SIPTA Summer School: Engineering Day

Scott Ferson, Institute for Risk and Uncertainty, University of Liverpool

SIPTA Society for Imprecise Probabilities: Theory and Applications

24-28 July 2018, Oviedo, Spain

Engineering

- Publication advice joke
- Practical applications
- Get things done, even if not perfectly
- Random variables, not just events
- Logic, arithmetic, FE and ODE models
- Continuous, infinite
- Quick & dirty is better than elegant & abstract
- Sloppy with notation, but big on computability

Installing R

- Locate the latest version by Googling "R" or going to http://www.r-project.org/
- Download and install the base software
  – For instance, get version 3.5.0 for Windows from
    https://cloud.r-project.org/bin/windows/base/R-3.5.0-win.exe
- Install R by executing this self-extracting file and invoke the program by double-clicking on the blue R icon it leaves on your desktop

Leaving R

- To leave R click File / Exit from the main menu or enter the command **quit()**
- R may ask whether you want to save your work
- If you do, your variables and functions will be available the next time you invoke R

Install the pba.r probability box library

- If you left R, invoke it again
- Enter **rm(list = ls())** to clear its memory
- Click File / Source R code on the main menu
- Locate and Open the file **pba.r**
- You'll get the message "`:pba> library loaded`"

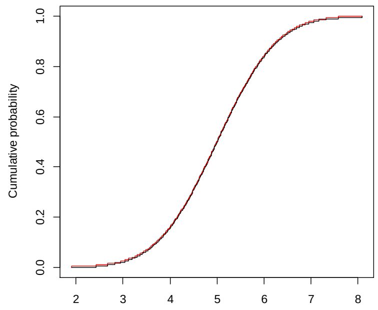
Die-hard RStudio users

- If, against advice, you must use RStudio…
- File / Open file **pbox.r**
- Source (Ctrl+Shift+S) **pbox.r**
- Enter the instruction **RStudio = TRUE**

pba.r probability bounds library

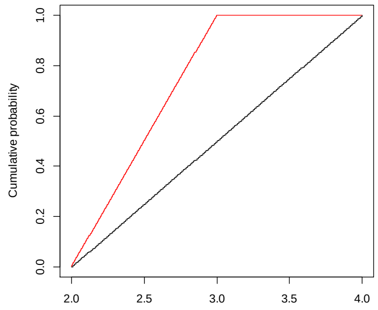
a = normal(5,1)

a

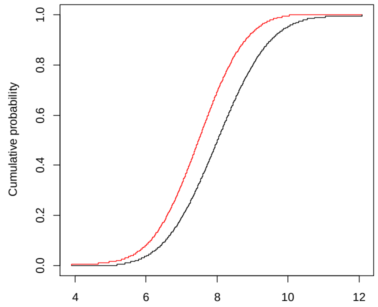


b = uniform(2, interval(3,4))

b



a + b



Generalized convolutions

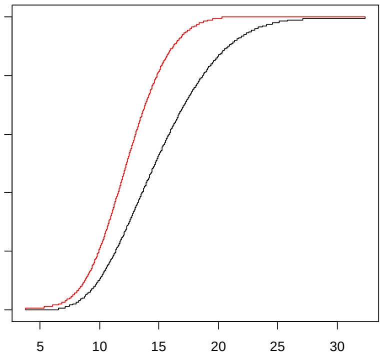
a * b

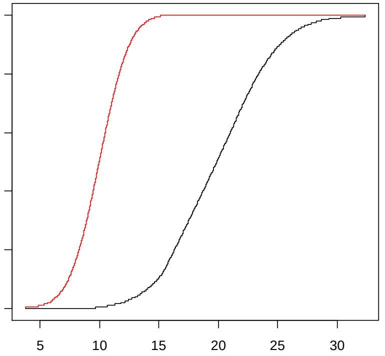
a %*% b

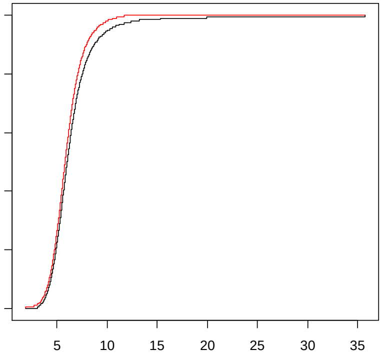
c = a + weibull(8,1) * beta(2,3) / b

c

mean(c)







Probability bounds analysis in pba.r

- Output
  - <enter variable name>, plot, lines, show, summary
- Characterize
  - mean, sd, var, median, quantile, fivenum, left, right, prob, cut, percentile, iqr, random, range
- Compute
  - exp, log, sqrt, abs, round, trunc, ceiling, floor, sign, sin, cos, tan, asin, acos, atan, atan2, reciprocate, negate, +, -, *, /, pmin, pmax, ^, and, or, not, mixture, smin, smax, complement

Probability bounds analysis in pba.r

- Construct
  - <distribution>
  - histogram
  - quantiles
  - pointlist
  - MM <tab> <tab
  - ME <tab> <tab>
  - ML <tab> <tab>
  - CB <tab> <tab>
  - NV <tab> <tab>

Supported named distributions

bernoulli, beta (B), binomial (Bin), cauchy, chi, chisquared, delta, dirac, discreteuniform, exponential, exponentialpower, extremevalue, F, f, fishersnedecor, fishertippett, fisk, frechet, gamma, gaussian, geometric, generalizedextremevalue (GEV), generalizedpareto, gumbel, histogram, inversegamma, laplace, logistic, loglogistic, lognormal (L), logtriangular, loguniform, negativebinomial, normal (N), pareto, pascal, powerfunction, poisson, quantiles, rayleigh, reciprocal, shiftedloglogistic, skewnormal (SN), student, trapezoidal, triangular (T), uniform (U), weibull

Nonparametric p-boxes

maxmean, minmax, minmaxmean, minmean, meanstd, meanvar, minmaxmode, minmaxmedian, minmaxmedianismode, minmaxpercentile, minmaxmeanismedian, minmaxmeanismode, mmms, mmmv, posmeanstd, symmeanstd, uniminmax, unimmmv, unimmms

Optional You don't need to test the installation. Just skip to the next slide

Testing the pba.r installation

- Start a new instance of R, and enter the command **rm(list = ls())**
- Click File / Source R code on the main menu; find and Open the file **pba.r**
- The R Console will display something like

```
> source("C:\\Users\\workshop sra\\pba.r")
:pba> library loaded
```

- Enter **plot(runif(100))** which will open a plot window with random points
- Click History / Recording on the main (if History is missing, click the plot)
- Click on the R Console, click File/Open script, and find and Open **test pba.r**
- Press Ctrl-A to select all its text, then press Ctrl-C to copy it to the clipboard, and close the R Editor window
- Click on the R Console, and press Ctrl-V to paste the text into the R Console for execution
- Click on the plot window, and press PgUp and PgDn to scroll through graphical results of the test calculations
- There should be no error messages on the R Console

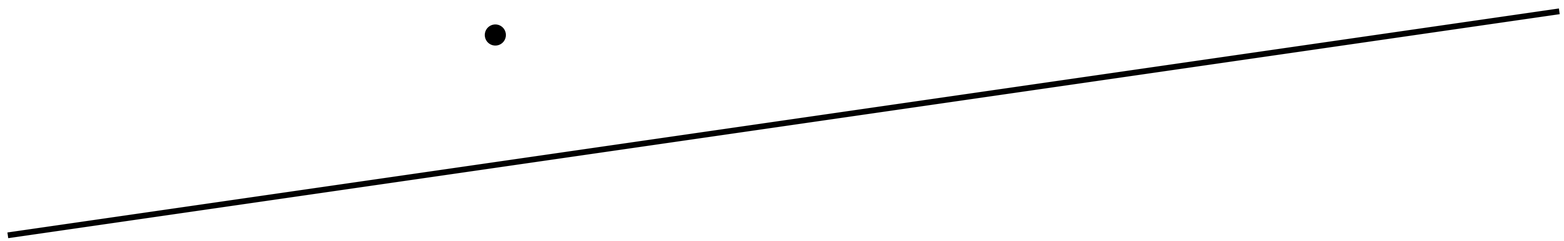
Some help text in the pba.r file

```
# Place this file on your computer and, from within R, select File/Source R code... from the main menu.  Select this file
# and click the Open button to read it into R.  You should see the completion message ":pbox> library loaded".  Once the
# library has been loaded, you can define probability distributions
#       a = normal(5,1)
#       b = uniform(2,3)
# and  p-boxes
#       c = meanvariance(0,2)
#       d = lognormal(interval(6,7), interval(1,2))
#       e = mmms(0,10,3,1)
# and perform mathematical operations on them, including the Frechet convolution such as
#       a  %+%  b
# or a traditional convolution assuming independence
#       a  %|+|%  b
# If you do not enclose the operator inside percent signs or vertical bars, the software tries to figure out how the
# arguments are related to one another. Expressions such as
#       a + b
#       a + log(a) * b
# autoselect the convolution to use.  If pba.r can’t tell what dependence the arguments have, it uses Frechet convolution.
# Variables containing probability distributions or p-boxes are assumed to be independent of one another unless one
# formally depends on the other, which happens if one was created as a function of the other. Assigning a variable
# containing a probability distribution or a p-box to another variable makes the two variables perfectly dependent.  To
# make an independent copy of the distribution or p-box, use the samedistribution function, e.g., c = samedistribution(a).
# By default, separately constructed distributions such as
#       a = normal(5,1)
#       b = uniform(2,3)
# will be assumed to be independent (so their convolution a+b will be a precise distribution).  You can acknowledge any
```

Two kinds of uncertainty

Euclid

Given a line in a plane, how many parallel lines can be drawn through a point not on the line?



For over twenty centuries, the answer was *one*

Relax one axiom

- Non-Euclidean geometries say *zero* or *many*
- At first, very controversial
- Mathematics richer and wider applications
- Used by Einstein in general relativity

Crossroads in uncertainty theory

- There is a kind of uncertainty that cannot be handled by traditional Laplacian probability

- Only one axiom changes
- Collision of different views about uncertainty
- Richer math, wider applications
- More reasonable and more reliable results

Two kinds of uncertainty

- Variability
- Aleatory uncertainty
- Type A uncertainty
- Stochasticity
- Randomness
- Chance
- Risk

- Incertitude
- Epistemic uncertainty
- Type B uncertainty
- Ambiguity
- Ignorance
- Imprecision
- True uncertainty

Variability = aleatory uncertainty

- Arises from natural stochasticity

- Variability arises from
  - spatial variation
  - temporal fluctuations
  - manufacturing or genetic differences

- Not reducible by empirical effort

Incertitude = epistemic uncertainty

- Arises from incomplete knowledge

- Incertitude arises from
  - limited sample size
  - mensurational limits (‘measurement uncertainty’)
  - use of surrogate data

- Reducible with empirical effort

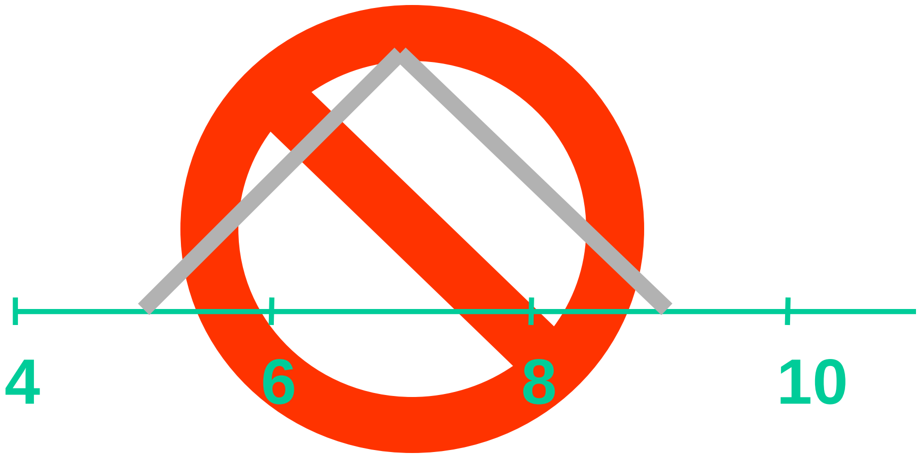
Propagating incertitude

Suppose

A is in [2, 4]

B is in [3, 5]

What can be said about the sum *A*+*B*?



The right answer for risk analysis is [5,9]

Must be treated *differently*

- Variability should be modeled as randomness with the methods of probability theory

- Incertitude should be modeled as ignorance with methods of interval or constraint analysis

- Probability bounding can do both at once

Probability bounds analysis

Bounding probability is an old idea

- Boole and de Morgan
- Chebyshev and Markov
- Borel and Fréchet
- Kolmogorov and Keynes
- Berger and Walley

Why bounding is a good idea

- Often sufficient to specify a decision
- Possible even when estimates are impossible
- Usually easy to compute and simple to combine
- Rigorous, rather than an approximation
- Bounding works with even the crappiest data

(after N.C. Rowe 1988)

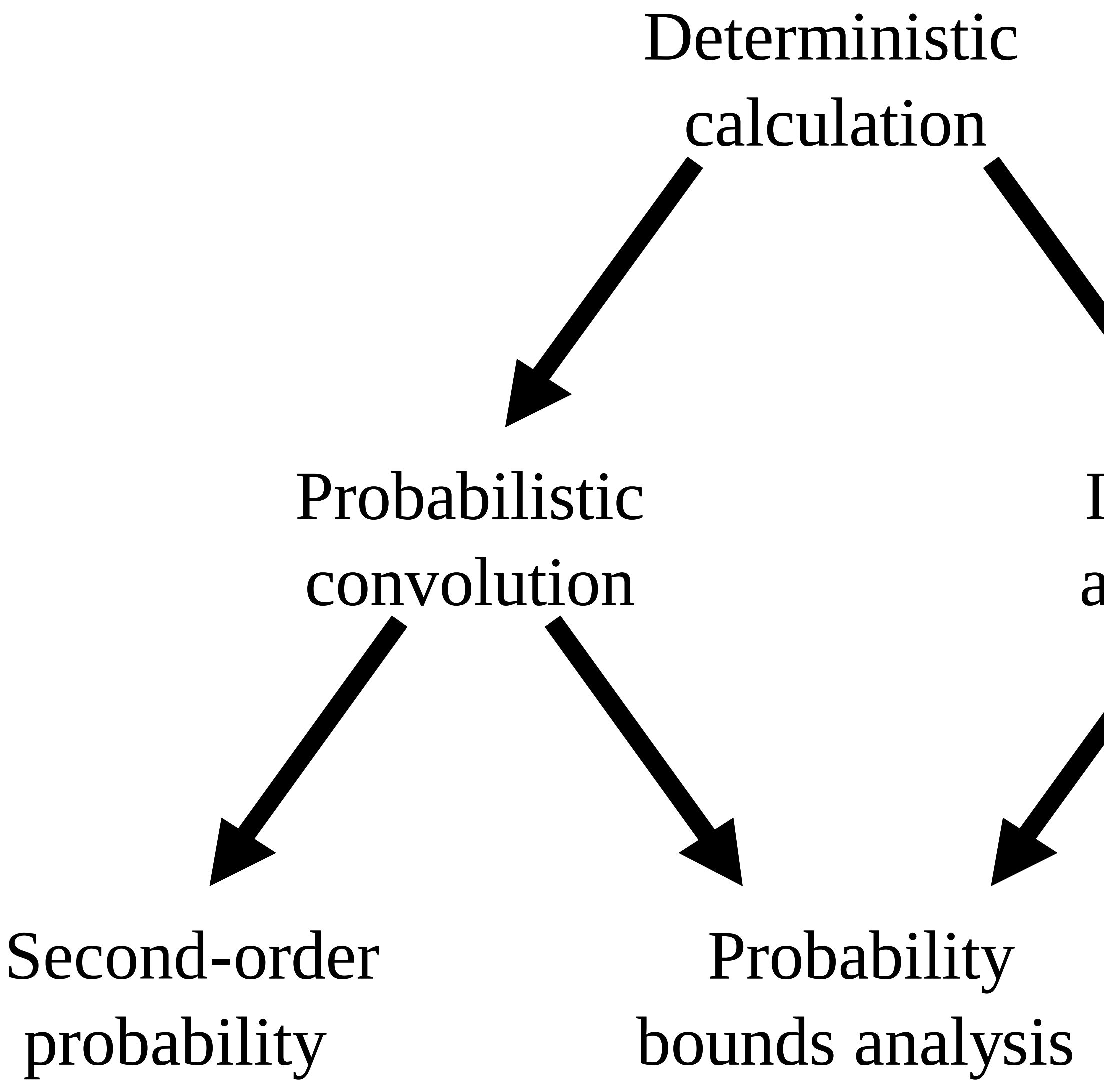
Rigorousness

- The computations can be guaranteed to enclose the true results (if the inputs do)
- “Automatically verified calculations”
- You can still be wrong, but the method won’t be the reason if you are

Closely related to other ideas

- ***Second-order probability***
  - PBA is easier to work with and more comprehensive
- ***Imprecise probabilities***
  - PBA is somewhat cruder, but a lot easier
- ***Robust Bayesian analysis***
  - PBA does convolutions rather than updating

Bounding approaches like PBA

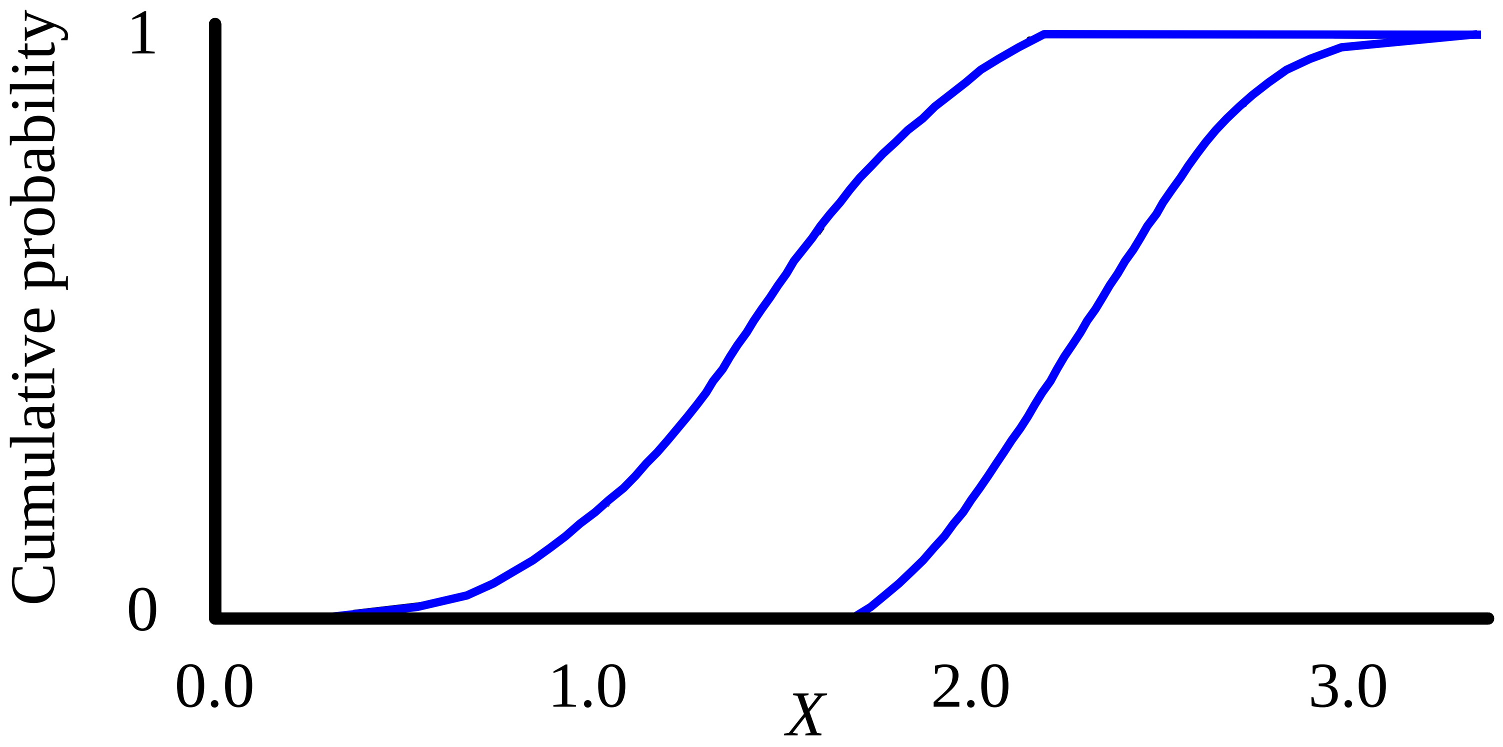

Deterministic calculation
Probabilistic convolution
Interval analysis
Second-order probability
Probability bounds analysis

Probability bounds

- Bridge between qualitative and quantitative
- When data are abundant, it works like probability theory
- When data are sparse, it yields both conservative and optimistic results
- Easy to set up, and cheap to implement

What is a probability box (p-box)?

Interval bounds on a cumulative distribution function (CDF)



P-box: mathematical definition

set of intervals

space of all distributions

$\{\overline{F}, \underline{F}, m, v, \mathbf{F}\}$ where $\overline{F}, \underline{F} \in \mathbb{D}$, $m, v \in \mathbb{I}$, $\mathbf{F} \subseteq \mathbb{D}$, denoting a set of distributions $F \in \mathbb{D}$ matching

$$\underline{F}(x) \leq F(x) \leq \overline{F}(x),$$

$$\int_{-\infty}^{\infty} x \, \mathrm{d}F(x) \in m,$$

$$\int_{-\infty}^{\infty} x^2 \, \mathrm{d}F(x) - \left(\int_{-\infty}^{\infty} x \, \mathrm{d}F(x)\right)^2 \in v, \text{ and}$$

$$F \in \mathbf{F}$$

$\mathbb{D} = \{D \mid D : \mathbb{R} \rightarrow [0,1], D(x) \leq D(y) \text{ whenever } x < y, \text{ for all } x, y \in \mathbb{R}\}$

$\mathbb{I} = \{i \mid i = [i_1, i_2], i_1 \leq i_2, i_1, i_2 \in \mathbb{R}\}$

Riemann–Stieltjes integrals don't depend on the differentiability of F

Probability bounds analysis

- It's *not* worst case analysis (distribution tails)
- Marries intervals with probability theory
- Distinguishes variability and incertitude
- Solves many problems in risk analysis
  - Input distributions unknown
  - Imperfectly known correlation and dependency
  - Large measurement error, censoring, small sample sizes
  - Model uncertainty

Calculations

- All standard mathematical operations
  - Arithmetic ($+$, $-$, $\times$, $\div$, ^, min, max)
  - Logical operations (and, or, not, if, etc.)
  - Transformations (exp, ln, sin, tan, abs, sqrt, etc.)
  - Backcalculation (deconvolutions, updating)
  - Magnitude comparisons ($<$, $\leq$, $>$, $\geq$, $\subseteq$)
  - Other operations (envelope, mixture, etc.)
- Quicker than Monte Carlo
- Guaranteed to bound the answer
- Optimal solutions often easy to compute

Example: uncontrolled fire

$$F = A \ \& \ B \ \& \ C \ \& \ D$$

Probability of ignition source

Probability of abundant fuel presence

Probability fire detection not timely

Probability of suppression system failure

Imperfect information

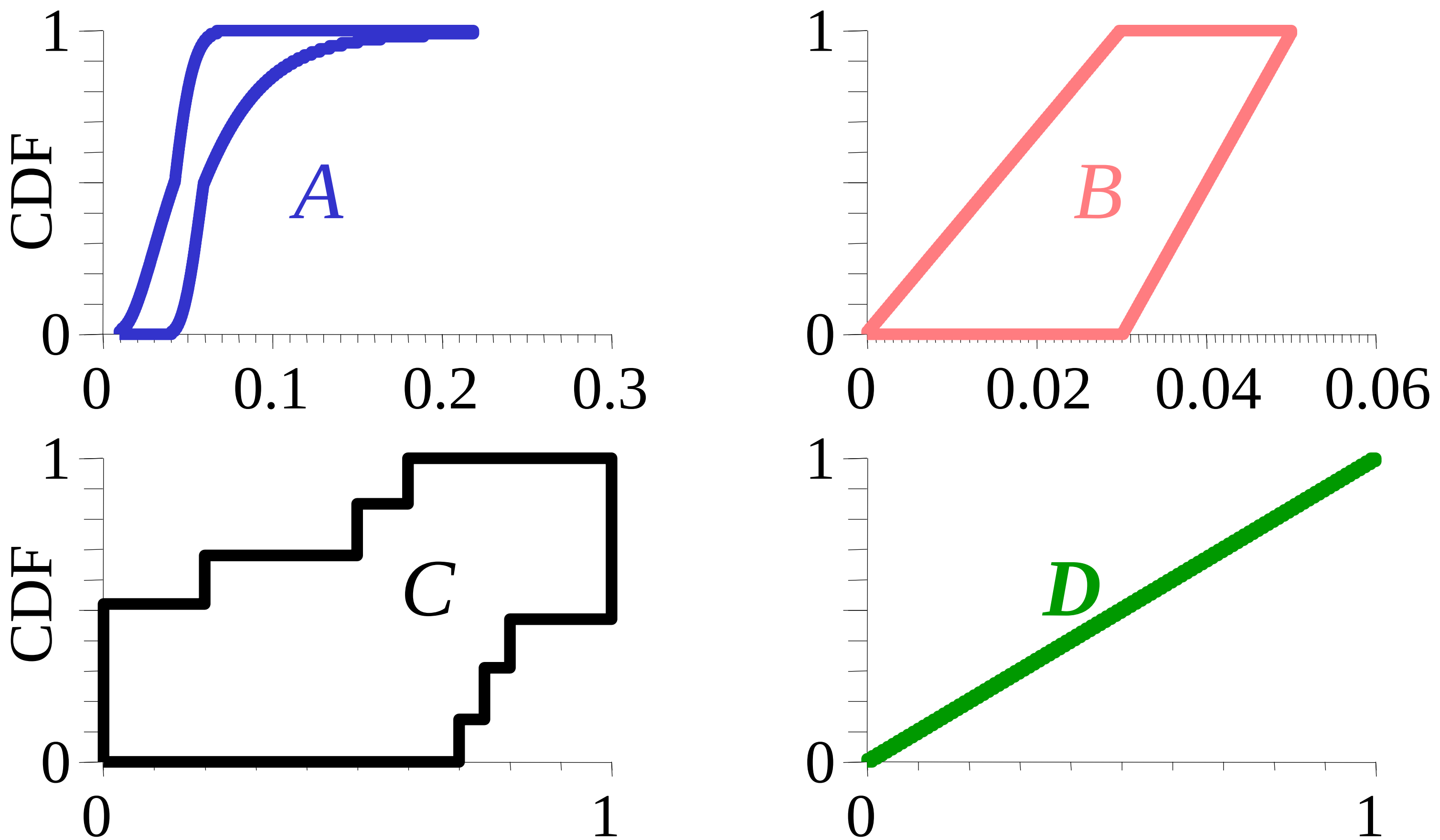
- Calculate $A \& B \& C \& D$, with partial information:
  - A's distribution is known, but not its parameters
  - B's parameters known, but not its shape
  - C has a small empirical data set
  - D is known to be a precise distribution
- Bounds assuming independence?
- Without any assumption about dependence?

A = {lognormal, mean = [.05,.06], variance = [.0001,.001])

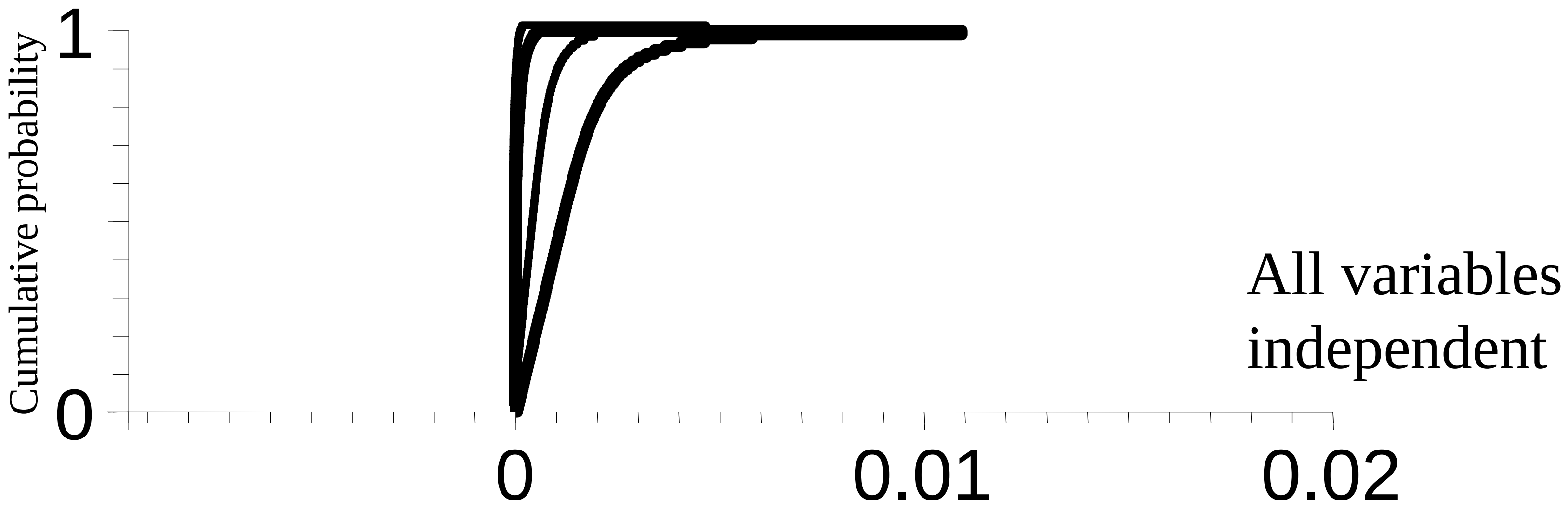
B = {min = 0, max = 0.05, mode = 0.03}

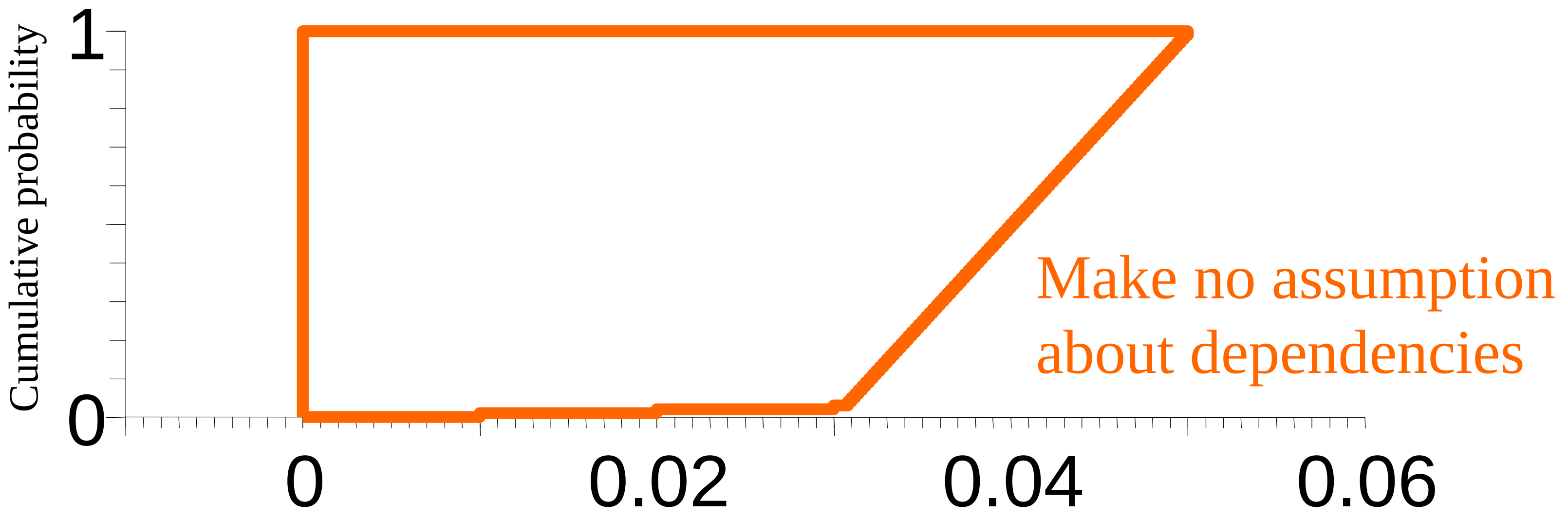
C = {sample data = 0.2, 0.5, 0.6, 0.7, 0.75, 0.8}

D = uniform(0, 1)



Resulting answers





Summary statistics

Independent	
Range	[0, 0.011]
Median	[0, 0.00113]
Mean	[0.00006, 0.00119]
Variance	[2.9×10^{-9}, 2.1×10^{-6}]
Standard deviation	[0.000054, 0.0014]
No assumptions about dependence	
Range	[0, 0.05]
Median	[0, 0.04]
Mean	[0, 0.04]
Variance	[0, 0.00052]
Standard deviation	[0, 0.023]

How to use the results

When uncertainty makes no difference

(because results are so clear), bounding gives confidence in the reliability of the decision

When uncertainty prevents a decision

(*i*) use other criteria within probability bounds, or

(*ii*) use results to identify inputs to study better

Can uncertainty swamp the answer?

- Sure, if uncertainty is huge
- *This should happen* (it's not "unhelpful")
- If you think the bounds are too wide, then put in whatever information is missing
- If there isn't any such information…
  do you want the results to mislead people?

Using the pba.r library

A = lognormal(i(.05,.06), sqrt(i(.0001,.001)))

B = minmaxmode(0,0.05,.03)

C = CBbinomial(2,4)

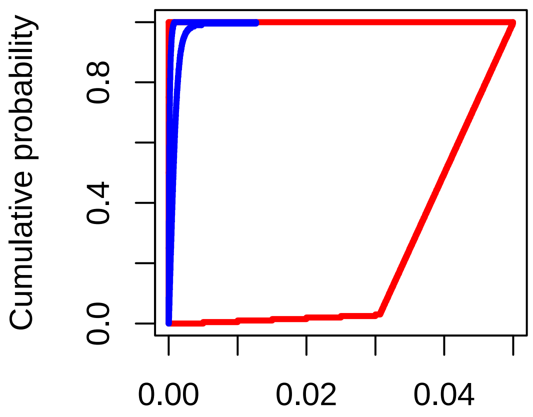
D = uniform(0,1)

fi = A %|&|% B %|&|% C %|&|% D

f = A %&% B %&% C %&% D

red(f)

blue(fi)



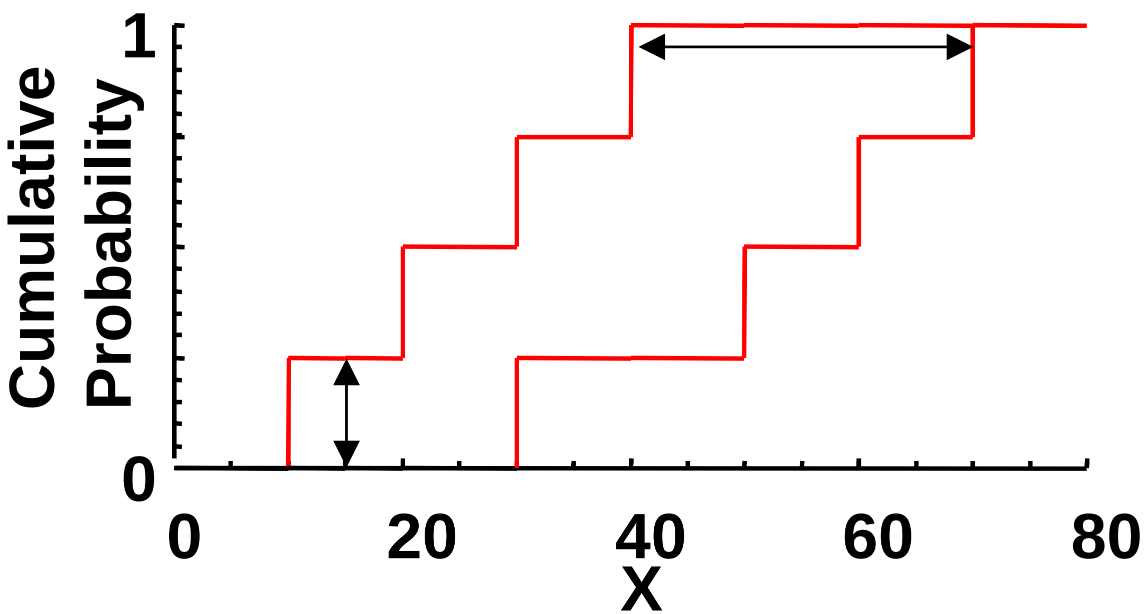
Duality

- Bounds on the probability at a value

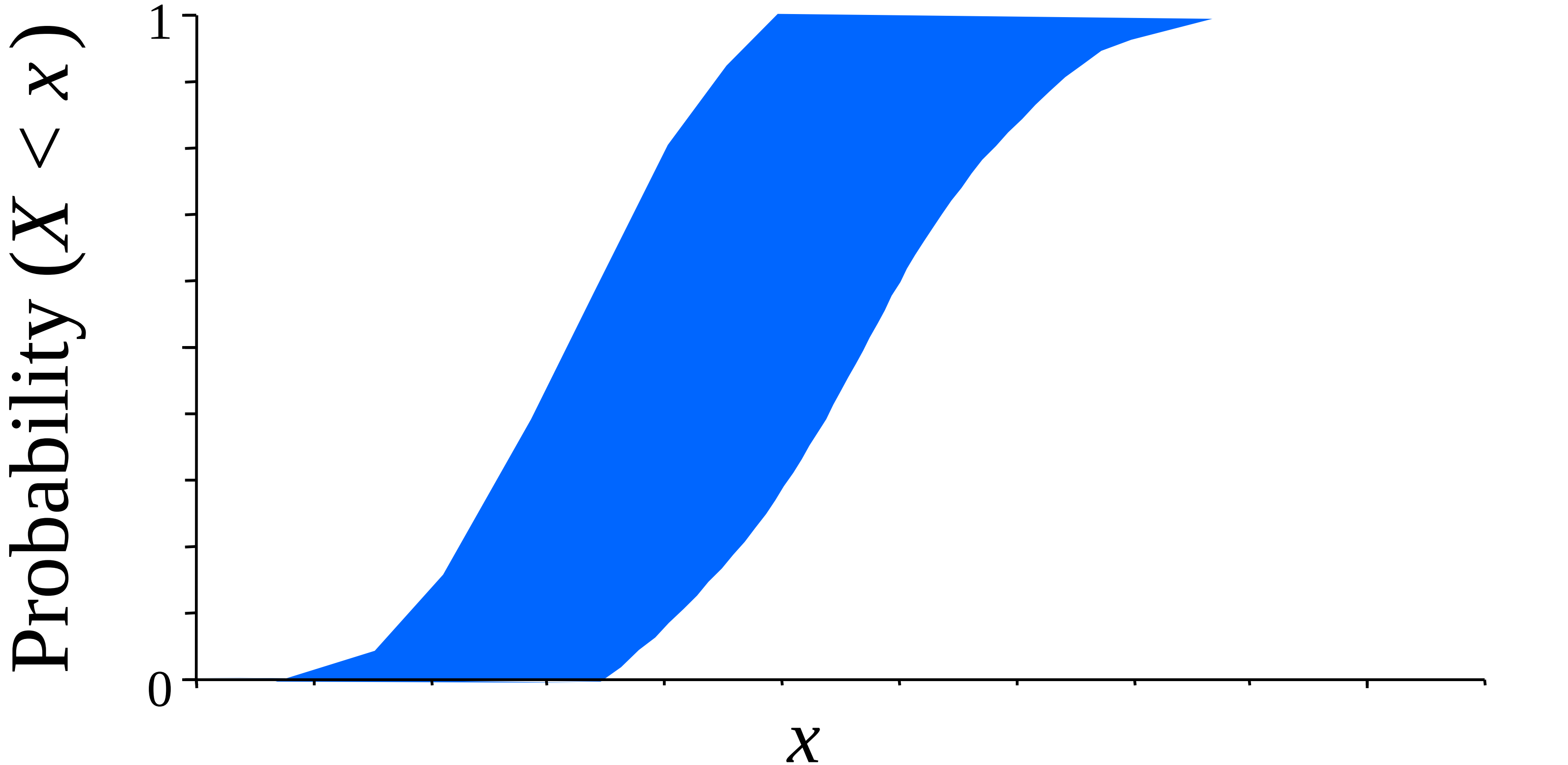
  Chance the value will be 15 or less is between 0 and 25%

