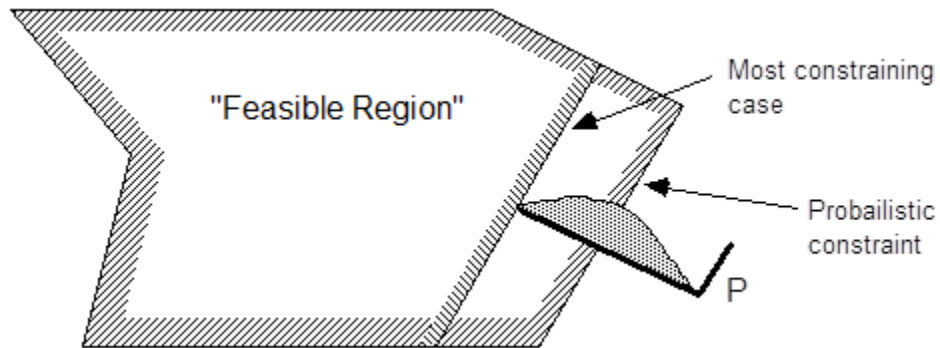
Probabilistic Constraints

Most deterministic non-linear programming methods find constraints on an ad hoc basis and do not allow the solution search to extend beyond the feasible region. Where probabilistic constraints are present, there will be some part of the "feasible region" that might not be feasible or might sometimes not be feasible.

What happens if a probabilistic constraint is violated? This can become a major problem. If the consequence of violating a constraint is catastrophic, then the probabilistic constraint can be replaced by an equivalent deterministic constraint representing the most constraining possibility for the constraint (assuming some feasible region is left under this case).

If violation of the probabilistic constraint is not catastrophic, then a penalty function for violation of the constraint needed to be incorporated into the problem. The penalty for violation of a "constraint" generally increases with the magnitude of the violation and the probability that violation occurs. This enlarges the "feasible region," but generally adds increasing penalties for approaching or going beyond

the probabilistic constraint.



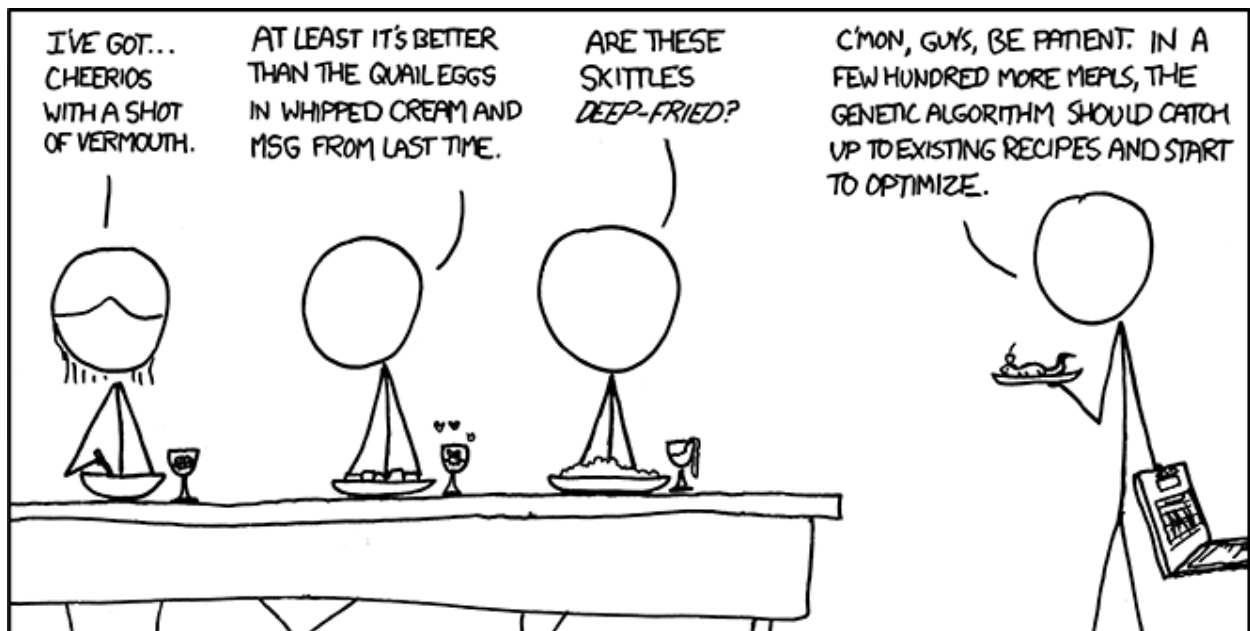
Ad Hoc Formulation and Solution Methods

Like most non-linear searches for deterministic problems, probabilistic optimal search methods are sort of an ad hoc affair and there are frequent compromises in finding a computationally feasible solution method that provides results meaningful for solving the real problem. There is much art to this and the most success seems to follow methodological flexibility.

In solving these problems, some combination of search methods, Monte Carlo evaluation, and more traditional solutions by enumeration, analytical methods, linear and dynamic programming, and probabilistic linear and dynamic programming are common. Selection of values for a few decision variables may require a probabilistic optimal search, but others might be better solved by some of these other methods.

References

Gates, T.K., R. R.J-B. Wets, and M.E. Grismer (1989), "[Stochastic Approximation Applied to Optimal Irrigation and Drainage Planning](#)," *Journal of Irrigation and Drainage Engineering*, ASCE, Vol. 115, No. 3, June, pp. 488-502.



WE'VE DECIDED TO DROP THE CS DEPARTMENT FROM OUR WEEKLY DINNER PARTY HOSTING ROTATION.

