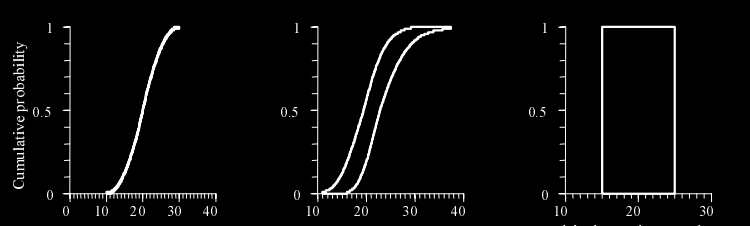
Where do the inputs come from?

Representation of empirical and theoretical information in probability boxes

Scott Ferson, Applied Biomathematics

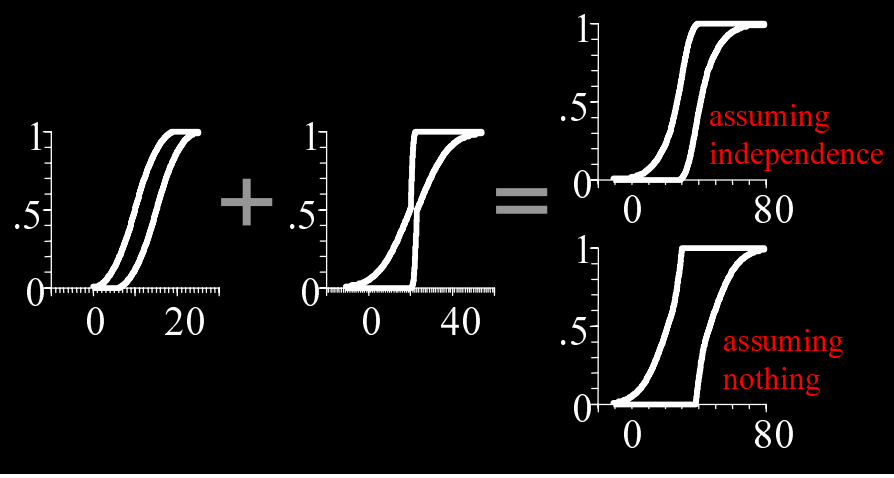
Probability box (p-box)

- Bounds on a cumulative distribution function (CDF)
- Envelope of a Dempster-Shafer structure
- Used in risk analysis and uncertainty arithmetic
- Generalizes probability distributions and intervals



This is an interval, *not* a uniform distribution

Probability bounds analysis (PBA)



PBA handles common problems

- Imprecisely specified distributions
- Poorly known or even unknown dependencies
- Non-negligible measurement error
- Non-detects or other censoring
- Small sample size
- Inconsistency in the quality of input data
- Model uncertainty
- Non-stationarity (non-constant distributions)

- Plus, it's much faster than Monte Carlo simulation

Where do p-boxes come from?

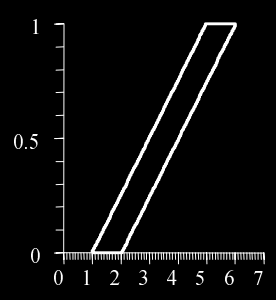
1. Assumption
2. Expert elicitation
3. Robust Bayes
4. Modeling
5. Constraints
6. Observations

Properties

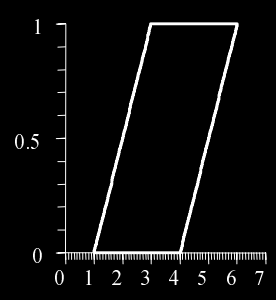
- **Rigor Preserving** (RP): the resultant p-box is rigorous so long as its specifications are
- **Best Possible** (BP): the resultant p-box could not be any tighter without more information
- **Sample Uncertainty** (SU): the resultant p-box represents a statistical confidence claim

1. Assumption

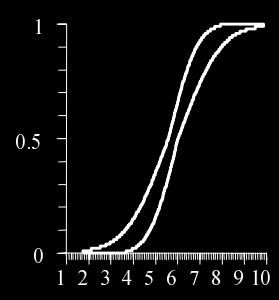
- Make it up
- Bertrand Russell's "Many advantages…" quip
- Precise: just need the CDF
- If parameters are uncertain: envelope (RP,BP)
  - Uniform, triangular, trapezoidal, etc.
  - Normal, Weibull, etc.
  - Wakeby, etc.



uniform
min = [1,2]
max = [5,6]



