Operations Research
The University of Texas
1 University Station, C2200
ETC 5.128D
512.232.8316
ebickel@mail.utexas.edu

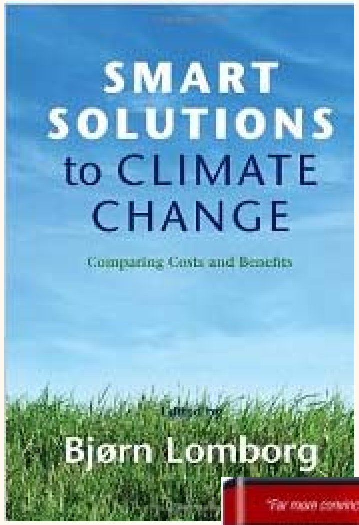
Discretization, Simulation, and Swanson's (Inaccurate) Mean

Presented by:
J. Eric Bickel
Graduate Program in Operations Research
Center for Petroleum Asset Risk Management
The University of Texas at Austin
ebickel@mail.utexas.edu

Presented to:
2011 DAAG
Interacting with Decision Makers
Chair: Tony Manzella

20 April 2011

My research and practice are focused on decision making under uncertainty.

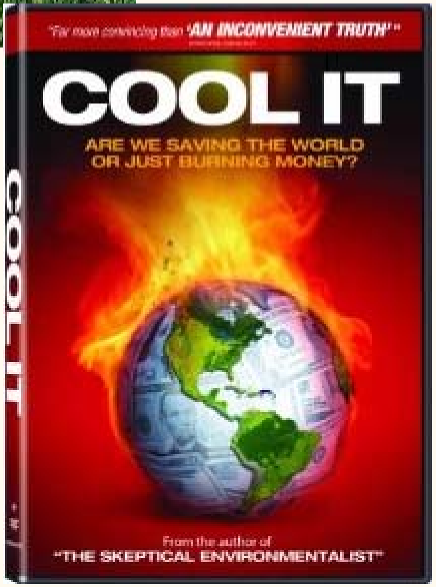


Methodological Research

- Efficient modeling of probabilistic dependence.
- Value of information.
- Strictly proper scoring rules.
- Auditing and scoring of expert forecasts.
- Corporate risk preference.

Application Areas

- Optimal development of unconventional gas reservoirs.
- Value of seismic information.
- Optimal sequencing of exploration wells.
- Risks of carbon capture and storage.
- Energy and climate policy.
- Baseball strategy.



http://www.coolit-themovie.com/

Bickel, J. Eric and Lee Lane, 2010. "Climate Engineering as a Response to Climate Change." In *Smart Solutions to Climate Change*. Edited by Bjørn Lomborg. Cambridge University Press.

Acknowledgements

This presentation is based upon:
Bickel, J. Eric, Larry W. Lake, and John Lehman. Forthcoming. Discretization, Simulation, and Swanson's (Inaccurate) Mean. *SPE Economics and Management*.

The authors thank the following organizations for their support of this research:

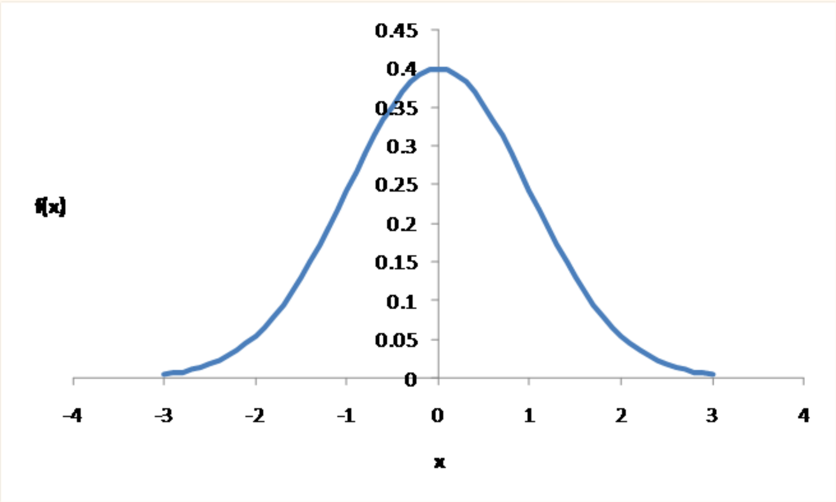
-- The Unconventional Gas program at the Research Partnership to Secure Energy for America (RPSEA).