21. Evolutionary Algorithms (Genetic Algorithms)

Evolutionary algorithms (EA) (formerly known as genetic algorithms) operate by mimicking Darwinian selection and evolution of better-performing solutions. A fun example: http://rednuht.org/genetic_cars_2/

- 1) Initialization of population of decision variable values.
 - a. Initial values of decisions for each solution in the initial population.
- 2) Fitness evaluation of population.
 - a. Run simulation for each individual solution in the population.
 - b. Assess performance and fitness of each individual solution.
- 3) Stop or continue?
 - a. Reached maximum number of iterations?
 - b. Reached a convergent solution? (No substantial improvement.)
 - c. IF yes, THEN present the best solutions from this last population, ELSE continue.
- 4) Selection.
 - a. Select better-performing (fitter) solutions from population, with randomness
 - b. Elitism (Keep best solutions from previous generation?)
- 5) Breed pairs of solutions from the performance-weighted individuals in the population.
 - a. Pairing – Select pairs to breed – various methods
 - b. Cross-over method - Which decisions are taken from each individual?
 - c. Elitism – Keeping high-performing individuals as they are, cloning.
- 6) Mutation – Randomizing a few of the decisions in some of the population
- 7) Go to step 2.

EA is often applied to multi-objective problems to generate Pareto-optimal solution set (Deb 2001). Use of crowding-based selection can help disperse the population along the Pareto-optimal frontier.

Evolutionary algorithms are essentially a stochastic solution method for deterministic optimization problems. The method can also be applied to stochastic optimization, for example, by using a Monte Carlo simulation model. Most parts of EA are quite amenable to parallel computing, which can significantly decrease run-times.

Applications

Evolutionary algorithms require a very fast simulation model, which must be run many times, and are easier if the decision variables can have discrete values which are easier to code. They are commonly applied in research and have been quite successful with water distribution system design and operations problems. (Nicklow, et al. 2010; Wardlaw and Sharif 1999; Kalman 1999)

References

- Deb, K. (2001), *Multi-objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, Ltd., N.Y., 508 pp.
- Kalman, Orit, “Stormwater Quality Management: Evaluation, Optimization, and Maintenance,” PhD dissertation, Department of Civil and Environmental Engineering, University of California, Davis, <http://cee.engr.ucdavis.edu/faculty/lund/students/OritKalmanDissertation.pdf>
- Nicklow, J, et al. (2010), “State of the art for genetic algorithms and beyond in water resources planning and management,” *Journal of Water Resources Planning and Management*, ASCE, Vol. 136, No. 4, pp. 412-432.
- Wardlaw, R. and M. Sharif (1999), “Evaluation of Genetic Algorithms for Optimal Reservoir System Operation,” *Journal Water Resources Planning and Management*, Volume 125, Issue 1, January/February, pp. 25-33.
- Journal: *Evolutionary Computation*

22. Some Thoughts on Probabilities in Engineering Calculations

Engineering in an Institutional Context

Most engineering analysis, despite its fundamentally rational and scientific basis, occurs within a social and economic context involving institutions, laws, economic incentives, expectations, resources and expertise available to conduct analysis, and individual personalities. Simple analysis can usually be more persuasive in such contexts, all else being equal. But even the most sophisticated analysis will involve many uncertainties and there will be demands for our analysis to address these uncertainties. In recent decades, probabilistic risk analysis has become more commonly accepted, at least in principle. Nevertheless, there remains a great deal of controversy regarding this approach, the many details of implementation, and its communication and use by policy makers and decision makers (Tracy 2008).

Some experiences with bringing probabilistic risk analysis into the US Army Corps of Engineers have been described (Davis et al. 2008; Gravens et al. 2008), Langsdale 2008). Dutch coastal and flood engineering has a longer tradition of probabilistic risk analysis, and has developed more comprehensive analysis methods (van Danzig 1956).

Reliability of Probability Formulations and Calculations

For a normal type of probability problem, what is the probability that each of the following will formulate the problem, mathematically solve, and then interpret the solution correctly?

An engineering professor – perhaps 90%

A practicing engineer who commonly works with probabilities – perhaps 75%

A typical agency engineer – perhaps 10%

A typical elected official or political appointee responsible for making a final decision – perhaps 1%

While all of these people are very smart, this should be very sobering to those who would introduce probabilities into everyday problem solving.

Some Roles for Probabilistic Thinking in Analysis

What are proper professional roles of probabilistic thinking in engineering design?

Some thoughts:

Conceptual role – Even if we never do any probability calculations, thinking in terms of

- relative probabilities of outcomes,
- how outcome probabilities are shaped by decisions, and
- how recourses can be arranged to respond to opportunities or problems

might be of great value in thinking through design problems and how to structure such problems. The “multiple-barrier” approach to sanitary engineering is a fine conceptual example of probabilistic thinking.

Computational role – Since most people, including engineers, are terrible at doing probability problems, it is important that applications of probability to real engineering problems be very well structured and that the calculations be performed by computer programs which are well tested and have been checked several times. If one engineering professor builds and tests the program, chances might be 90% that it works. If a second engineering professor verifies this work, chances are $1-(0.1^2) = 99\%$, that it works.

Professional role – As illustrated by the calculations immediately above, there is a role for probability thinking within a profession which is interested in product reliability. By arranging how engineering work is developed, checked, and tested (with probabilities and reliabilities in mind), we can improve engineering products and productivity.