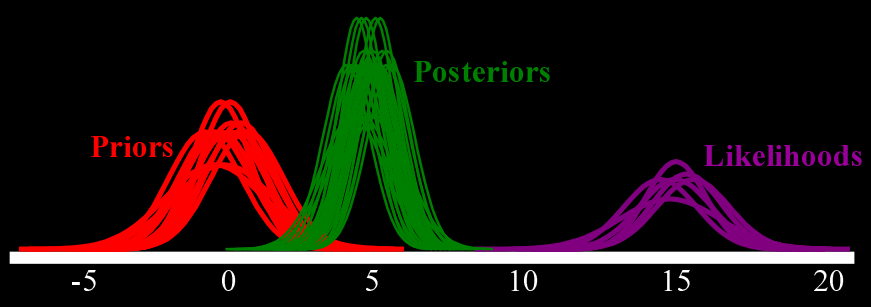
uniform
min = [1,4]
max = [3,6]



normal
mean = [5.6,6]
variance = [1,1.5]

Here the min and the max intervals overlap.
Try *that* in a two-dimensional Monte Carlo!

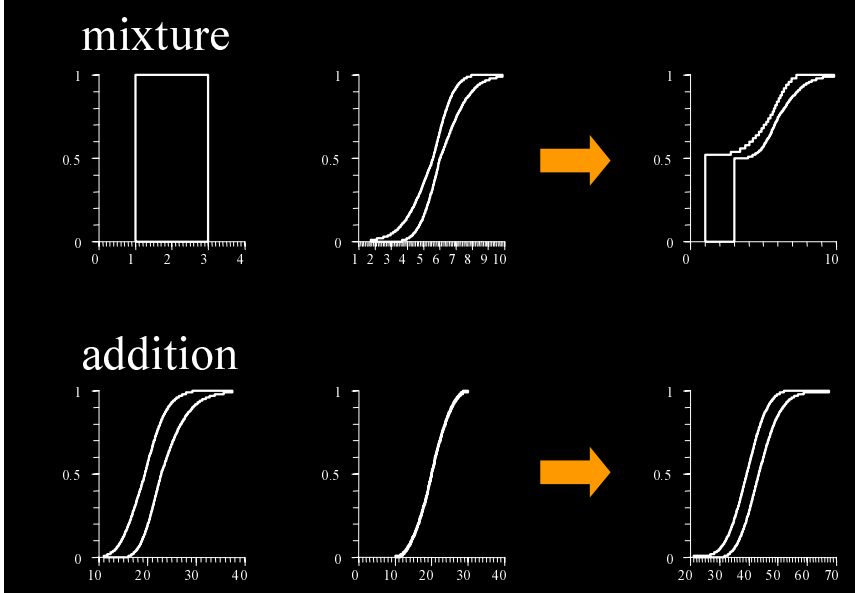
3. Robust Bayes



Class of priors, class of likelihoods ⇒ class of posteriors
Then cumulate the posteriors and take their envelope

4. Modeling

- Break it into subproblems
- Envelope (RP,BP)
- Imposition (RP, BP)
- Mixture
  - Weights precise (RP, BP)
  - If weights uncertain
- Convolution (RP, BP if no repeated parameters)
- Composition
- Deconvolution (generally not BP)



5. Constraints

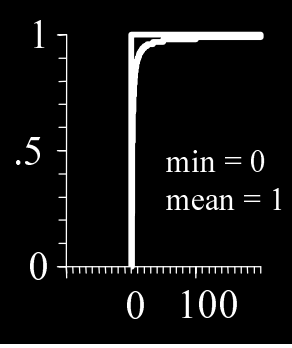
- What's known about the random variable?
- Similar to use of maximum entropy criterion
- Limits on risks (medians and other percentiles)
- Information about other statistics
- Qualitative information on distribution shape
- Assuring convergence (p-box finite)

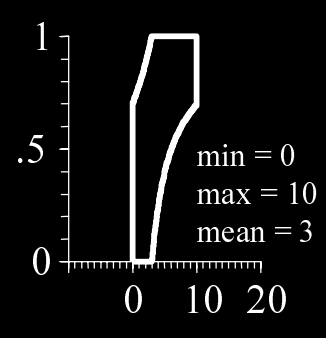
Assuring convergence

- Range limits
- Variance limits
- Density limits

Constraints: range limits

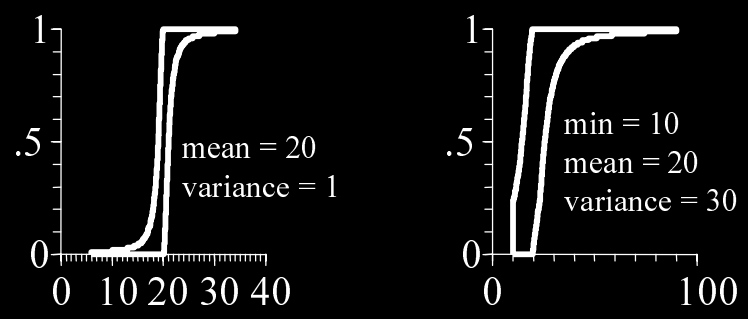
- Markov inequality (RP, BP)
- Rowe's range-mean inequality (RP, BP)