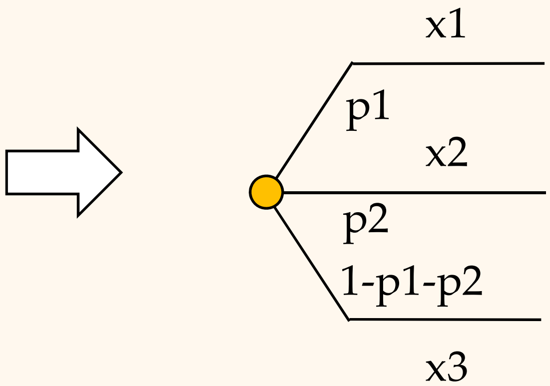
-- The Center for Petroleum Asset Risk Management (CPARM) at The University of Texas at Austin.

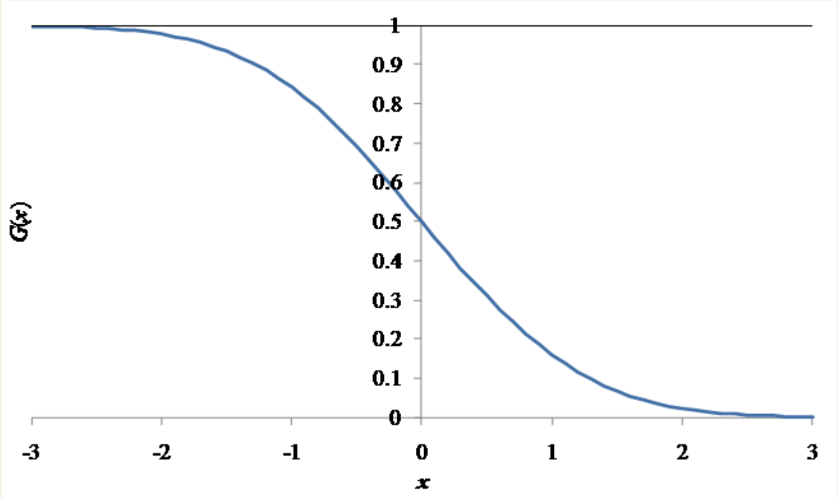
-- Strategic Decisions Group (SDG).

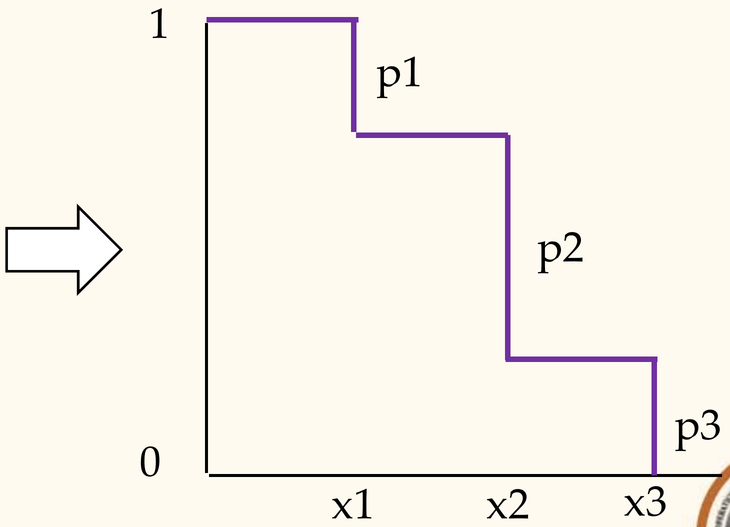
Thank you to DAAG for allowing us to present this work.

We often need to approximate continuous uncertainties (e.g., to include them in a decision tree).