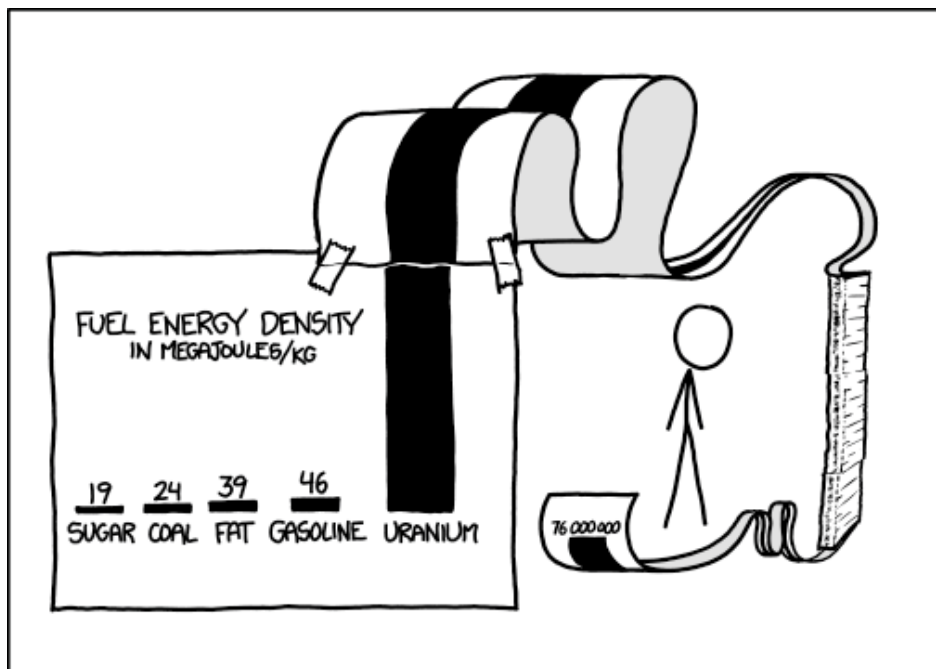
When to use probabilities in engineering?

We should use probability a) only when we have lots of data for estimating probabilities or b) when we have no data and so have the greatest uncertainty. These are two schools of thought. Obviously, I cautiously subscribe to the second school of thought – it is more interesting and holds out hope for being able to better structure and advance on truly difficult problems.

We should be careful. Probability formulations are not necessarily reliable, but they can be more reliable than other forms of analysis or non-analysis.

References

- Davis, D., B.A. Faber, and J.R. Stedinger (2008), “USACE Experience in Implementing Risk Analysis for Flood Damage Reduction Projects,” *Journal of Contemporary Water Research and Education*, Issue 140, September, pp. 3-14.
- Gravens, M.B. K.D. Hofeth, R.M. Males, D.A. Moser, and C.M. Rogers (2008), “Integrating Complexity and Uncertainty in Planning Models for Engineering: Economic Analysis within the U.S. Army Corps of Engineers,” *Journal of Contemporary Water Research and Education*, Issue 140, September, pp. 15-23.
- Langsdale, S. (2008), “Communications of Climate Change Uncertainty to Stakeholders Using the Scenario Approach,” *Journal of Contemporary Water Research and Education*, Issue 140, September, pp. 24-29.
- Lund, J.R. (2008), “A Risk analysis of Risk Analysis,” *Journal of Contemporary Water Research and Education*, Issue 140, September, pp. 53-60.
- Tracy, J.C. (2008), “Complexity and Uncertainty in Water Resources Management,” *Journal of Contemporary Water Research and Education*, Issue 140, September. Contains several useful papers.
- Van Danzig, D. (1956), “Economic decision problems for flood prevention,” *Econometrica*, Vol. 24, No. 3, pp. 276-287.



SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT PROPERLY.

Selected Bibliography of Probabilistic Design and Optimization

Books

General

- Bedford, T. and R. Cooke (2001), *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge University Press, Cambridge, UK, 390 pp.
- Birge, John R. and François Louveaux (1997), *Introduction to Stochastic Programming*, Springer, N.Y., NY, 421 pp.
- Wagner, Harvey (1975), *Principles of Operations Research*, Prentice-Hall.
- Hillier and Lieberman (1991), *Introduction to Operations Research*, McGraw-Hill.
- Sahinidis, N.V. (2004), "Optimization under uncertainty: state-of-the-art and opportunities," *Computers and Chemical Engineering*, Vol. 28, pp. 971-983.
- Taha (1991), *Operations Research: An Introduction*, Macmillan.
- Winston, W.L. (1994), *Operations Research*, Duxbury Press, Belmont, CA.

Probability and Decision Theory

- Benjamin, Jack R. and C. Allin Cornell (1970), *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill Book. Co., N.Y.
- Jaynes, E.T. (1996), *Probability Theory - The Logic of Science*, Cambridge University Press, pre-published manuscript at <http://omega.albany.edu:8008/JaynesBook.html> and other writings at <http://bayes.wustl.edu>
- Morgan, M.G. and M. Henrion (1990), *Uncertainty*, Cambridge University Press, N.Y.
- Parzen, Emanuel (1960), *Modern Probability Theory and Its Applications*, John Wiley & Sons, N.Y.
- Pratt, John W., Howard Raiffa, and Robert Schlaifer (1996), *Introduction to Statistical Decision Theory*, MIT Press, Cambridge, MA.
- Raiffa, Howard A. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, MA.
- Tribus, M. (1969), *Rational Descriptions, Decisions, and Designs*, Pergamon Press, New York.

Water and Environmental Problems

- Loucks, Stedinger, and Haith (1981), *Water Resources Systems Planning and Analysis*, Prentice-Hall.
- Mays and Tung (1992), *Hydrosystems Engineering and Management*, McGraw-Hill.

Some good journals

Journals are listed in rough order of the frequency of interesting papers over the last 20 years or so. Many other journals have interesting papers, but with lesser frequency.