- Bounds on the value at a probability

  95^{th} percentile is between 40 and 70

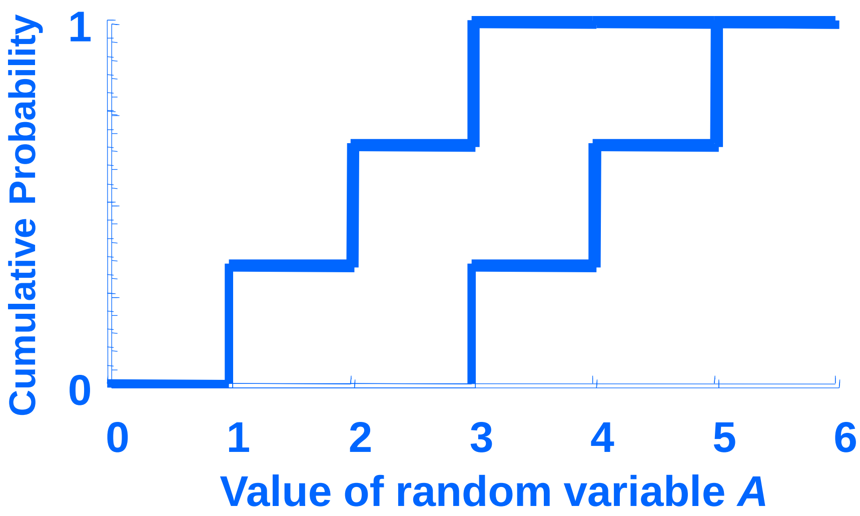


Breadth (incertitude) v. tilt (variability)

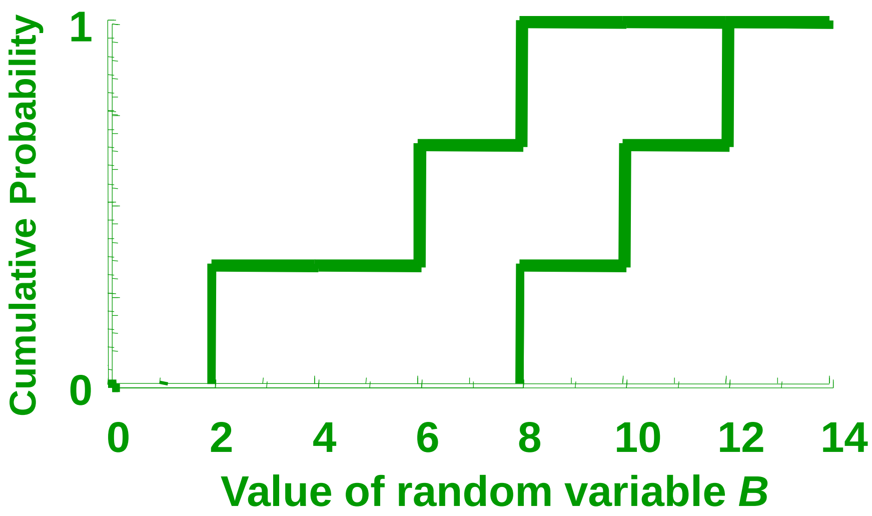


Probability bounds arithmetic

P-box for random variable *A*



P-box for random variable *B*

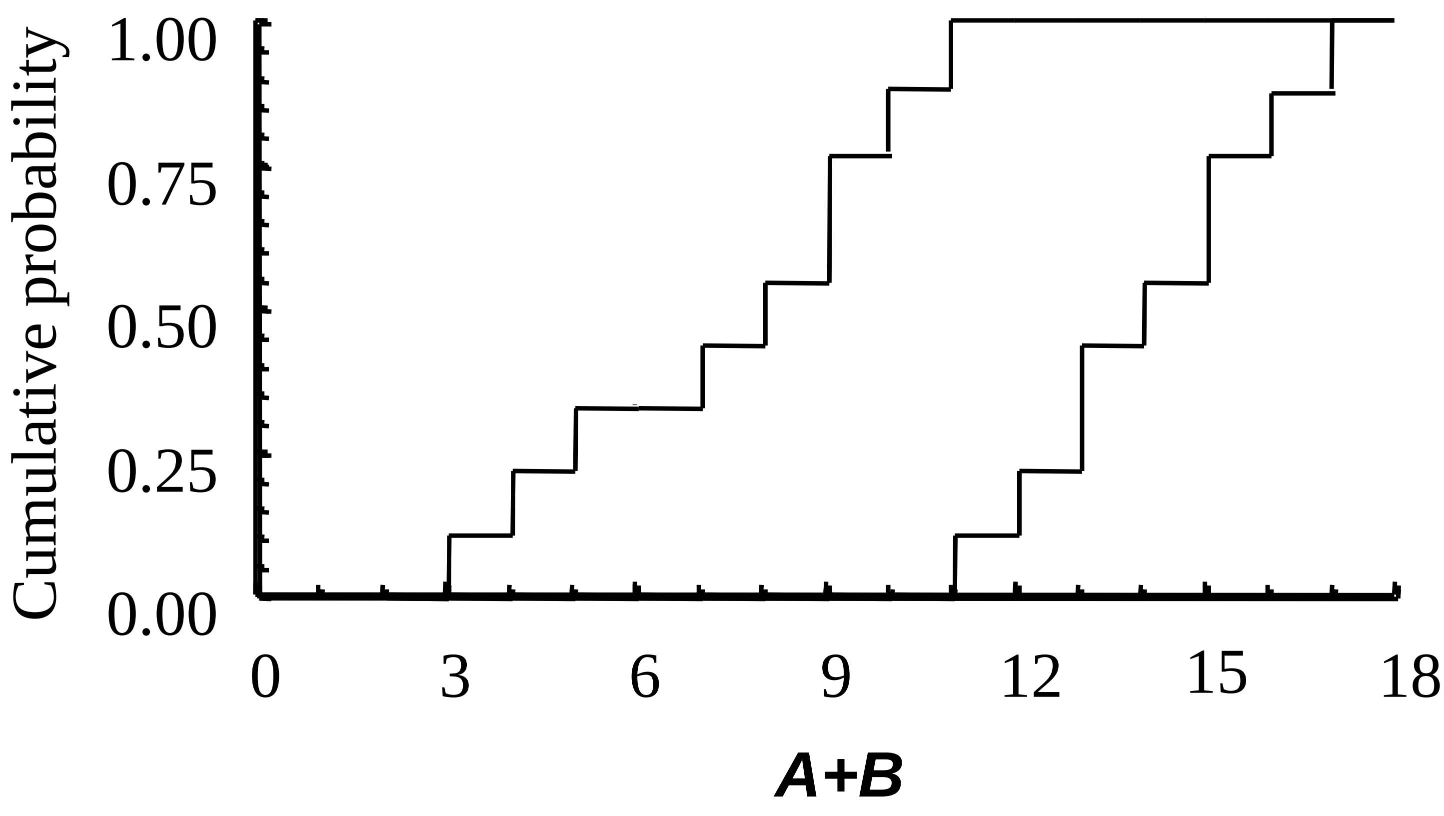


What are the bounds on the distribution of the sum of *A*+*B*?

Cartesian product

$A+B$ independence	$A\in[1,3]$ $p_1 = 1/3$	$A\in[2,4]$ $p_2 = 1/3$	$A\in[3,5]$ $p_3 = 1/3$
$B\in[2,8]$ $q_1 = 1/3$	$A+B\in[3,11]$ prob=1/9	$A+B\in[4,12]$ prob=1/9	$A+B\in[5,13]$ prob=1/9
$B\in[6,10]$ $q_2 = 1/3$	$A+B\in[7,13]$ prob=1/9	$A+B\in[8,14]$ prob=1/9	$A+B\in[9,15]$ prob=1/9
$B\in[8,12]$ $q_3 = 1/3$	$A+B\in[9,15]$ prob=1/9	$A+B\in[10,16]$ prob=1/9	$A+B\in[11,17]$ prob=1/9

A+*B* under independence

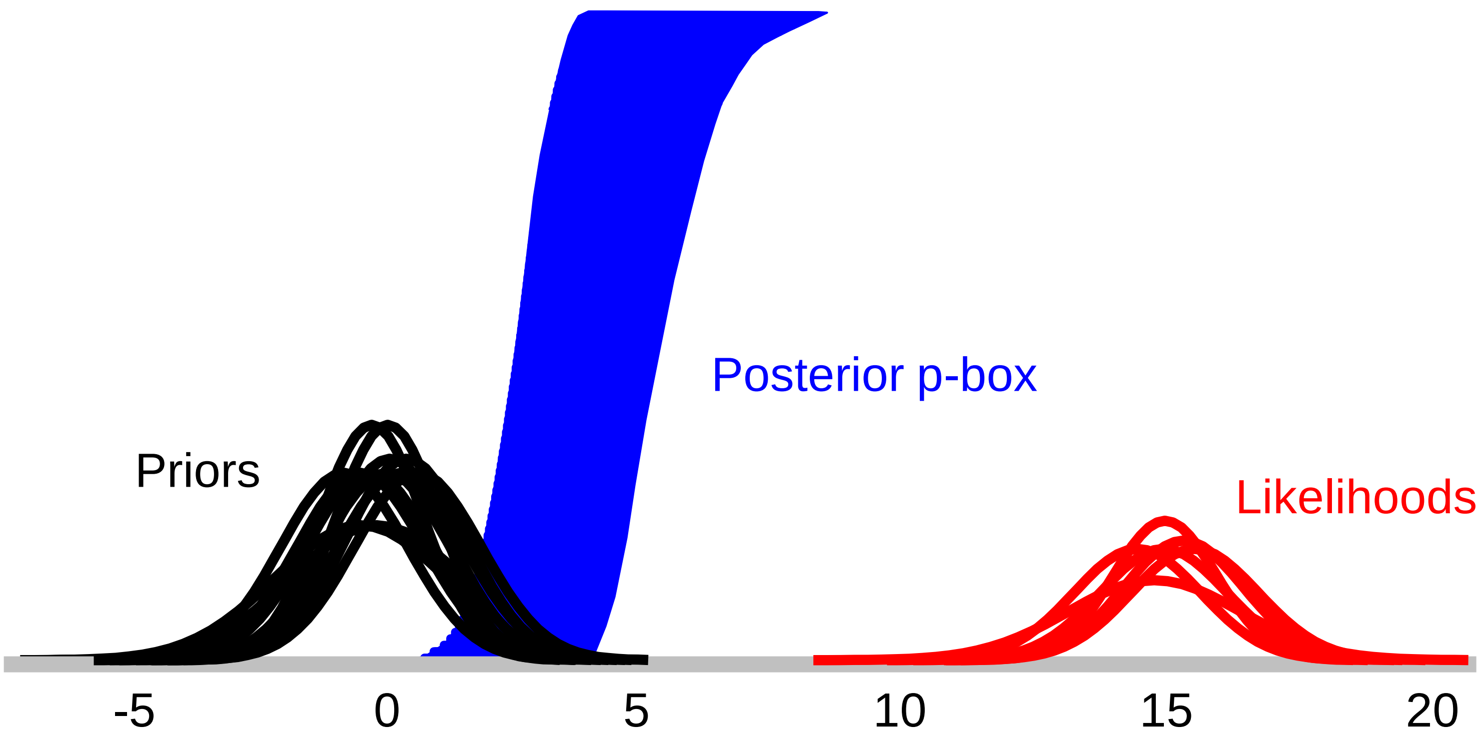


Where do inputs come from?

Where do input p-boxes come from?

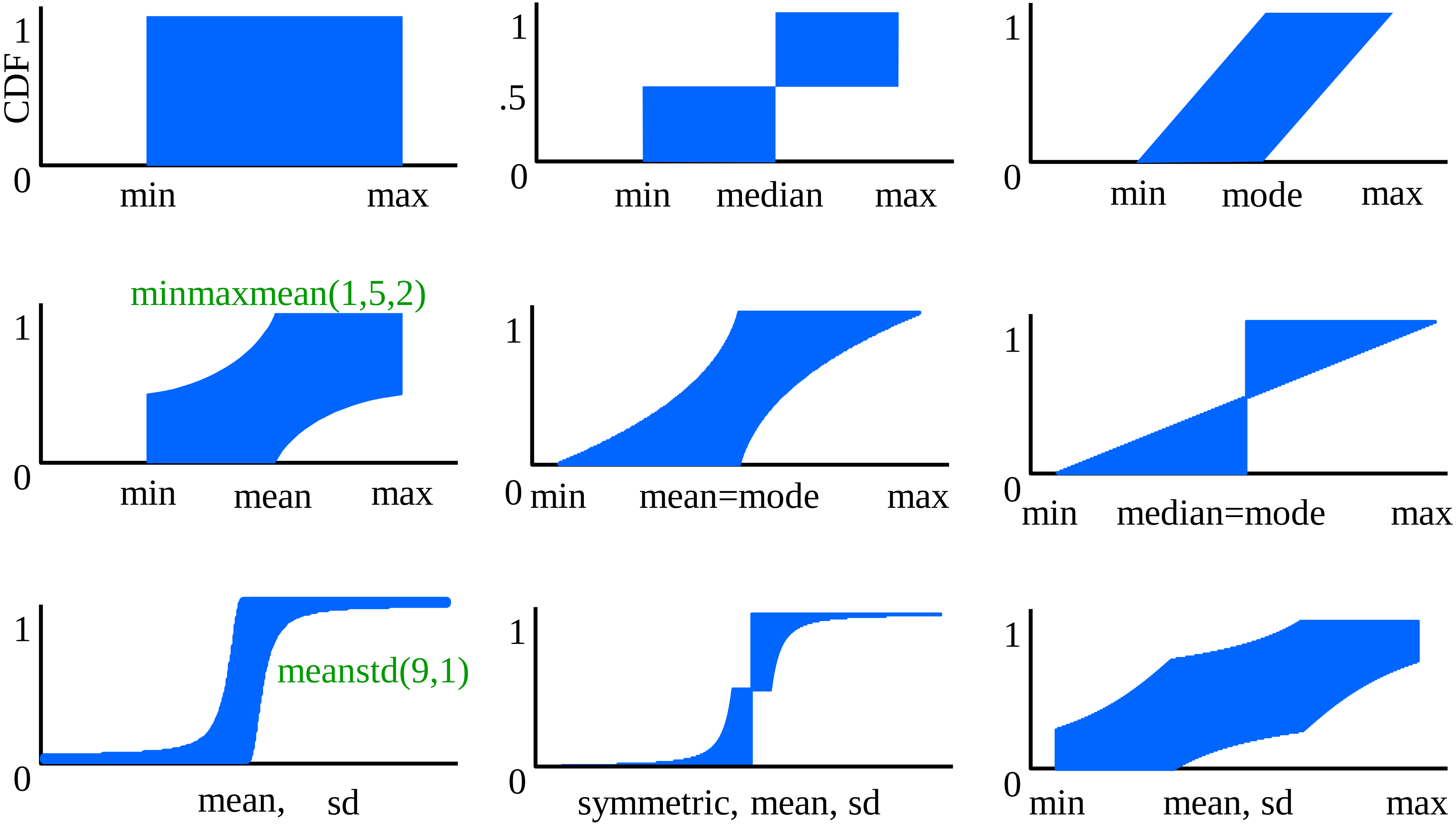
- Prior modeling
  - Uncertainty about dependence
  - Robust Bayes analysis
- Constraint information
  - Summary publications lacking original data
- Sparse or imprecise data
  - Shallow likelihood functions to maximize
  - Measurement uncertainty, censoring, missing data
  - Inferential (sampling) uncertainty

Robust Bayes can make a p-box

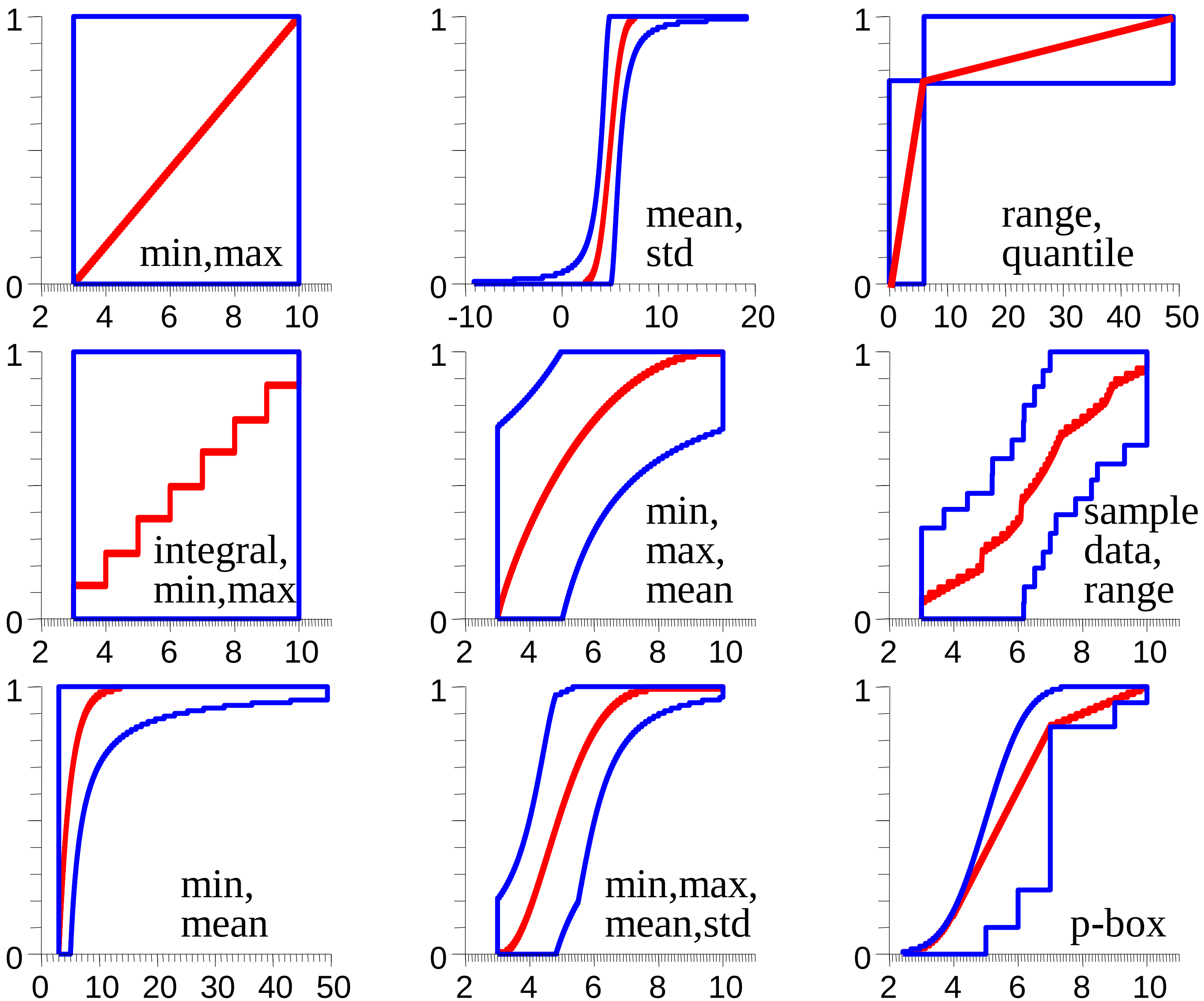


class of priors, class of likelihoods $\Rightarrow$ class of posteriors

Constraint propagation



Comparing p-boxes with maximum entropy distributions



Maximum entropy's problem

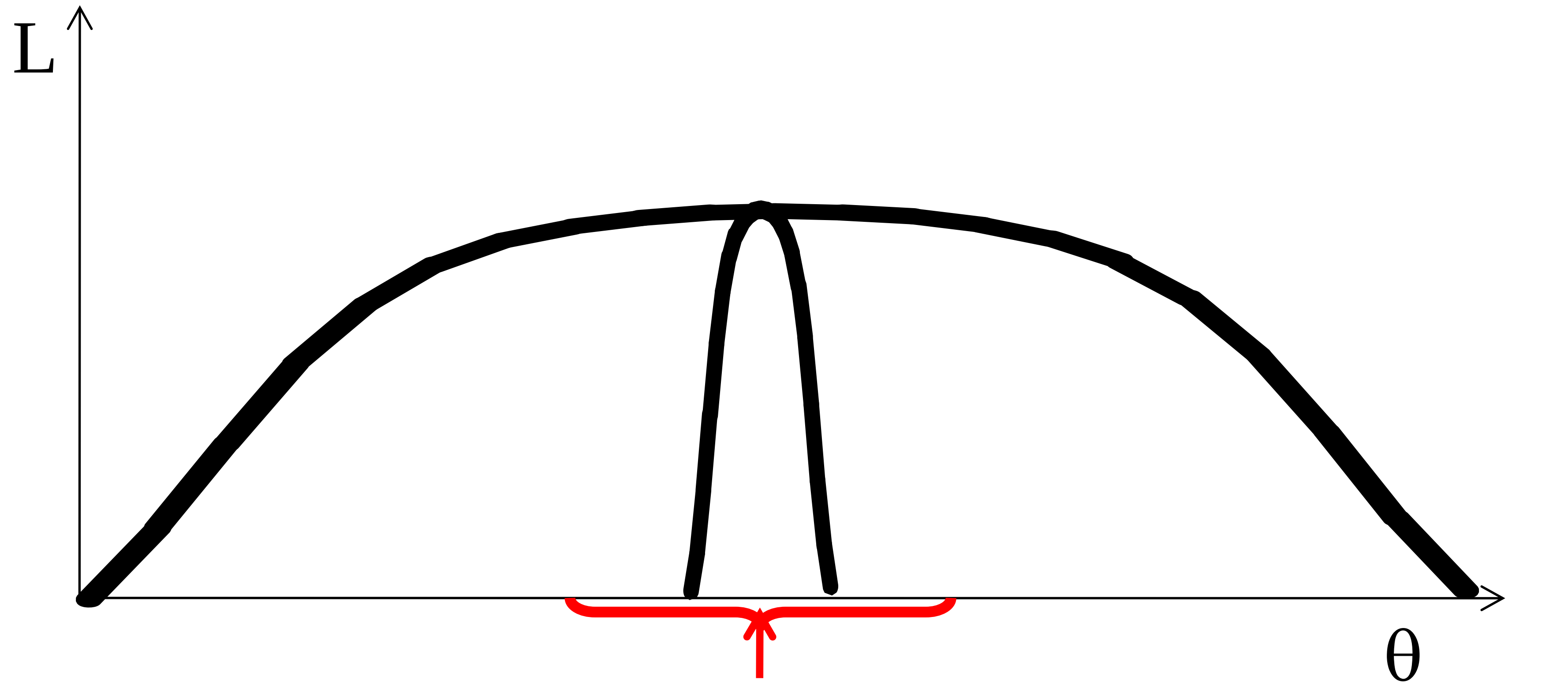
- Depends on the choice of scale

- A solution in terms of degradation rate is incompatible with one based on half life even though the information is equivalent

- P-boxes are the *same* whichever scale is used

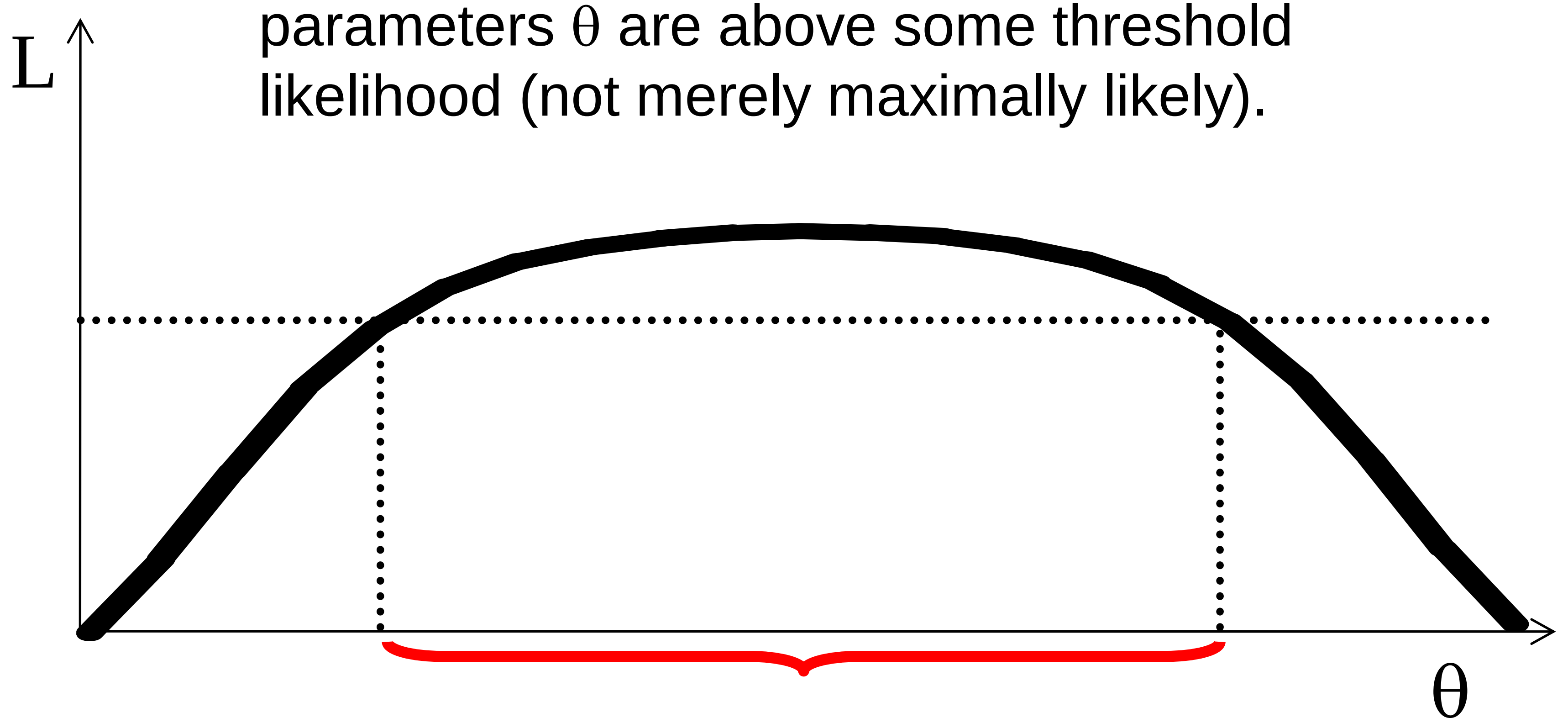
Warner North interprets Ed Jaynes as saying that "two states of information that are judged to be equivalent should lead to the same probability assignments". Maxent doesn't do this! But PBA does.

Sparse data yield shallow likelihoods



Treelining likelihood makes a p-box

The p-box encloses all distributions whose parameters θ are above some threshold likelihood (not merely maximally likely).

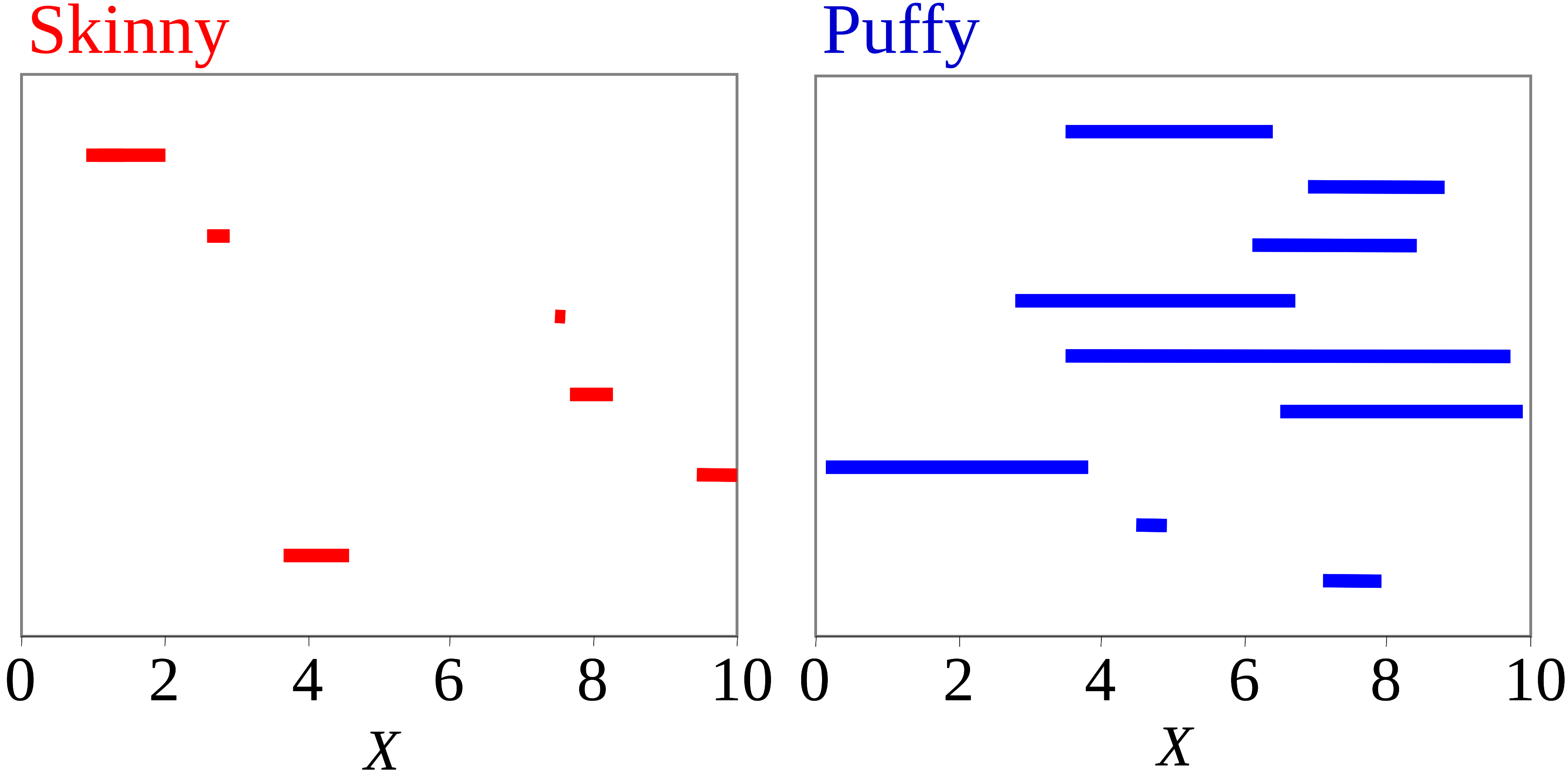


Sources of incertitude in data

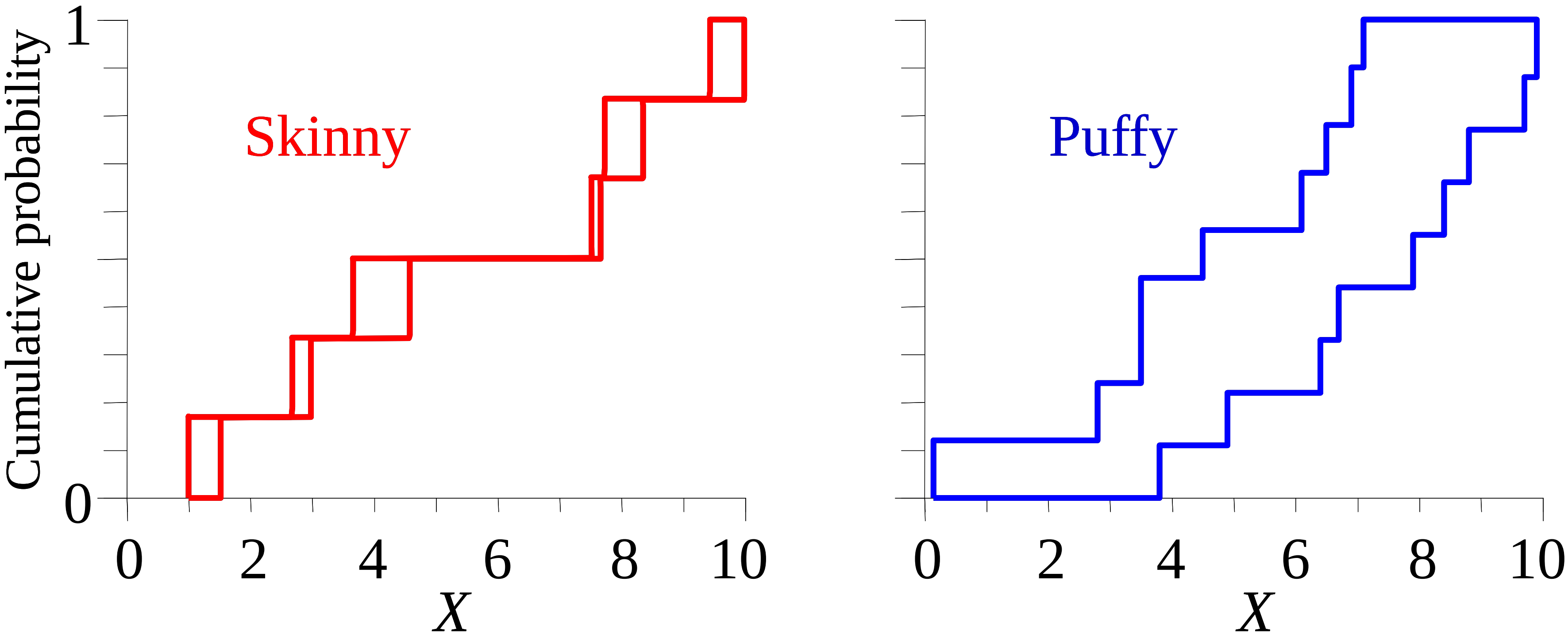
- Periodic observations
  When did the fish in my aquarium die during the night?
- Plus-or-minus measurement uncertainties
  Coarse measurements, measurements from digital readouts
- Non-detects and data censoring
  Chemical detection limits, studies prematurely terminated
- Privacy requirements
  Epidemiological or medical information, census data
- Theoretical constraints
  Concentrations, solubilities, probabilities, survival rates
- Bounding studies
  Presumed or hypothetical limits in what-if calculations

Imprecise sample data

Skinny data	**Puffy data**
[1.00, 2.00]	[3.5, 6.4]
[2.68, 2.98]	[6.9, 8.8]
[7.52, 7.67]	[6.1, 8.4]
[7.73, 8.35]	[2.8, 6.7]
[9.44, 9.99]	[3.5, 9.7]
[3.66, 4.58]	[6.5, 9.9]
	[0.15, 3.8]
	[4.5, 4.9]
	[7.1, 7.9]

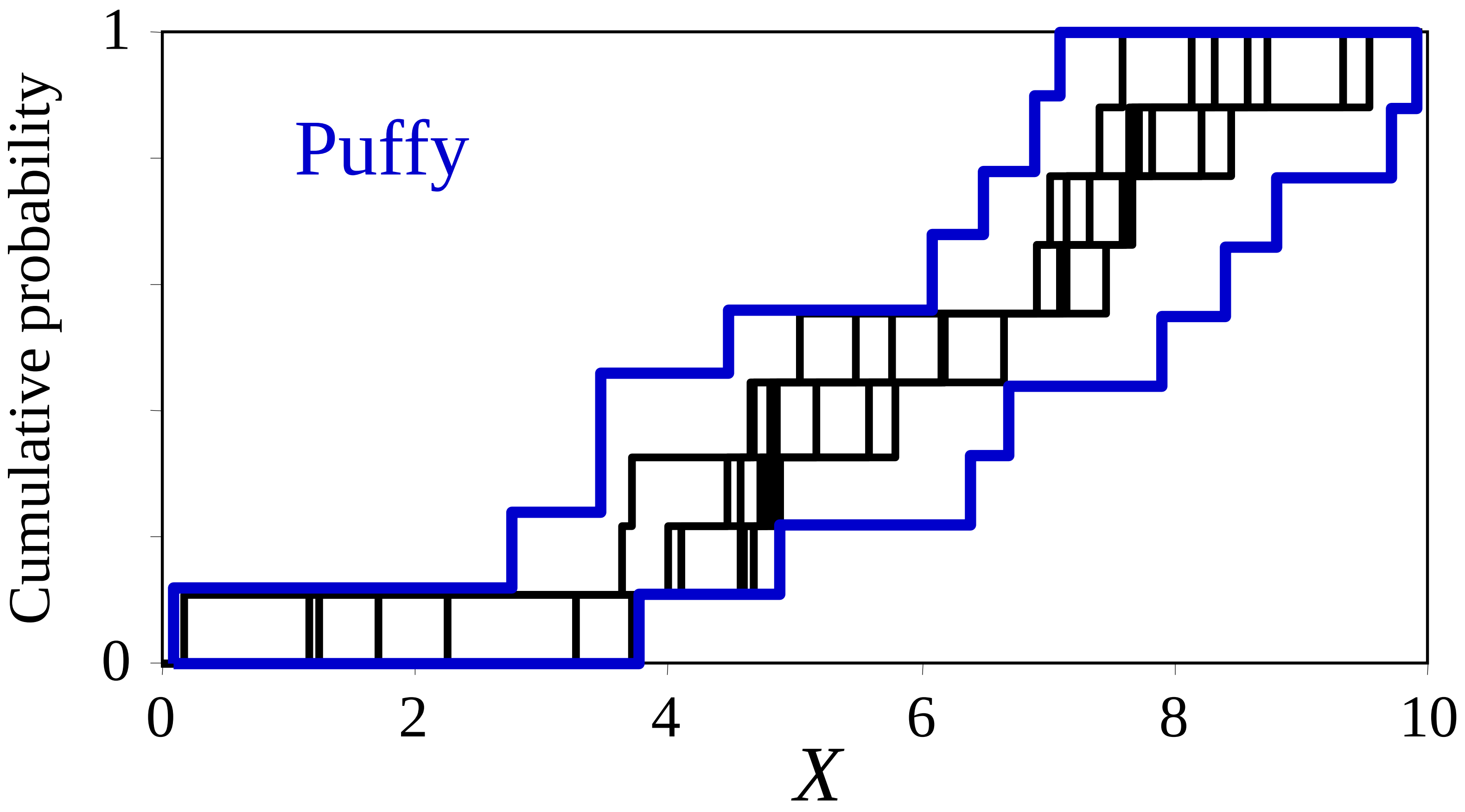

Skinny
Puffy
0
2
4
6
8
10
X

Empirical distribution of intervals

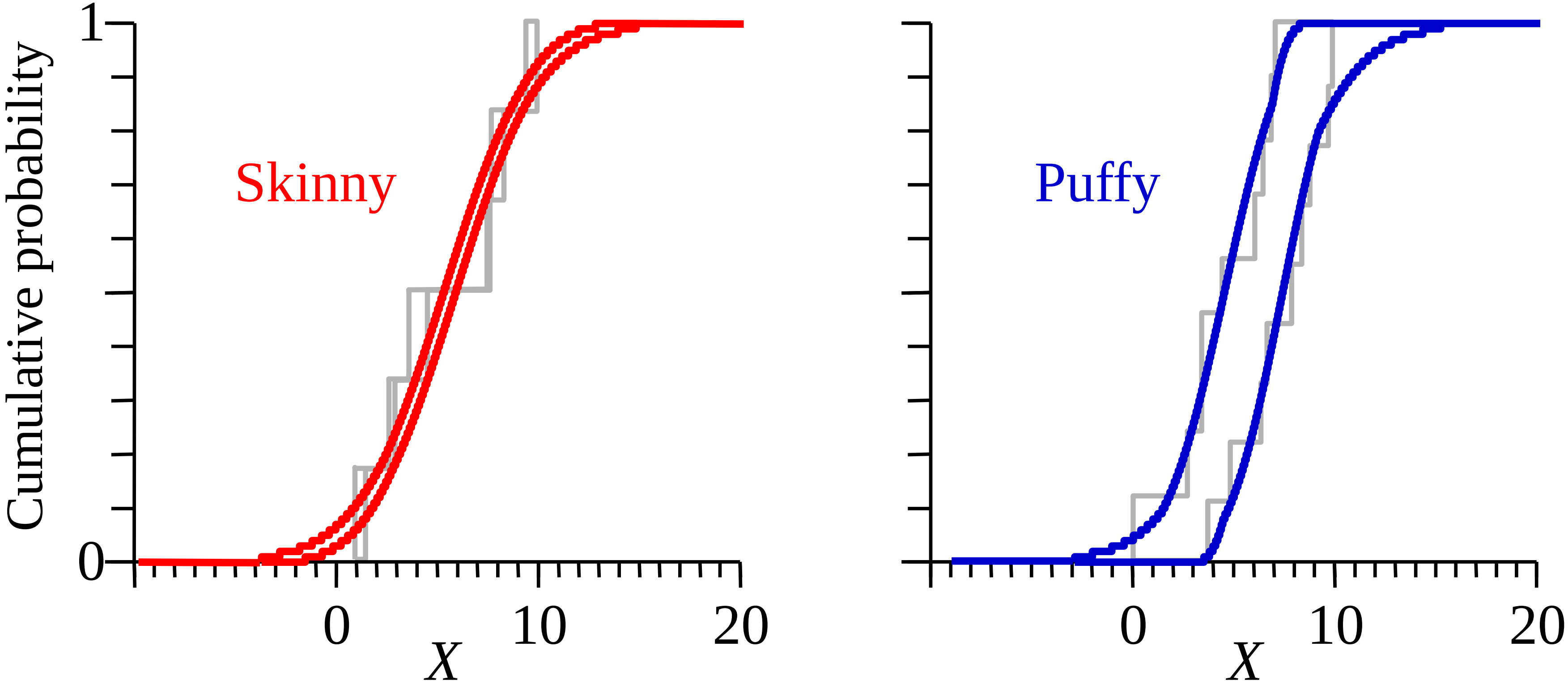


- Each side is cumulation of respective endpoints
- Represents both uncertainty and variability