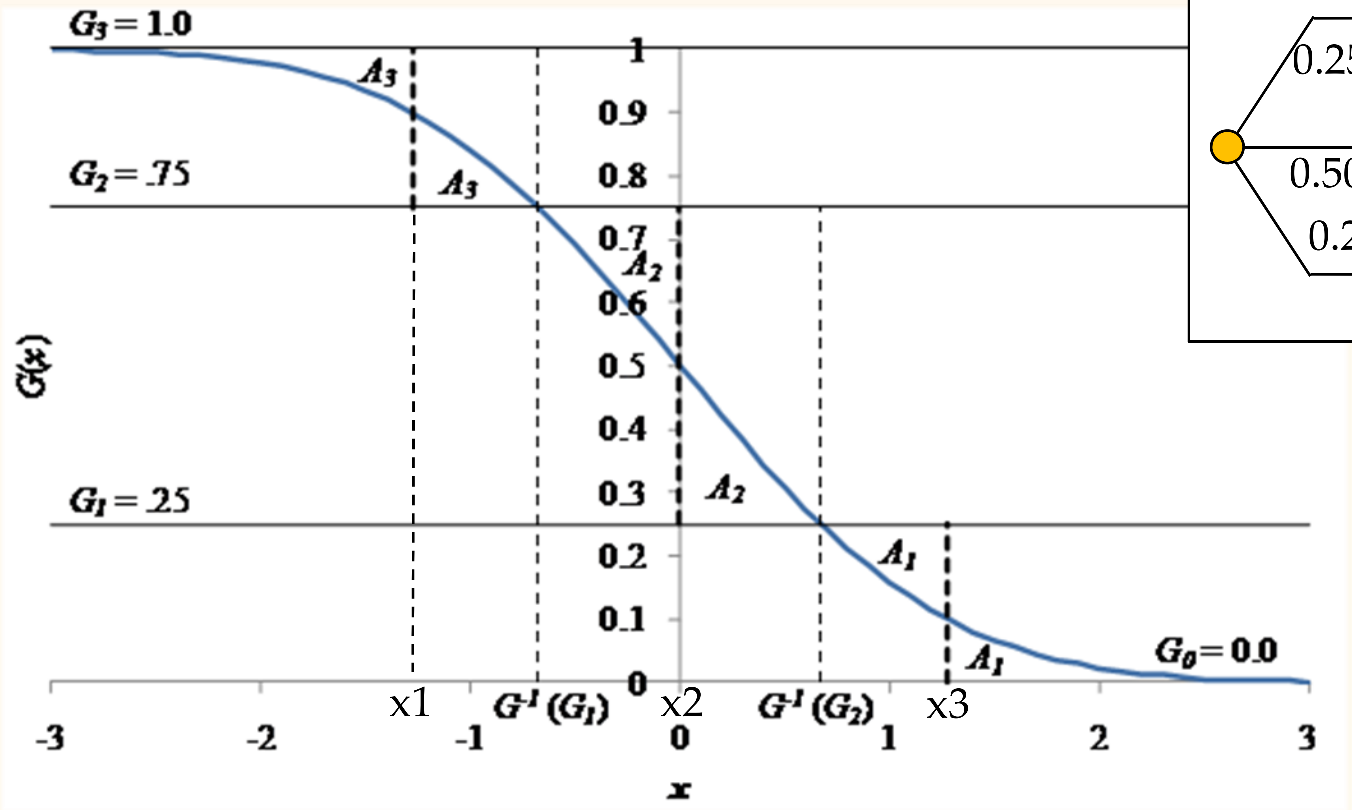


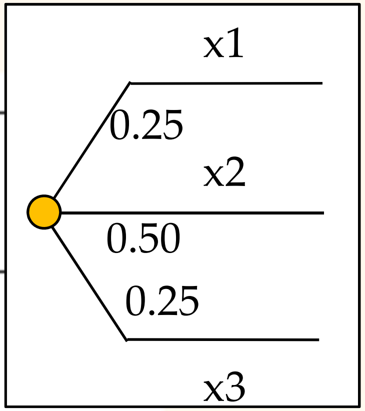




One frequently used method is known as "Bracket Mean" or "Equal Areas."

Developed by Jim Matheson and colleagues at Stanford Research Institute (later SDG) in the late 1960s.





Notice that the 25-50-25 approximation is very close, in this case, to the P90-P50-P10 points. This is the 25-50-25 shortcut popularized by SDG.