General

Operations Research
Management Science
Interfaces
European Journal of Operational Research
Risk Analysis
IIE Transactions

Water and Environmental Problems

Water Resources Research
Journal of Water Resources Planning and Management, ASCE
Journal of Environmental Engineering, ASCE
Journal of Infrastructure Systems, ASCE

Transportation Problems

Transportation Science, Naval Research Logistics Quarterly

Appendix - Worked examples

Worked probabilities from Tse fly decision problem with an imperfect test.

	P(disease test result)		Test Result
	SS	NSS	Probability
+	0.1915	0.8085	0.235
-	0.0065	0.9935	0.765

For decision-making, the following table of results can be used: (again to be filled in by students)

Test	Overall Probability of Death	
	Treated	Untreated
No Test	$0.05095 = 0.05 + 0.95(0.05)(0.1)(0.2)$	0.01
+	0.0537	0.0383
-	0.05123	0.0013

Probability of Death	
Test	$0.01 = 0.0383(0.235) + 0.0013(0.765)$
No Test	0.01

Appendix - Institutional Models for Integrating Knowledge and Management (adapted from NRC 2007)

The purpose of applied quantitative modeling is to provide information and insights to individuals and groups with decision-making and management responsibilities (Geoffrion 1976). These decision-makers exist within (sometimes competing) institutions which make and support decisions and operations. The purpose of model results and insights is to improve decisions and provide decision-makers with greater confidence in the likely effectiveness of their decisions.

Modeling and model results can enter decision-making in several ways.

- a) Directly determine a decision. Direct adoption of solutions suggested by a model is rare. In a few narrow cases, model results directly determine decisions, such as selecting the operation of particular hydropower turbines over short periods. A few water distribution systems are also largely operated by model results over short periods, mostly as an aid to system operators.
- b) Provide technical support, along with monitoring data and experience, for operating decisions. This is common for the operation of most large water systems. One or more computer models will be tailored to provide specific information to system operators and managers for hourly, daily, monthly, and longer-term operational decisions.
- c) Model is a major direct part of the negotiating/decision-making environment. Especially for routine technical decision-making, model use is common. Models can be tailored for such situations and provide results, which, while imperfect, provide consistent and insightful results for decision-makers experiences with a routine problem. For non-routine decision-making, where conflicts are more common and the models are less well-tailored to the problem, models have less of a role. The use of models in negotiations is discussed in more detail below, and has sometimes been successful.
- d) Model results inform the background of decision-making and decision issue. More commonly, model results provide background information for decision-makers, much as any background technical study provides useful information.
- e) Decision makers, their staffs, and ultimately the public are educated in general through long use of model and model results. For most major water systems, agency staff become educated through development and use of models as well as through direct experience with the system. In the course of such exercises, staff develop an understanding of how the system would perform under a wider variety of circumstances than have been directly experienced. Staff also become familiar with the models, as well as their strengths and weaknesses. Modeling staff often are promoted to middle or senior management, where their reliance on models is less direct, but was foundational for their understanding of the system.

The design and execution of modeling efforts should consider the decision-making environment which it is intended to inform. Several sorts of decision-making environments are discussed below with their implications for modeling.

Technical decision-making

A basic difference exists in the use of models in technical versus adversarial approaches to discourse and decision-making. Technical and scientific decision-making ideally examines a wide range of solutions, eliminating solutions with unpromising performance until a final small set of alternatives remains, from which one promising solution is chosen. Among technical people, uncertainty always exists in all but the most fundamental knowledge (such as conservation of mass, energy, and momentum). Because most applied models are based on assumptions beyond fundamental knowledge (empirical, and often professional judgment), almost all models are imperfect and will err somewhat. There is no such thing as a “scientifically valid” model unless it is based on fundamental principles (Konikow and Bredehoeft 1992). Like other scientific hypotheses, one can only invalidate a model. A model can never be completely validated. In an applied context, all model results must be interpreted and judged.

Uncertainty always exists (Oreskes 2003).

Where there is consensus on objectives, technical-scientific decision making is quite successful. Quantitative models are routinely used and trusted for major water and environmental decisions every day. National Weather Service models of storms and floods have been tremendously effective for reducing loss of lives and property from storms, even though they are imperfect and their results have significant uncertainties. Outside of environmental applications, quantitative models are relied on for making buildings and bridges more reliable, airlines run more on time, and countless other practical applications. All of these models retain important uncertainties, but they do provide insights and a logical basis for conclusions which unaided humans cannot provide.

Adversarial decision-making

Where conflict on objectives exists, typical decision-making processes ask more of quantitative models.

Adversarial decision-making, which dominates legal and political discourse, is a contest among alternatives or for and against a particular proposed alternative. In such a contest, models and model results supporting an alternative are presented by proponents. Proponents attack or discredit models and model results which do not support the proposition. Adversaries to a proposal take the opposite view. In such contests, an uncertain model or imperfect model results are often easily discredited. Adversarial decision-making has trouble making use of models and model results without a preponderance of scientific support (Jackson 2005). Communication of model results is especially important in an adversarial process.

In adversarial environments, proponents of the status quo will often call for additional study, detailed modeling, or long periods of data collection, particularly if they are financed by others. For status quo proponents, more studies and modeling are always needed. One productive use of modeling in adversarial situations is to help re-shape understanding of a problem and solutions over a long period. This poses little urgent threat to the status quo, and allows improved understanding and solutions to be crafted for a future time when the political environment is more fluid, as in the aftermath of a major drought, flood, or lawsuit.

Negotiations

It is often proposed to use models and model results as part of negotiated solutions. The original adaptive management (Hollings 1978), “shared vision modeling” (Palmer et al. 1999; Lund and Palmer 1997), and gaming approaches all have in common the use of computer models to represent trade-offs, certainties and uncertainties in negotiations among conflicting parties. This decision-making environment lies between pure technical-scientific and adversarial decision-making. Where broad motivation exists to come to a consensus agreement with realization that technical support is needed for such an agreement, models can have a useful, even central, role in negotiated decision-making.