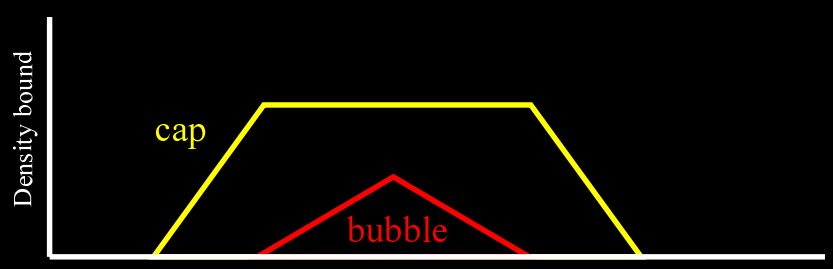
Constraints: variance limits

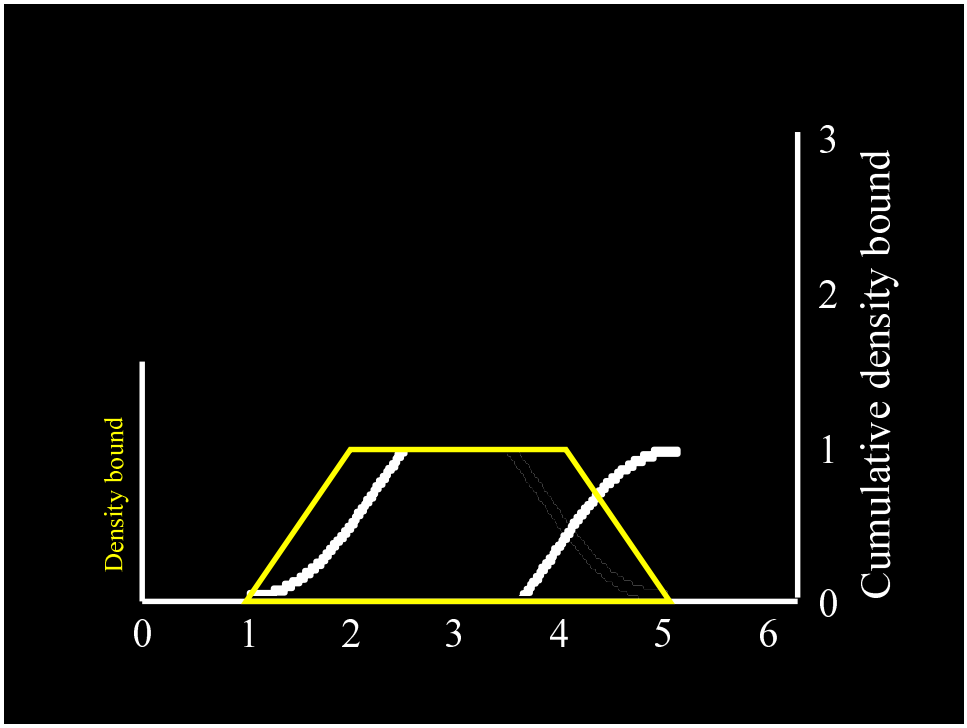
- Chebyshev inequality (RP, BP)
- Cantelli inequalities (RP, BP)
- Karlin and Studden's Chebyshev systems