Uncertainty about the EDF



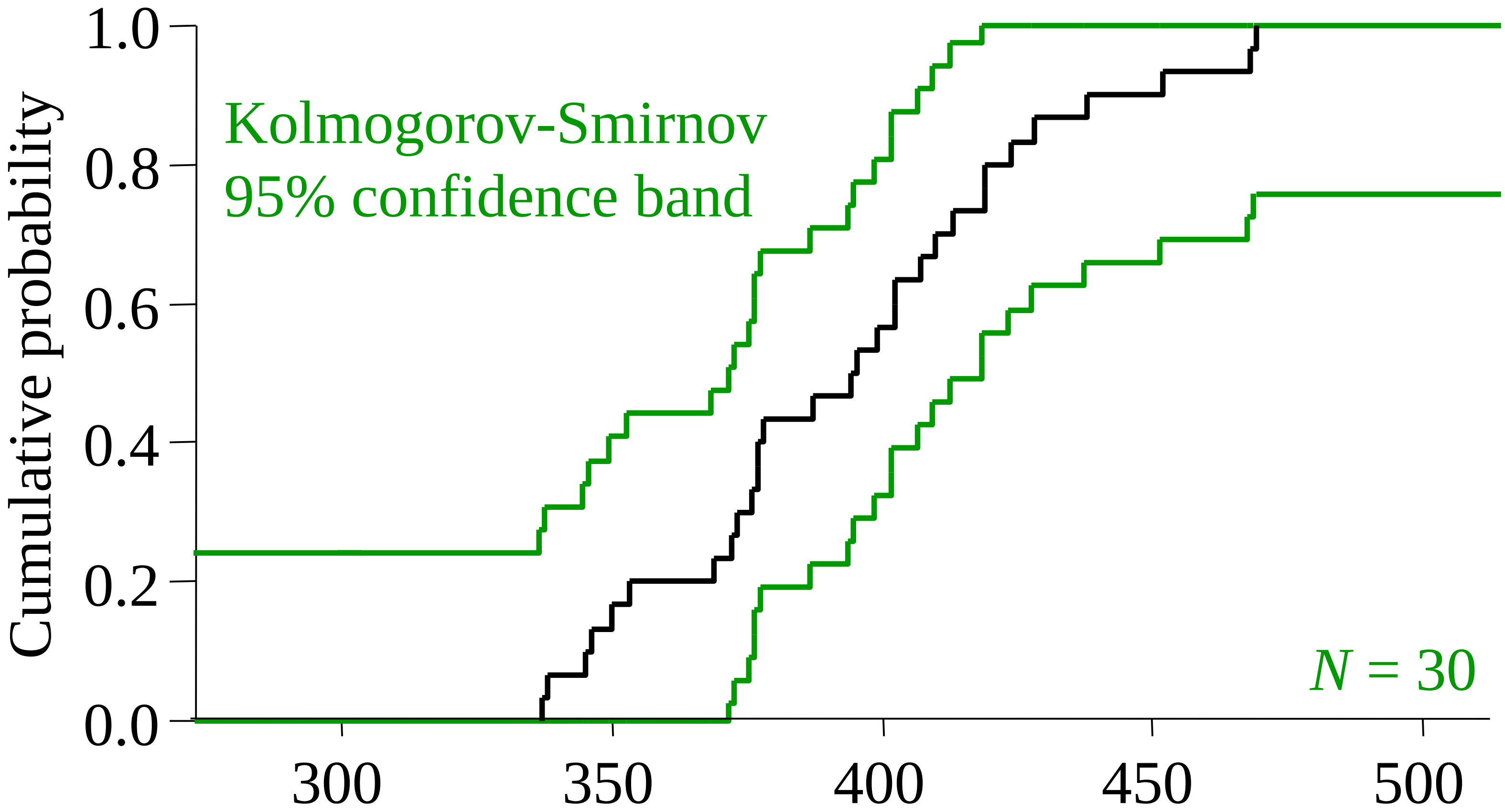
Fitted to normals



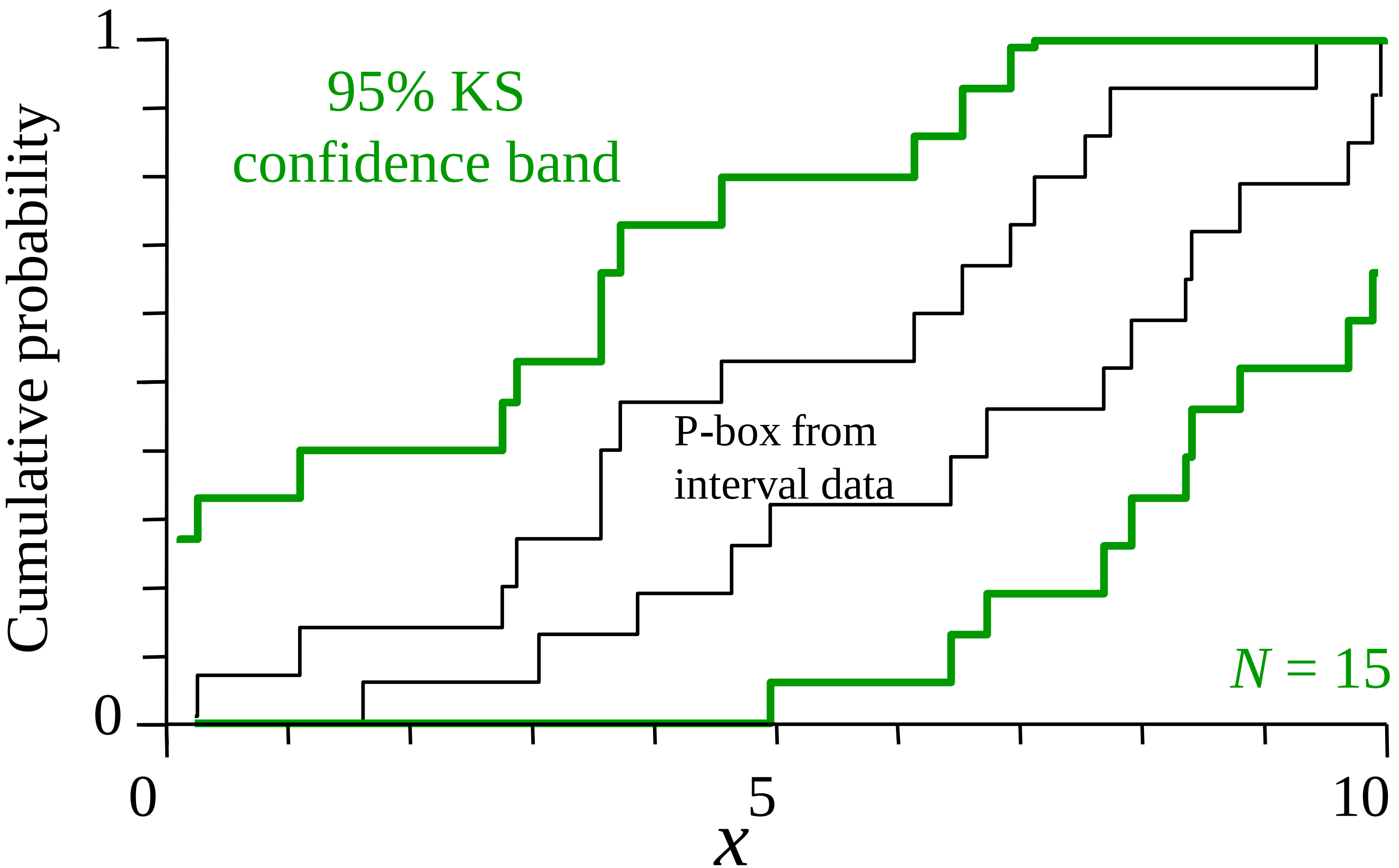
Distributional uncertainty

- Can we trust the raw data or any distribution fitted to them?
- Should account for *sampling uncertainty* about a probability distribution given sampled values
- Kolmogorov-Smirnov confidence procedure
  - 95% (or whatever) confidence for distribution
  - Assumes continuous distribution
  - Random samples, which come from random inputs

Distributional confidence limits



Confidence band about a p-box

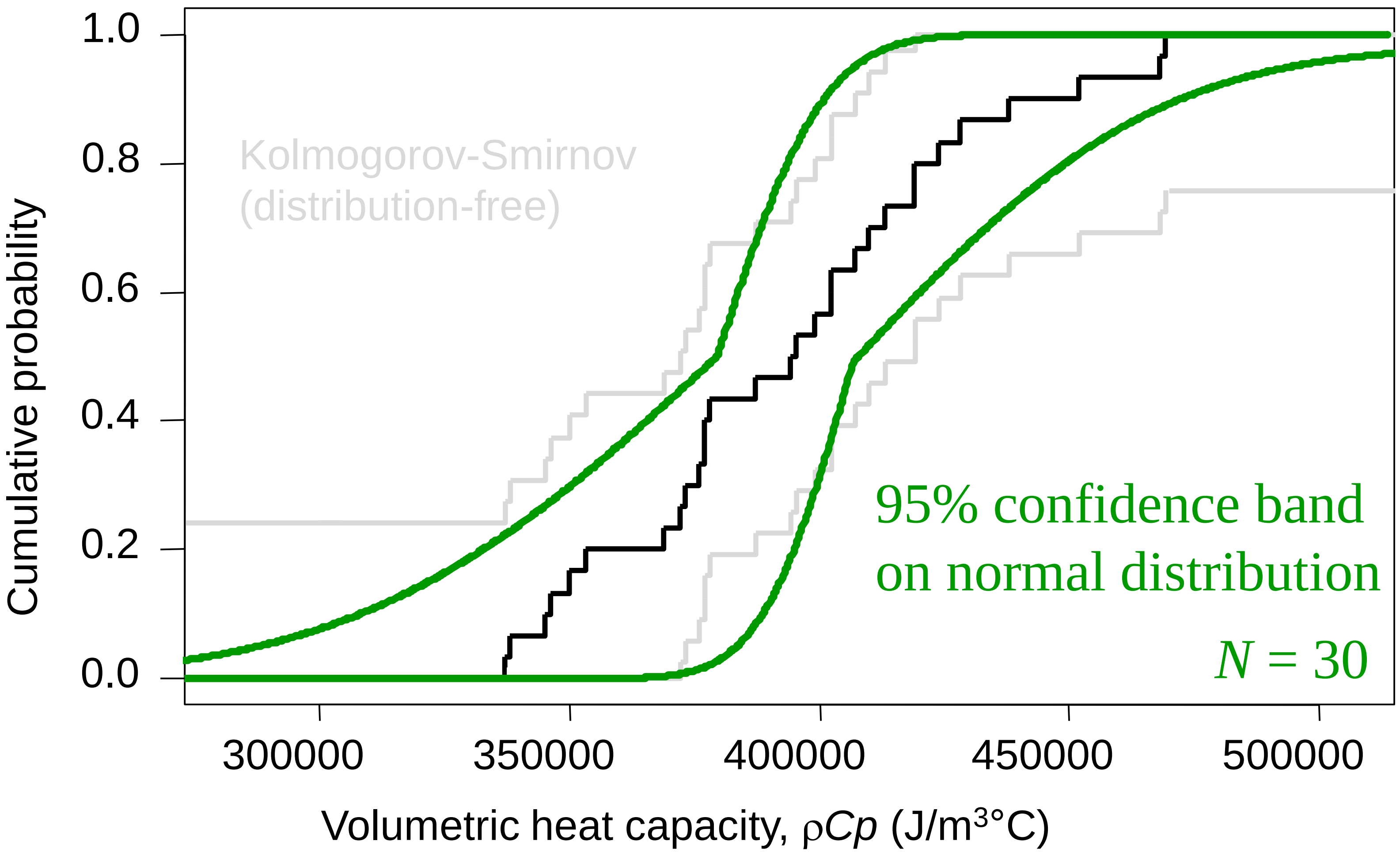


KS generalizes well

- Works with arbitrary input distribution shapes
- Can handle incertitude too
- Generalizes to multivariate outputs
- Conclusion is distribution-free
- Assuming output shape could tighten bounds

Distributional confidence limits

(Chen and Iles; Basu)



SKIP

Single-sample confidence band

- Rodríguez described confidence intervals for mean and standard deviation assuming normality from *only one random sample*

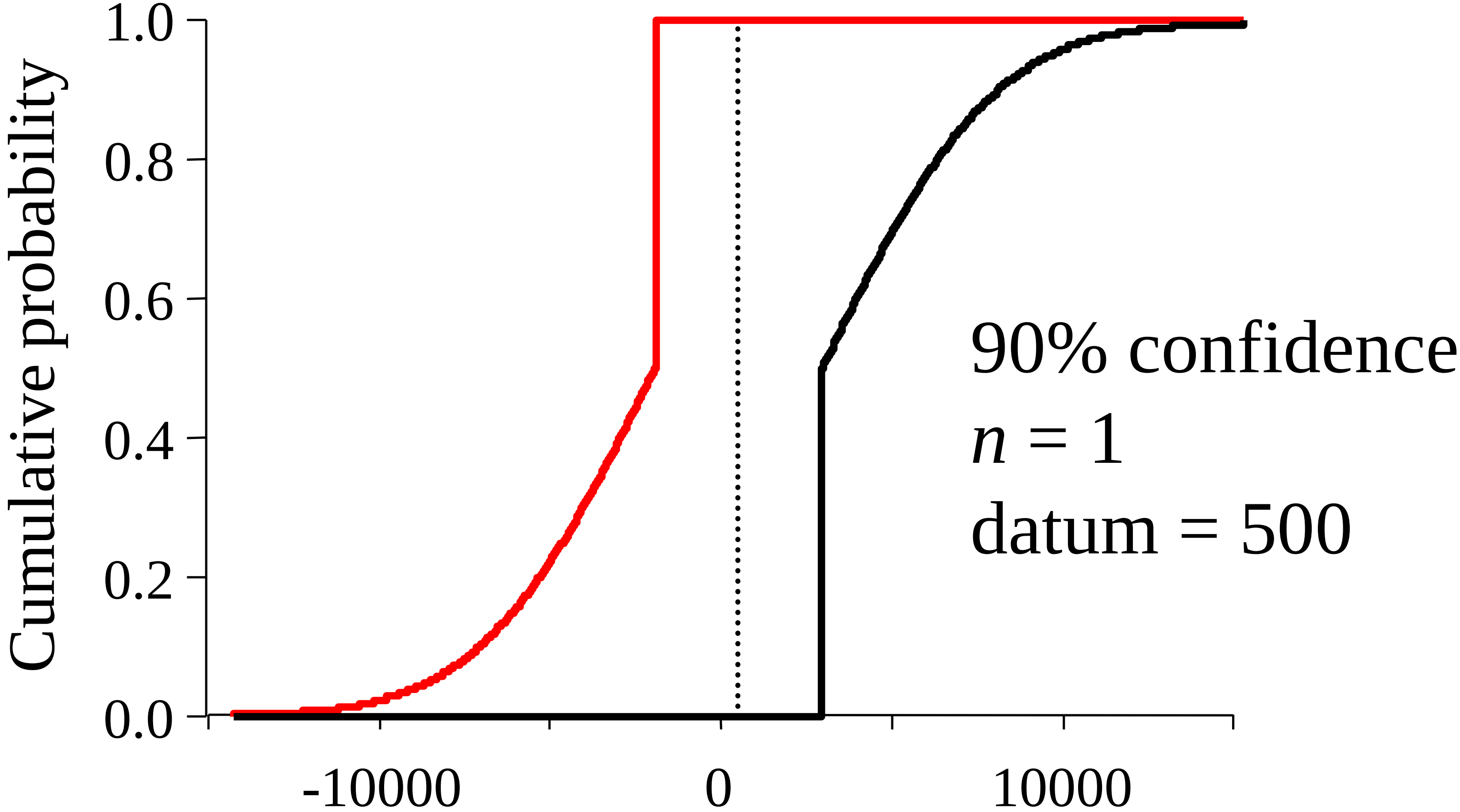
- Combine them to get confidence band

```
onepointconfidenceband.normal = function(x, c=0.95) {
  stopifnot(length(x)==1)
  tm = c(4.83952, 9.678851, 48.39413)
  ts = c(8,17,70)
  k = which(c==c(0.9, 0.95, 0.99))
  normal(x+tm[[k]]*abs(x)*interval(-1,1),interval(0,ts[[k]]*abs(x)))
  }
```

creates a true confidence band

SKIP

Very wide, even assuming normality



Caveat

Confidence bands aren't rigorous

- Not compatible with interval arithmetic
  - You can't intersect or compute with them
  - Not rigorous…only statistical
- Could artfully use the Fréchet inequality
  - Combining two bands at 95% yields a 90% band
  - Two 97.5% → 95%, three 95% → 85%, three 98.35% → 95%
- Could *assume* the confidence band is rigorous
  - Assumption often used (e.g., EPA uses a UCL as the EPC)

Every p-box represents assumptions

- Constraint p-boxes assumed you knew those parameters

- Nothing intrinsically different about the assumption that converts a confidence band to a p-box
  – You’re assuming, given the data, that the distribution is entirely inside the bounds

Confidence boxes

- Structures that let you infer confidence intervals (at *any* confidence level) for a parameter

- Extend confidence distributions (like Student's *t*)

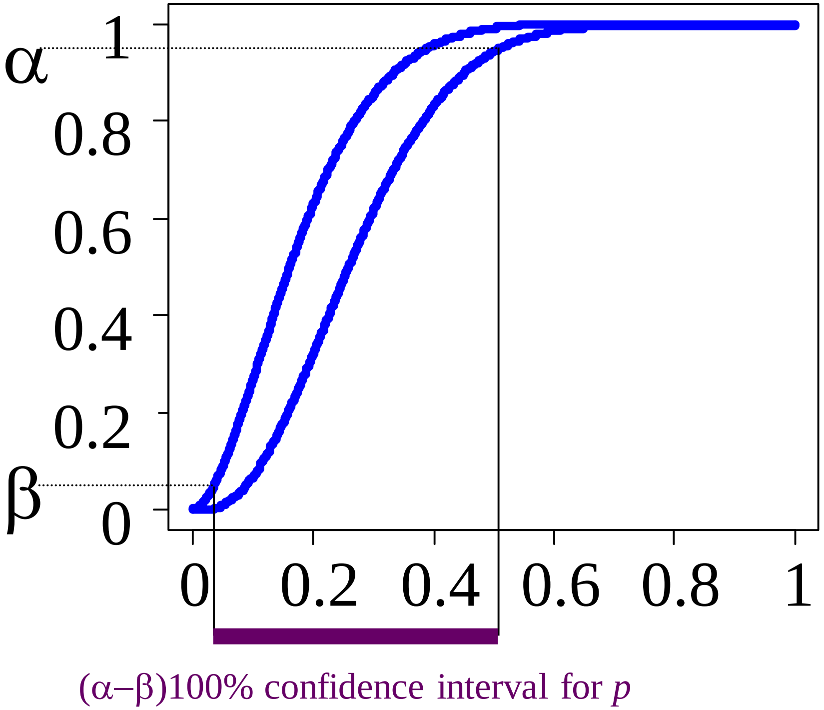
- Different from confidence bands

- Can be propagated as ordinary probability bounds

Example: binomial rate

- Probability, given *k* successes out of *n* trials

Example
$k = 2$
$n = 10$



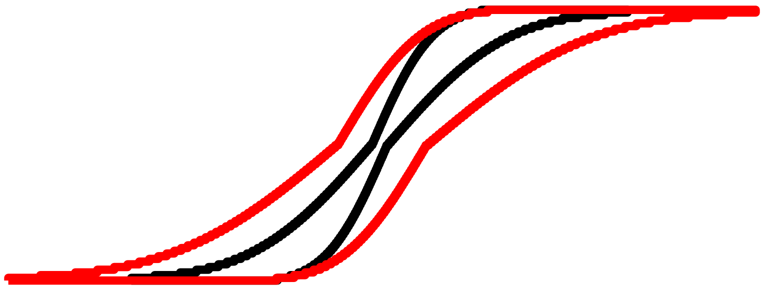
- Like robust Bayes but no assumption of prior

Uncertainties expressible with p-boxes

- Sampling uncertainty (from small *N*)
  - Distributional confidence bands, confidence boxes
- Measurement incertitude
  - Plus-minus ranges, censoring intervals
- Uncertainty about distribution shape
  - Constraints (non-parametric p-boxes)
- Surrogacy uncertainty (have *X* but want *Y*)
  - Modeling

Modeling for surrogacy

- Sometimes you *correct* the best estimate
  - Subtract the weight of the dish when they charge for salad in the cafeteria (+)
  - Sulfur hexafluoride is worth so much CO_2 (×)
- If you're not sure exactly how, then you should *widen the uncertainty*



Example: Lobascio's travel time

$$T = \frac{\left(n + BD \times foc \times Koc\right)L}{K \times i}$$

Parameter	**Units**	**Min**	**Max**	**Mean**	**Stdv**
L source-receptor distance	m	80	120	100	11.55
i hydraulic gradient	m/m	0.0003	0.0008	0.00055	0.0001443
K hydraulic conductivity	m/yr	300	3000	1000	750
n effective soil porosity	-	0.2	0.35	0.25	0.05
BD soil bulk density	kg/m^3	1500	1750	1650	100
foc fraction organic carbon	-	0.0001	0.005	0.00255	0.001415
Koc organic partition coefficient	m^3/kg	5	20	10	3

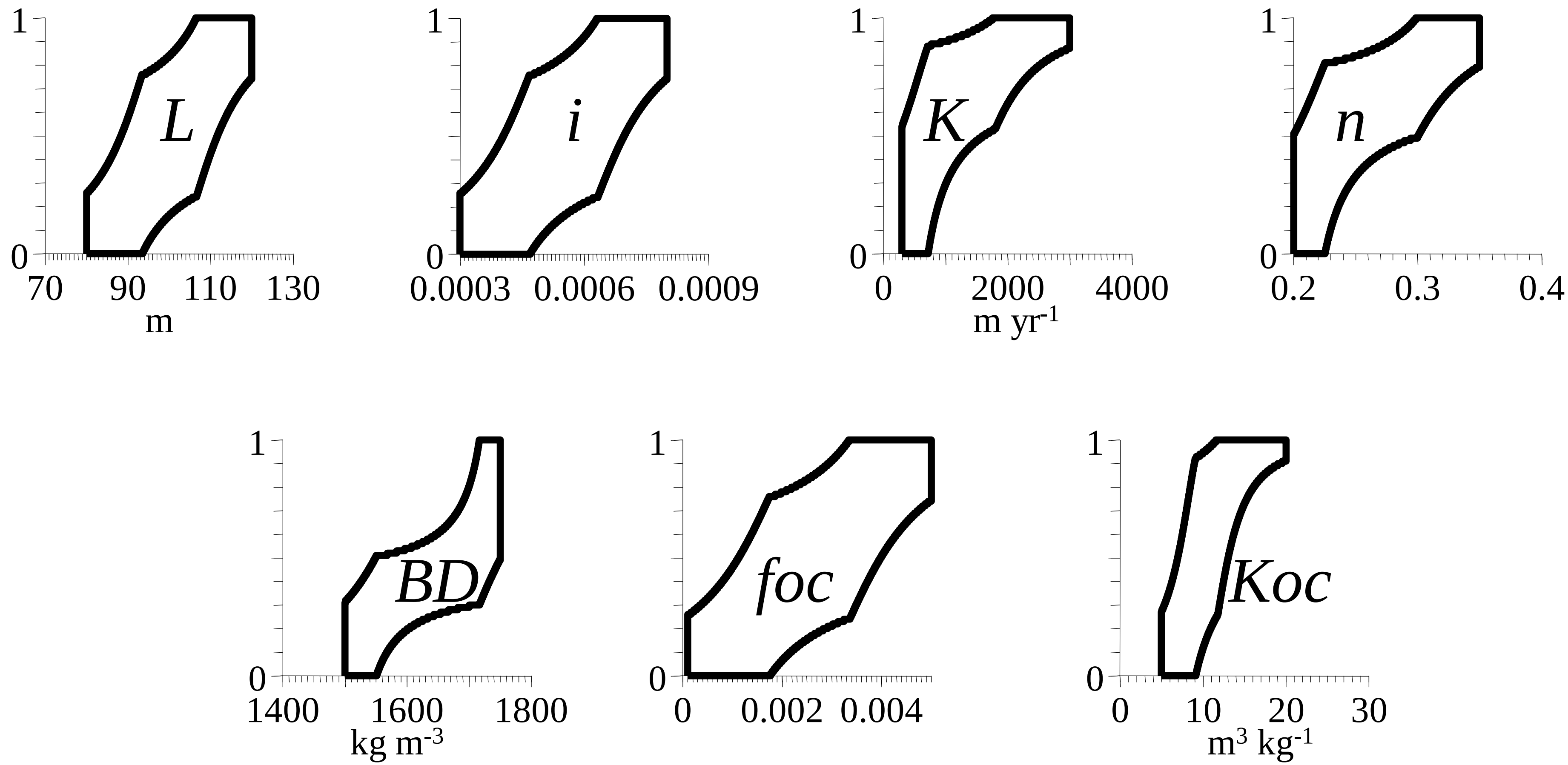
SKIP

pba.r

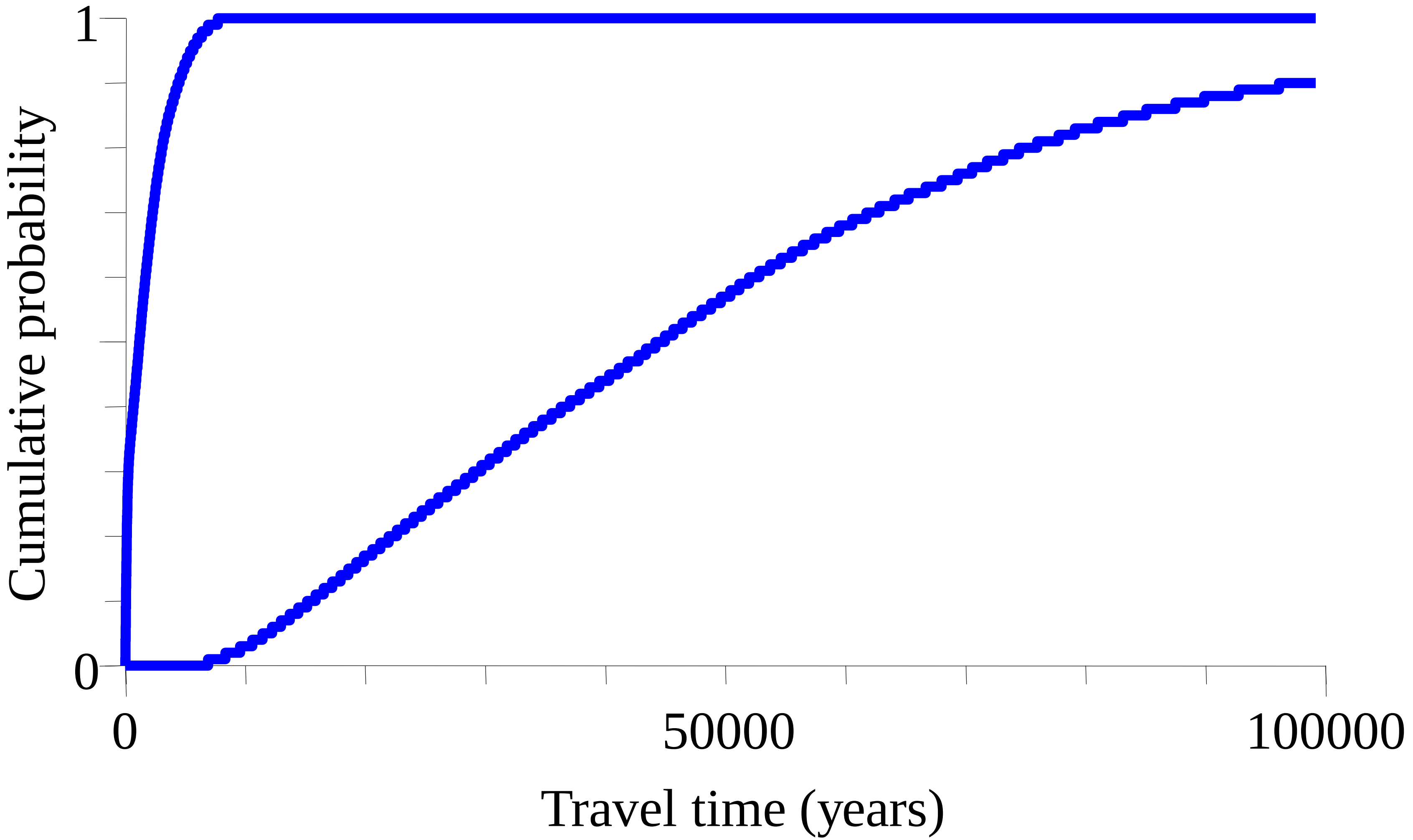
```
L   = mmms(80,120,100,11.55)                        # source-receptor distance
i   = mmms(0.0003,0.0008,0.00055,0.0001443)         # hydraulic gradient
K   = mmms(300,3000,1000,750)                       # hydraulic conductivity
n   = mmms(0.2,0.35,0.25,0.05)                      # effective soil porosity
BD  = mmms(1500,1750,1650,100)                      # soil bulk density
foc = mmms(0.0001,0.005,0.00255,0.001415)           # fraction organic carbon
Koc = mmms(5,20,10,3)                               # organic partition coefficient

up = 100000                                         # detail
Tind  = (n + BD * foc * Koc) * L / (K * i)
Tind = pmin(Tind, up)
```

N.B. The implementation of mmms in pba.r is incomplete so, while its results are bounds, they are not best possible bounds

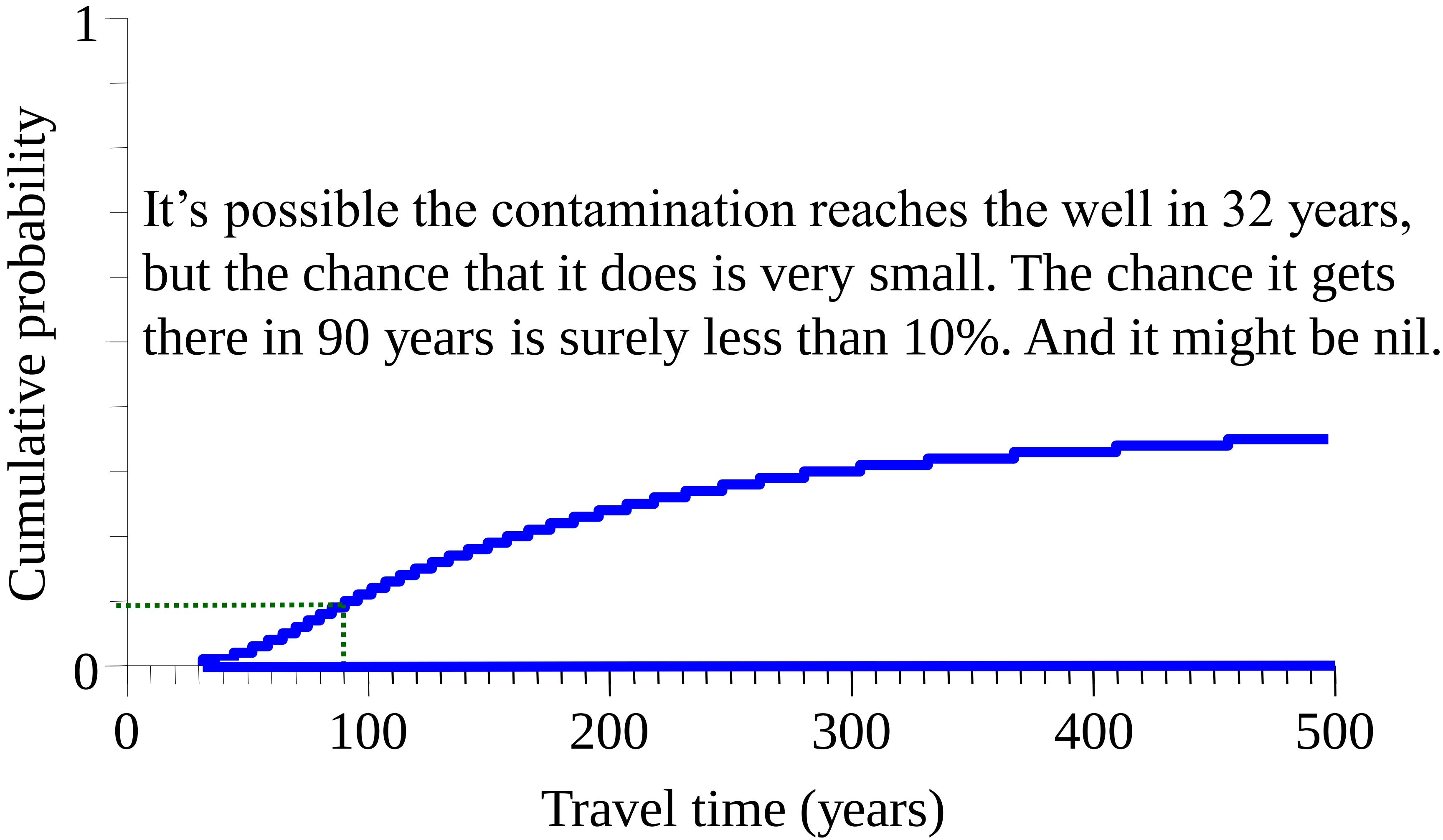
Inputs as mmms p-boxes



Output p-box


1
0
Cumulative probability
0
50000
100000
Travel time (years)

Detail of left tail



Example: mercury in wild mink

Location: Bayou d'Inde, Louisiana

Receptor: generic piscivorous small mammal

Contaminant: mercury

Exposure route: diet (fish and invertebrates)

Based loosely on the assessment described in "Appendix I2: Assessment of Risks to Piscivorus [sic] Mammals in the Calcasieu Estuary", *Calcasieu Estuary Remedial Investigation/Feasibility Study (RI/FS): Baseline Ecological Risk Assessment (BERA),* prepared October 2002 for the U.S. Environmental Protection Agency. See http://www.epa.gov/earth1r6/6sf/pdffiles/appendixi2.pdf.

Total daily intake from diet

$$TDI = FMR | \times | \left(\frac{C_{\text{fish}} \, | \times | \, P_{\text{fish}}}{AE_{\text{fish}} \times GE_{\text{fish}}} + \frac{C_{\text{inverts}} \, | \times | \, P_{\text{inverts}}}{AE_{\text{inverts}} \, | \times | \, GE_{\text{inverts}}} \right)$$

FMR	normalized free metabolic rate of the mammal
AE_{fish}	assimilation efficiency of dietary fish in the mammal
AE_{inverts}	assimilation efficiency of dietary invertebrates in the mammal
GE_{fish}	gross energy of fish tissue
GE_{inverts}	gross energy of invertebrate tissue
C_{fish}	mercury concentration in fish tissue
C_{inverts}	mercury concentration in invertebrate tissue
P_{fish}	proportion of fish in the mammal’s diet
P_{inverts}	proportion of invertebrates in the mammal’s diet

What is known about the variables

FMR, free metabolic rate
Studied in terms of mammal body mass; $FMR = a\, BW^b$, regression intervals a, b

BW, body mass of the mammal
Empirically well studied; Normal with highly precise mean and dispersion

AE_{fish}, AE_{inverts}, assimilation efficiencies for different diet components
A few field measurements; Mean and upper and lower values

GE_{fish}, GE_{inverts}, gross energy of fish and invertebrate tissues
Many measurements; Normal distribution with precise mean and dispersion

C_{fish}, C_{inverts}, mercury concentration in fish or invertebrate tissue
Dictated by EPA policy; Range between sample mean and 95% UCL

P_{fish}, P_{inverts}, proportions of fish and invertebrates in the mammal's diet
Assumed by analyst; Constant

Input assignments

BW = normal(608 gram, 66.9 gram)

FMR = [0.412±0.058]*mag(*BW*)^[0.862±0.026]* 1 Kcal over kg day

*AE*fish = minmaxmean(0.77, 0.98, 0.91)

*AE*inverts = minmaxmean(0.72, 0.96, 0.87)

*GE*fish = normal(1200 Kcal per kg, 240 Kcal per kg)

*GE*inverts = normal(1050 Kcal per kg, 225 Kcal per kg)

*C*fish = [0.1,0.3] mg per kg

*C*inverts = [0.02, 0.06] mg per kg

*P*fish = 0.9

*P*inverts = 0.1

SKIP

pba.r

BW = normal(608, 66.9)

pmi = function(m, pm) pbox(interval(m-pm, m+pm))

FMR = pmi(0.412,0.058) * *BW* ^ pmi(0.862,0.026)

*AE*fish = minmaxmean(0.77, 0.98, 0.91)

*AE*inverts = minmaxmean(0.72, 0.96, 0.87)

*GE*fish = normal(1200, 240)

*GE*inverts = normal(1050, 225)

*C*fish = interval(0.1,0.3)

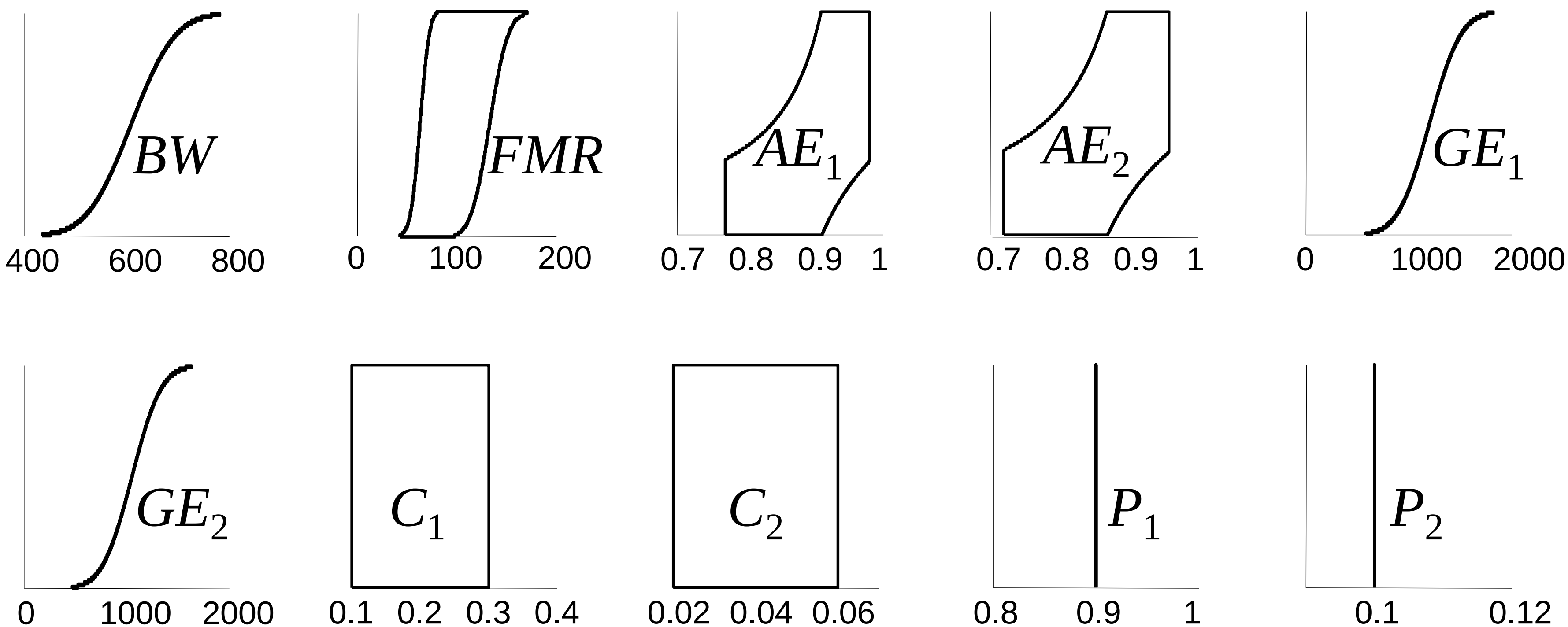
*C*inverts = interval(0.02, 0.06)

*P*fish = 0.9

*P*inverts = 0.1

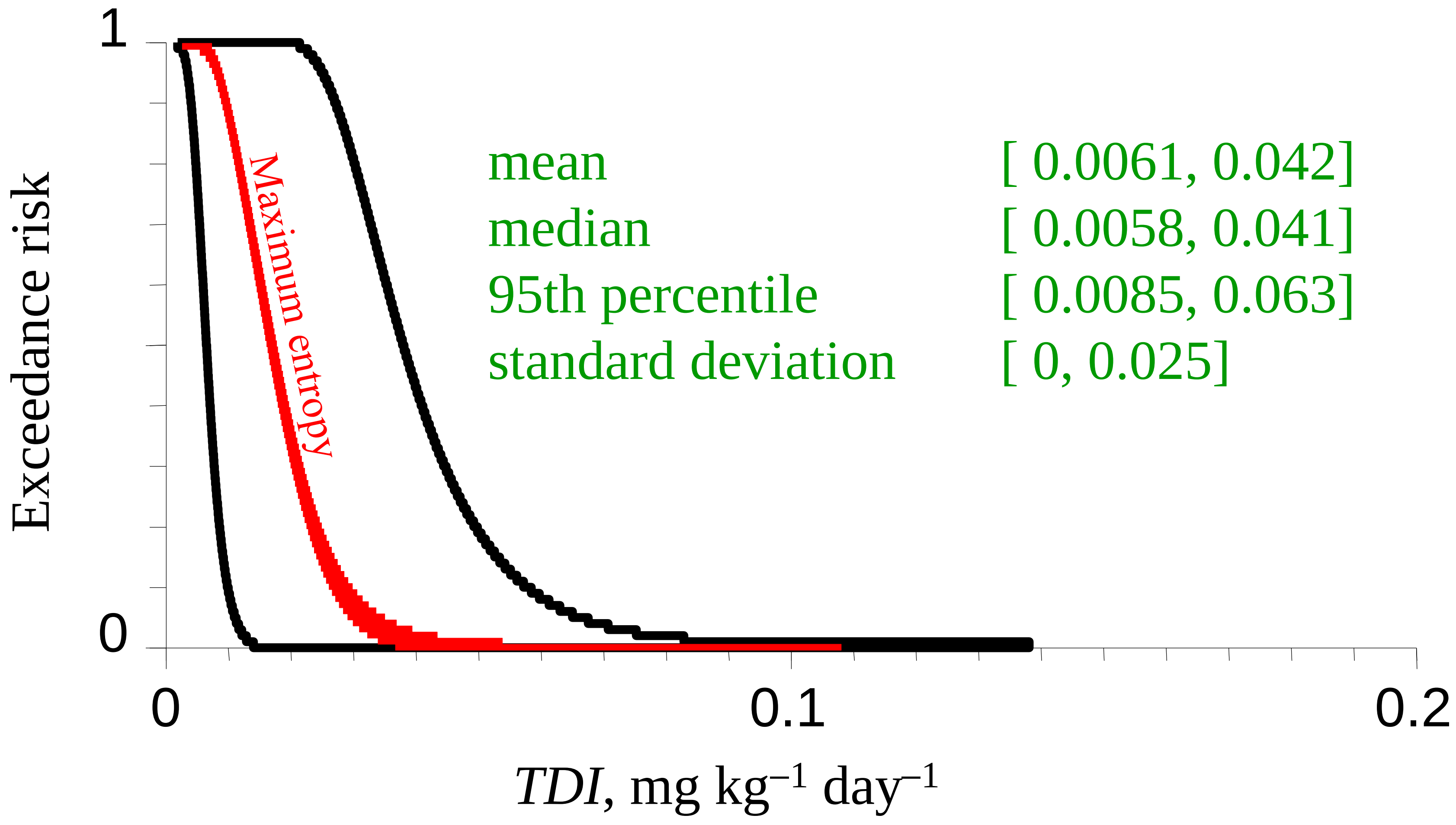
TDI = *FMR* * (*C*fish%*%*P*fish%/%(*AE*fish**GE*fish) %+% *C*inverts%*% *P*inverts%/%(*AE*inverts**GE*inverts))

Input p-boxes



Subscript 1 denotes fish, 2 denotes inverts

Results


1
0
Exceedance risk
Maximum entropy
mean [0.0061, 0.042]
median [0.0058, 0.041]
95th percentile [0.0085, 0.063]
standard deviation [0, 0.025]
0
0.1
0.2
TDI, mg kg–1 day–1

2MC simulations may not fill p-boxes

- 2^{nd} order Monte Carlo is not comprehensive
  - Inadequate model of ignorance
  - Dependence among parameters of a distribution
  - Uncertainty about dependence (Fréchet)
  - Non-denumerable model uncertainty

- Probability bounds analysis is not optimal
  - Independence between parameters of a distribution
  - Ternary (and higher) Fréchet operations

Probability bounds analysis

- Combines interval and probability methods
- Shows when uncertainty is (or isn't) important
- Won't underestimate risks of extremes, yet isn't hyperconservative like worst case
- Solves many risk analysis problems
  - Input distributions unknown
  - Large measurement error, censored data, and small samples
  - Correlation and dependency ignored
  - Model uncertainty

Advantages

- Fewer assumptions
  - Not just *different* assumptions, *fewer* of them
  - Distribution-free methods

- Rigorous results
  - Built-in quality assurance
  - Automatically verified calculation

What p-boxes can’t do

- Show what’s most likely within a p-box
- Express second-order information
- Get best-possible bounds on non-tail risks
- Get best-possible bounds for any data set
  - When dependencies are intricate
  - When information about modes are availabe

Don’t know the input distributions

Solved

- Don’t *have* to specify the distributions
- Shouldn’t use a distribution without evidence
- Maximum entropy criterion erases uncertainty rather than propagates it
- Sensitivity analysis is very hard since it’s an infinite-dimensional problem
- P-boxes easy, but should use all information

Probability bounds analysis in pba.r

- Output
  - <enter variable name>, plot, lines, show, summary
- Characterize
  - mean, sd, var, median, quantile, fivenum, left, right, prob, cut, percentile, iqr, random, range
- Compute
  - exp, log, sqrt, abs, round, trunc, ceiling, floor, sign, sin, cos, tan, asin, acos, atan, atan2, reciprocate, negate, +, -, *, /, pmin, pmax, ^, and, or, not, mixture, smin, smax, complement

Probability bounds analysis in pba.r

- Construct
  - <named distribution>
  - histogram
  - quantiles
  - pointlist
  - MM <tab> <tab>
  - ME <tab> <tab>
  - ML <tab> <tab>
  - CB <tab> <tab>
  - NV <tab> <tab>

Supported named distributions

bernoulli, beta (B), binomial (Bin), cauchy, chi, chisquared, delta, dirac, discreteuniform, exponential, exponentialpower, extremevalue, F, f, fishersnedecor, fishertippett, fisk, frechet, gamma, gaussian, geometric, generalizedextremevalue (GEV), generalizedpareto, gumbel, histogram, inversegamma, laplace, logistic, loglogistic, lognormal (L), logtriangular, loguniform, negativebinomial, normal (N), pareto, pascal, powerfunction, poisson, quantiles, rayleigh, reciprocal, shiftedloglogistic, skewnormal (SN), student, trapezoidal, triangular (T), uniform (U), weibull

Computing with confidence

Many ways to fit distributions to data

- Maximum entropy
- Maximum likelihood
- Bayesian inference
- Method of matching moments ⟵ still most common
- Goodness of fit (KS, AD, χ^2, etc.)
- PERT
- Regression techniques
- Empirical distribution functions

…in fact there are even more methods…

Little coherence in practice

- Disparate methods used across risk analysis
- Common to mix and match distributions with different justifications
- Analyses are thus based on no clear criterion or standard of performance
- Is this okay?