Quantitative models can have several roles in policy negotiations:

A decision-support-system for negotiations. Here, computer models form the central venue and technical arbiter for negotiations, constituting a substantially agreed-upon technical basis for discussions and comparison of performance for proposed or crafted alternatives. This typically “neutral” technical-scientific party creates the model support for negotiations or a process where technical representatives from major stakeholders come to an agreed model representation of the problem.

Model results used directly in negotiations. Here, model results are used in negotiations, as would any technical study or document. This approach does not require as much consensus on the technical merits of the work, and allows the modeling to have a more peripheral role in the negotiation deliberations.

Preparation for negotiations. Models and modeling results are often used to prepare for negotiations.

Parties often perform internal modeling studies to investigate options from their perspective and those of other parties. These studies can help form the basis of proposals and critiques offered during the negotiation process. Sometimes such internal modeling studies are performed during the course of negotiations.

Models used to train technical people who advise negotiations. Actual negotiations often are on time-frames short enough that new modeling studies might not be possible. In such cases, past model studies, often accumulated over decades, provide negotiators or technical advisors to negotiators with considerable knowledge of promising and unpromising alternatives, as well as insights and concerns worthwhile during negotiations.

An adversarial process often follows such a period of negotiation. Even if a negotiation leads to a formal agreement, opportunities arise for further negotiation and adversarial decision-making in implementing any agreement.

Regulatory environments

Often agencies are tasked with enforcing environmental and water regulations, laws, and property rights. In an ideal world, field monitoring data would be abundant, precise, and accurate. However, field data are imperfect, typically sparse, and unavailable for hypothetical future conditions. Thus, for routine regulatory proceedings, field data are often unavailable or insufficient alone to make permitting or enforcement assessments. Here, quantitative models can have two roles. First, models can often effectively interpolate or extrapolate from existing field data (which are often used to calibrate the model and establish boundary conditions). This saves the agency and the permittee considerable expense and delay for data collection. Models are used to assess the likely environmental or resource effects and the effectiveness of any proposed mitigation actions. These applications all employ model results as information to a human regulatory decision.

Another role of modeling is for more formal accounting of environmental effects. Here, the model is effectively designated as an accounting standard, eliminating human assessment. For water rights allocations, models are almost the only practical means to assess water availability, however imperfect. Using quantitative models as a basis for TMDLs and TMDL allocations is a more modest example of the model developing into a standardized understanding of a system. To some degree, the automation of model-based accounting can provide greater transparency and predictability of regulatory decisions, as any party can presumably run the model.

The particular type of resource or environmental regulation also can affect the use of quantitative modeling. Where environmental regulation is based on traditional command and control, including specification of required technology, such as specifying particular wastewater treatment processes or so-called “best management practices”, routine model use is less important, although models might be useful for determining which technologies should be required. Where regulations specify only a performance standard, those regulated have greater flexibility to use potentially more economical means achieve the standard, but monitoring or modeling requirements are increased, to make the regulations enforceable. For modeling to trade-off well against monitoring, monitoring must be expensive and/or models of performance relatively good. For market-based regulations (such as water markets or markets for TMDL), the use of models as an accounting mechanism becomes attractive, as it is often onerous to have enough density or accuracy of field monitoring to enforce property rights.

Overall discussion

Modeling can be used to support or attack a particular decision or it can be used in a technical way to broaden the range of solutions, improve currently-proposed solutions, eliminate solutions that perform poorly, and select more promising solutions. Negotiated and regulatory decision environments lie somewhere between these extremes. Each of these environments asks more or less of modeling,

particularly in terms of establishing credibility and communication of results.

The Modeling Process

“All models are wrong, but some are useful.” G.E.P. Box

Modeling is a fallible art of trying to quantitatively represent the complexity and processes of real systems. Scientifically, such representations provide an ability to assemble more complex understandings of complex real systems than would be possible without such aids. They can be used to develop hypotheses which integrate many aspects of complex phenomena. Moreover, application of models commonly allows better predictions of the outcomes of proposed actions. This use of models allows solutions to practical problems to be developed virtually, in a more rapid, less costly, and less risky manner than by direct experimentation with the real system.

As such, models, and particularly computer models, have become indispensable for managing complex systems ranging from transportation systems (including most airline scheduling), large building structures, as well as routine wholesaling, retailing, and commercial systems by engineers, business managers, and economists. Scientifically, mathematical and computer models have been central to the physical sciences, and have become increasingly used in all areas of scientific research.

Historically, the scientific use of mathematical models began as early as Galileo’s time with engineering applications becoming established in France before the revolution. Modeling now is the accepted approach for improving the efficiency and effectiveness of efforts to understand and solve complex problems. To improve the likelihood that modeling will deliver on such promise, model development and use commonly follows a fairly standardized process, described in this section. Scientific progress results when the hypothetical understanding of the system represented by the model diverges from field observations, leading to improvements in the model, our understanding of the modeled system, and the model’s predictive powers.

The Modeling Process

While many authors present somewhat different versions, most presentations support a process of developing and applying a model with elements and steps similar to those below (Gass 1983; Satkowski et al., 2000; USEPA 2003).

1. Problem and purpose definition
2. Model development
3. Calibration and testing
4. Sensitivity analysis
5. Results interpretation and communication

Like other problem-solving activities, the development and application of a model benefits greatly from having a planned approach (Polya 1957).