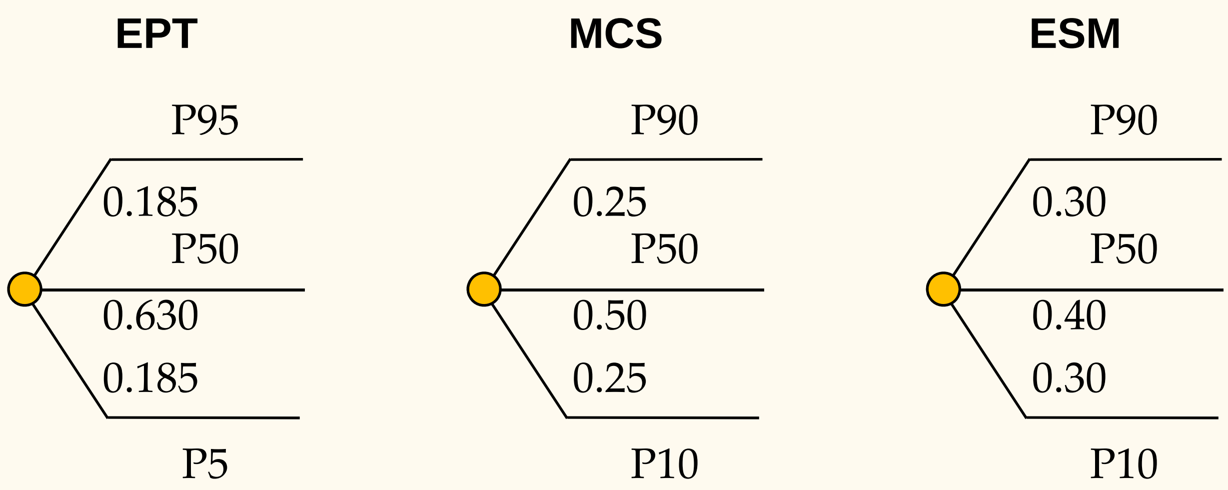
To ease practice, several three-point shortcuts have been developed; I focus on only three of these.

Shortcuts (in order of development)

Extended Person-Tukey (EPT)

McNamee-Celona Shortcut (MCS)

Extended Swanson-Megill (ESM)



A more precise method is known as "Gaussian Quadrature."

- Methods that can be used to match the moments of distributions were developed by Gauss in the early 1800s.
- Miller and Rice (1983) introduced them to decision analysts almost 30 years ago.
- Here are P10-P50-P90 approximations that match the first three moments of the listed distributions:

Uniform	Normal	Exponential	Triangular*
$f(x)=1$	$f(x)=(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$	$f(x)=e^{-x}$	$f(x)=\begin{cases}\frac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \le b \\ 0 & \text{otherwise}\end{cases}$
$x \quad x \in [0,1]$	$x \in (-\infty, \infty)$	$x \ge 0$	$x \in [a=0, b=1, c=0.5]$
0.260	0.304	0.465	0.273
0.480	0.392	0.175	0.454
0.260	0.304	0.360	0.273

* Note: Triangular discretization is a function of c, but this relationship is weak enough to ignore.

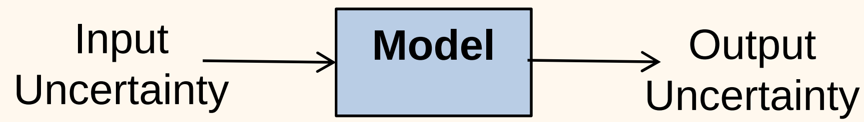
One interesting question is how accurate the differing methods are in approximating the underlying distribution.

What do we mean by "accurate"?

- Preserving the shape of the distribution? What do we mean by "shape"?
- Preserving particular fractiles of the distribution?
- Preserving the moments of the distribution?
- Preserving the "decision structure" of the underlying problem?
- Are we trying to preserve the input distribution or the output?

All of the discretization methods discussed herein are intended to preserve the moments of the underlying <u>input</u> distributions; as we will see, some do a better job than others.