Frequentist confidence intervals

- Favored by many engineers
- Guarantees statistical performance over time
- But difficult to employ consistently in analyses
- Not clear how to propagate them through mathematical calculations

Bayesian approaches

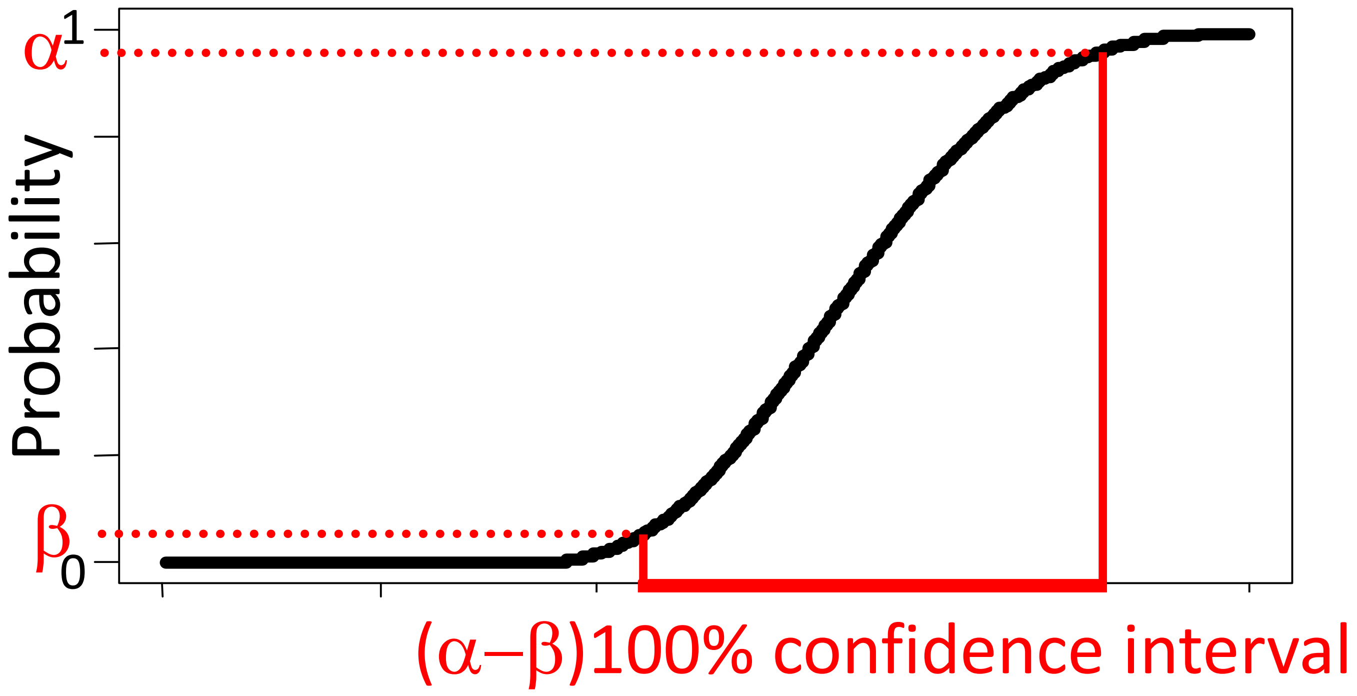
- Permit mathematical calculations
- But lack guarantees ensuring long-run statistical performance
- Many engineers are reluctant to use Bayesian methods

Confidence distributions

- Not widely used in statistics
- Introduced by Cox in the 1950s
- Closely related to well known ideas
  - Student's *t*-distribution
  - Bootstrap distributions

Confidence distributions

- Distributional estimators of (fixed) parameters
- Give confidence interval at *any* confidence level



Confidence interval

- A confidence interval with coverage α

  In replicate problems, a proportion α of computed confidence intervals will enclose the true value

- Using methods to compute confidence intervals thus ensures statistical *performance*

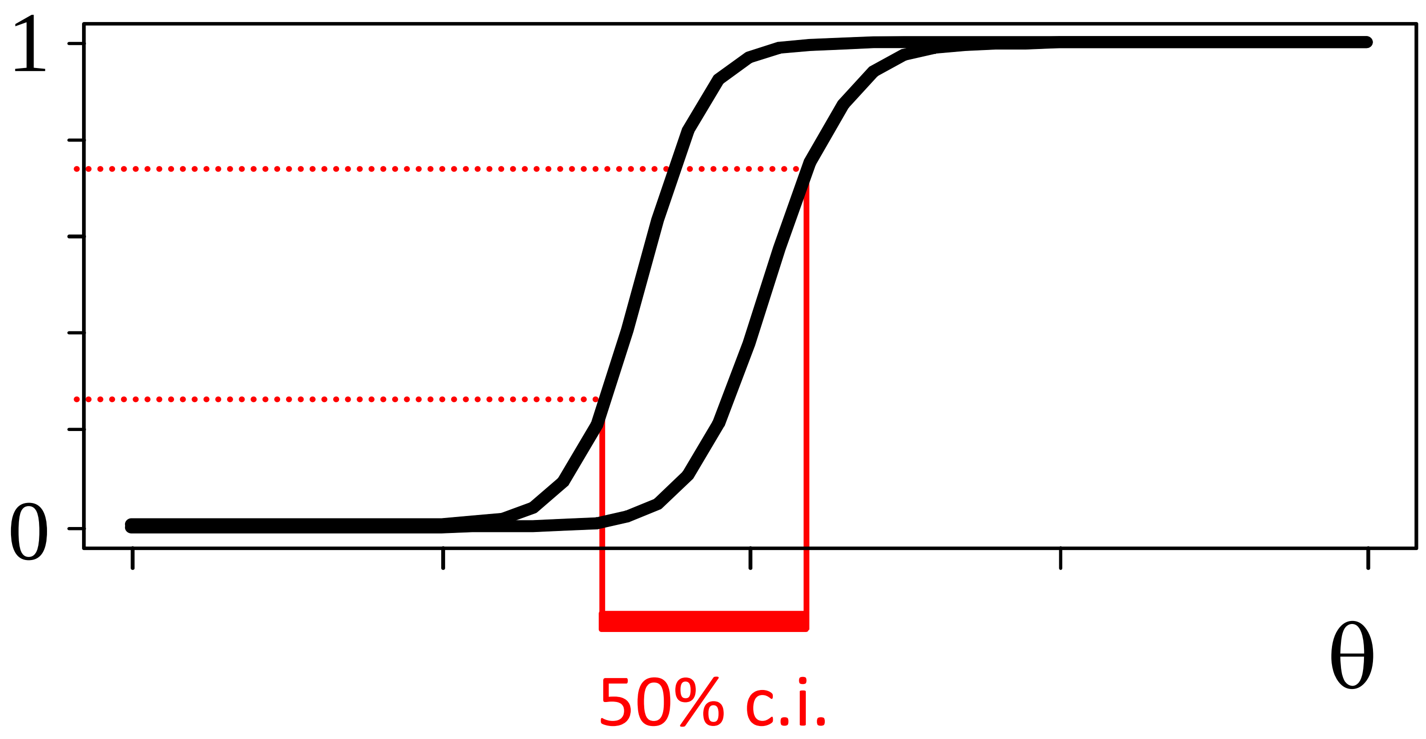
Confidence distributions

- Have the *shape* of a distribution
- But correspond to no random variables
- Not supposed to compute with them
- Don’t always exist (e.g., for the binomial rate)

Confidence structures (c-boxes)

- Generalization of confidence distributions
- Reflect inferential uncertainty about parameter
- Known for many cases
  - binomial rate and other discrete parameters
  - normals, and many other problems
  - non-parametric case
- Still have performance/confidence interpretation

Confidence interpretation



Estimators

- Point estimates (e.g., sample mean)
- Interval estimates (e.g., confidence intervals)
- Distributional estimates (Bayesian posteriors)
- P-box estimates (e.g., c-boxes)

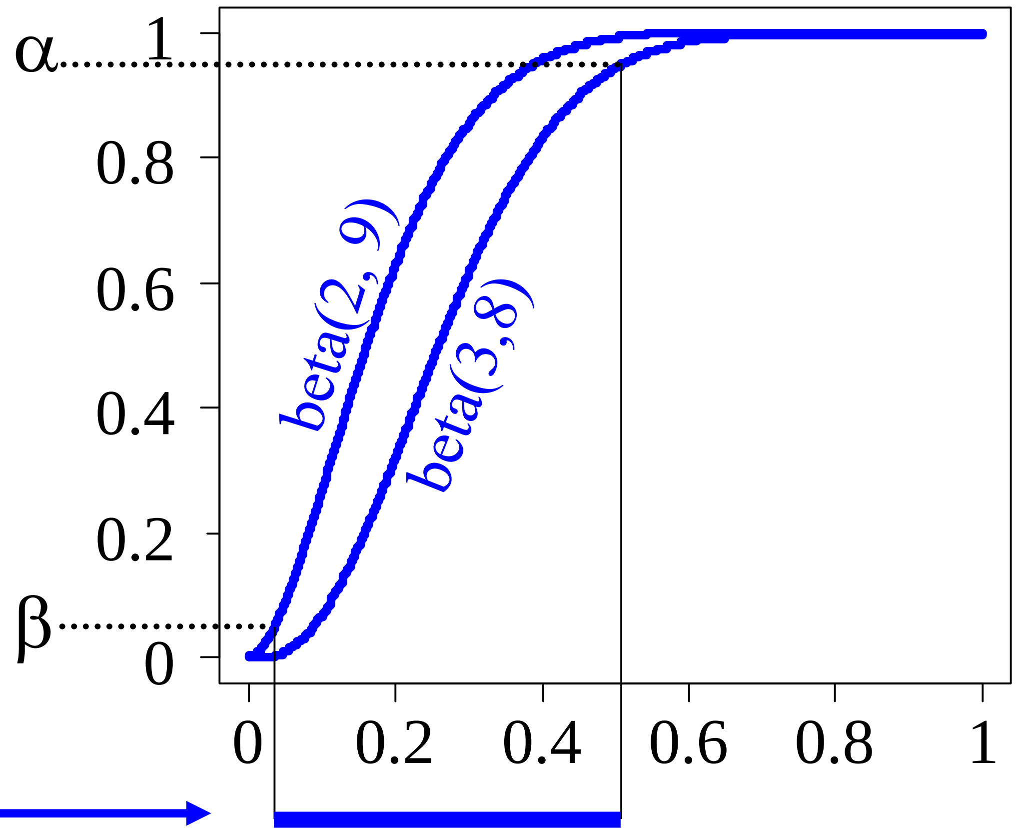
Binomial rate p for k of n trials

$p \sim \mathrm{env}(\mathrm{beta}(k, n-k+1), \mathrm{beta}(k+1, n-k))$

Data

$k = 2$

$n = 10$

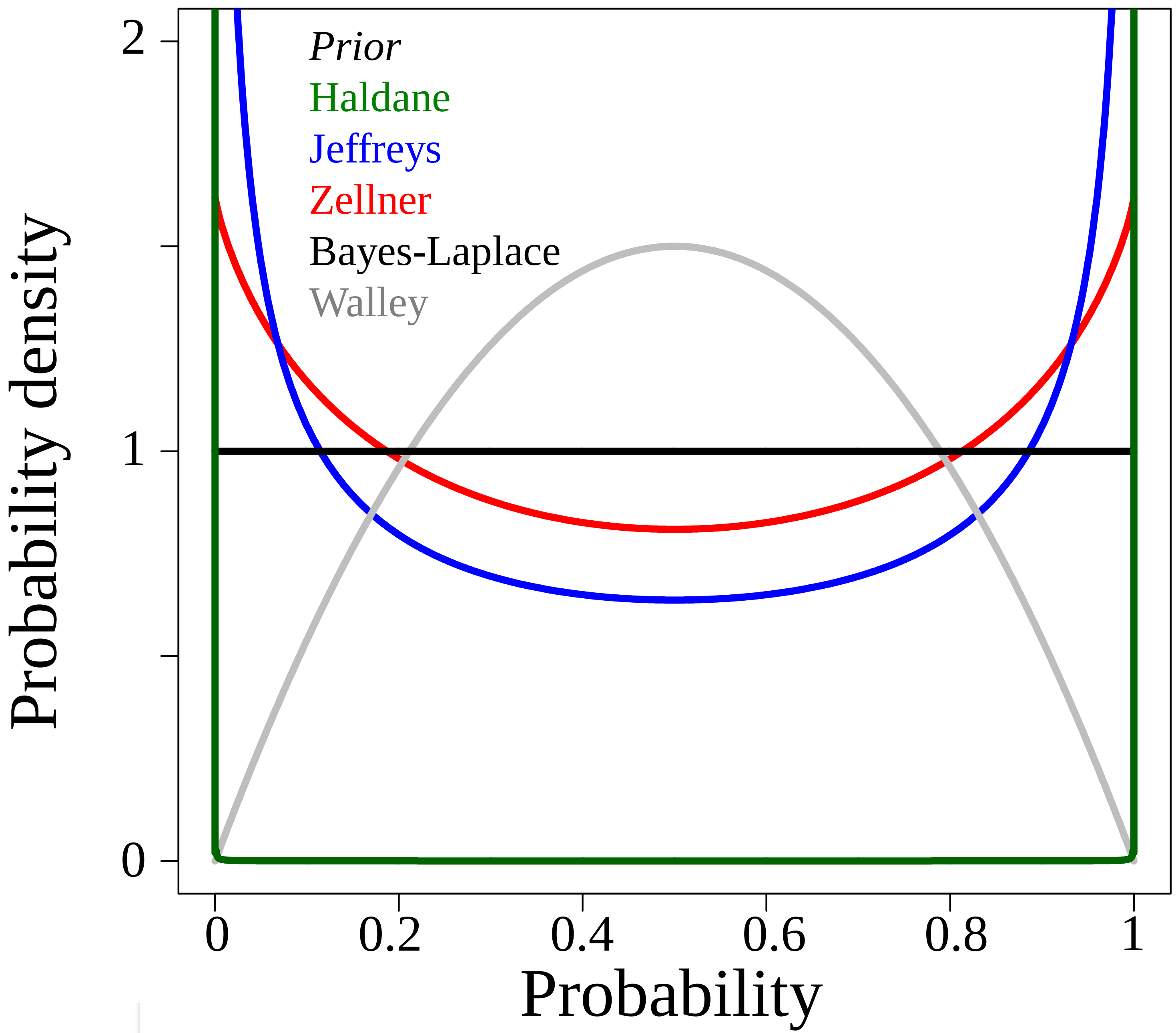


$(\alpha-\beta)100\%$ confidence interval for p

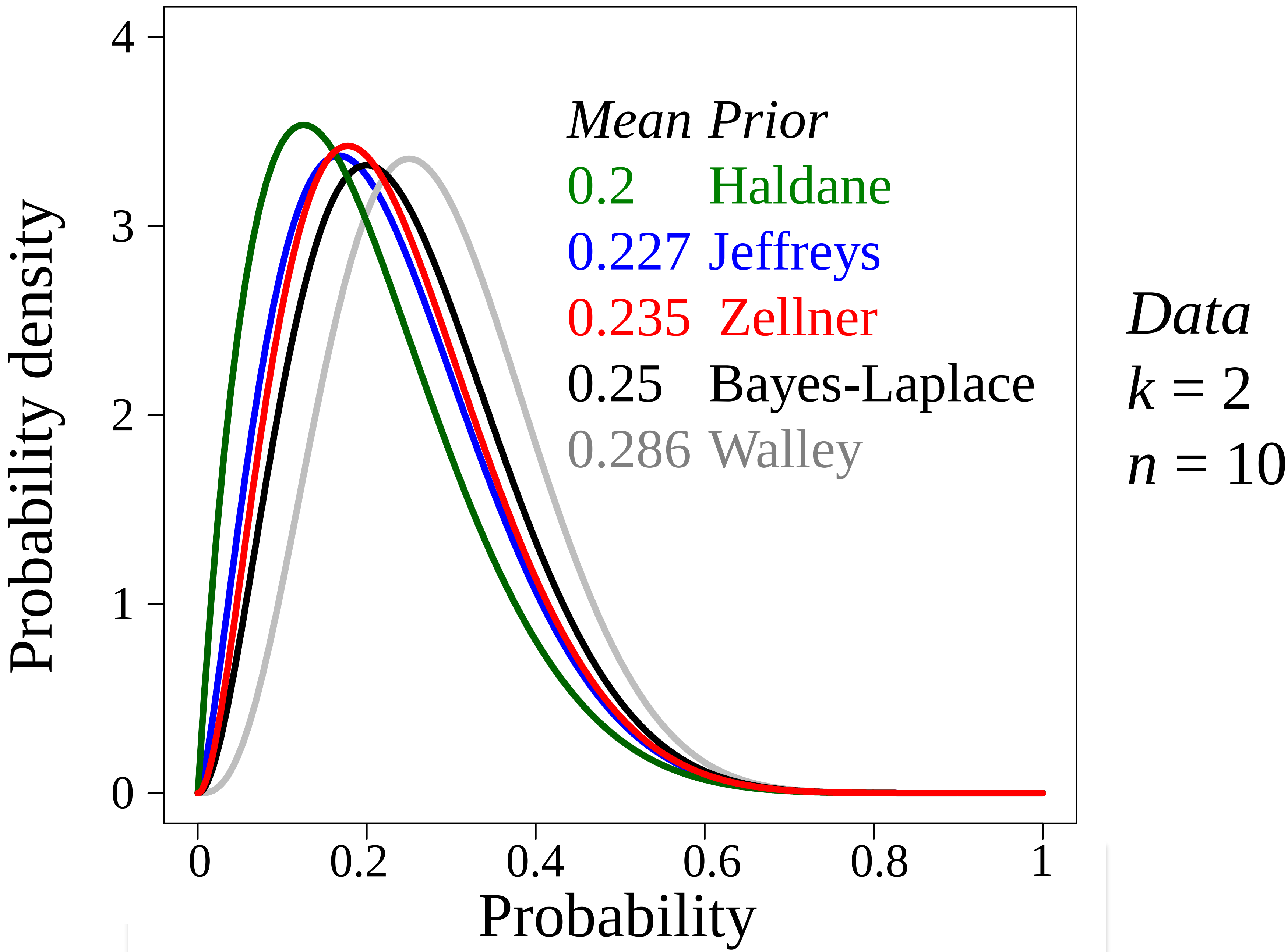
If $1-\alpha = \beta$, result is identical to classical Clopper–Pearson interval

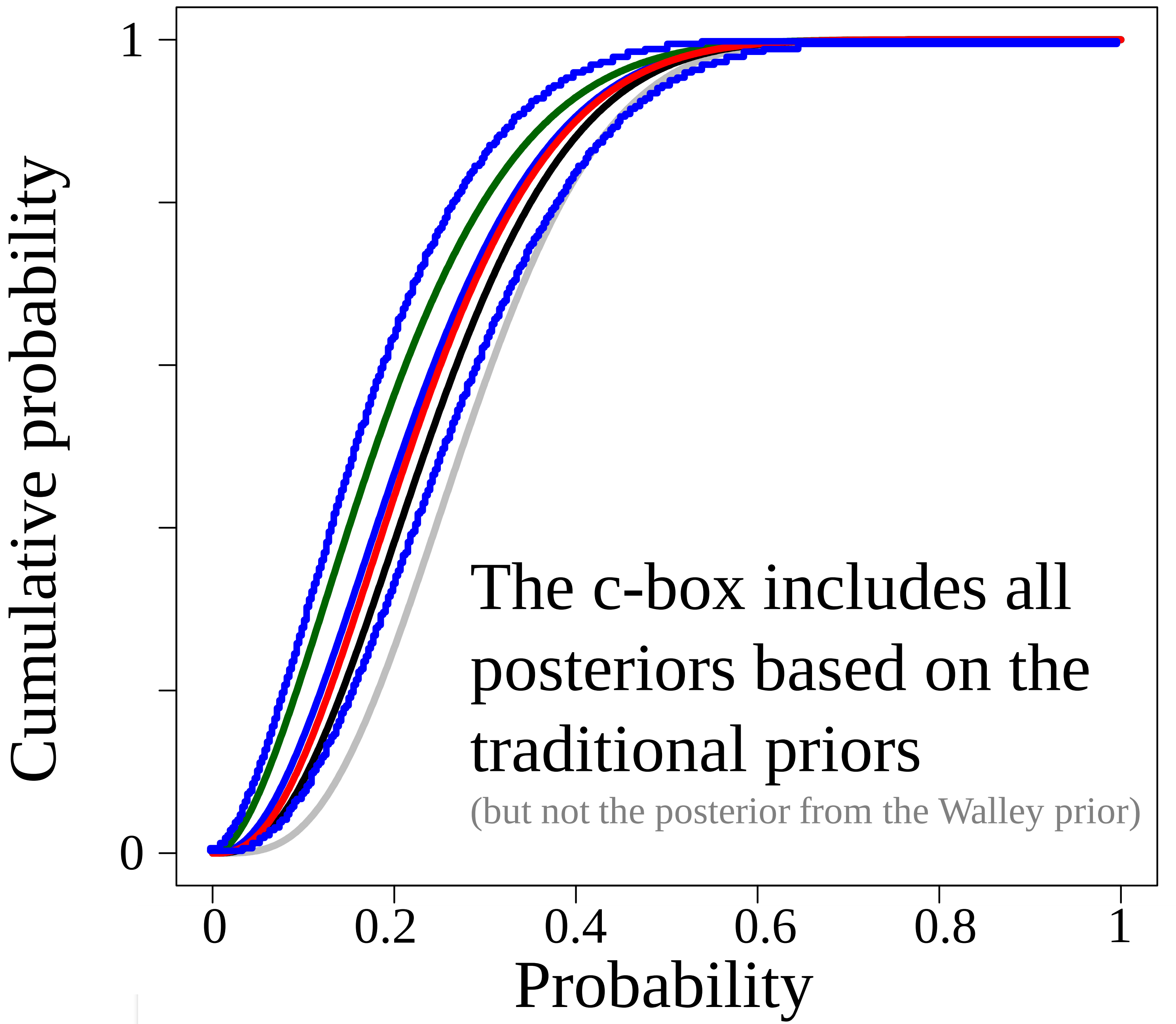
How does the Bayes analysis compare?

- No such thing as *the* Bayes analysis
- There are always many possible analyses
  - Different priors, which yield different answers
  - When data sets are small, the differences are big
- For binomial rate there are four or five priors Bayesians have not been able to chose among



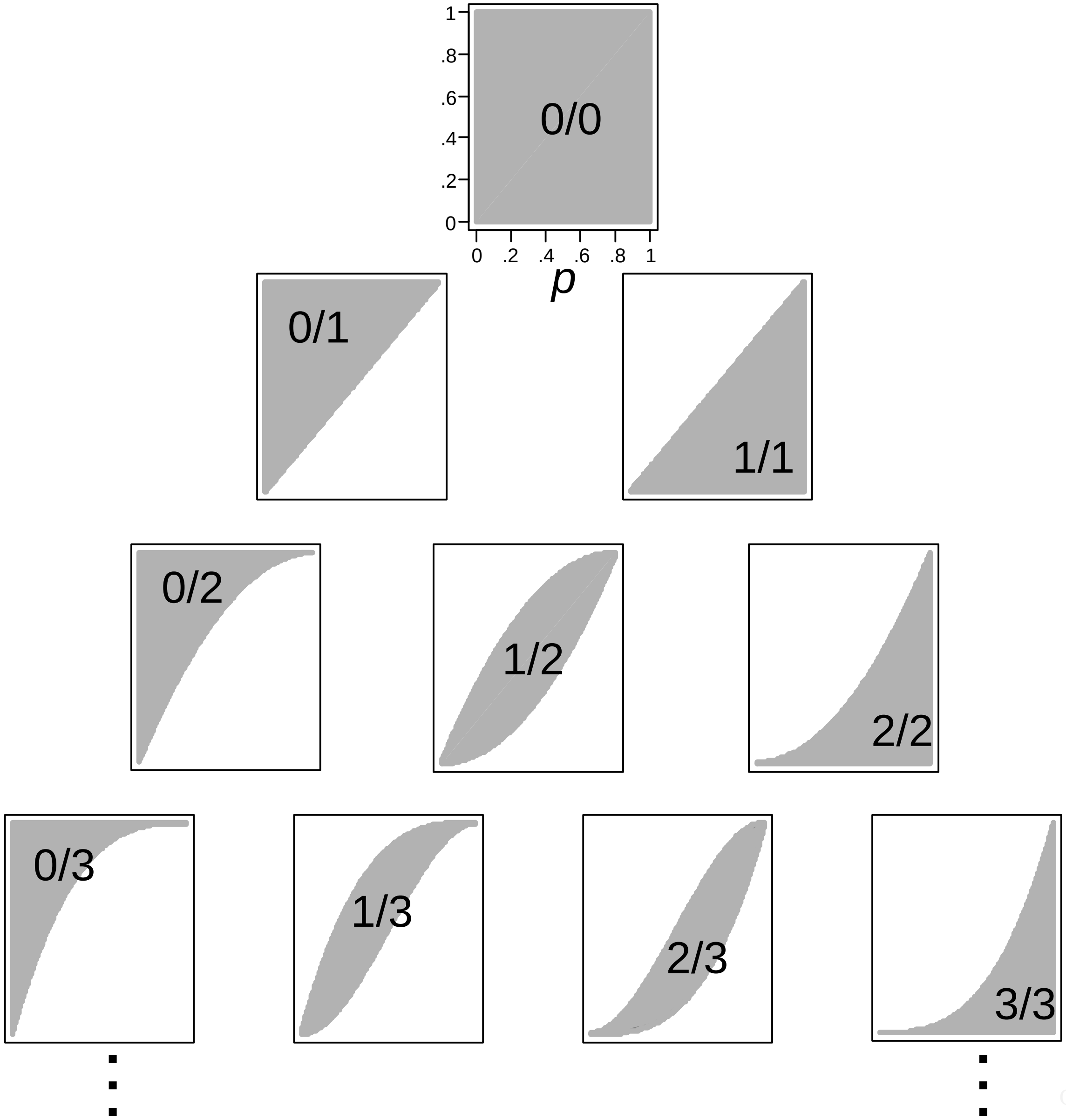
Prior
Haldane
Jeffreys
Zellner
Bayes-Laplace
Walley
Probability density
Probability
0
1
2
0
0.2
0.4
0.6
0.8
1


Mean Prior
0.2 Haldane
0.227 Jeffreys
0.235 Zellner
0.25 Bayes-Laplace
0.286 Walley
Data
k = 2
n = 10
Probability density
Probability
0
1
2
3
4
0
0.2
0.4
0.6
0.8
1



1
0
0
0.2
0.4
0.6
0.8
1
Probability
Cumulative probability
The c-box includes all
posteriors based on the
traditional priors
(but not the posterior from the Walley prior)

C-boxes partition the vacuous square

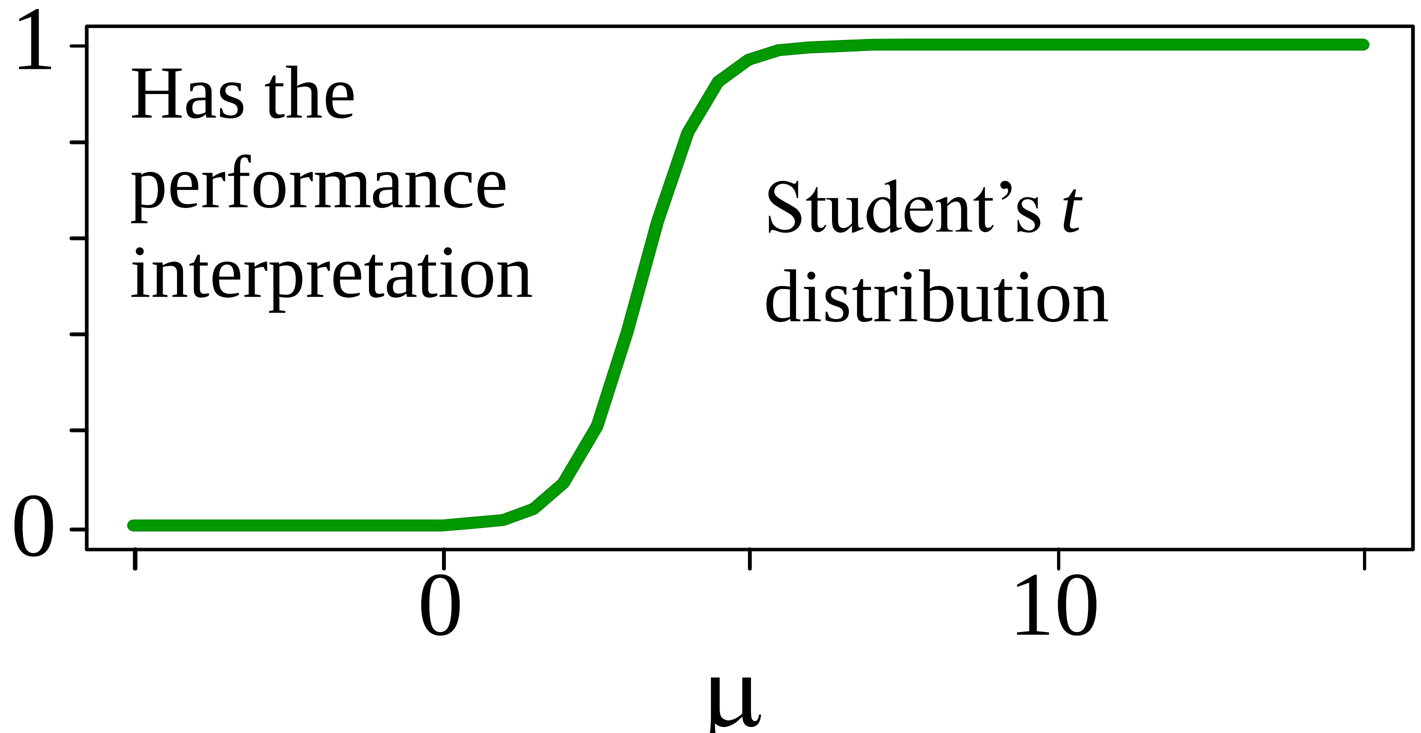


Example: normal mean

$$\mu \sim \bar{x} + s \cdot T_{n-1} / \sqrt{n}$$

Data

8
5.5
–1.3
3.5
0.8
2.8
1.8
2.2
3.5
5.3

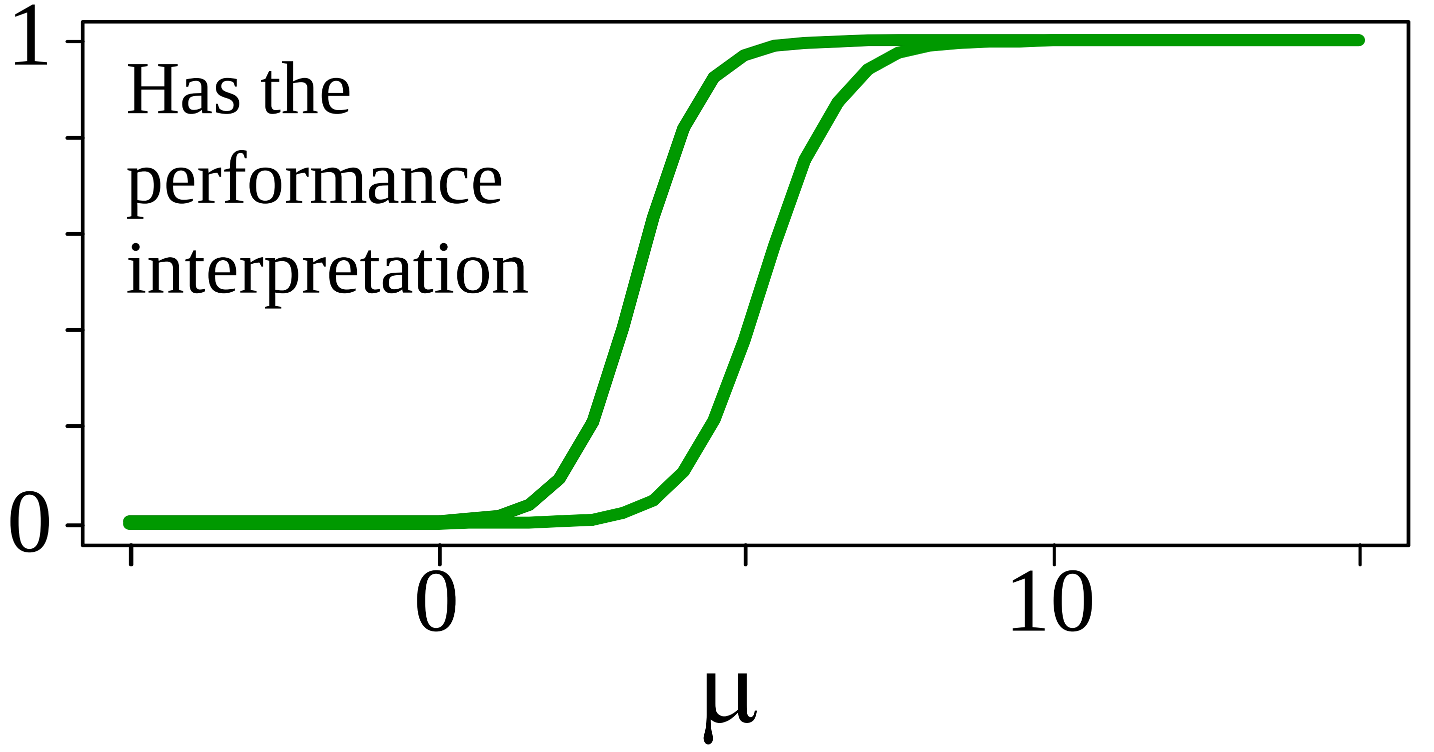


Example: normal mean

$$\mu \sim \bar{x} + s \cdot T_{n-1} / \sqrt{n}$$

Data

[8,11]
[5.5,6.9]
[–1.3,0.3]
[3.5,7.5]
[0.8,1]
[2.8,4.2]
[1.8,5.2]
[2.2,5.2]
[3.5,5.7]
[5.3,6.1]



Deriving c-boxes

- Have to be derived for each special case
- Traditional approaches based on pivots
- Many solutions have been worked out