1. Problem statement and purpose

“A problem well posed is a problem half-solved.” Attributed to John Dewey

Most models are developed for a particular scientific or problem-solving purpose. Defining a purpose for the model provides a focus and rationale for including factors and processes which are more likely to be important and excluding others. Modeling the world in its entirety is never an option. A well-defined applied model not only describes the real system, but also contributes to the problem-solving situation and institutional framework as well.

Defining the purpose of the model is the most important step in modeling. For applied modeling, a common error is modeling the wrong problem. Such errors can arise from adapting an older model –

developed for a different purpose – to a new purpose, from professional narrow-mindedness where an expert seeks to apply this specialty to unsuitable situations, from political concentration on one aspect of a problem when another dominates, and frequently when the nature of the problem changes in the years of model development (a lag between model development and use of model results). Sometimes it is impossible to tailor a model to a particular purpose without important simplifications; a “simple but accurate” model of a real complex system is often an impossible modeling purpose. Thus, it is common for models not to be perfectly suitable for the practical problems to which they are applied. Such problems often can be mitigated by changes in the model (perhaps re-calibration) and careful interpretation and communication of model results and insights, described later.

2. Model development

Following a statement of the model’s purpose, knowledge believed to be relevant is assembled. This knowledge can take the form of empirical relationships, observed locally or in similar circumstances, and relationships derived from fundamental and well-proven principles. Conservation of mass, energy, and momentum are examples of fundamental principles from which relationships can be derived. Empirical relationships are inferred from field data by regression or other types of fitting to equations. Field data can be obtained locally or for situations deemed similar. Examples of empirical relationships include several forms of hydraulic equations (such as the Manning or Chezy equation) as well as regression equations fit to field data. Mathematical forms of these empirical and fundamental relationships are then organized into a coherent mathematical representation of the system for the purposes of the problem. Simplifications of the problem are often required to do this, neglecting or simplifying some parts of the problem.

This simplified mathematical representation of the problem must sometimes be further simplified to allow solution of the mathematical problem, or approximate solution to the mathematical problem. Numerical methods, such as finite element or finite difference techniques are often used to solve relatively complex mathematical representations. Frequent checks on the stability and accuracy of the numerical solution are often required.

At the end of this step, the model of the system is twice simplified from the original real problem, first to create a mathematical representation of the problem and then to create a solvable mathematical representation. Nevertheless, the result is commonly a far more complex and transparent representation of the problem than would be possible without mathematical aid, and a representation which allows integration of diverse types of scientific knowledge and understandings of the system.

3. Calibration

A further empirical phase of model development is model calibration. Calibration consists of adjusting some of the more empirical parameters in the mathematical model to fit data observed from the field. Sometimes parameters in component sub-models are adjusted against field data and sometimes parameters in several model components are adjusted together against field data. Sometimes calibration is based on data observed in field conditions elsewhere, if local data is unavailable. It is greatly preferred to have local field data under problem-relevant conditions to calibrate empirical parameters. However, field data are rarely available to the extent desired within a time frame relevant for the problem.

The adjustment of parameters is often done by experts in modeling the type of system being modeled. Such adjustments are sometimes aided by automated algorithms, particularly when calibration parameters are numerous. Since there are often many possible sets of parameter values which “fit” field data, the background and understanding of the modeling experts has an important role in calibration. Usually parameter calibration is limited within a “reasonable” range based on field and modeling experience for a range of similar conditions.

The residual differences between observed field data and the calibrated model represents how well the

model “fits” the field data, and provides a form of model test. Calibration residuals are a fairly weak form of model test as the modeler had an opportunity to “fit” or adjust the parameter values to these data. Thus, when the number of parameters in the model is large or similar to the number of field observations, the utility of calibration residuals for model testing can be small.

4. Model testing and evaluation

Model testing can consist of a wide variety of techniques intended to evaluate and demonstrate the strengths and limitations of a model for particular purposes (Gass 1983; Kleijnen 1995; Beck 2002; Parker et al. 2002). Ideally, model testing procedures and protocols are established early in the modeling process (Kauffman, et al. 2001). Some common forms of model testing include:

Software tests

Software tests can occur at several levels and by several means (Kauffman et al. 2001). Parts or components of the model can be tested separately, in functional units, and then together as a modeling system. These code tests are ideally done by people other than the authors and can be done by a designated “librarian”, a peer-review process, by parallel development teams, or formal individual or group “walk-throughs” of the code. When programmers understand that others will inspect and test their code, coding tends to be more reliable.

Numerical tests

Numerical tests are used to ensure that the model’s calculations are stable and correct for some well-known cases and solutions. Complex models can be numerically unstable for some situations and numerical tests can help establish these limits (Sobey 2001). Routine model applications of common software often rely on software and numerical tests done by the model developer and prior applications of the model.

Empirical tests

Comparisons with field data at the component or system scales are useful tests of a model. Such tests are stronger if they are done with data sets different than those used for model calibration and over a wide range of field conditions (wet and dry years, for example). Unfortunately, field data are often sparse and unavailable for complete empirical testing over a wide range of conditions. Such empirical tests against independent field data are often called “model validation” studies, but the sparseness of field data usually means that such tests do not fully demonstrate the “validity” of the model for all relevant field conditions. Empirical model testing is never directly available for model applications for non-existing conditions, such as conditions in the future with alternative solutions (Gass 1983). An additional problem is the quality of field data; difficulties and errors in field observations make empirical tests of a model less accurate.

Model comparison tests

Often a large system model must simplify components or the overall representation of a system relative to detailed models which might exist of the system or system components. Where the detailed model or model components provide greater confidence in the representation (sometimes they don’t), then comparison between the complex and simplified models can provide some insights and understanding of the relative limitations of the two models. Model comparisons can often be made over a wide range of virtual field conditions, and so avoids the limitations and expense of comparisons of model and field results. However, model comparisons are weaker tests than good empirical tests. Model comparison results also are often used to assess the numerical errors in the model solution method.