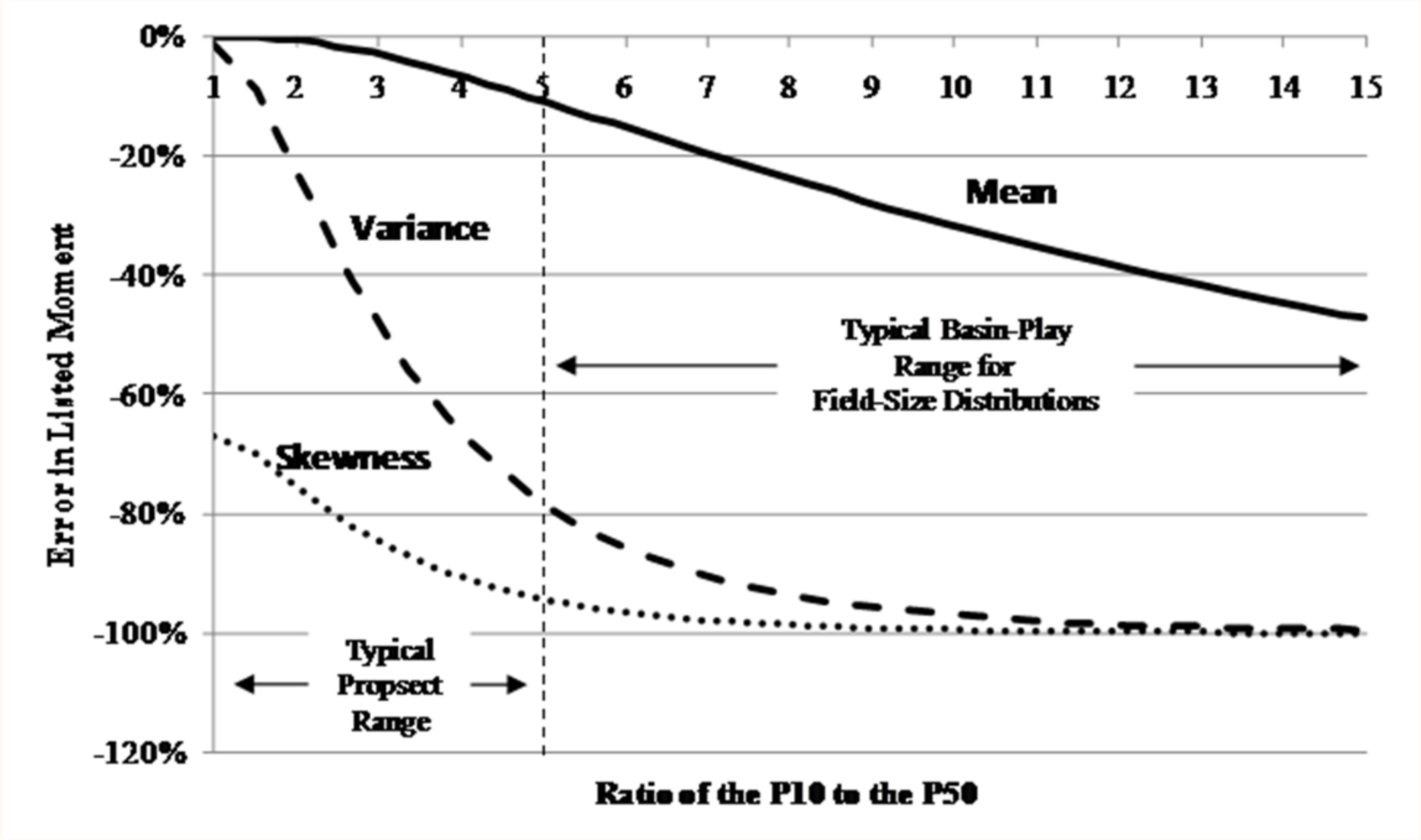
Why moments?



- The mean of our model output (e.g., NPV) depends on all the moments of the input uncertainties.
- Thus, it is not enough, as is often assumed, to simply match the mean of the input uncertainties.
- For example, suppose your input uncertainty, X, is normally distributed with 0 mean and unit variance.
  - If NPV(x) = x^2 then expected NPV is 1.
  - However, the NPV evaluated at the expected value of X is 0.

Implication for Practice: We often discretize input distributions, combining them in a tree, and produce a cumulative distribution. What does this distribution mean? The discretizations were not designed to preserve this cdf. Reading particular fractiles off of this cdf (e.g., the probability that NPV is less than zero) may not be justified.