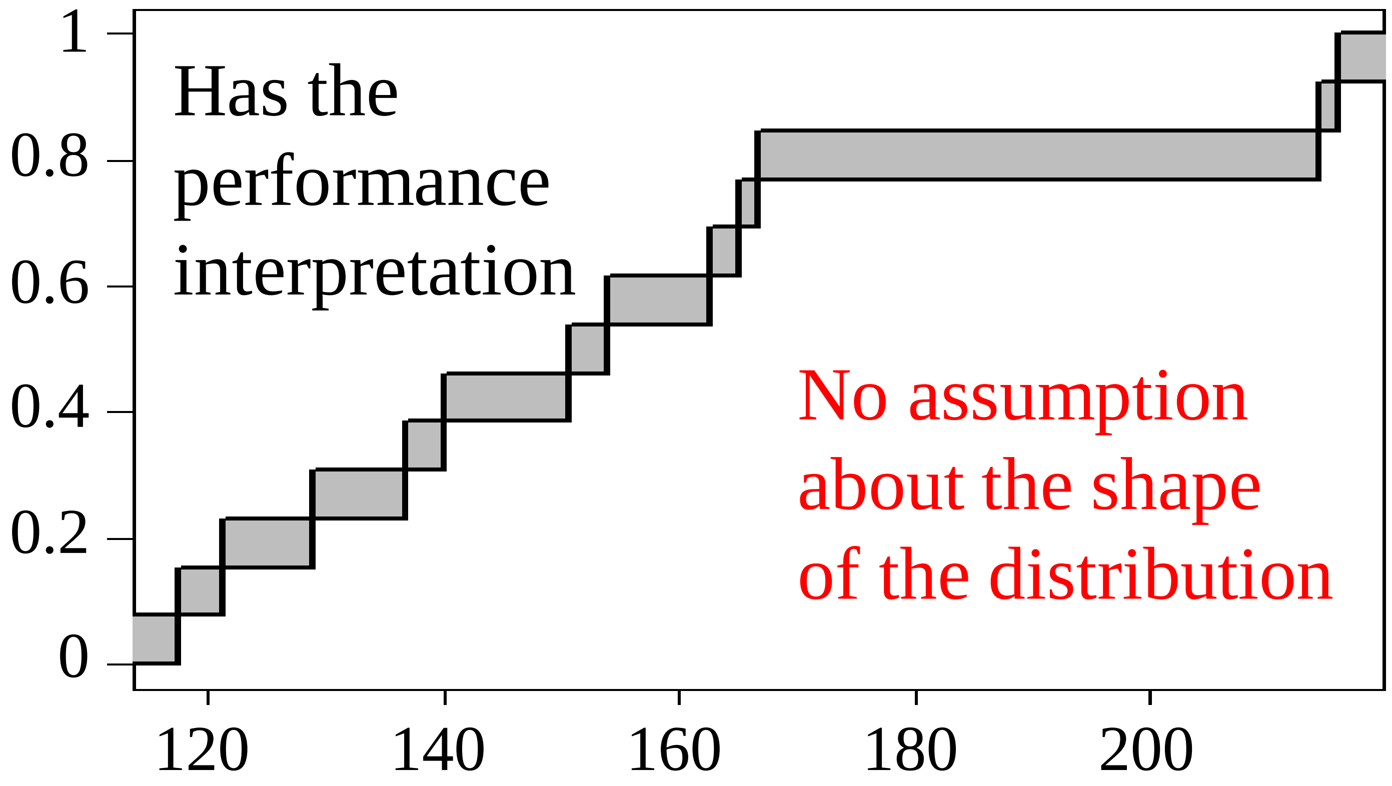
binomial(p, n), given n	normal(μ, σ)
binomial(p, n), given p	lognormal(μ, σ)
binomial(p, n)	gamma(a, b)
Poisson(p)	exponential(λ)
⋮	⋮

Example: non-parametric problem

$X \sim [(1+C(x))/(1+n), C(x)/(1+n)]$

where $C(x) = \#(X_i \leq x)$

Data
140.2
121.2
154
162.6
136.9
215.9
117.5
166.7
165.2
128.9
150.6
214.3



Captured uncertainties

- Uncertainty about distribution shape
- Sampling uncertainty (from small *n*)
- Measurement incertitude (±, censoring)
- Demographic stochsticity (integrality of data)

Propagated as probability boxes

- C-boxes can be combined in mathematical expressions using the p-box technology

- *Results also have performance interpretations*

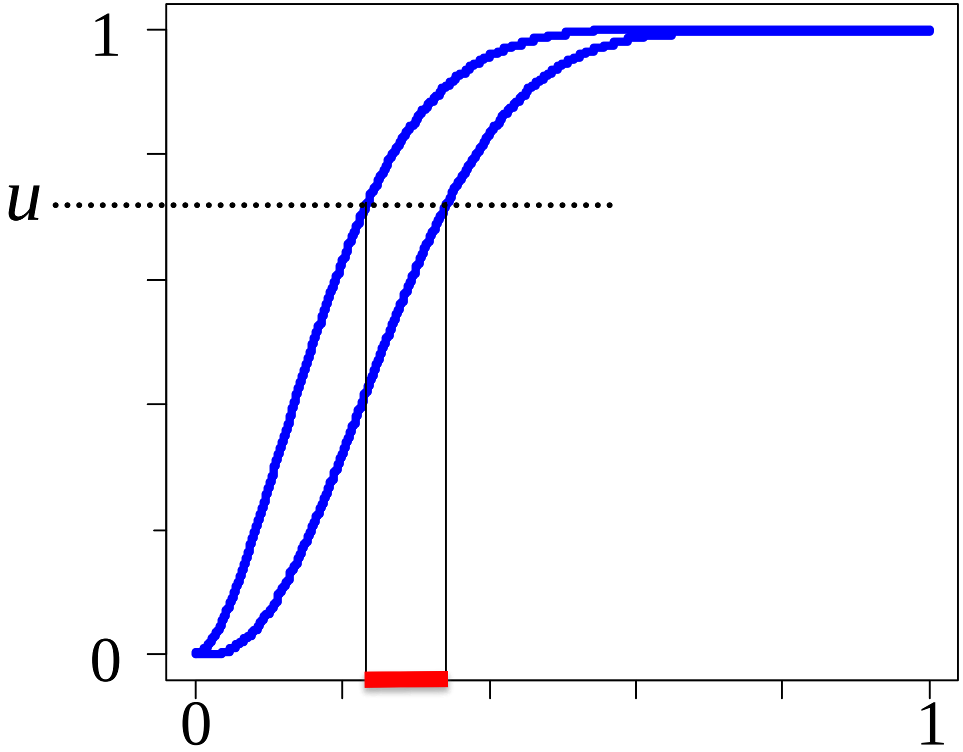
- C-boxes can also make predictive p-boxes
  - Analogous to frequentist prediction distributions
  - Or Bayesian posterior predictive distributions

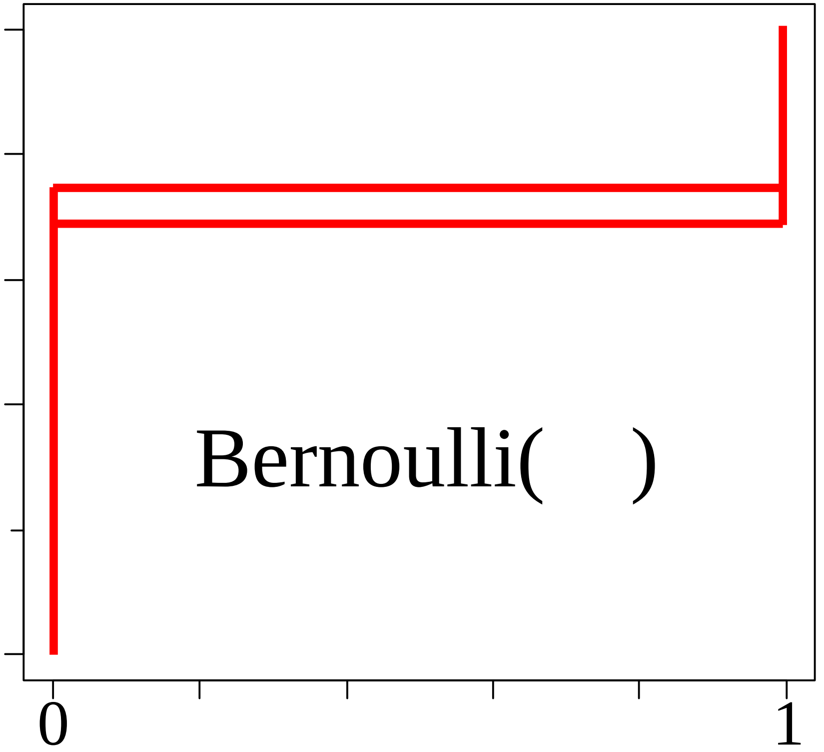
Say whaaaa?

Prediction structures

- C-boxes can model the uncertainty about the *underlying distribution* that generated the data
- Stochastic mixture of p-boxes from interval parameters specified by slices from the c-box
- This is a *composition* of the c-box through the probability model

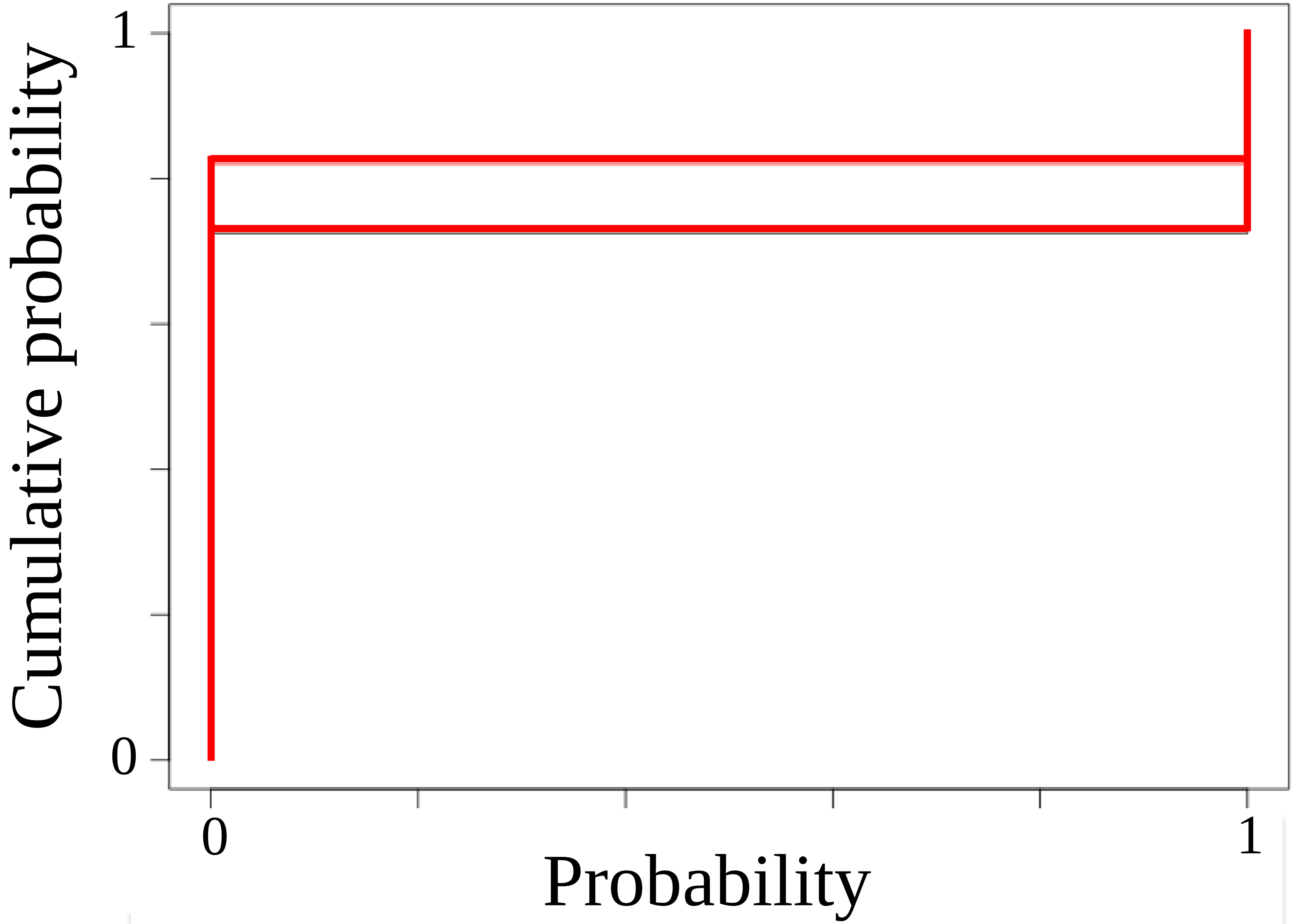
Example: Bernoulli distribution





Each interval slice defines a p-box for the underlying distribution (rather than a precise distribution)

Average all such p-boxes



Predictive p-boxes for *observables*

Bernoulli trials $b_i \sim \mathrm{B}(p)$, $i=1,\ldots,n$

$b_{n+1} \sim \mathrm{Bernoulli}([k, k+1]/(n+1))$, $k = \Sigma_i b_i$

Random binomial sample $k \sim \mathrm{binomial}(p, n)$

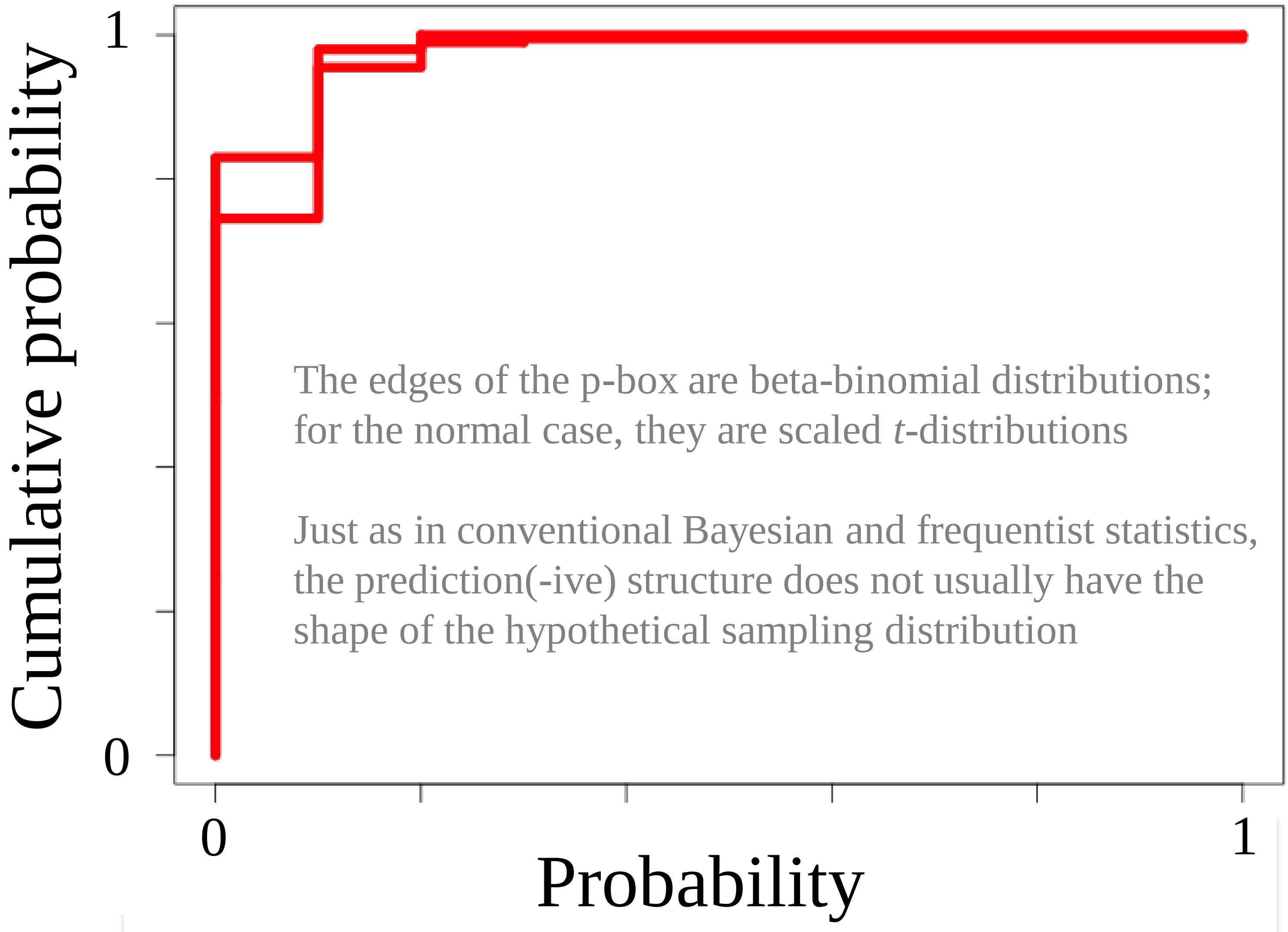
$k_2 \sim [\mathrm{BB}(k, nN-k+1, N), \mathrm{BB}(k+1, nN-k, N)]$

Random normal samples $X_i \sim \mathrm{N}(\mu, \sigma)$, $i=1,\ldots,n$

$X_{n+1} \sim \mathrm{mean}(X_i) + \mathrm{stdev}(X_i) \cdot \mathrm{sqrt}(1+1/n) \cdot \mathrm{T}(n-1)$

BB denotes beta-binomial; T denotes Student's *t*

Example: Binomial distribution



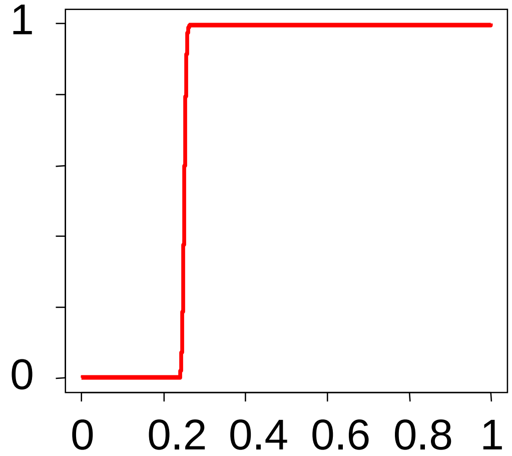
The edges of the p-box are beta-binomial distributions; for the normal case, they are scaled *t*-distributions

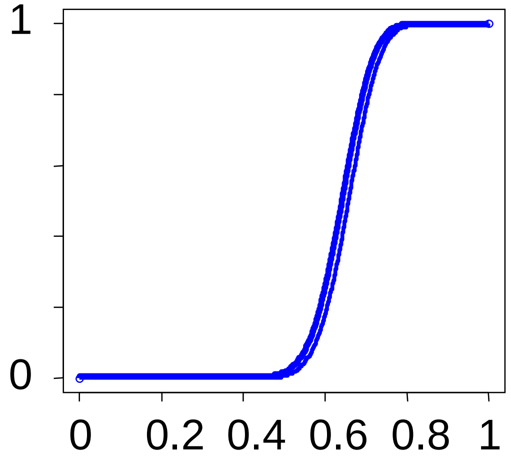
Just as in conventional Bayesian and frequentist statistics, the prediction(-ive) structure does not usually have the shape of the hypothetical sampling distribution

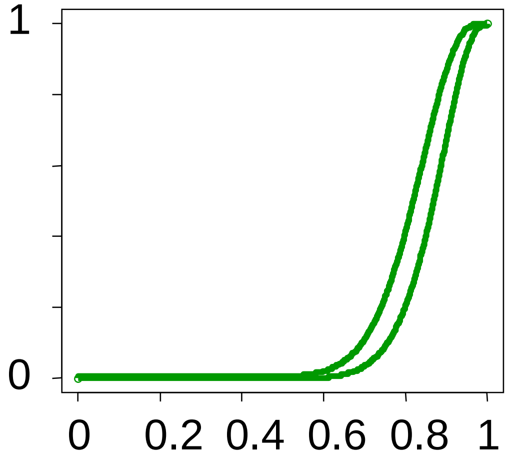
Prediction structures are p-boxes

- Results also have the confidence interpretation, but the intervals are *prediction intervals* rather than confidence intervals
- Prediction intervals enclose a specified percentage of observable values, on average
- It is also possible to derive analogous tolerance structures, which encode *tolerance intervals* (X% sure to enclose Y% of the population)

Computing with c-boxes







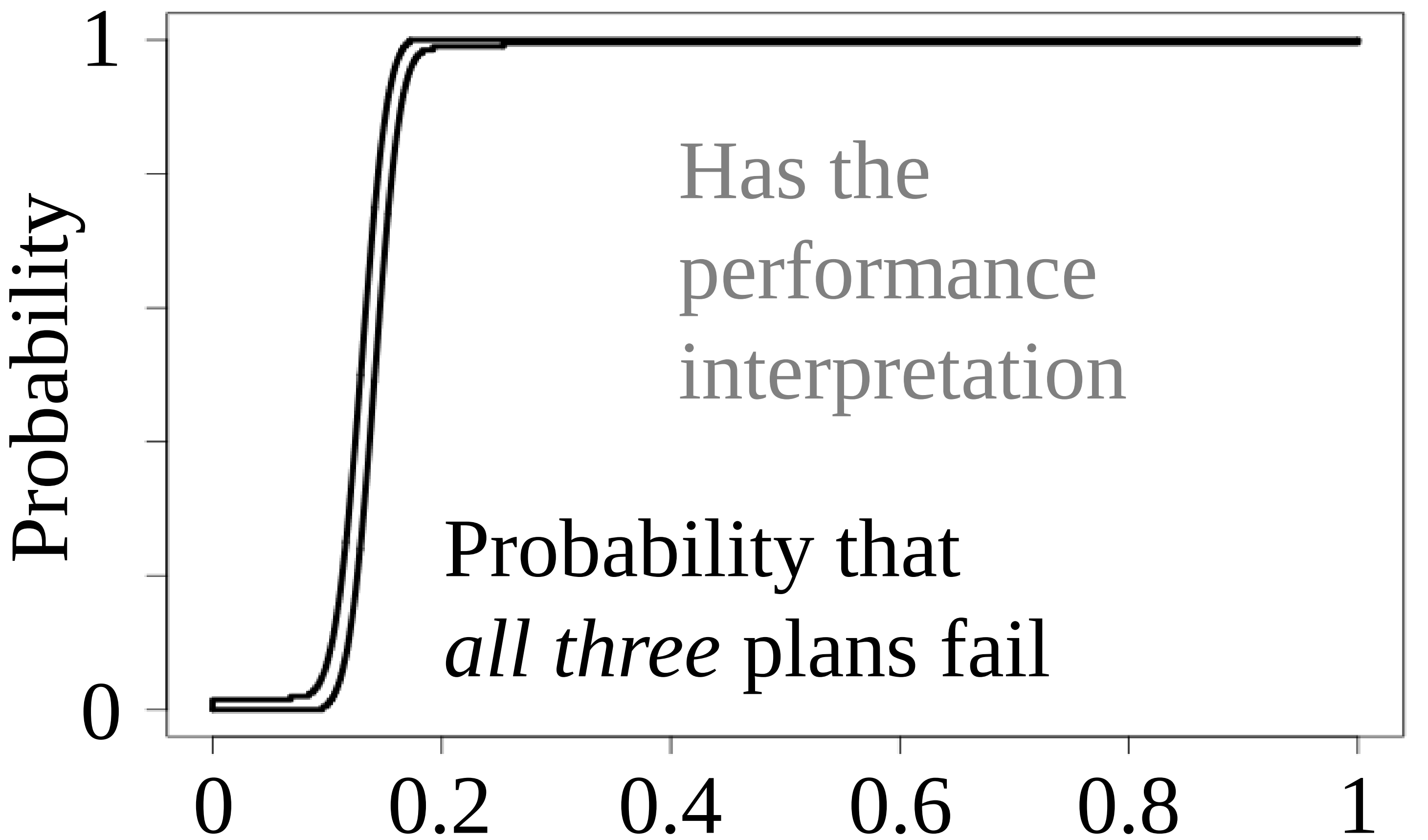
Plan A
25% fail

Plan B
39 out of 60 failed

Plan C
17 out of 20 failed

What if we used all three plans independently?

Conjunction (AND)



C-boxes are fully Bayesian

Bayesian sensitivity analysis

- Under robust Bayes approach, c-boxes can be thought of as Bayesian posteriors
  - Don't require specification of a unique prior
  - Have added feature of statistical performance
  - Imply posterior predictive distributions
  - Compatible with specifying a robust or precise prior when that's desirable

C-boxes are fully frequentist too

- Frequentists like confidence intervals but cannot use them in subsequent calculations

- Bayesians can compute with posteriors, but they don’t guarantee statistical performance

- C-boxes take the best from both worlds

Summary for c-boxes

- Confidence boxes carry inferential uncertainties through mathematical operations
- Give confidence intervals on results at any α level
- Defined by performance, so not unique
  - Just as confidence intervals are not unique
  - May create some flexibility
- Don’t seem to be overly conservative
  - Elaborate simulation studies have so far not found this

Conclusions

- C-boxes characterize risk analysis inputs given limited sample or constraint information
- Reasonable answers when data and tenable assumptions don’t justify particular distributions
- C-boxes don’t optimize; they *perform*
- C-boxes could serve as the lexicon in a language of risk analysis

More information

https://sites.google.com/site/confidenceboxes/

- Papers
- Slide presentations
- Free software

Google “confidence boxes” [plural…singular is a blog on teenage self-esteem & self-empowerment]

Little data, or none at all

Dramatic technocratic failures

- Shuttle risk estimated at 1/10,000 per flight, it was actually 1 per 70 flights
- 100-year flood risks underestimated and uncertainties not communicated
- Grossly understating risks in the financial industry precipitated the 2008 recession
- Failure cascades in electricity distribution systems are more common than they are forecasted to be (RAWG 2005; USCPSOTF 2006)

Dramatic technocratic failures

- Kansai International Airport
- Diablo Canyon Nuclear Power Plant
- Vioxx withdrawal missed decision context
- Fukushima non-redundancy
- Mars Climate Orbiter units
- Google Flu
- Ariane 5 integer overflow
- Minneapolis bridge collapse

These are not noble failures

- They are not simply the unfortunate but expected extreme tail events
- They occur more often than our risk estimates predict
- Analysts are doing risk estimations wrong
- Managers never see the problems coming

Engineers cannot always get data

- New systems may have no performance history
  - spacecraft of new design or in a new environment
  - biological control strategies using novel genetic constructs that have never existed before

- Two ways to estimate probabilities without data
  - disaggregation into parts whose probabilities are easier to estimate (i.e., breaking it into subproblems)
  - expert elicitation (i.e., guessing)

Rare event probabilities

- Hardly ever any actual data
- Sometimes we have experts
- But how should we model bald assertions:
  - “1 in 10 million”
  - “about 1 in 1000”
  - “never been seen in 100 years”

In my experience

People using expert elicitation often take the pronouncements made by their informants *exactly as they are specified*

– Sometimes even as point values

(If your experience is different, I'd like to chat)

Is there no uncertainty?

- What is the uncertainty in estimates like
  - "1 in 10^7",
  - "about 1 in 1000", or
  - "never been seen in over 100 years of observation"?

- How should this uncertainty be captured and projected in computations?

Rare events

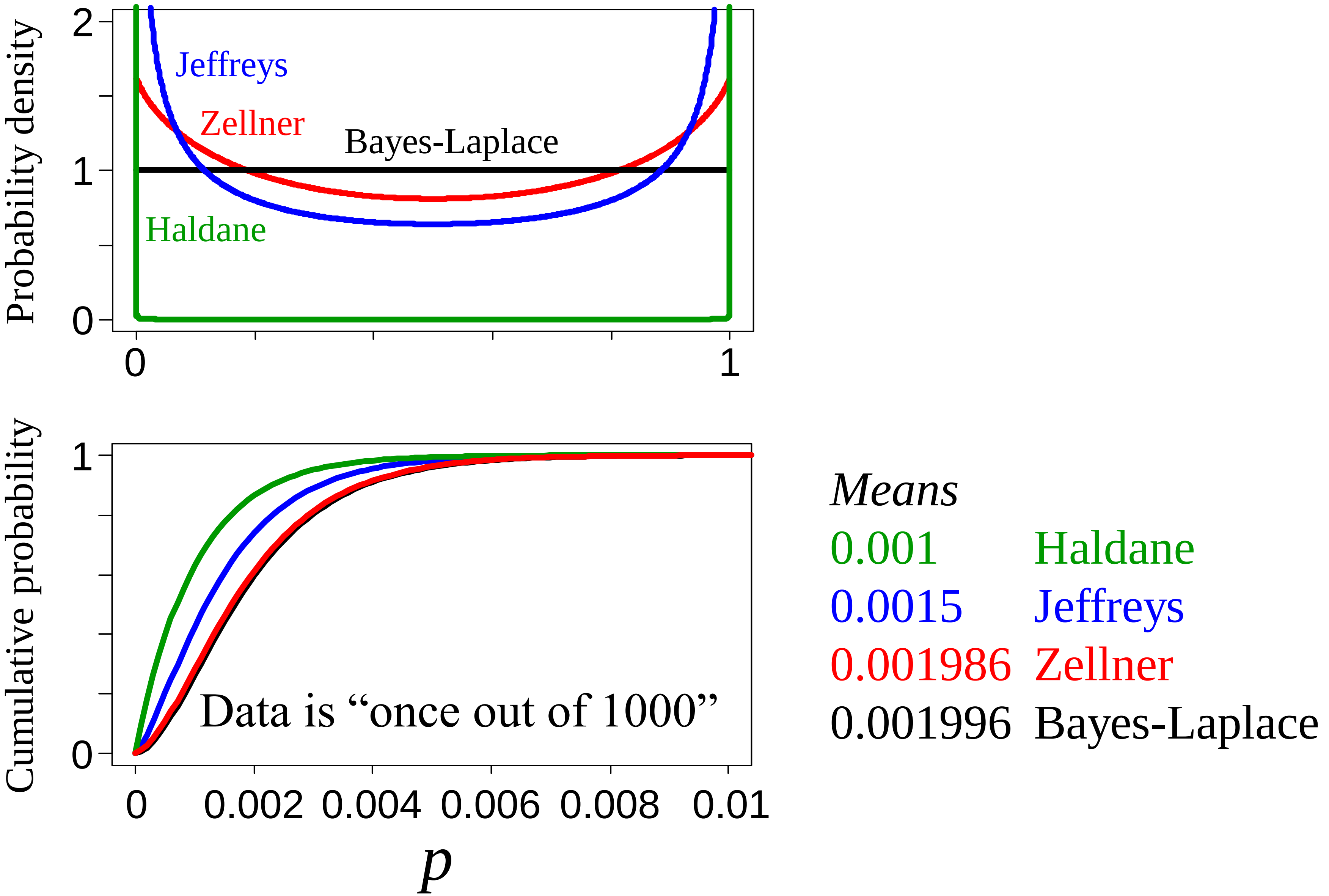
- Often the driving concern in analyses
- Typically big consequences
- Hardly ever characterized by good data
- Perhaps never seen, or seen only once

Random sample data

maximum likelihood

- ML says *p* is zero for never-seen events
  - Nobody believes this is a reasonable estimate
- Bayesian estimator is more reasonable…

Just kidding



Probability density
2
1
0
0
1
Jeffreys
Zellner
Bayes-Laplace
Haldane
Cumulative probability
1
0
0
0.002
0.004
0.006
0.008
0.01
p
Data is “once out of 1000”
Means
0.001 Haldane
0.0015 Jeffreys
0.001986 Zellner
0.001996 Bayes-Laplace

Random sample data

- ML says *p* is zero for never-seen events
  - Nobody believes this is a reasonable estimate
- The Bayesian estimator is many things
  - 'Reasonable' only in that *it's whatever you want*
- Modern estimators
  - Imprecise beta (Dirichlet) models
  - Confidence structures

Confidence structure (c-box)

- P-box-shaped estimator of a (fixed) parameter
- Gives confidence interval at *any* confidence level
- Can be propagated just like p-boxes
- Allow us to *compute with confidence*

Probability of rare event

- Inference about probability from binary data, k successes out of n trials

$$p \sim [\mathrm{beta}(k, n-k+1), \mathrm{beta}(k+1, n-k)]$$

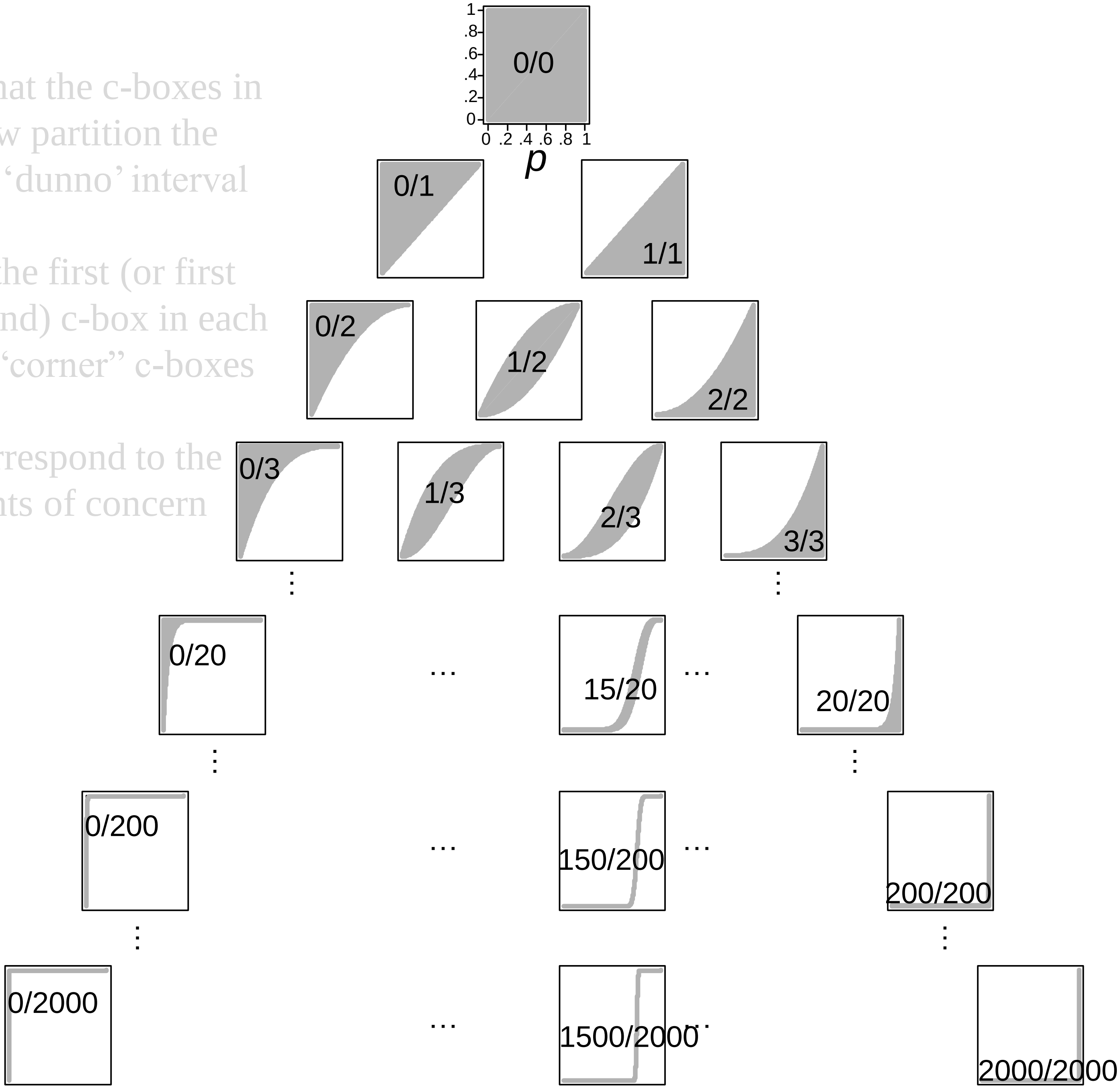
Notation extends the use of tilda

- Identical to Walley's Imprecise Beta Model with s=1, but needs no prior

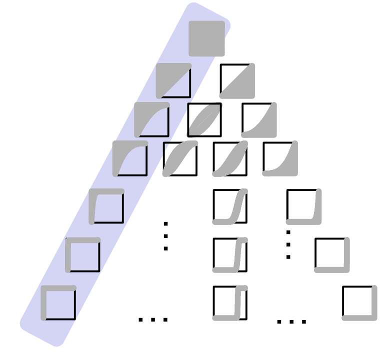
Notice that the c-boxes in every row partition the vacuous 'dunno' interval

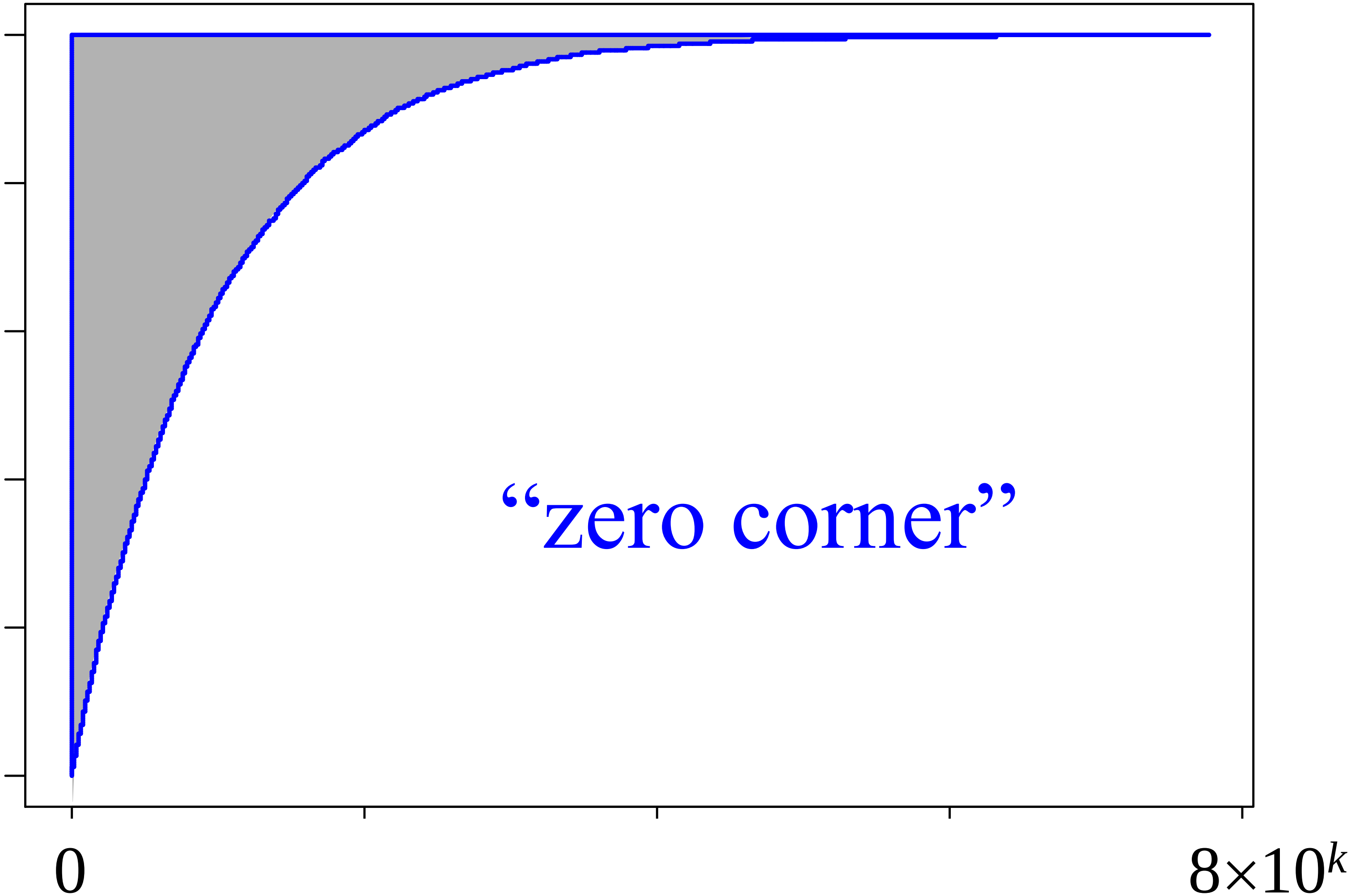
We call the first (or first and second) c-box in each row the "corner" c-boxes