Sensitivity analysis

Sensitivity studies quantify the effects of small changes in model assumptions on model results. Such sensitivity results provide insights as to the likely range of error in model results from such causes. Sensitivity results can be useful for interpreting model results and assessing the data quality needed or

desirable from field investigations. (Frey and Patil 2002). Linear programming methods often produce some forms of sensitivity analysis as a by-product of their numerical solution algorithms.

Expert evaluation

Almost all model results are evaluated by experts in the problem being modeled. This occurs in model development, calibration, and application. Errors are frequent in modeling complex systems, and expert inspections are often the most readily-available and capable means to identify potential errors.

Expert review is commonly done internally by the modeling team. Sometimes this internal review is informal and sometimes it is quite structured. Additional external review by local problem experts or distant academic, consulting, or agency experts on the general type of problem of modeling also can be employed.

Overall, as noted by Quade (1980), “A particularly dangerous myth is the belief that a policy model can be fully validated – that is proved correct. Such models can, at best, be invalidated. ... Thus the aim of the validation [testing] (or rather invalidation) attempts is to increase the degree of confidence that the events inferred from the model will, in fact, occur under the conditions assumed.. When you have tried all the reasonable invalidation procedures you can think of, you will not, of course, have a valid model (and you may not have a model at all). You will, however, have a good understanding of the strengths and weaknesses of the model, and you are able to meet criticisms of omissions by being able to say why something was left out and what difference including it would have made. Knowing the limits of the model’s predictive capabilities will enable you to express proper confidence in the results obtained from it.”

Every decision maker has a mental model or understanding of the problem (Gass 1983). However, these mental models are only tested very indirectly by political election or appointment processes which place an individual in a decision-making capacity. It should be possible for quantitative models based on scientific and technical information to demonstrate greater levels of credibility to supplement, aid, or improve on decision-makers’ mental models, and ultimately improve the consideration and selection of decisions.

5. Results interpretation and communication

Even a perfect model will be useless if its results are not trusted and employed for understanding or solving a problem. Model results and their implications must be interpreted and communicated for non-modeling specialists in the context of the applied problem. The communications of results must often address to issues: 1) communication and support of insights and results and 2) demonstration of the credibility and limitations of the model and its results.

Documentation and External Review

Documentation facilitates training of model users, supports the credibility and transparency of a model (allowing the work to be externally reviewed), and furthers the education of the water management and modeling community regarding the problem being modeled. Documentation also has an important internal quality control function. Documenting a model and the work and thought that goes into documentation helps ensure that a model works, its limitations are well understood and can be communicated, and future improvements are identified.

Peer or external review can be useful for communicating and establishing model credibility. However, a credible model review will almost always find some real or potential flaws, so in an adversarial environment, external reviews are risky. ...

Establishing Model Credibility

A primary aspect of model development, testing, and application is establishing the credibility of the modeling effort (Gass 1983). Credibility can be established based on:

- model's agreement with specialist or popular notions regarding the system (face validity),
- credentials of the modeler or modeling organization,
- technical procedures and protocols followed in model development,
- model documentation produced,
- tests conducted on the model and its results,
- qualifications of advisers or reviewers of the effort,
- a (long) period over which the model has been employed,
- current model use,
- diversity of situations for which the model has been employed, and
- authoritative (agency) sponsorship of the model or modeling effort.

Some of these factors which bolster the perceived credibility of a model may have little to do with its actual technical reliability. But perception of credibility is credibility.

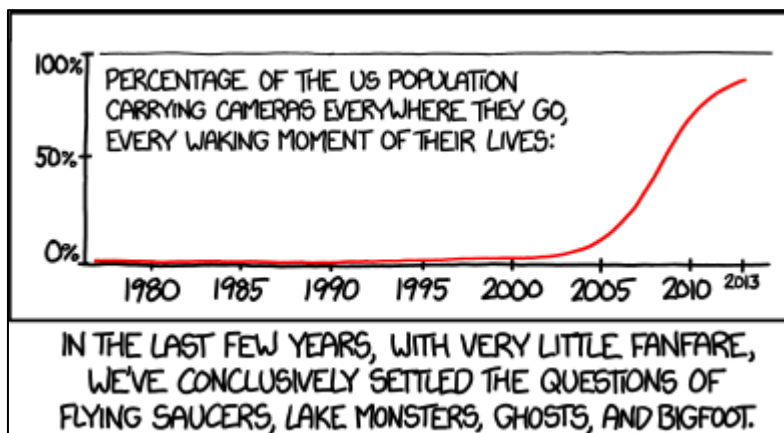
Models developed for applications in an adversarial environment must be pursued with particular care. Where a model or its results can be expected to enter into legal or political proceedings, an especially systematic, tested, transparent, and articulate modeling effort is required.

While no amount of effort can ensure that a model is perfect. However, following the kind of systematic model development and application process described above can greatly increase the likelihood that a model will be useful for understanding or developing solutions for problems.