Swanson-Megill fails to faithfully represent the moments of a lognormal probability distribution function, which is the distribution it was “designed” for.



There is no justification for using ESM for any distribution other than the normal.

We have found that the following four-point approximations better represent the lognormal (these points will match the mean and variance).

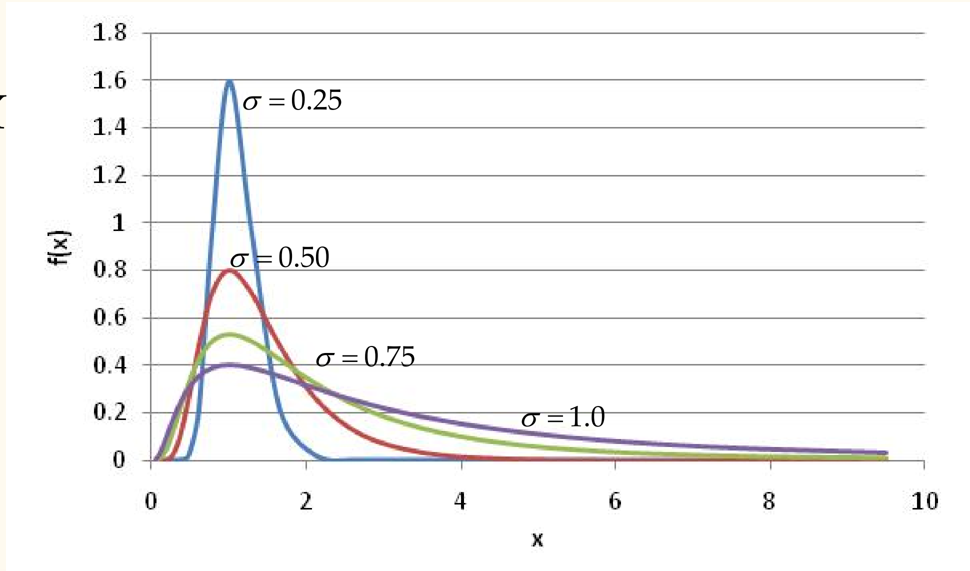
$\sigma = .25$, $\gamma_3 = 0.78$		$\sigma = .50$, $\gamma_3 = 1.75$		$\sigma = .75$, $\gamma_3 = 3.26$		$\sigma = 1.0$, $\gamma_3 = 6.18$	
p_i	α_i	p_i	α_i	p_i	α_i	p_i	α_i
0.203	90.0	0.151	90.0	0.186	90.0	0.436	90.0
0.288	59.9	0.265	69.1	0.235	77.3	0.274	84.1
0.303	50.0	0.378	50.0	0.357	50.0	0.017	50.0
0.206	5.0	0.206	5.0	0.222	5.0	0.272	5.0