They correspond to the rare events of concern

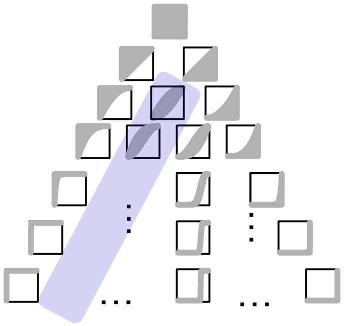


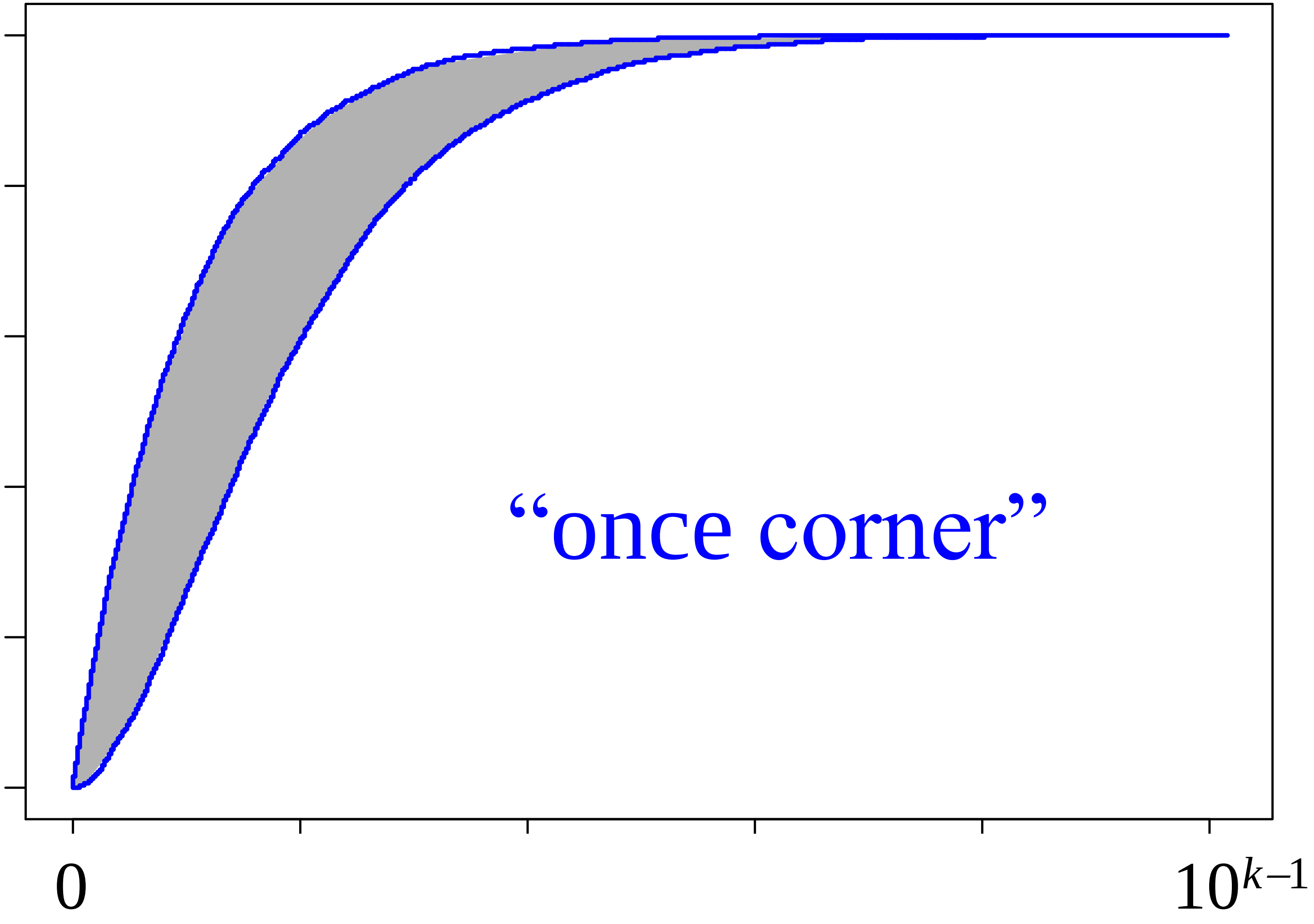
Zero out of 10^k trials





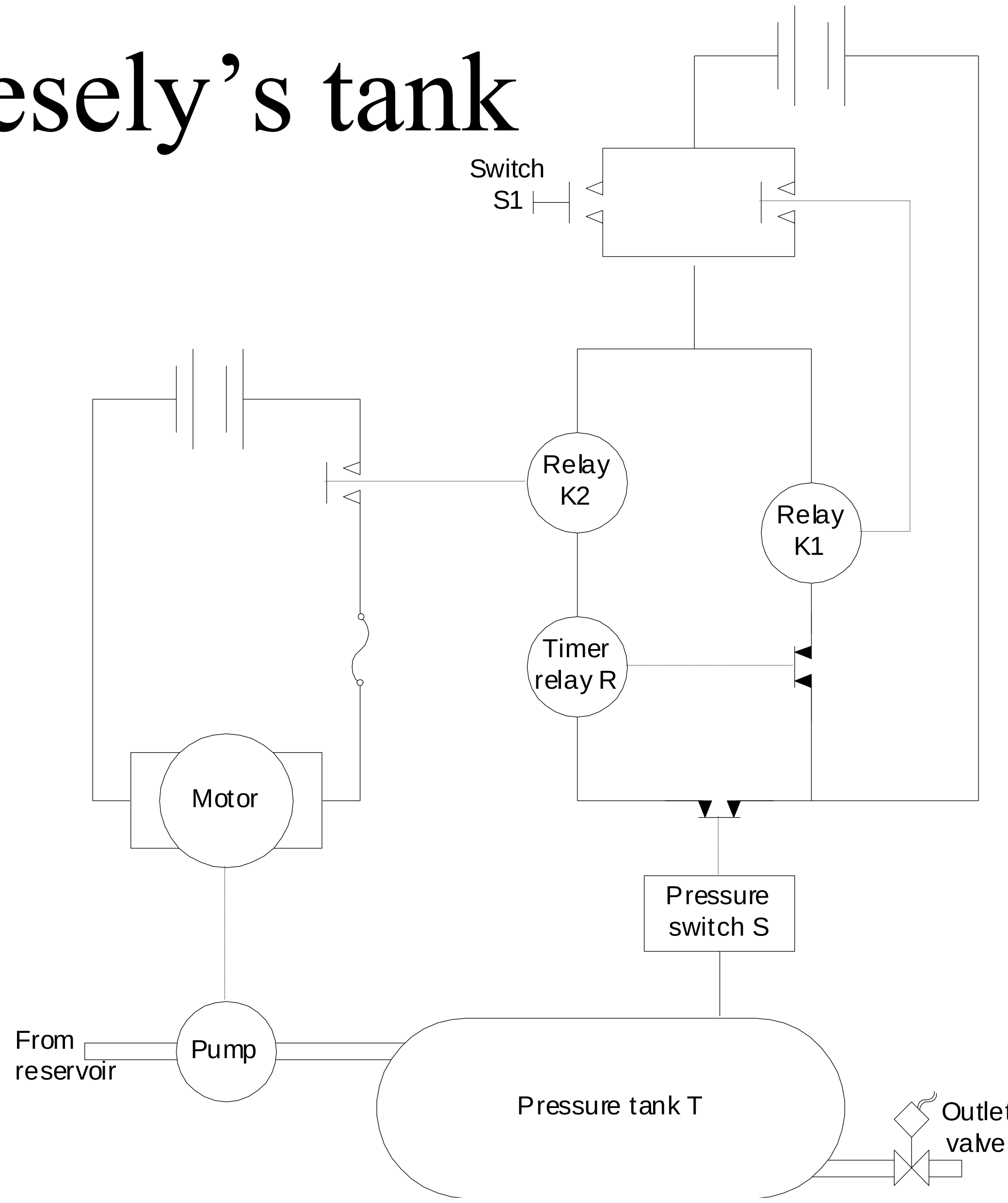
One out of 10^k trials





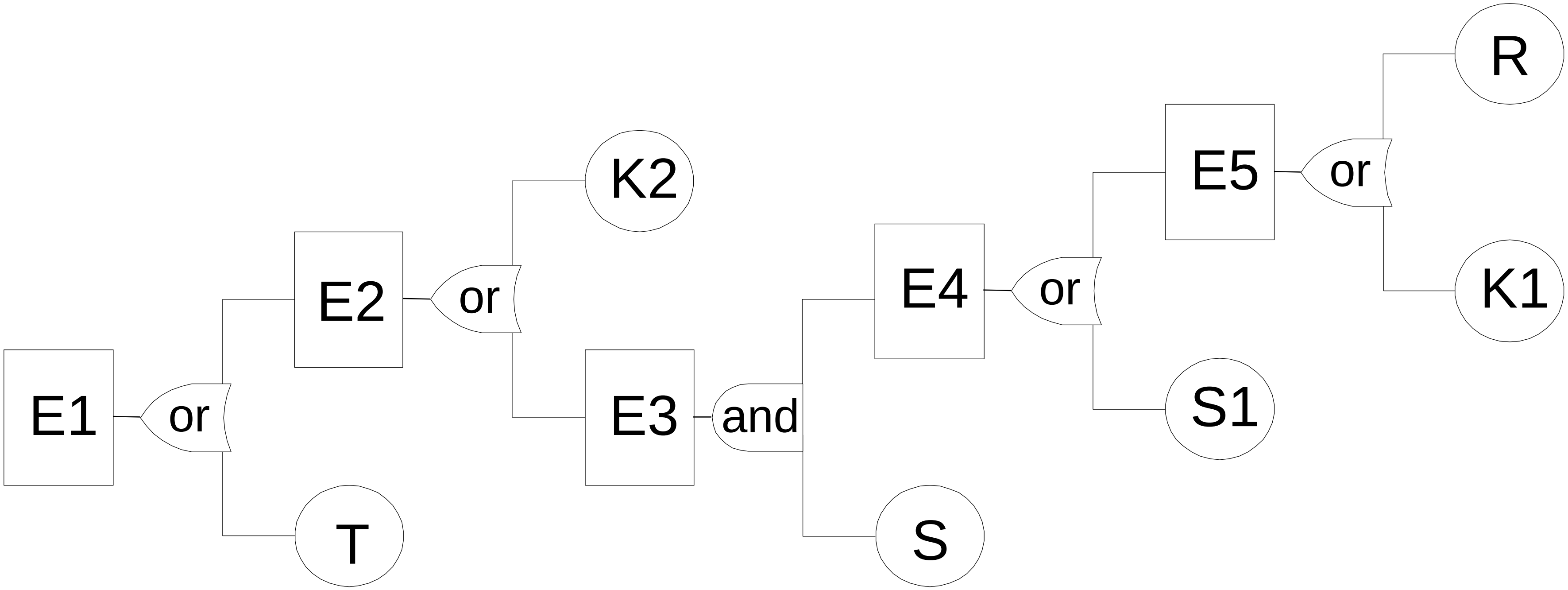
Vesely's tank

What's the chance the tank ruptures under pumping?



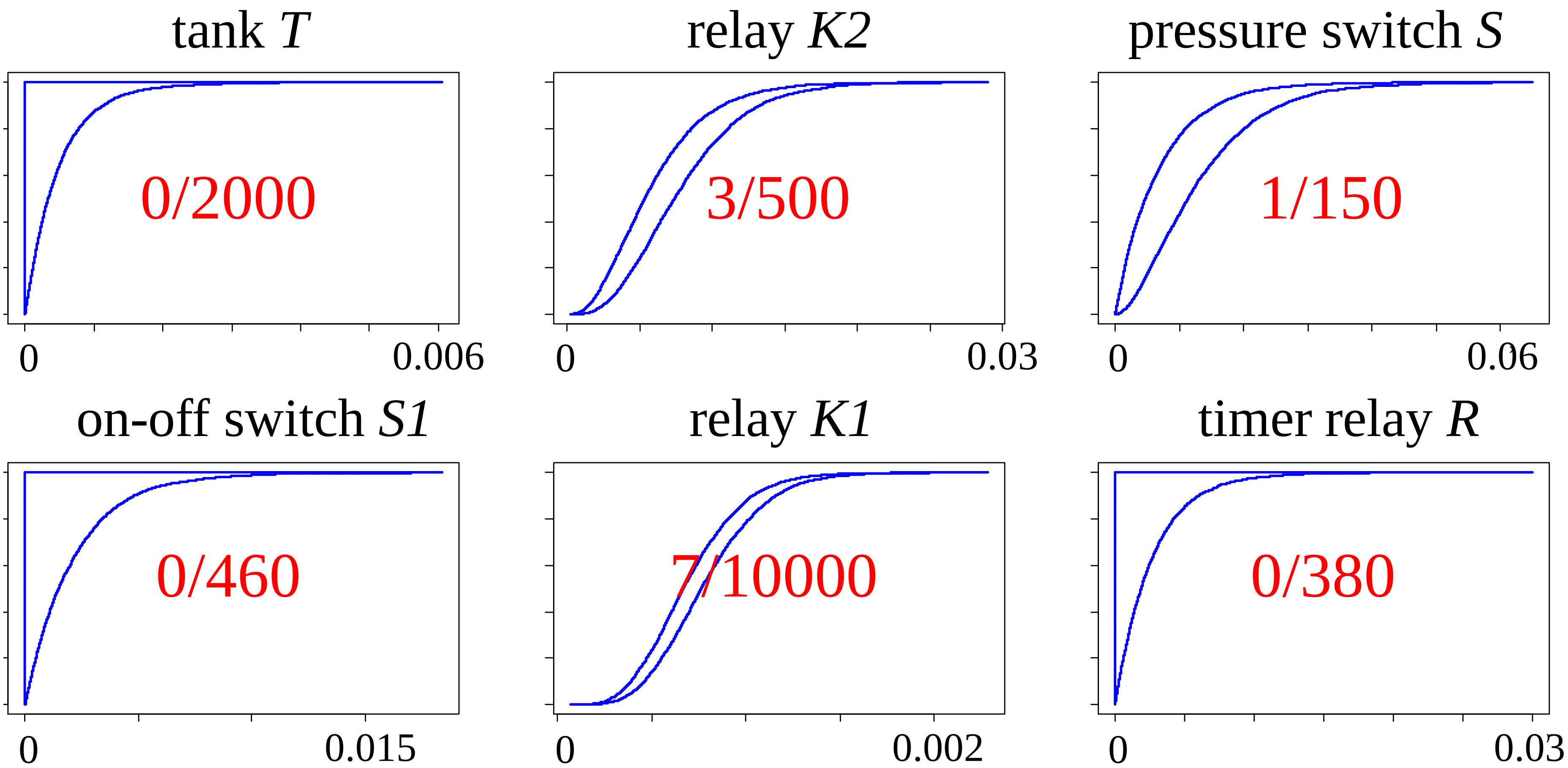
Vesely et al. 1981

Fault tree

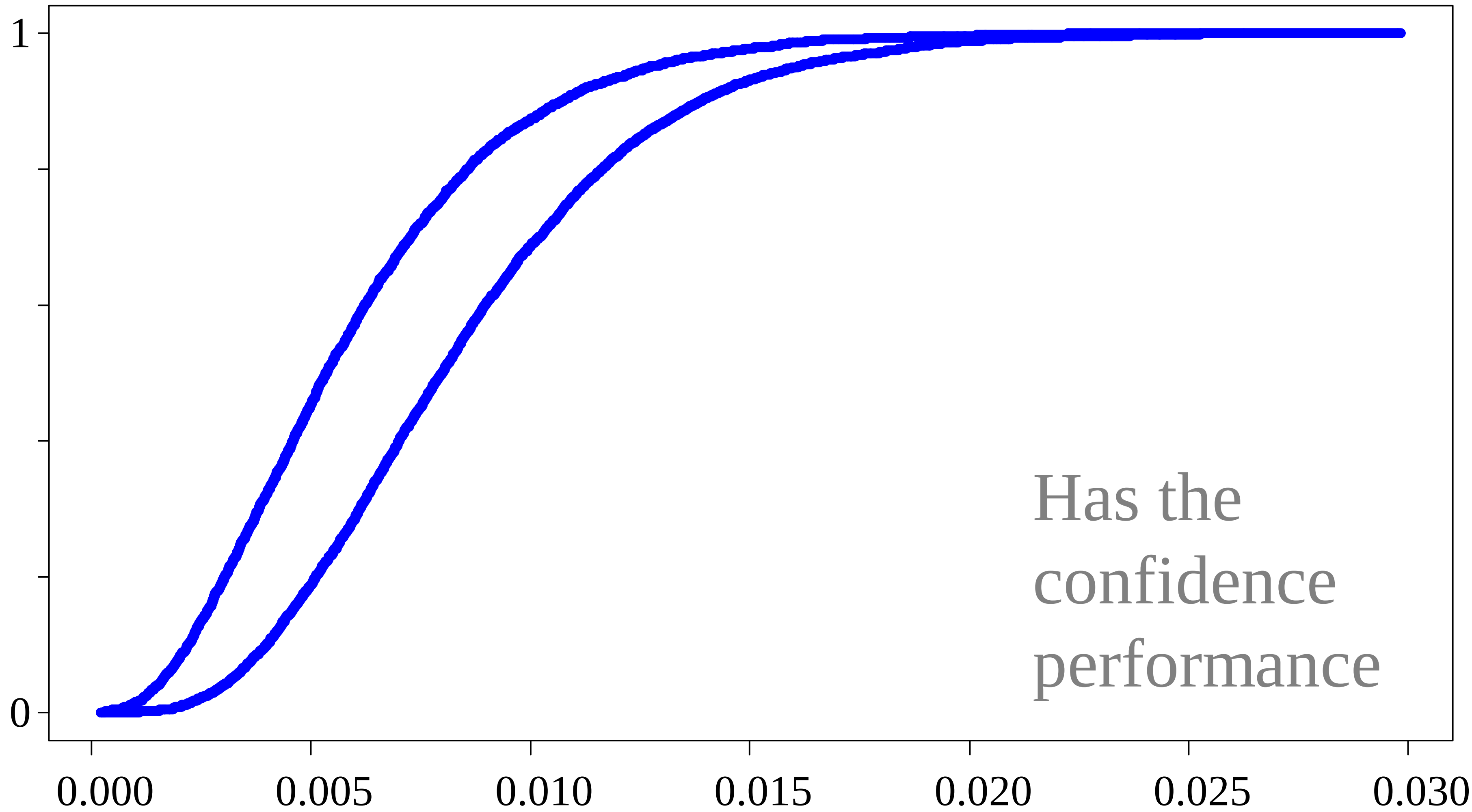


$$E1 = T \vee (K2 \vee (S \,\&\, (S1 \vee (K1 \vee R))))$$

Vesely's pressurized tank



Top event tank rupture under pressurization *E1*



Try it in pba.r

T = CBbinomial(0, 2000)

K2 = CBbinomial(3, 500)

.

.

.

E1 = T %|% (K2 %|% (S %&% (S1 %|% (K1 %|% R))))

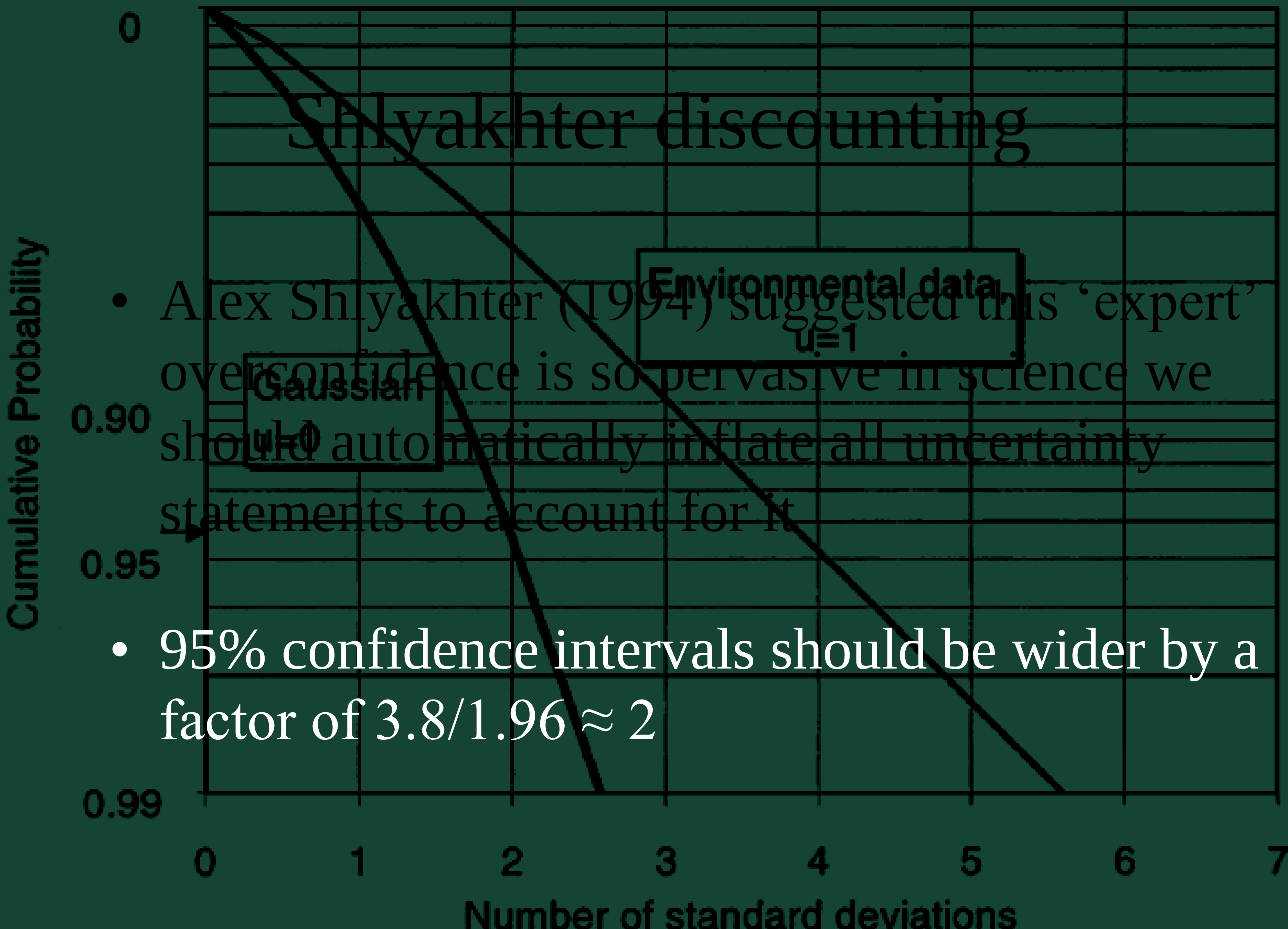
Overconfidence

- People, including scientists and engineers, systematically understate their uncertainty
  - 90% confidence intervals ought to enclose the true value 90% of the time, but do only about 40% of time
  - Overconfidence "has been found to be almost universal in all measurements of physical quantities"
- Likely to be at least as important in expert elicitation when nothing is being measured

Shlyakhter discounting

- Alex Shlyakhter (1994) suggested this ‘expert’ overconfidence is so pervasive in science we should automatically inflate all uncertainty statements to account for it

- 95% confidence intervals should be wider by a factor of 3.8/1.96 ≈ 2



We need an elicitation penalty

- Expert opinions aren't like random *data*
- Their uncertainty should be inflated
- The penalty size should be derived empirically from validation studies of prior elicitations

Who's doing such studies?

Here's a *placeholder* for the penalty

- The numbers in "1 in 300" are not counts but estimates with imprecision implied by sigdigs

"1 in 300" = [0.5, 1.5] / [250, 350]
= [0.5 / 350, 1.5 / 250]

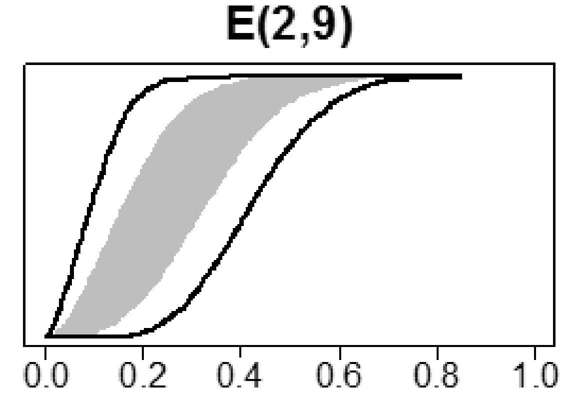
- The *envelope* of all corresponding c-boxes

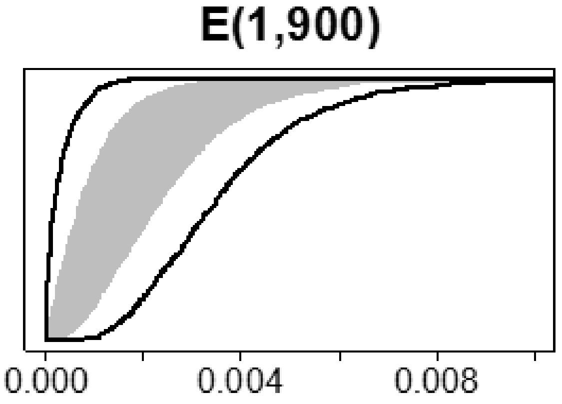
Significant-digit intervals

- Significant digits imply an interval (± half the magnitude of the last significant decimal place)

x	s(x)
0	[0, 0.5]
1	[0.5, 1.5]
9	[8.5, 9.5]
10	[5, 15]
300	[250, 350]
8150	[8145, 8155]
1×10^7	[5×10^6, 1.5×10^7]

Examples for elicitations “*k* in *n*”





Gray c-boxes B(*k*, *n*), and black c-box envelopes E(*k*, *n*) = B(s(*k*), s(*n*))

More uncertainty for round numbers

- Doubles (or more than doubles) the uncertainty
- Why *more than* doubles?

  s(9) = [8.5, 9.5]   unit width

  s(10) = [5, 15]   width of 10

- Presumes greater uncertainty when round numbers are used to characterize a probability

Maximum likelihood

Maximum entropy

Bayesian inference

Estimation

Expert elicitation

“the great frontier of making things up”

Method of moments

PERT

C-boxes

- Don’t optimize anything; they *perform*
- Characterize inputs from limited or even no data
- With an uncertainty penalty, seems to capture some uncertainty in experts’ bald assertions like

  “1 in 1000”

  “1 in ten million”

  “never been seen in 100 years”

C-boxes are compatible p-boxes

- P-boxes have a performance interpretation too
- If constraints are known that specify a rigorous p-box, then it encodes prediction intervals
- So our performance interpretation applies for
  - Parametric problems
  - Nonparametric problems
  - No data problems
  - Constraint problems

Appears to provide a unification of asymptotically large sample estimates, small sample estimates, expert elicitation, and constraint analysis, as well as a union of Bayesian and frequentist criteria.

Topics

- Why bounding's not enough 15
- Dependence 35
- Sensitivity analysis 30
- Validation 30
- Backcalculation 25
- Spacecraft design at NASA 20
- Statistics for interval data 35
- Neuroscience of uncertainty 40
- Model uncertainty 20

Imprecise probabilities

Probability of an event

- Imagine a gamble that pays one dollar if an event occurs (but nothing otherwise)
  - How much would you pay to buy this gamble?
  - How much would you be willing to sell it for?

- Probability theory requires the same price for both
  - By asserting the probability of the event, you agree to buy any such gamble offered for this amount or less, *and* to sell the same gamble for any amount equal to or more than this ‘fair’ price…and to do so for *every* event!

- IP just says, sometimes, your highest buying price might be smaller than your lowest selling price

Why is yet another method needed?

- Some statements of uncertainty can't be expressed with p-boxes or interval probability
- Need to express uncertainty of *all kinds*
- Sometimes mere bounding isn't good enough

What bounding probability can’t do

- Represent comparative probability judgments, e.g., event *A* is at least as likely as event *B*
- Give unique expectations needed for making decisions
- Give unique conditional probabilities needed for making inferences
- Maintain best possible bounds through updating

(Walley 2000)

Imprecise probabilities

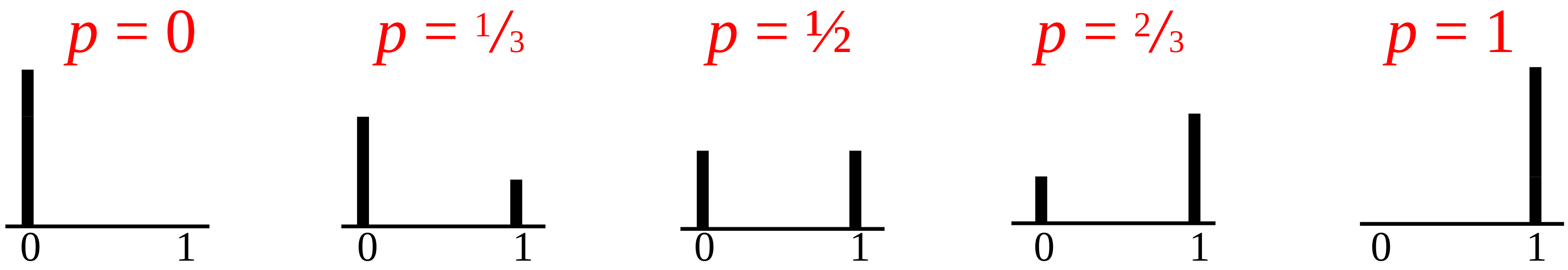
- Generic term for any theory that doesn't assume a unique underlying probability
- Often expressed in terms of closed, convex sets of probability distributions (not the same as a p-box)
- General case v. special case

    set v. interval

    imprecise probabilities v. p-boxes, probability intervals

Sets of distribution functions

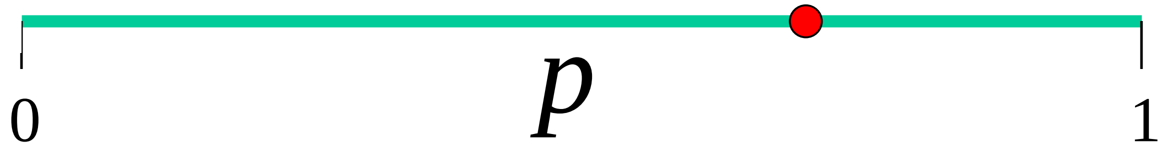
- Consider the set of all Bernoulli distributions (which are discrete with mass at only 0 and 1)



- Clearly, there's a one-dimensional family of such distributions, parameterized by how the mass is distributed between the two points

Space of distributions

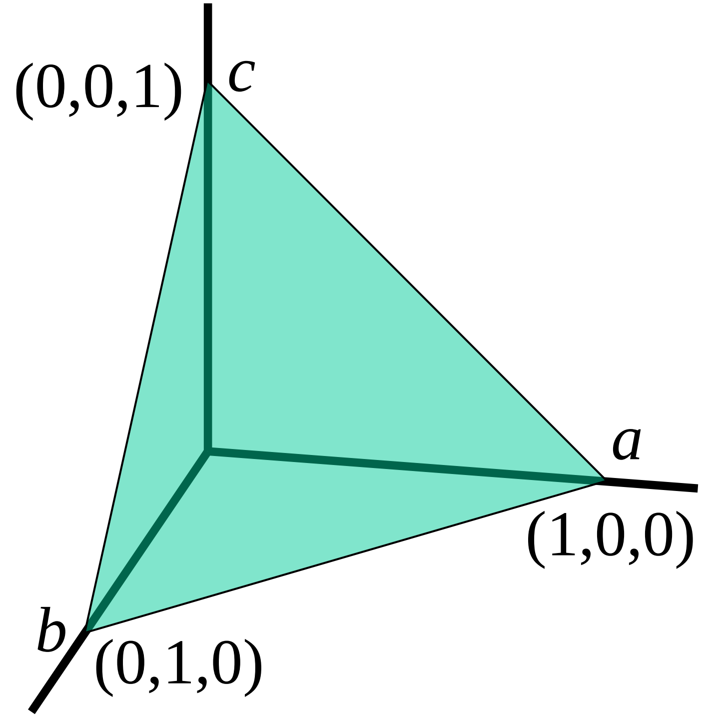
- This one-dimensional family constitutes a space of distributions in which each point represents a distribution



Three-dimensional case

- When the distributions have 3 point masses, the space becomes two-dimensional and has a triangular shape

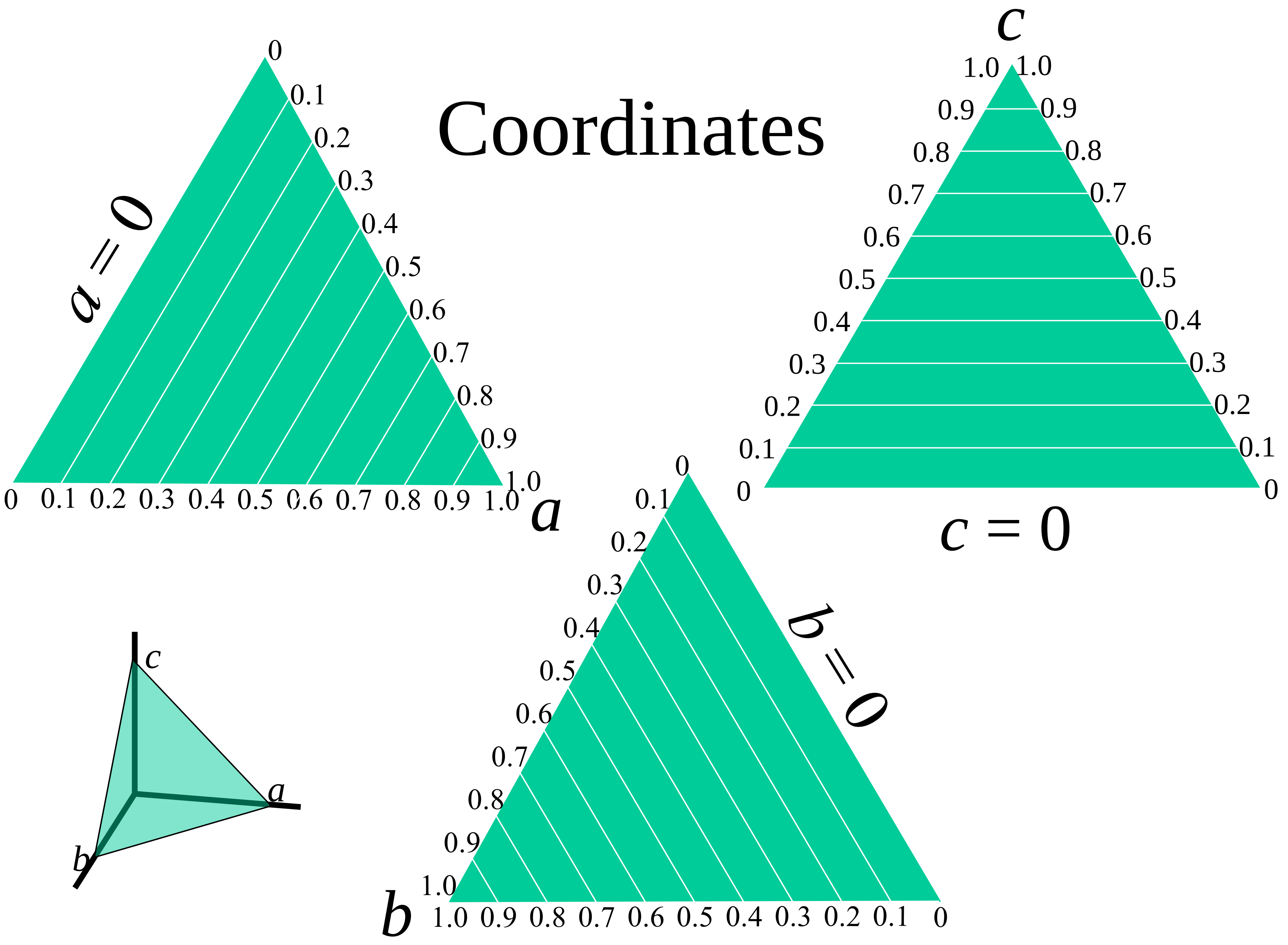
- The points on this surface are those whose coordinates add to one



Simplex

- For discrete distributions with n masses, the space, called a *simplex*, has $(n-1)$-dimensions
- One degree of freedom is lost to the constraint that probabilities sum to one
- For the continuous case, the space becomes infinite-dimensional, or you could be content to use discrete approximations

Coordinates



Walley’s (2000) football game

- 3 possibilities for our team: win, draw, loss
- Suppose we have qualitative judgments:

  ‘Not win’ is at least as probable as win

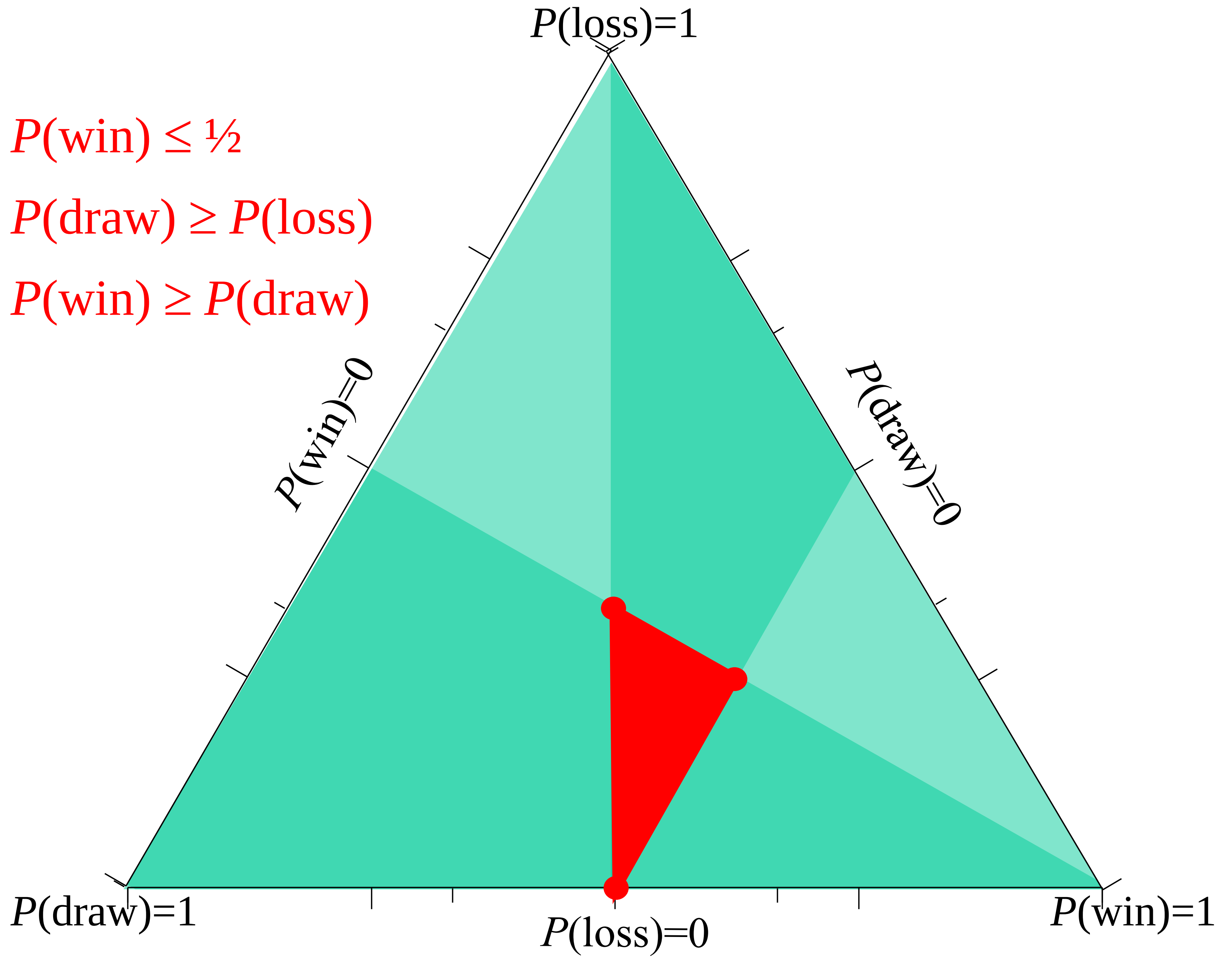
  Win is at least as probable as draw

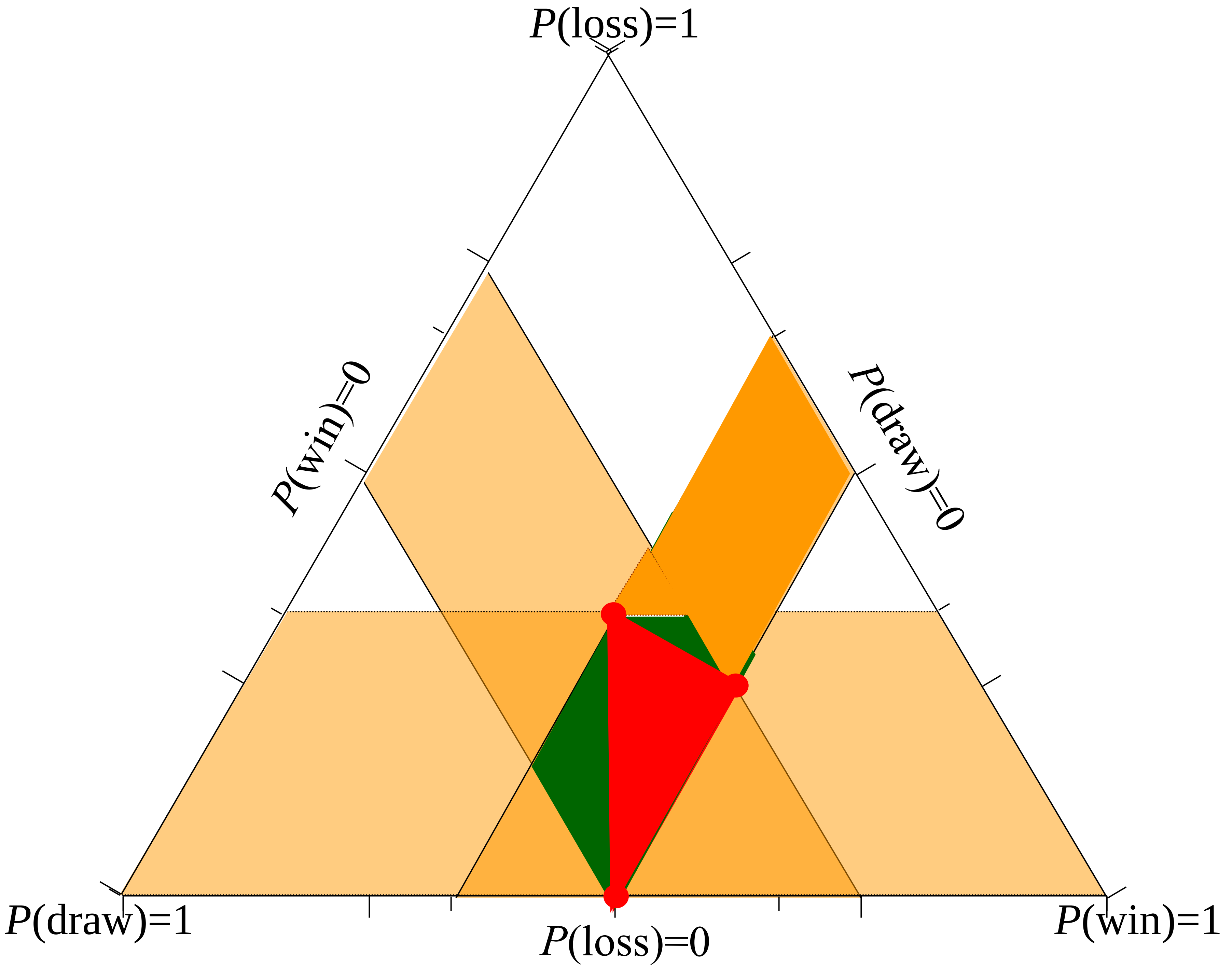
  Draw is at least as probable as loss
- These constrain the probability distribution P

  $P(\text{win}) \leq ½$

  $P(\text{win}) \geq P(\text{draw})$

  $P(\text{draw}) \geq P(\text{loss})$


P(loss)=1
P(win) ≤ ½
P(draw) ≥ P(loss)
P(win) ≥ P(draw)
P(win)=0
P(draw)=0
P(draw)=1
P(loss)=0
P(win)=1



P(loss)=1
P(win)=0
P(draw)=0
P(draw)=1
P(loss)=0
P(win)=1

So what?