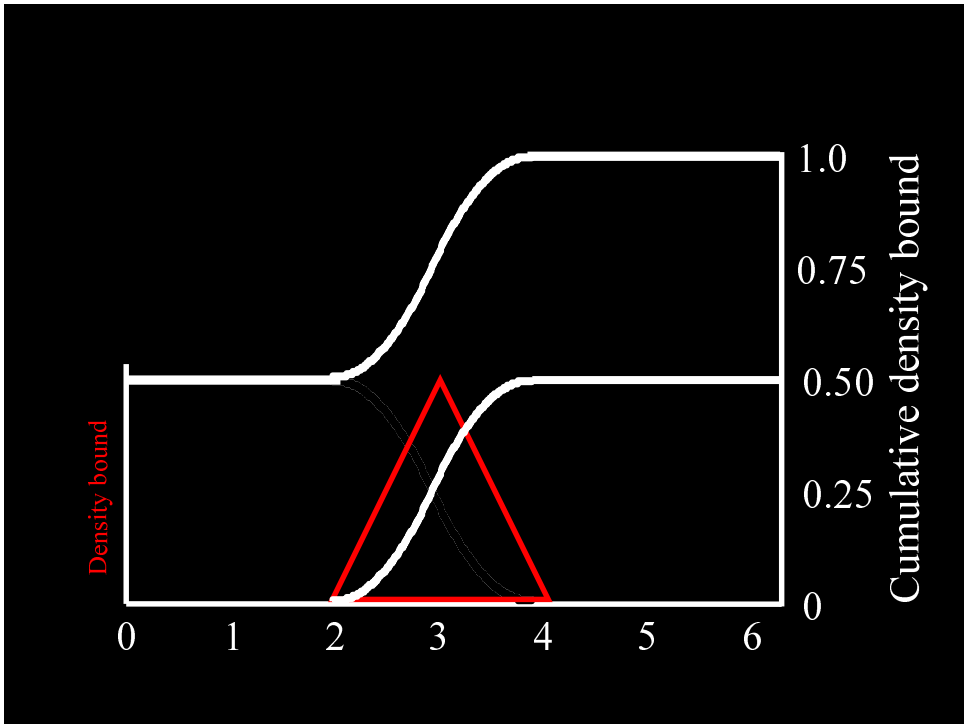


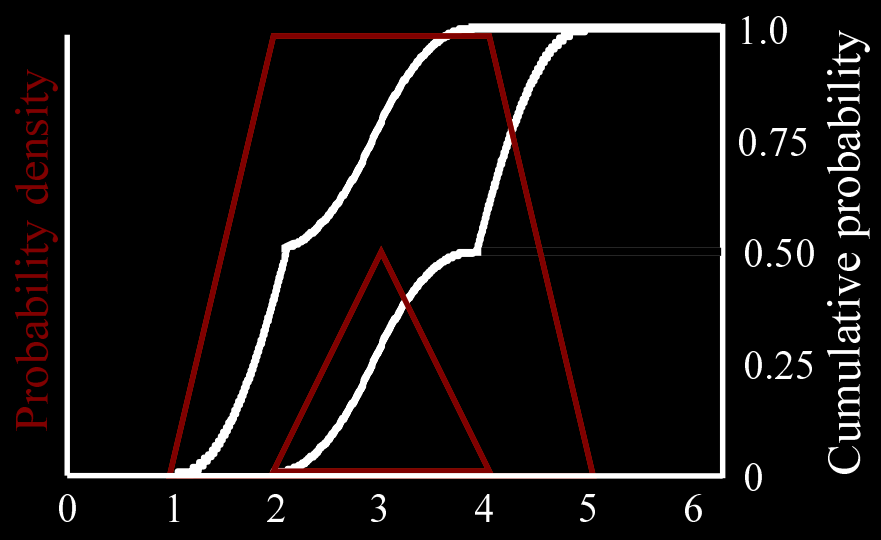
Constraints: density limits

- Upper or lower bounds on probability *density*
- Area of cap must be greater than one
- Area of bubble must be less than one