$\sigma \equiv$ standard deviation of $\ln X$

$\gamma_3 \equiv$ skewness of X

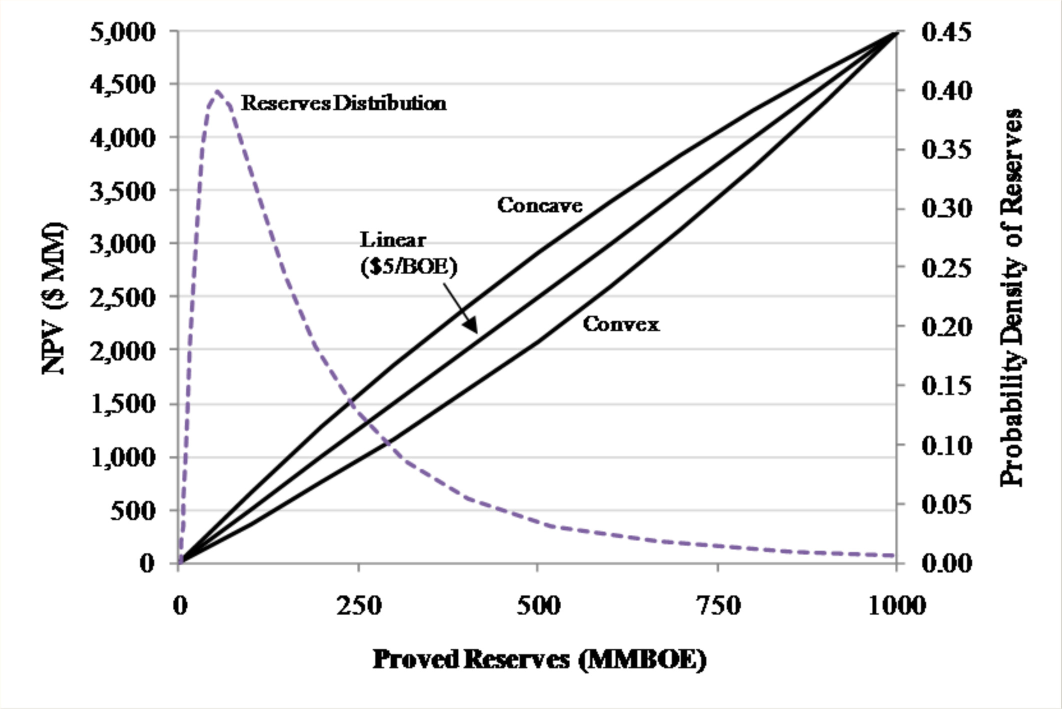
$p_i \equiv$ probability of point i

$\alpha_i \equiv$ fractile i



To understand the accuracy of each method, let's consider a simple example:

- An oil company is considering the purchase of a prospect, whose reserves are believed to be lognormally distributed with a mean of 90 MMBOE and a standard deviation of 118 MMBOE (this yields a standard deviation of log-reserves of 1.0).
- Reserves transact on the market for $5/BOE, but may differ from this.



The accuracy of the method depends upon the shape of the value function.

- In the linear case, the accuracy of the output is the same as the accuracy of the input.

- In the non-linear cases, the failure of the approximations to faithfully represent the underlying moments (including the higher moments) introduces additional error.

- The bracket mean (equal areas) methods outperform all but the EPT approximation.

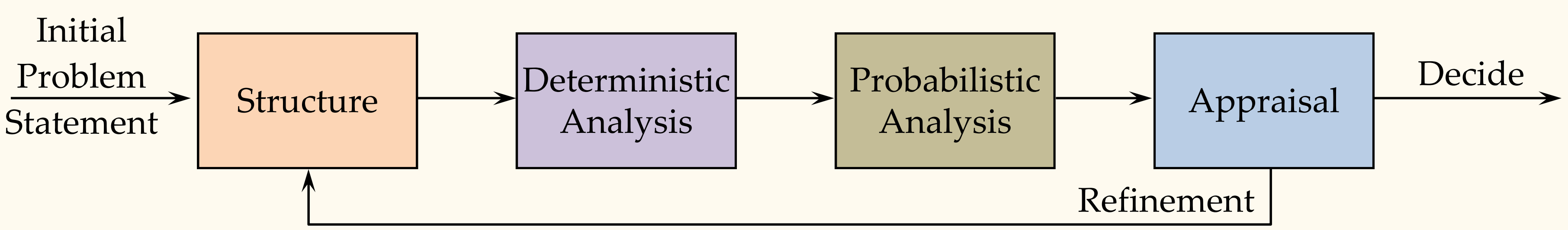
	Error in Mean			Error in Variance		
Approximation	concave	linear	convex	concave	linear	convex
4-pt LN	-1%	0%	-1%	30%	0%	-57%
EPT	0%	-1%	-4%	-18%	-36%	-72%
ESM	-3%	-5%	-9%	-45%	-60%	-83%
MCS	-8%	-11%	-15%	-53%	-66%	-86%
3-pt Bracket Mean	2%	0%	-4%	-38%	-54%	-80%
4-pt Bracket Mean	1%	0%	-3%	-21%	-37%	-71%

MCS was never intended to be used in a final analysis.

McNamee and Celona (1990, pp. 32-33), cautioned that this shortcut should be used in the "early stages" of analyzing a decision and that one needs to carefully assess the distribution and develop a full discretization (using Equal Areas) "more carefully later on!" [emphasis in original].

Implication for Practice: Over time, this guidance has been widely forgotten and today MCS is commonly applied without regard for the shape of the underlying distribution and is not followed with a secondary and more careful assessment and discretization.