- The closed, convex set of probability distributions (the red triangular region) expresses the uncertainty
- This set of distributions is smaller than the set implied by bounds on the three probabilities (the green area enclosing the triangle)
- This difference can affect expectations of functions that depend on the events, and conditional probabilities

Credal set

- Knowledge and judgments can be used to define a set of possible probability measures

Kinds of information

  - All distributions within bounds are possible
  - Only distributions having a given shape
  - Probability of an event is within some interval
  - Event *A* is at least as probable as event *B*
  - Nothing is known about the probability of *C*

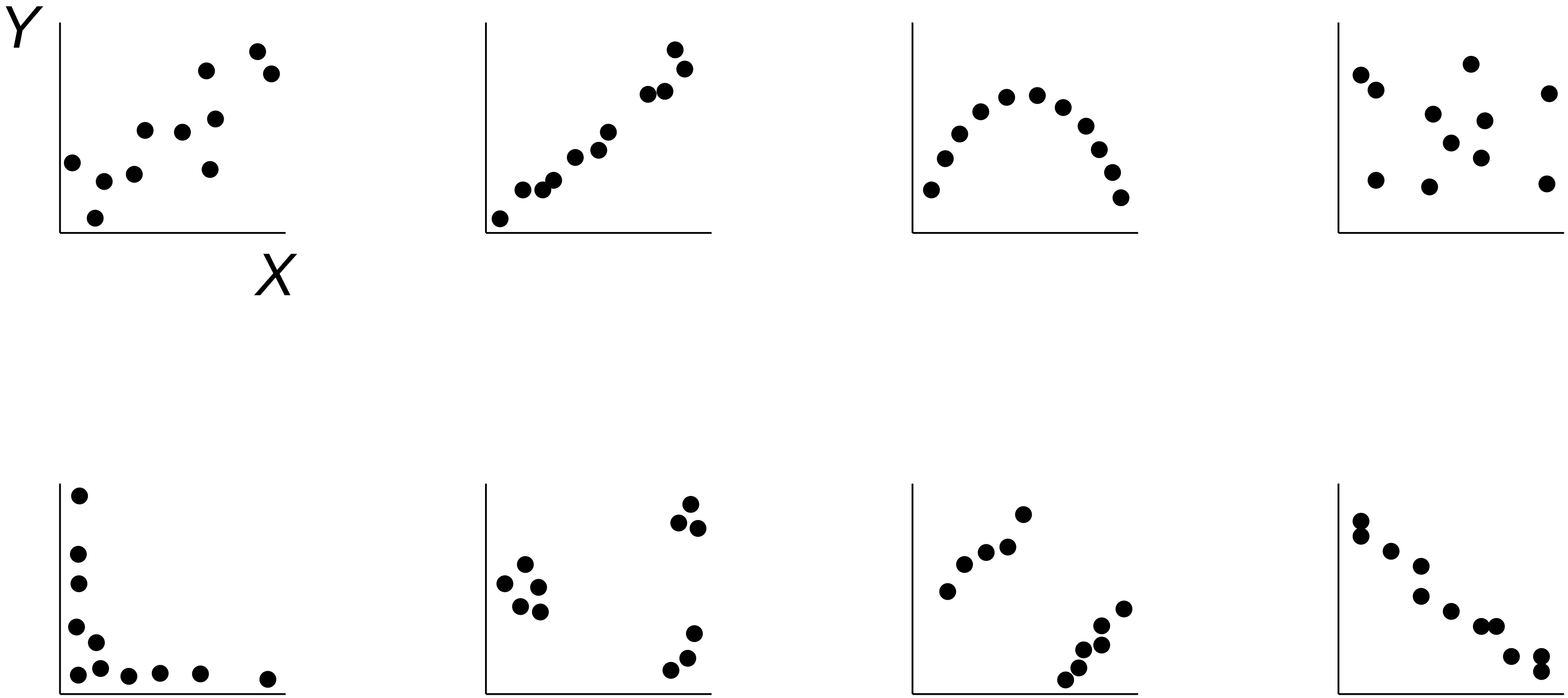
- Computing with credal sets usually requires mathematical programming

Correlation and dependency

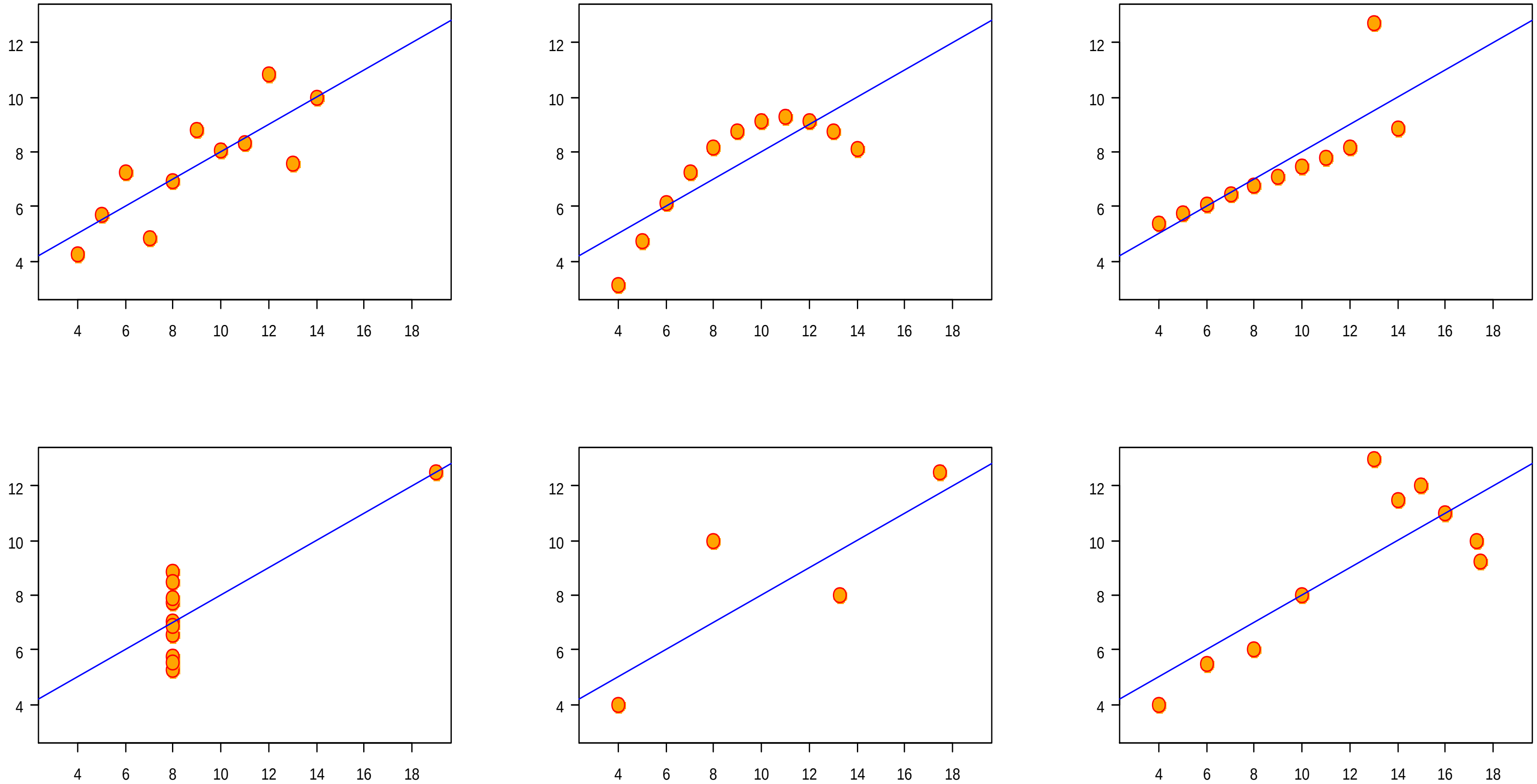
Dependence

- Most variables assumed independent
- But some variables clearly aren’t
  - Density and porosity
  - Rainfall and temperature
  - Body size and skin surface area

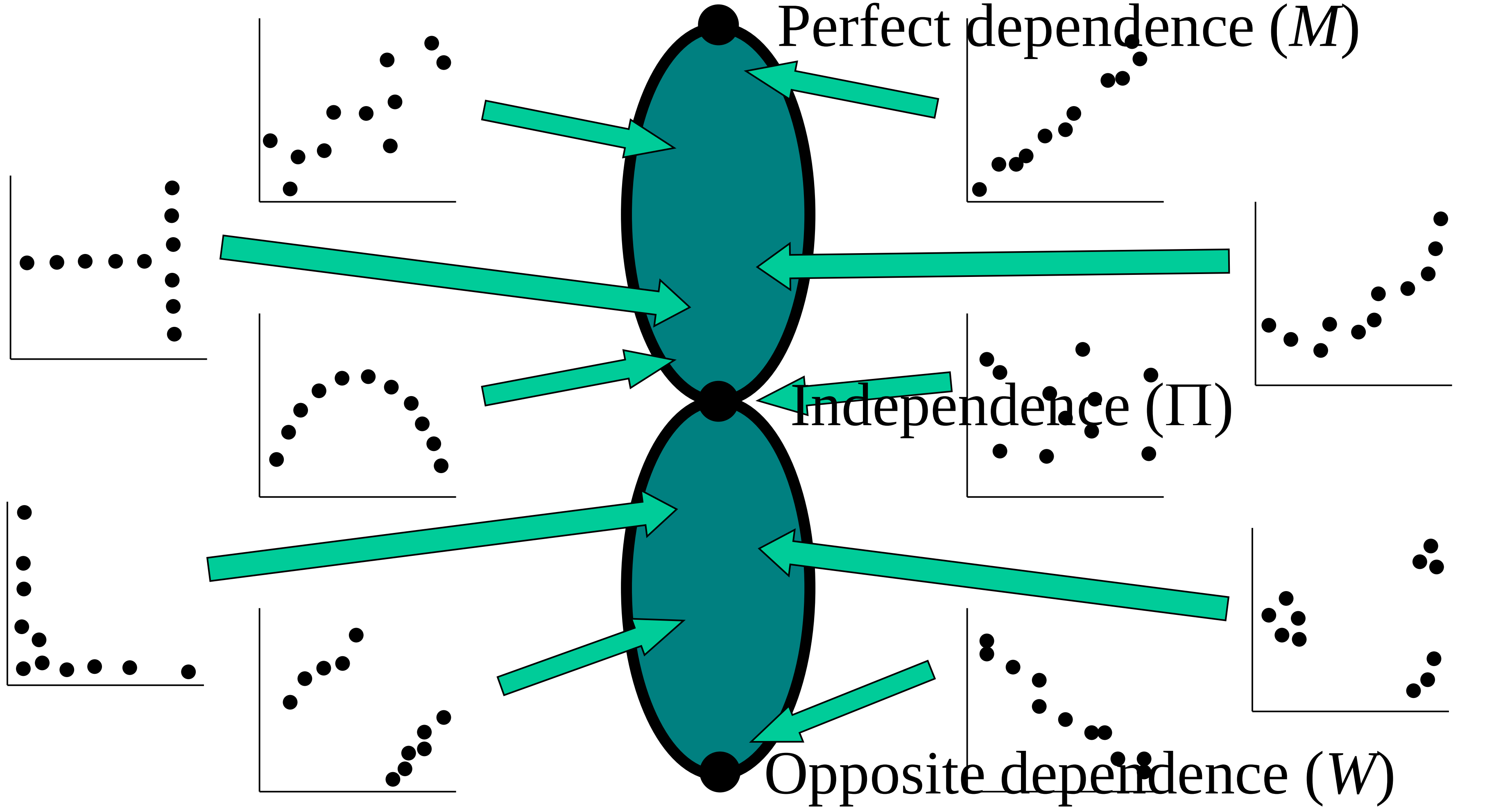
Dependence can be complex



Dependence for fixed correlation



Every scattergram has (Pearson) correlation 0.816



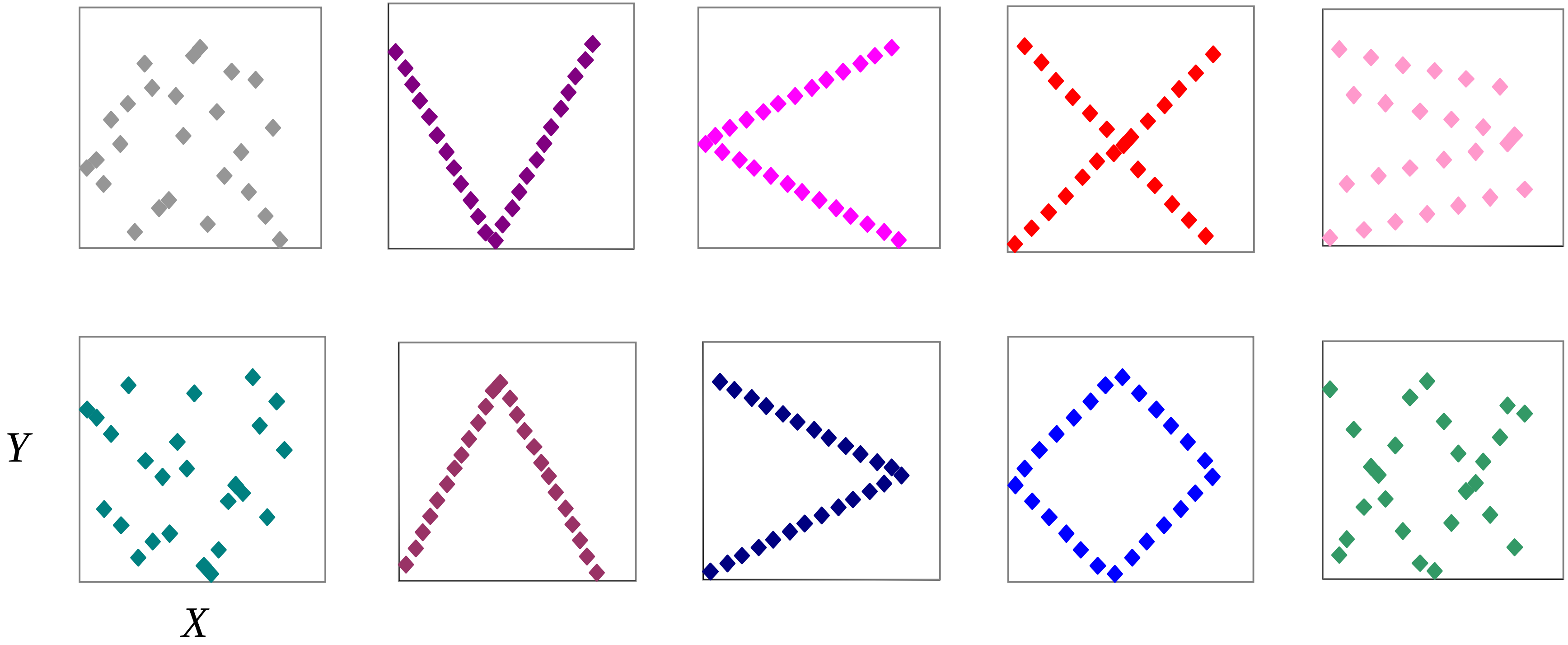
Perfect dependence (M)
Independence (Π)
Opposite dependence (W)

Dependencies

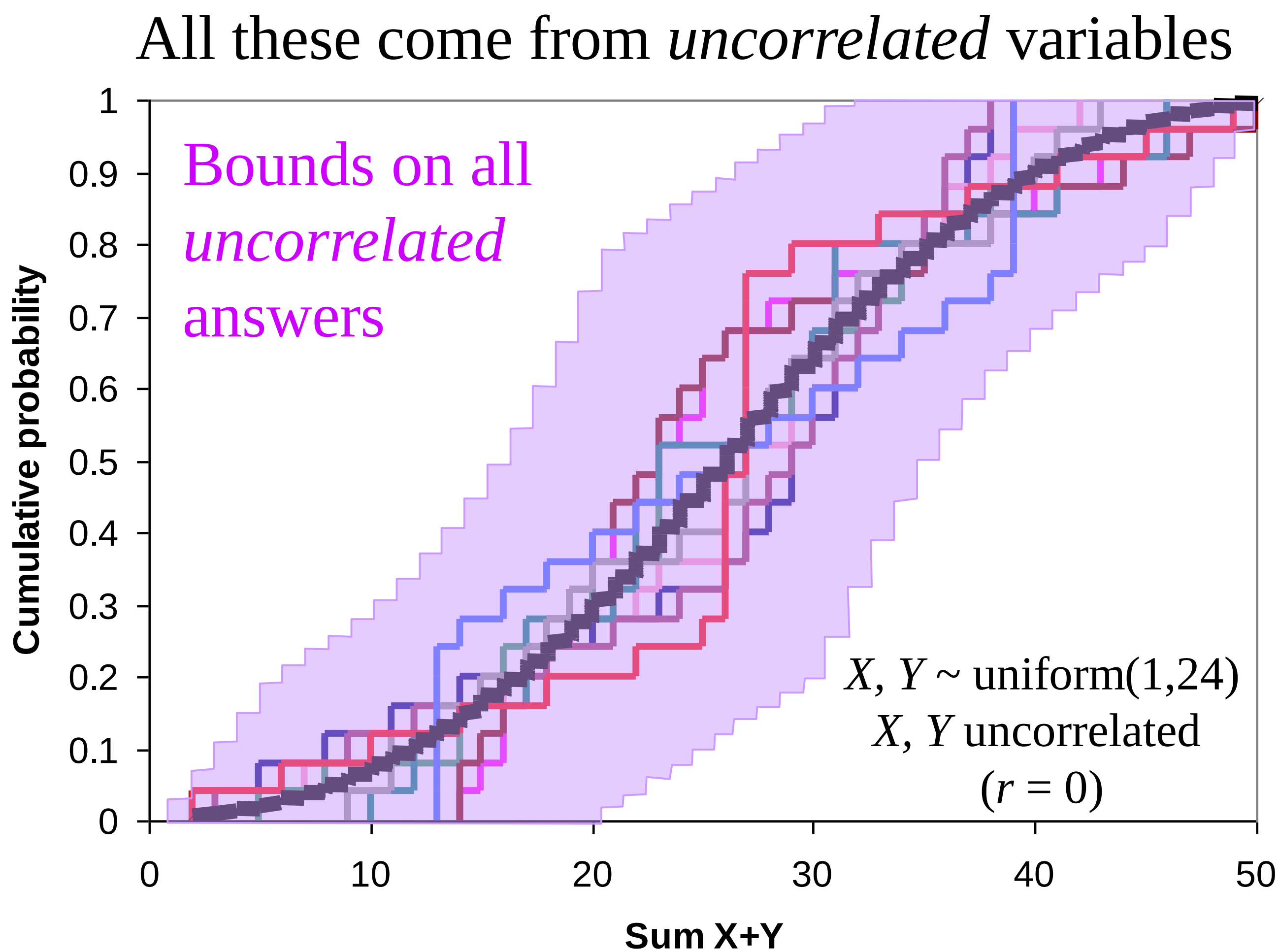
- Independence (knowing $X \sim F$ tells nothing about $Y \sim G$)
- Perfect dependence $Y = G^{-1}(F(X))$
- Opposite dependence $Y = G^{-1}(1 - F(X))$
- Complete dependence $Y = z(X)$
- Linearly correlated $Y = mX + b + \varepsilon$
- Ranks linearly correlated $Y = G^{-1}(mF(X) + b + \varepsilon)$
- Functional modeling $Y = z(X) + \varepsilon$
- Complex dependence (anything else!)

Uncorrelatedness is not independence

Scattergrams with *zero* correlation



X+Y depends on dependence



Can’t always assume independence

- If you know the mechanism of dependence, you can model it
- Else, have to reproduce the statistical patterns
- If dependence unknown, can’t use either way
- Even small correlations can have a big effect on convolutions

What about other dependencies?

- Independent
- Perfectly positive (comonotonic)
- Opposite (countermonotonic)
- Positively or negatively associated
- Specified correlation coefficient
- Nonlinear dependence (copula)
- Unknown dependence

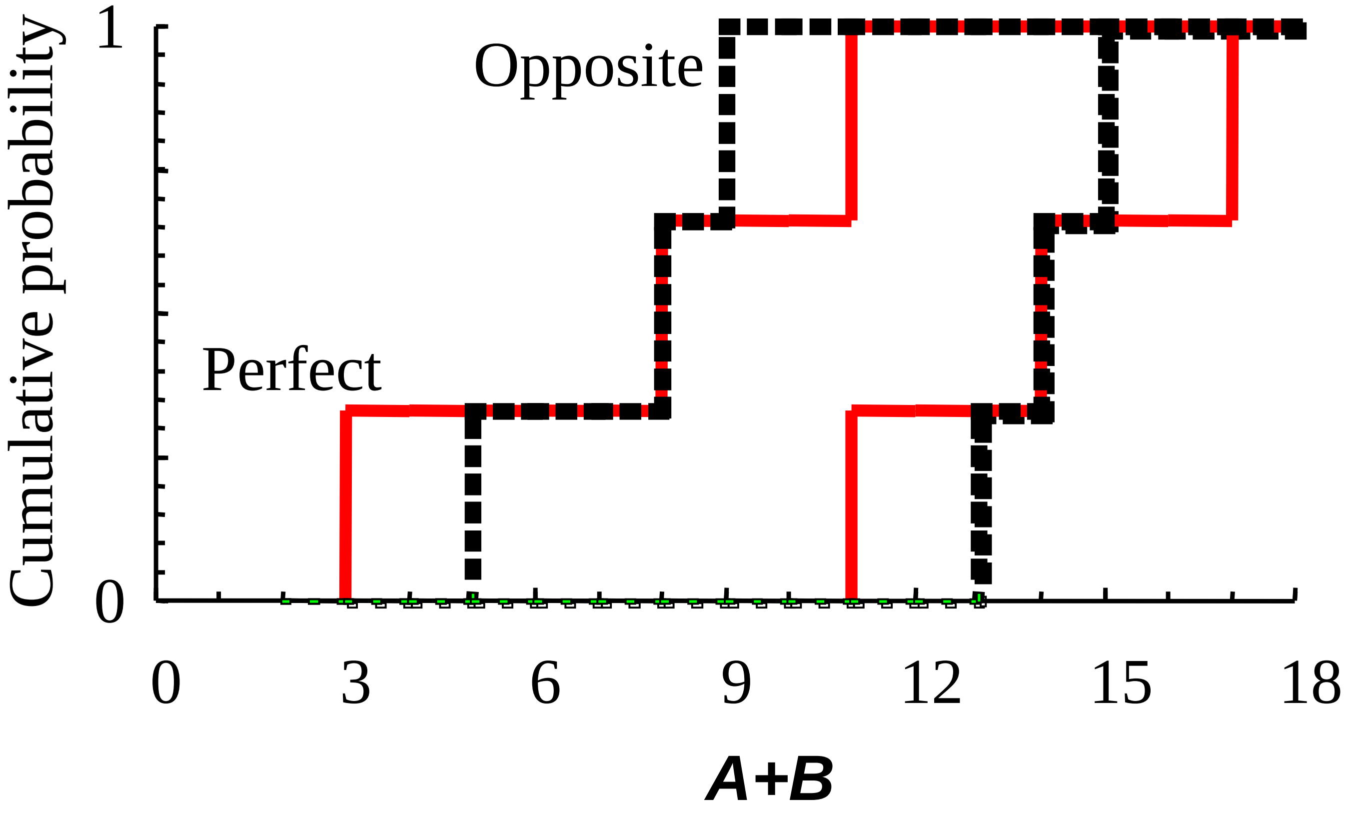
Perfect dependence

$A+B$ perfect positive	$A \in [1,3]$ $p_1 = 1/3$	$A \in [2,4]$ $p_2 = 1/3$	$A \in [3,5]$ $p_3 = 1/3$
$B \in [2,8]$ $q_1 = 1/3$	$A+B \in [3,11]$ prob=1/3	$A+B \in [4,12]$ prob=0	$A+B \in [5,13]$ prob=0
$B \in [6,10]$ $q_2 = 1/3$	$A+B \in [7,13]$ prob=0	$A+B \in [8,14]$ prob=1/3	$A+B \in [9,15]$ prob=0
$B \in [8,12]$ $q_3 = 1/3$	$A+B \in [9,15]$ prob=0	$A+B \in [10,16]$ prob=0	$A+B \in [11,17]$ prob=1/3

Opposite dependence

$A+B$ opposite positive	$A \in [1,3]$ $p_1 = 1/3$	$A \in [2,4]$ $p_2 = 1/3$	$A \in [3,5]$ $p_3 = 1/3$
$B \in [2,8]$ $q_1 = 1/3$	$A+B \in [3,11]$ prob=0	$A+B \in [4,12]$ prob=0	$A+B \in [5,13]$ prob=1/3
$B \in [6,10]$ $q_2 = 1/3$	$A+B \in [7,13]$ prob=0	$A+B \in [8,14]$ prob=1/3	$A+B \in [9,15]$ prob=0
$B \in [8,12]$ $q_3 = 1/3$	$A+B \in [9,15]$ prob= 1/3	$A+B \in [10,16]$ prob=0	$A+B \in [11,17]$ prob=0

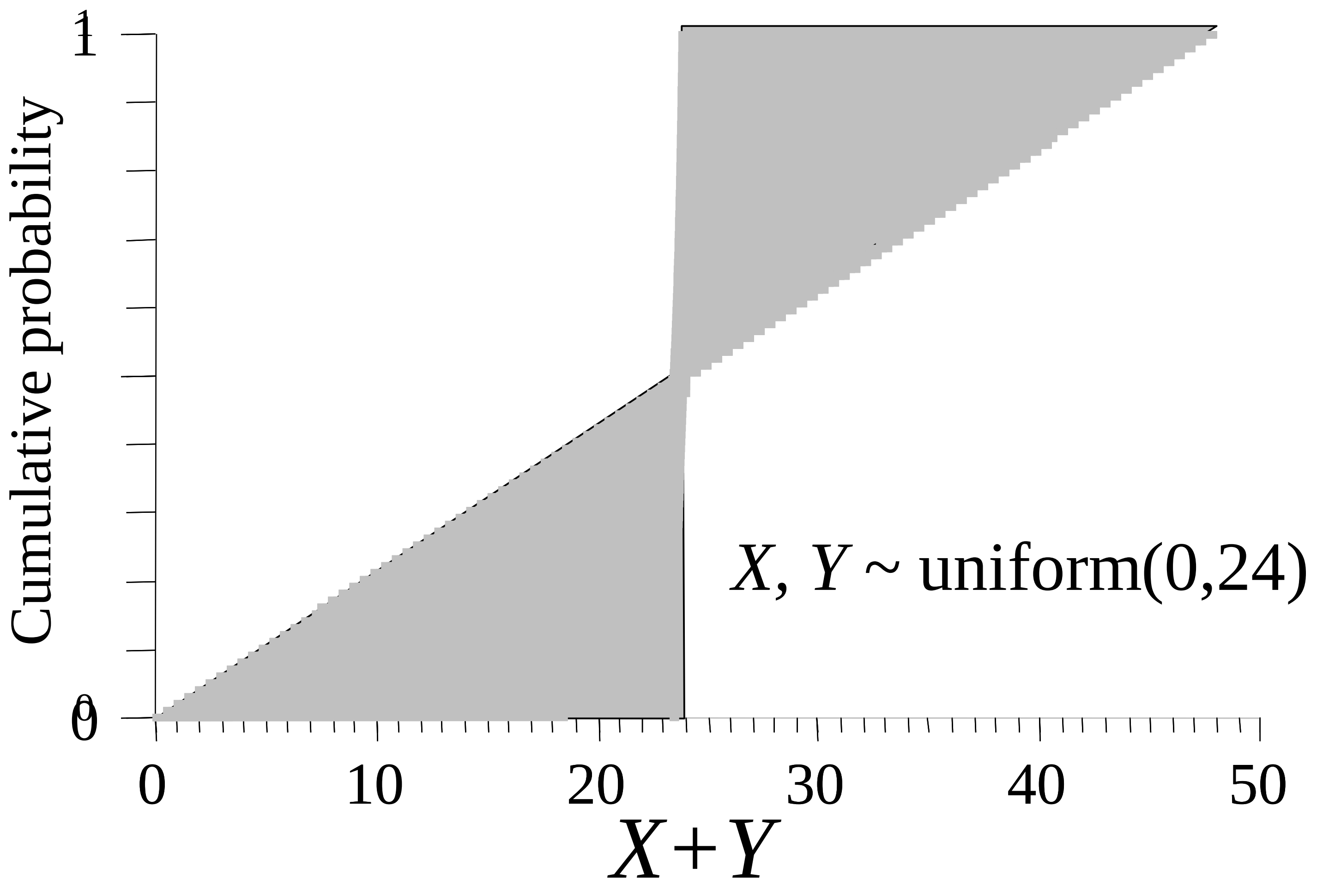
Perfect and opposite dependencies



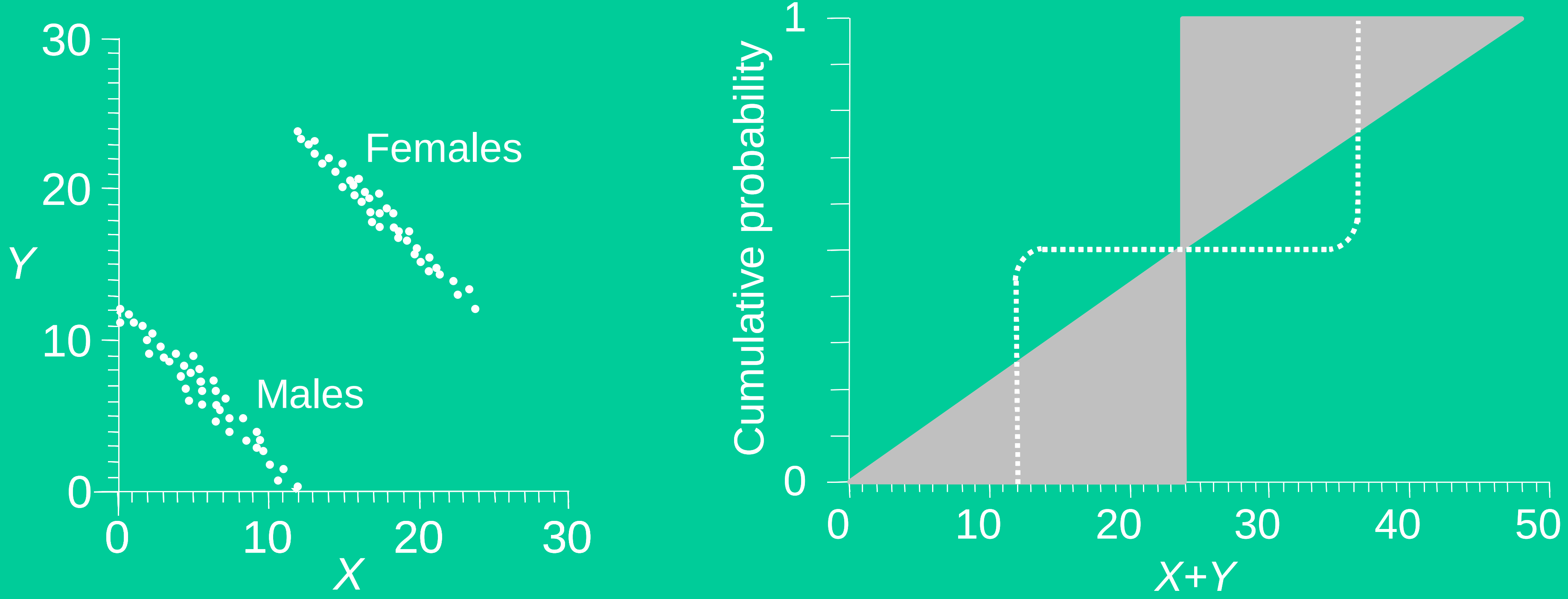
Uncertainty about dependence

- Sensitivity analyses usually used
  - Vary correlation coefficient between –1 and +1
- But this *underestimates* the true uncertainty
  - Example: suppose X, $Y \sim$ uniform(0,24) but we don't know the dependence between X and Y

Varying the correlation coefficient



Counterexample: outside the cone!

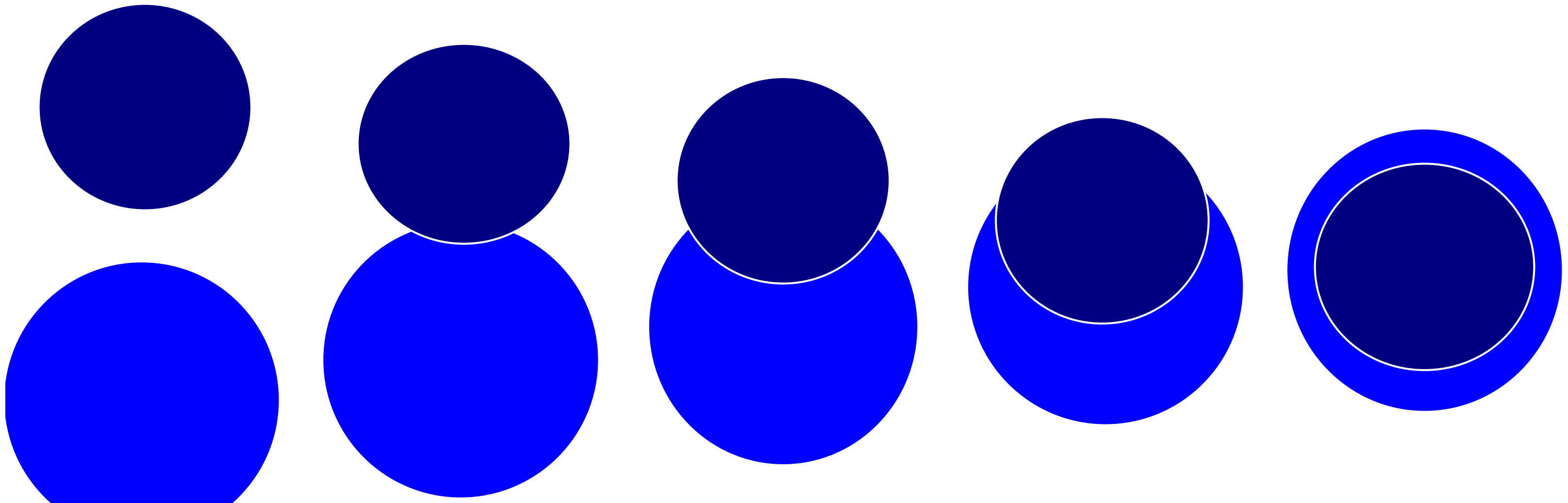


Fréchet inequalities

They make *no assumption* about dependence (Fréchet 1935)

$$\max(0, P(A)+P(B)-1) \le P(A \,\&\, B) \le \min(P(A), P(B))$$

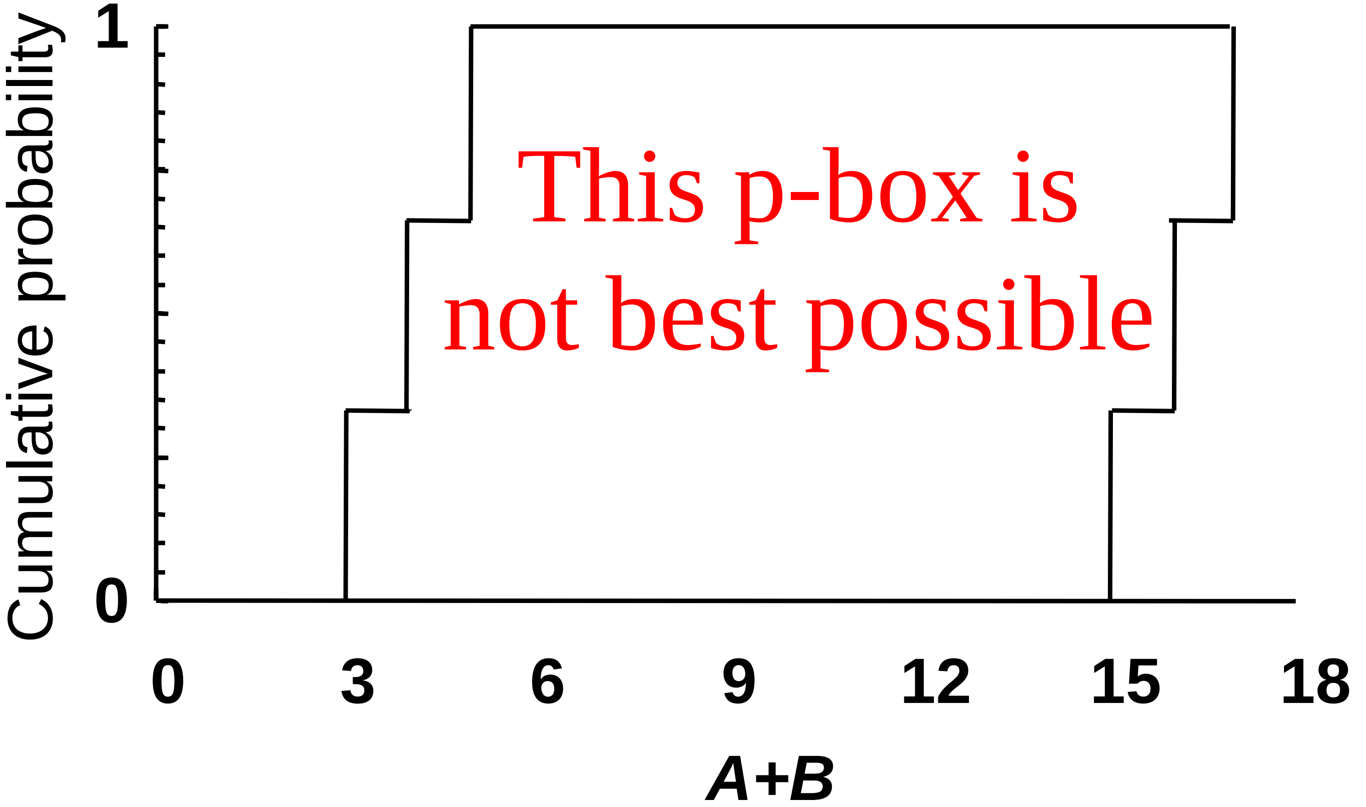
$$\max(P(A), P(B)) \le P(A \vee B) \le \min(1, P(A)+P(B))$$



Fréchet case (no assumption)

$A+B$ Fréchet case	$A\in[1,3]$ $p_1 = 1/3$	$A\in[2,4]$ $p_2 = 1/3$	$A\in[3,5]$ $p_3 = 1/3$
$B\in[2,8]$ $q_1 = 1/3$	$A+B\in[3,11]$ prob=[0,1/3]	$A+B\in[4,12]$ prob=[0,1/3]	$A+B\in[5,13]$ prob=[0,1/3]
$B\in[6,10]$ $q_2 = 1/3$	$A+B\in[7,13]$ prob=[0,1/3]	$A+B\in[8,14]$ prob=[0,1/3]	$A+B\in[9,15]$ prob=[0,1/3]
$B\in[8,12]$ $q_3 = 1/3$	$A+B\in[9,15]$ prob=[0,1/3]	$A+B\in[10,16]$ prob=[0,1/3]	$A+B\in[11,17]$ prob=[0,1/3]

Naïve Fréchet case



Fréchet can be improved

- Interval estimates of probabilities don’t reflect the fact that the sum must equal one
- Resulting p-box is too fat
- Linear programming needed to get the optimal answer using this approach

- Frank, Nelsen and Sklar (1987) gave a way to compute the optimal answer directly

Frank, Nelsen and Sklar (1987)

Suppose $X \sim F$ and $Y \sim G$.
If X and Y are independent, then the distribution of $X+Y$ is