References

- Beck, M.B. (2002), "Model evaluation and performance," in A.H. El-Shaarawi and W.W. Piegorsch (eds.), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd, Chichester, Volume 3, pp 1275–1279. http://www.cgrer.uiowa.edu/cleaner/italy/Beck_Venicem.pdf
- Beck, MB (1991), "Principles of Modelling", *Water Science and Technology* WSTED4, Vol. 24, No. 6, p 1-8.
- Box GEP, "Robustness in the strategy of scientific model building." in: Launer RL & Wilkinson GN *Robustness in Statistics*. New York: Academic Press, 1979:pp. 202.
- Frey, H.C. and S.R. Patil (2002), "Identification and review of sensitivity analysis methods," *Risk Analysis*, Vol. 22, No. 3, pp. 553 – 578.
- Gass, S.I. (1983), Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis, *Operations Research*, Vol. 31, No. 4 (Jul. - Aug., 1983), pp. 603-631
- Geoffrion, A.M. (1976), "The purpose of mathematical programming is insight, not numbers," *Interfaces*, 7(1):81--92, 1976.
- Holling, C.S. (ed.) (1978), *Adaptive Environmental Assessment and Management*, John Wiley & Sons, N.Y.
- Jackson, S. (2006), "Water models and water politics: design, deliberation, and virtual accountability," ACM International Conference Proceeding Series; Vol. 151, *Proceedings of the 2006 international conference on Digital government research*, San Diego, California, pp. 95 – 104.
- Kauffman, B., T. Bettge, L. Buja, T. Craig, C. DeLuca, B. Eaton, M. Hecht, E. Kluzek, J. Rosinski, and M. Vertenstein (2001), *Community Climate System Model Software Developer's Guide*, June http://www.cesm.ucar.edu/working_groups/Software/dev_guide/dev_guide/dev_guide.html
- Kleijnen J.P.C. (1995), "Verification and validation of simulation models," *European Journal of Operational Research*, Vol. 82, pp. 145-162.
- Klemes, V. (1982), "Empirical and causal models in hydrology," *Scientific Basis of Water Resources Management*, National Academy of Sciences, Washington, D.C., pp. 95-104.
- Konikow, L.F. and J.D. Bredehoeft (1992), "Ground-water models cannot be validated," *Advances in Water Resources*, Vol. 15, pp. 75-83.
- Lund, J.R. and R.N. Palmer (1997), "Water Resource System Modeling for Conflict Resolution," in

- Water Resources Update*, UCOWR, Issue 108, Summer, pp. 70-82.
- National Research Council Committee on Hydrology, Ecology, and Fishes of the Klamath River (2007), *Hydrology, Ecology, and Fishes of the Klamath River Basin*, National Academy Press, Washington, DC, November.
- Oreskes, N. (2003), "The Role of Quantitative Models in Science," in *Models in Ecosystem Science*, edited by Charles D. Canham, Jonathan J. Cole, and William K. Lauenroth (Princeton: Princeton University Press), pp. 13-31.
- Palmer, R.N., W.J. Werick, A. MacEwan, and A.W. Woods (1999), "Modeling Water Resources Opportunities, Challenges and Trade-offs: The Use of Shared Vision Modeling for Negotiation and Conflict Resolution,"
- Parker, P., R. Letcher, A. Jakeman, M.B. Beck, G. Harris, R.M. Argent, M. Hare, C. Pahl-Wostl, A. Voinov, M. Janssen, P. Sullivan, M. Scocciarro, A. Friend k, M. Sonnenshein, D. Barker, L. Matejicek, D. Odulaja, P. Deadman, K. Lim, G. Larocque, P. Tarikhi, C. Fletcher, A. Put, T. Maxwell, A. Charles, H. Breeze, N. Nakatani, S. Mudgal, W. Naito, O. Osidele, I. Eriksson, U. Kautsky, E. Kautsky, B. Naeslund, L. Kumblad, R. Park, S. Maltagliati, P. Girardin, A. Rizzoli, D. Mauriello, R. Hoch, D. Pelletier, J. Reilly, R. Olafsdottir, and S. Bin (2002), "Progress in integrated assessment and modeling," *Environmental Modelling & Software*, Vol. 17, pp. 209-217.
- Polya G. (1957), *How to Solve It*, 2nd ed., Princeton University Press.
- Quade, E.S. (1980), "Pitfalls in formulation and modeling," in Majone, G. and E.S. Quade, *Pitfalls of Analysis*, John Wiley and Sons, N.Y.
- Satkowski, R., J.R. Lund, H. Morel-Seytoux, A. Nelson, and T. Roefs, *Protocols for Water and Environmental Modeling*, Bay-Delta Modeling Forum, Richmond, CA, Report No. BDMF 2000-1, 38 pp, January 2000.
- Sobey R.J. (2001), "Standardized Evaluation of Estuarine Network Models," in *Coastal Engineering 2000*, Proc 27 the International Conference on Coastal Engineering, ASCE, Reston, VA, pp. 3683
- USEPA (2003), *Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models*, USEPA Office of Science Policy, Office of Research and Development, Washington, DC.
- 3696.



xkcd.com/1235/

Appendix - Buckingham π theorem (adapted from Wikipedia, the free encyclopedia)

More formally, the number of dimensionless terms that can be formed, p , is equal to the [nullity](#) of the [dimensional matrix](#), and k is the [rank](#). For the purposes of the experimenter, different systems which share the same description in terms of these [dimensionless numbers](#) are equivalent.

In mathematical terms, if we have a physically meaningful equation such as

$$f(q_1, q_2, \dots, q_n) = 0$$

where the q_i are the n physical variables, and they are expressed in terms of k independent physical units, then the above equation can be restated as

$$F(\pi_1, \pi_2, \dots, \pi_p) = 0$$

where the π_i are dimensionless parameters constructed from the q_i by $p = n - k$ equations of the form

$$\pi_i = q_1^{m_1} q_2^{m_2} \dots q_n^{m_n}$$

where the exponents m_i are rational numbers (they can always be taken to be integers: just raise it to a power to clear denominators).

The use of the π_i as the dimensionless parameters was introduced by [Edgar Buckingham](#) in his original 1914 paper on the subject from which the theorem draws its name.

See also examples on the Wikipedia web site.