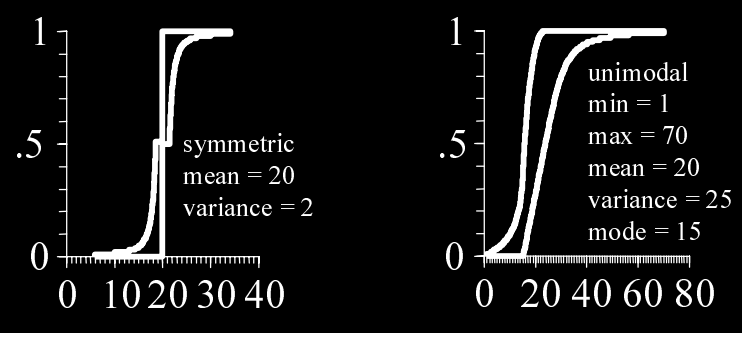

Density bound
Cumulative density bound
0
1
2
3
4
5
6
0
1
2
3


Density bound
Cumulative density bound
0
1
2
3
4
5
6
0
0.25
0.50
0.75
1.0



Qualitative shape information

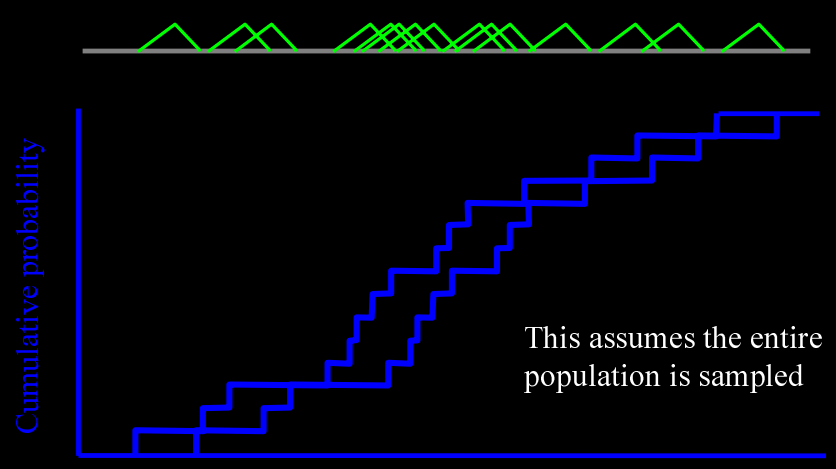
- Positive, symmetric, etc.
- Unimodal (Khinchine's theorem, mode known)
- Increasing or decreasing hazard function



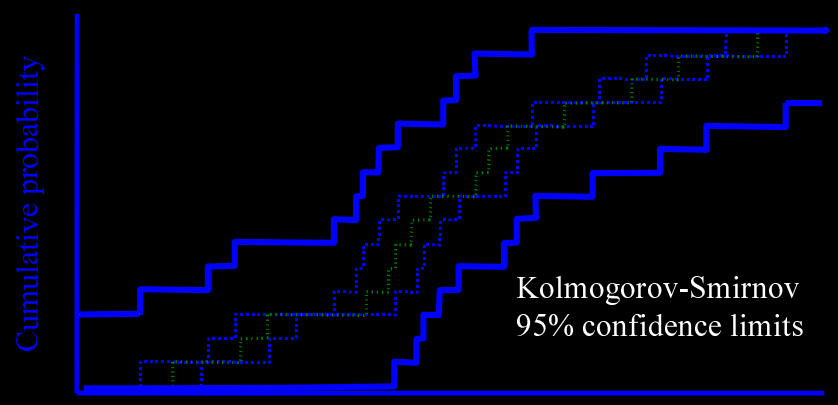
6. Observations

- Empirical information ("just use the data")
- Interval measurement error (RP, BP)
- Censoring (RP, BP)
- Sampling uncertainty with 95% KS limits (SU)
- Extreme value theory to specify tails (SU)
- Verdonck's bootstrapping approach (SU)

Accounting for measurement error



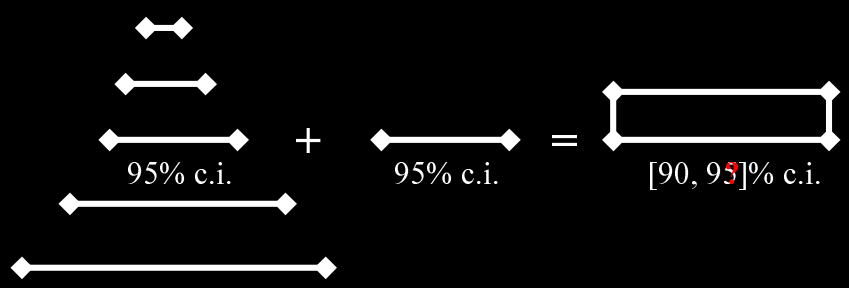
Sampling uncertainty



The problem

- Assuming input intervals enclose their respective values, interval calculations guarantee that the result will also
- In fact, this guarantee rises to the level of mathematical proof
- However, if we use statistical confidence procedures to come up with input intervals, then the condition doesn't hold and we get no guarantee

A simple example



A single confidence level is an incomplete depiction of sample uncertainty. What about all the other levels?

This Christmas tree turns out to be a fuzzy number.

Sampling theory for p-boxes

- Kolmogorov-Smirnov
- Extreme tails à la Gumbel
- Saw et al.'s generalization of Chebyshev

- But need to check whether the propagation methods still make sense when the bounds are not rigorous
- Hybrid numbers (which incorporate both fuzzy and probabilistic uncertainty) may be part of the solution

END