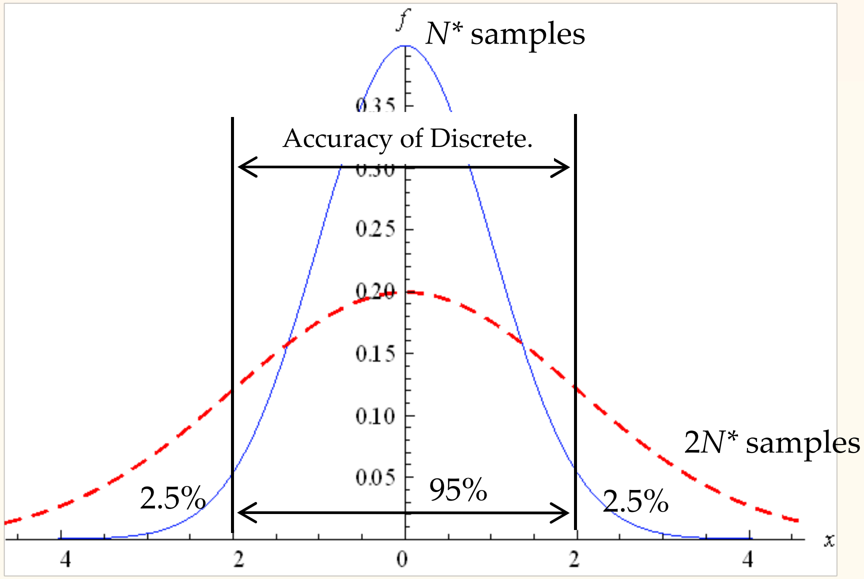
The Decision Analysis Cycle



Swanson-Megill is being used as its advocates have suggested.

We don't have to worry about this, because we simulate everything! Really?

- Discretizations result in *approximation* error. Monte Carlo methods have *sampling* error.
- When we simulate, we compute a mean (the average of all the trials). According to the Central Limit Theorem*, for a large number of samples, this mean is normally distributed.



* The Edgeworth expansion must be used for higher moments and when the underlying pdf is skewed.

The best discretization methods are equivalent to tens of thousands (or more) of Monte Carlo trails.

Raw Moment	EPT				
	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	∞	∞	∞	>1MM	29,499
Second	4,830	>1MM	67,654	57,470	2,781
Third	1,940	∞	24,388	3,918	48,591

Raw Moment	ESM				
	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	∞	∞	∞	>1MM	2,495
Second	2,128	36,165	9,745	1,674	941
Third	855	∞	3,508	498	34,475

Raw Moment	MCS				
	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	∞	∞	∞	1,451	560
Second	30,732	240	14,068	407	676
Third	12,347	∞	5,064	207	35,520

Note: >1MM = "more than 1 million"

Recommendations for Practice

- If possible, use Gaussian Quadrature or moment matching.

- The Equal Areas method (Bracket Mean) is a reasonable approach.

- Direct application of 25-50-25 or 30-40-30 to all distributions, without regard for their shape, should not be considered acceptable practice.
    - Recall, 25-50-20 was never intended to be used this way; 30-40-30, however, was.

- Use of 30-40-30 for lognormal distributions is especially error prone. If you must, please apply it to the LOG of the uncertainty.
    - In this case 30-40-30 is nearly exact.

Are you looking to hire very smart students with deep training in decision analysis?

Look no further!

The Graduate Program in Operations Research at UT has very good (domestic) students looking for summer and full-time jobs in decision analysis.

Robert Hammond is here today. Please spend some time with him!

References

Bickel, J. Eric, Larry W. Lake, and John Lehman. Forthcoming. Discretization, Simulation, and Swanson's (Inaccurate) Mean. *SPE Economics and Management*.

McNamee, P. and Celona, J. 1990. Decision Analysis with Supertree, Second edition. South San Francisco, California: The Scientific Press.

McNamee, P. and Celona, J. 1991. Decision Analysis with Supertree: Instructor's Manual, Second edition. South San Francisco, CA: The Scientific Press.

Miller, A.C., III and Rice, T.R. 1983. Discrete Approximations of Probability Distributions. *Management Science* **29** (3) 352-362.