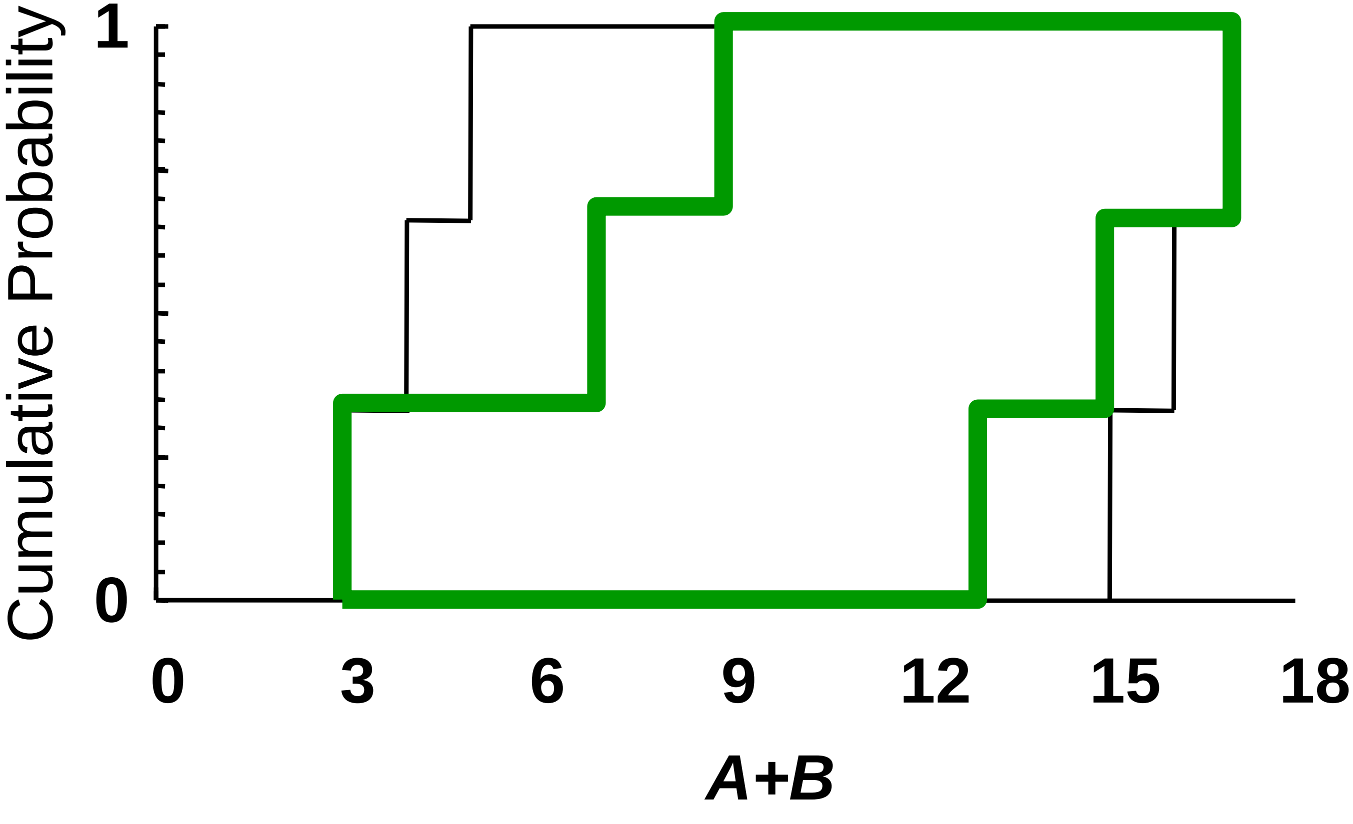
$$\sigma_{+,C}(F,G)(z) = \int_{x+y<z} dC\big(F(x), G(y)\big)$$

In any case, and irrespective of their dependence, this distribution is bounded by

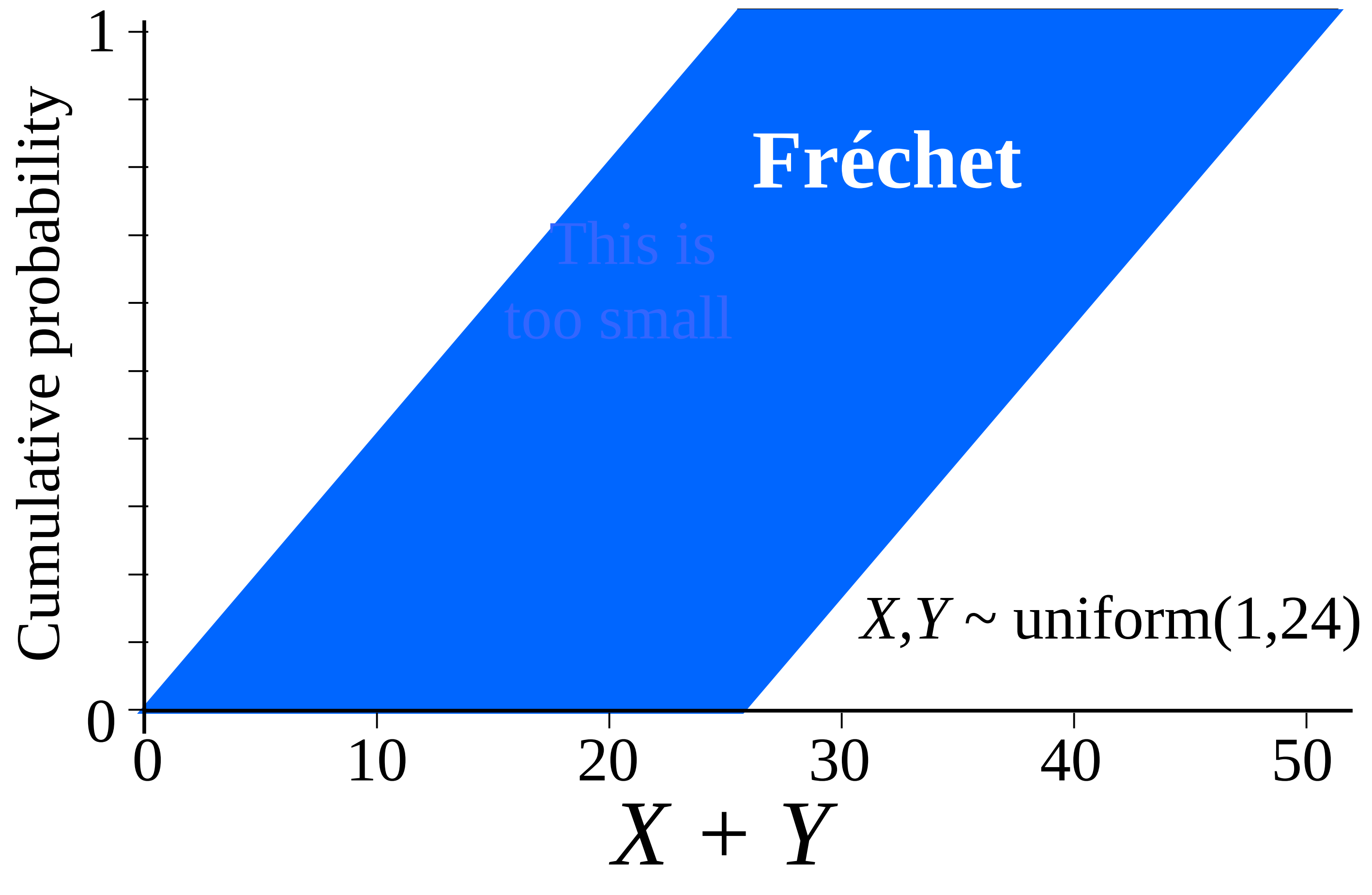
$$\left[\sup_{z=x+y} \max\big(F(x)+G(y)-1,\, 0\big),\ \inf_{z=x+y} \min\big(F(x)+G(y),\, 1\big) \right]$$

This formula can be generalized to work with bounds on F and G.

Best possible bounds



Unknown dependence



Fréchet dependence bounds

- Cannot be obtained by sensitivity studies
- Guaranteed to enclose results no matter what correlation or dependence there may be between the variables
- Best possible (couldn’t be any tighter without saying more about the dependence)
- Can be combined with independence assumptions between other variables

Between independence and Fréchet

- Some information may be available by which the p-boxes could be tightened over the Fréchet case without specifying the dependence perfectly, e.g.,

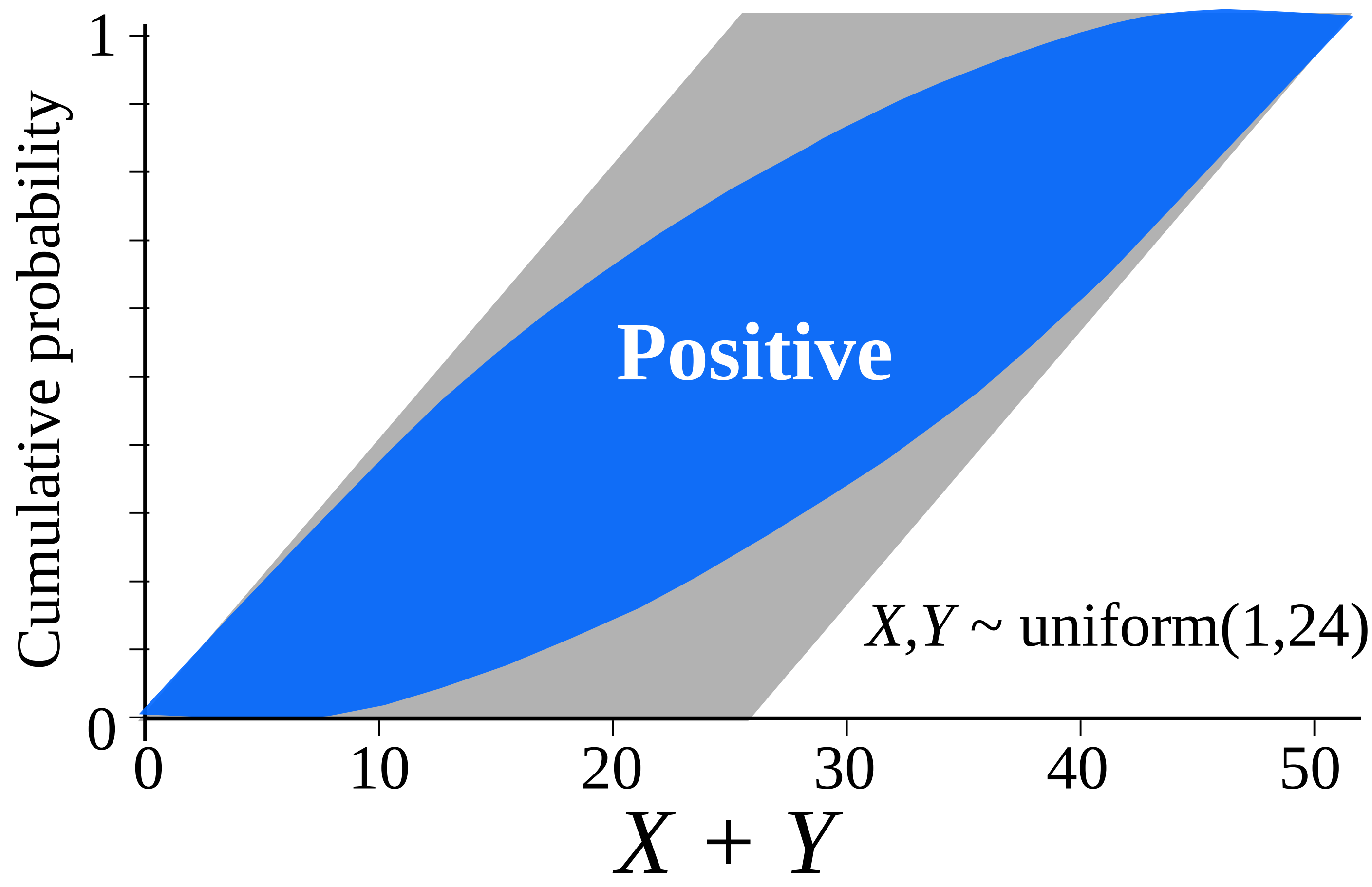
- Dependence is positive (PQD)

  $P(X \leq x, Y \leq y) \geq P(X \leq x)\, P(Y \leq y)$ for all x and y

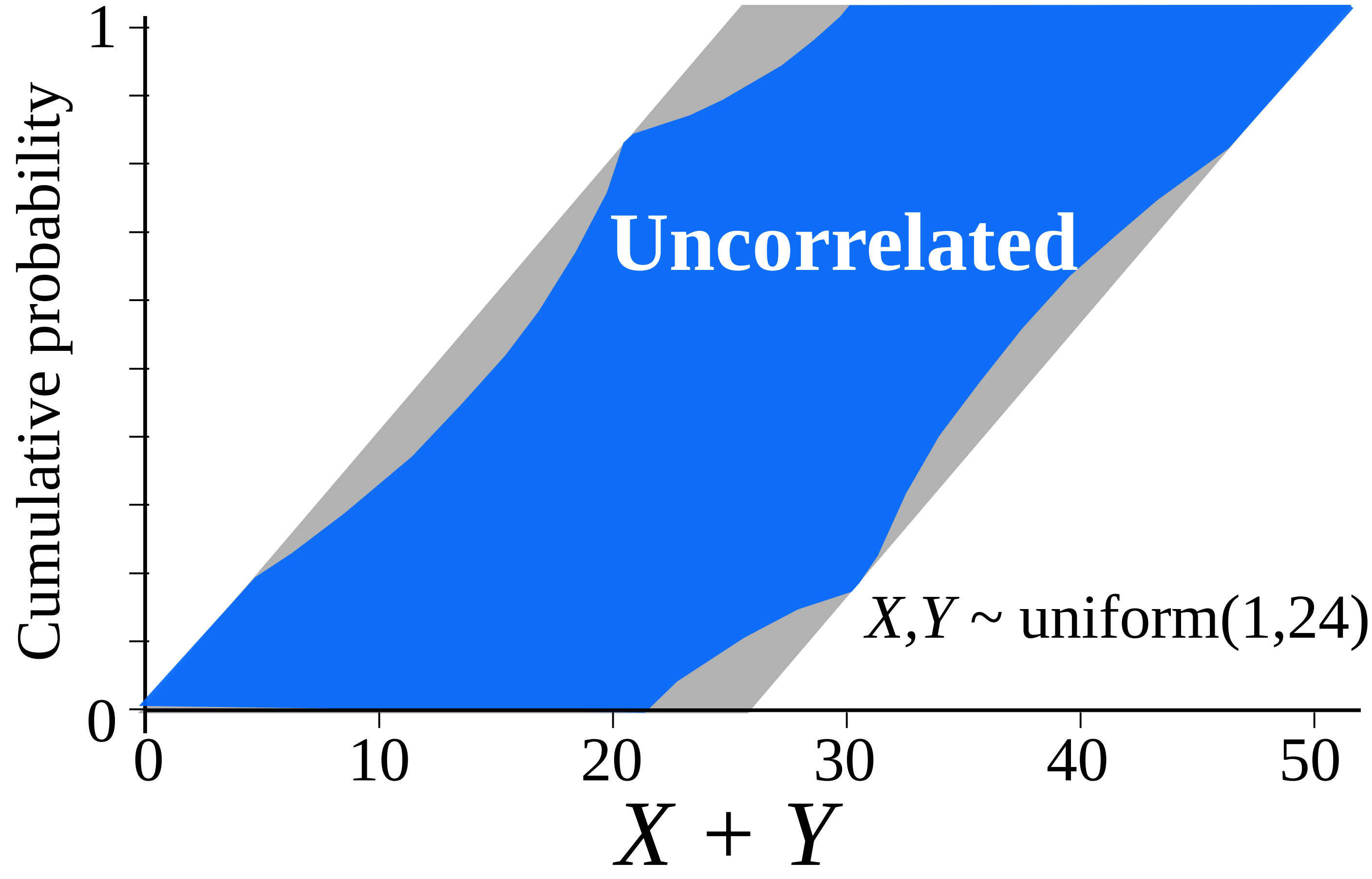
- Variables are uncorrelated

  Pearson correlation r is zero

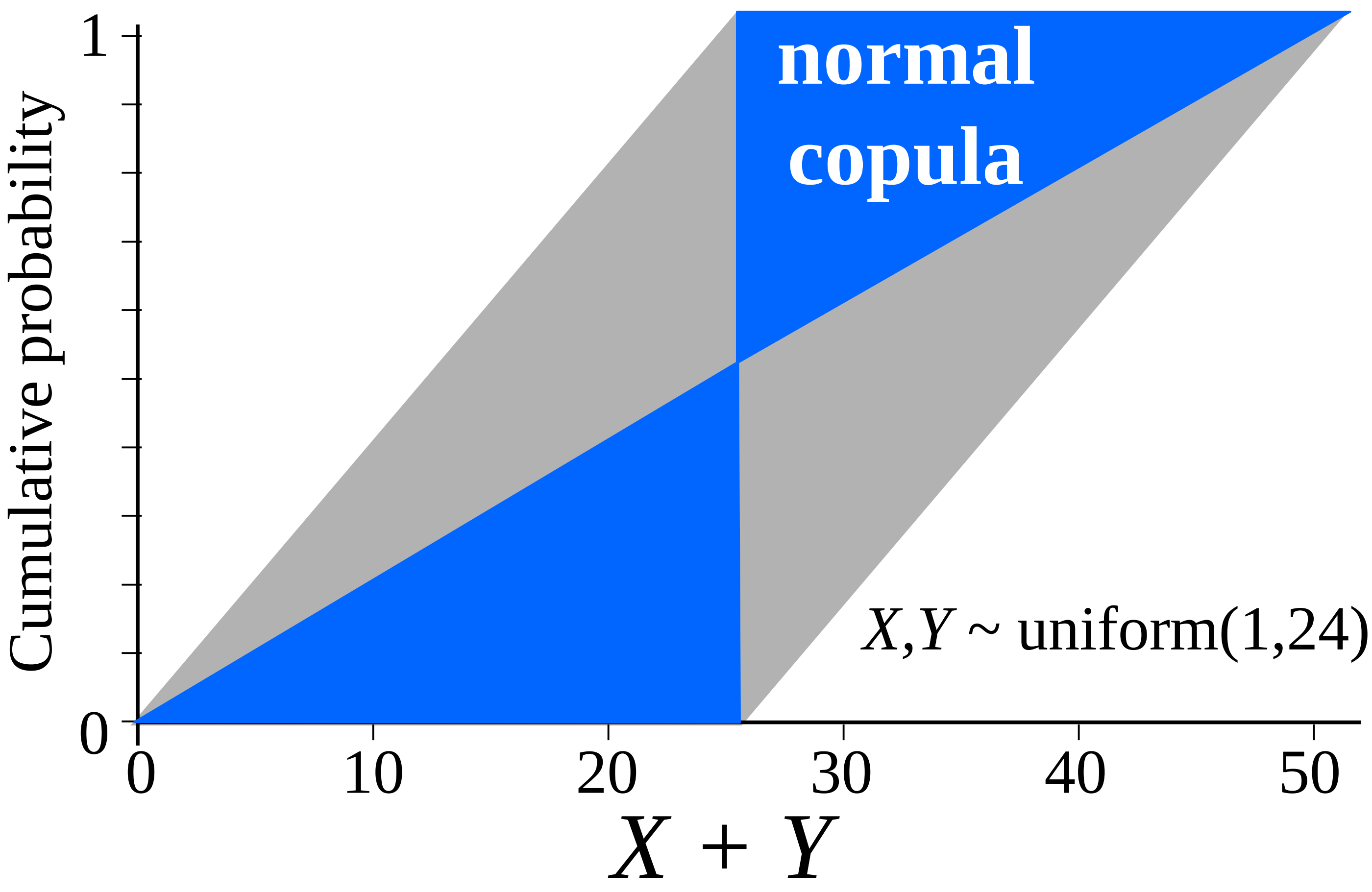
Unknown but positive dependence



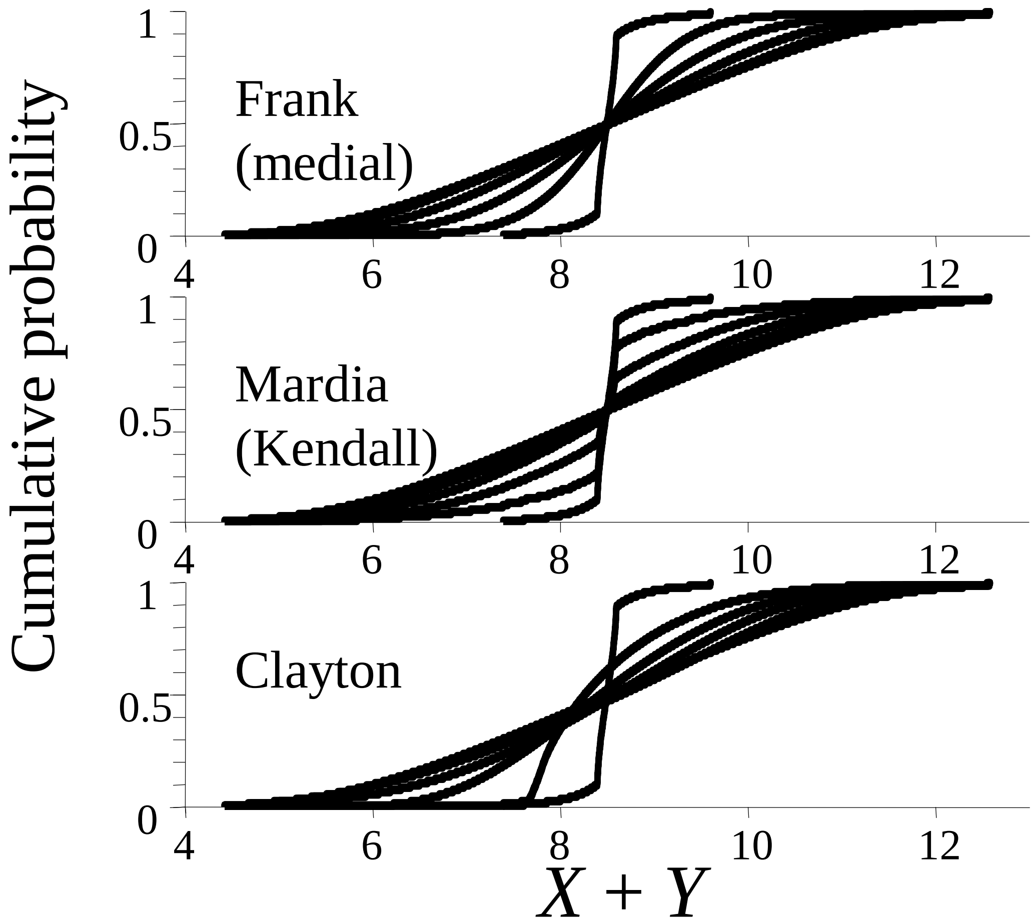
Uncorrelated variables



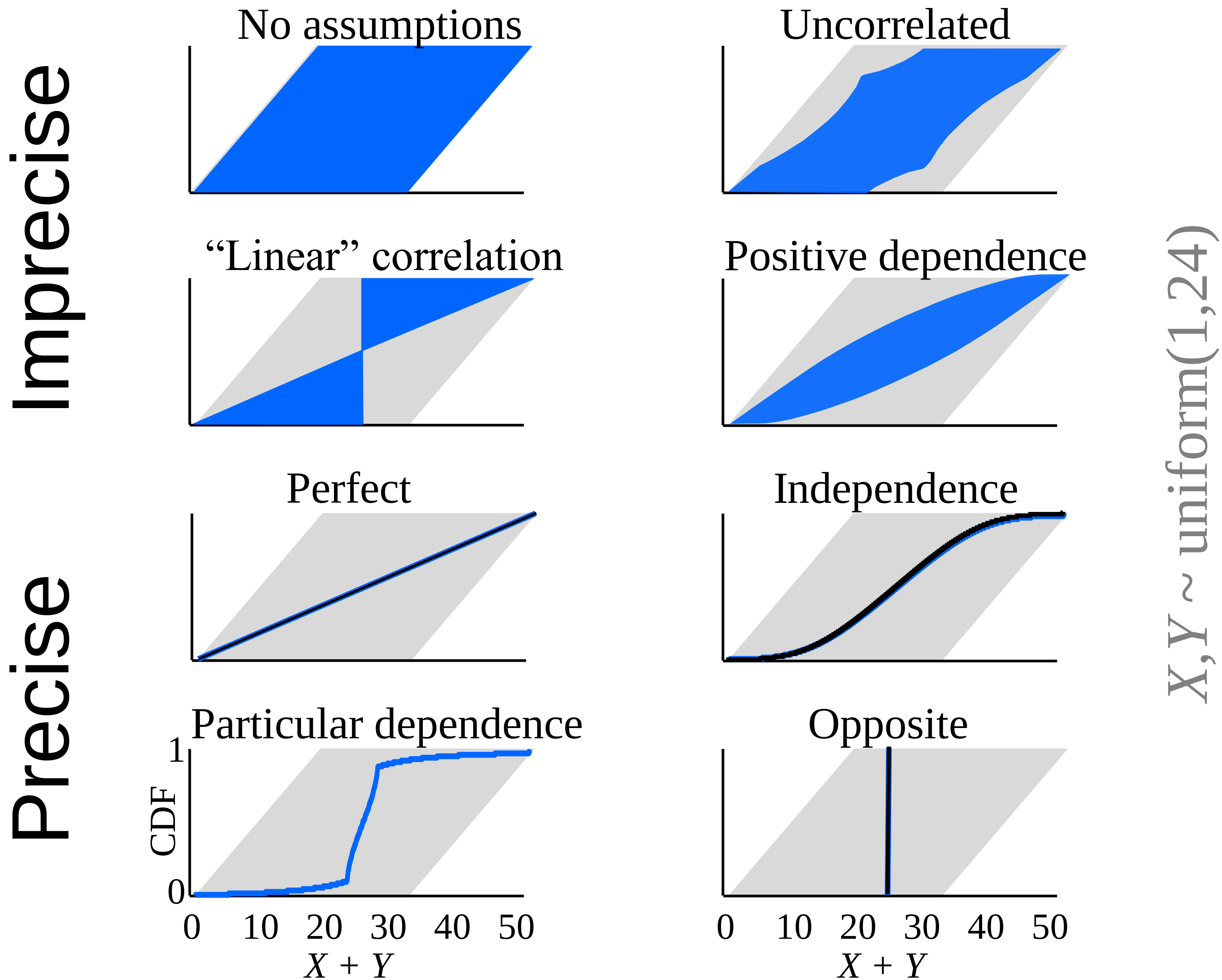
“Linear” correlation



Can model dependence exactly too



X ~ normal(5,1)
Y ~ uniform(2,5)
various correlations
and dependence
functions (copulas)



Imprecise
Precise
No assumptions
Uncorrelated
“Linear” correlation
Positive dependence
Perfect
Independence
Particular dependence
Opposite
1
0
CDF
0 10 20 30 40 50
X + Y
0 10 20 30 40 50
X + Y
X, Y ~ uniform(1,24)

Example: dioxin inhalation

Location: Superfund site in California

Receptor: adults in neighboring community

Contaminant: dioxin

Exposure route: inhalation of windborne soil

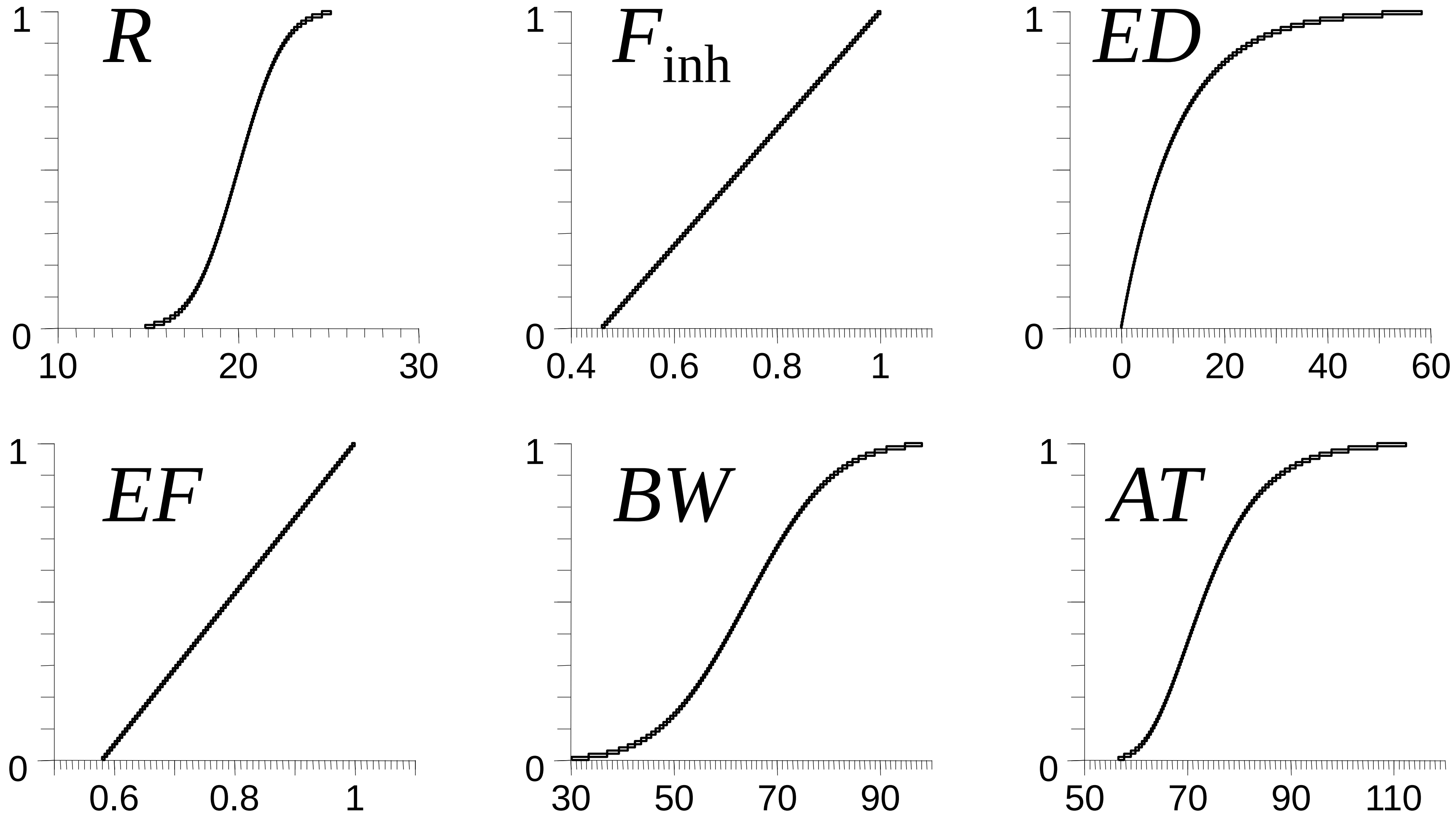
Modified from Table II and IV in Copeland, T.L., A.M. Holbrow, J.M Otani, K.T. Conner and D.J. Paustenbach 1994. Use of probabilistic methods to understand the conservativism in California's approach to assessing health risks posed by air contaminant. *Journal of the Air and Waste Management Association* 44: 1399-1413.

Total daily intake from inhalation

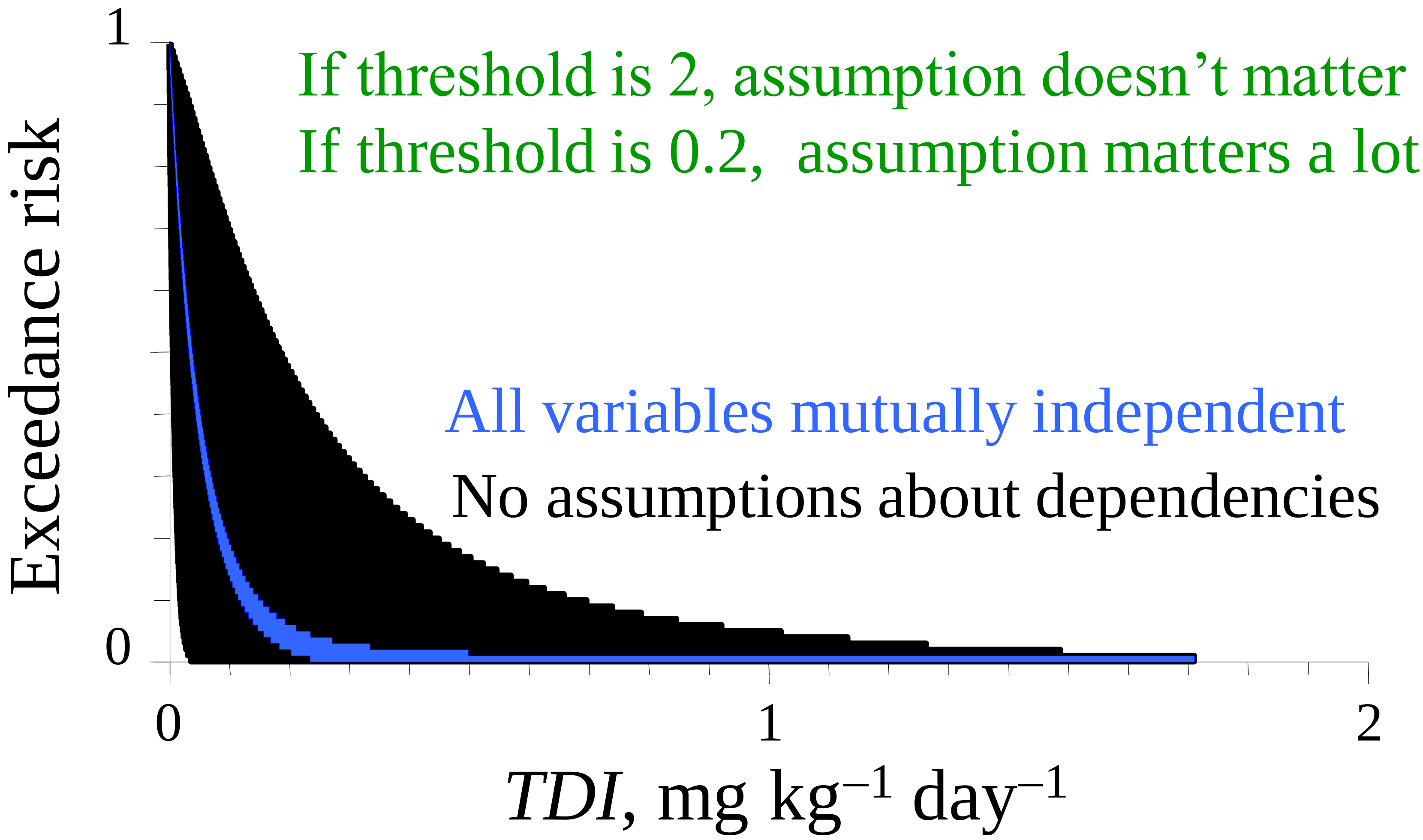
$$TDI = \left(\frac{R \times C_{\mathrm{GL}} \times F_{\mathrm{inh}} \times ED \times EF}{BW \times AT} \right)$$

R = normal(20, 2)	respiration rate, m^3/day
C_{GL} = 2	concentration at ground level, mg/m^3
F_{inh} = uniform(0.46,1)	fraction of particulates retained in lung, [unitless]
ED = exponential(11)	exposure duration, years
EF = uniform(0.58, 1)	exposure frequency, fraction of a year
BW = normal(64.2, 13.19)	receptor body weight, kg
AT = gumbel(70, 8)	averaging time, years

Input distributions



Results



1
0
Exceedance risk
If threshold is 2, assumption doesn't matter
If threshold is 0.2, assumption matters a lot
All variables mutually independent
No assumptions about dependencies
0
1
2
TDI, mg kg–1 day–1

Uncertainty about dependence

Solved

- Impossible with sensitivity analysis since it’s an infinite-dimensional problem

- Kolmogorov-Fréchet bounding lets you be sure

- Sometimes there’s a big difference, sometimes it’s negligible

Try it in pba.r

	Auto	Frechet	Perfect	Opposite	Independent
Plus	+	%+%	%/+/%	%o+o%	%|+|%
Minus	-	%-%	%/-/%	%o-o%	%|-|%
etc.					

SKIP

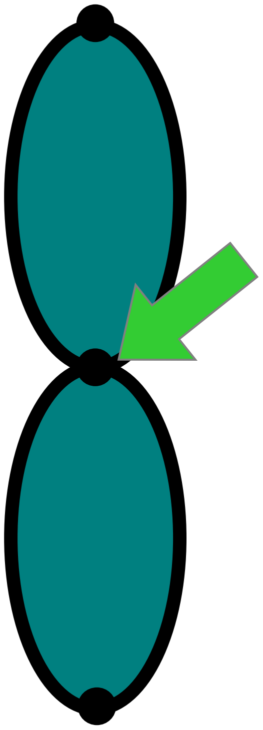
Independence

- In the context of precise probabilities, there was a unique notion of independence
- In the context of imprecise probabilities, however, this notion radiates into several distinct ideas
- The different kinds of independence behave differently in computations

SKIP

Imprecise probability independence

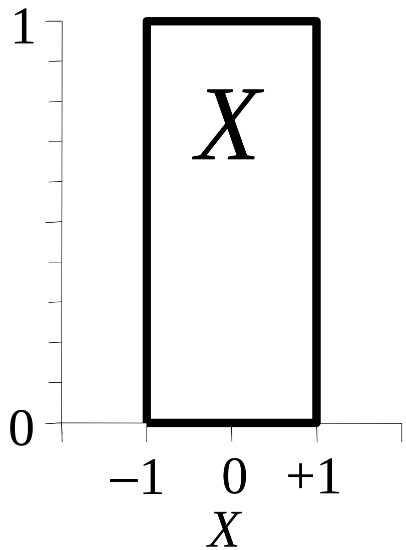
- Random-set independence
- Epistemic irrelevance (asymmetric)
- Epistemic independence
- Strong independence
- Repetition independence
- Others?

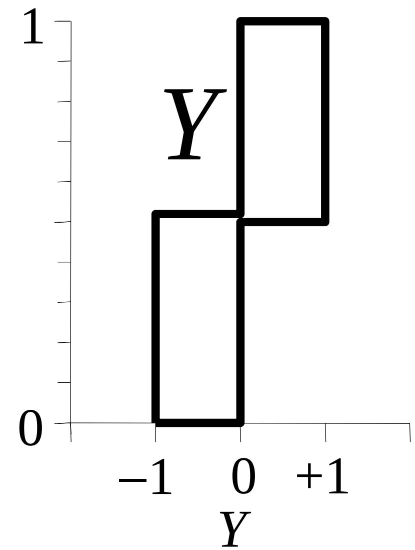


SKIP

Interesting example

- $X = [-1, +1]$, $\quad Y = \{([-1, 0], ½), ([0, 1], ½)\}$





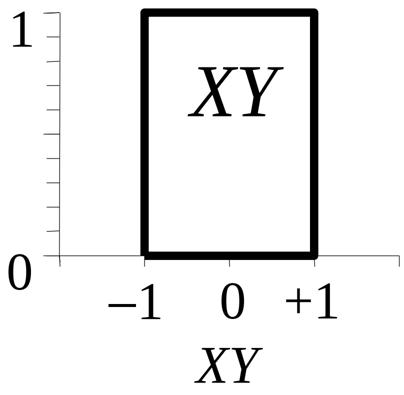
- If X and Y are "independent", what is $Z = XY$?

SKIP

Compute via Yager's convolution

		Y	
		([–1, 0], ½)	([0, 1], ½)
X	([–1, +1], 1)	([–1, +1], ½)	([–1, +1], ½)

The Cartesian product with one row and two columns produces this p-box



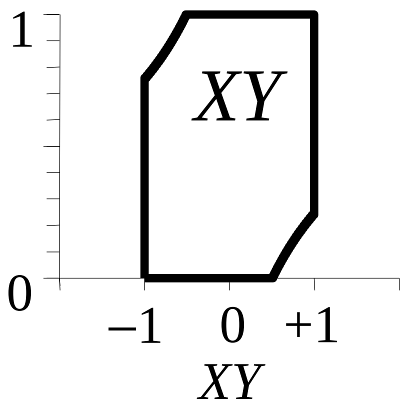
SKIP

But consider the means

- Clearly, EX = [–1,+1] and EY=[–½, +½]

- Therefore, E(XY) = [–½, +½]

- But if this is the mean of the product, and its range is [–1,+1], then we know better bounds on the CDF

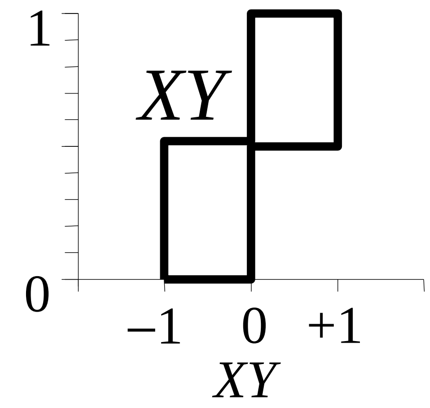


SKIP

And consider the quantity signs

- What's the probability P_Z that $Z < 0$?
- $Z < 0$ only if $X < 0$ or $Y < 0$ (but not both)
- $P_Z = P_X(1-P_Y) + P_Y(1-P_X)$, where

  $P_X = P(X < 0)$, $P_Y = P(Y < 0)$
- But P_Y is ½ by construction
- So $P_Z = ½P_X + ½(1-P_X) = ½$
- Thus, zero is the median of Z
- Knowing median and range improves bounds



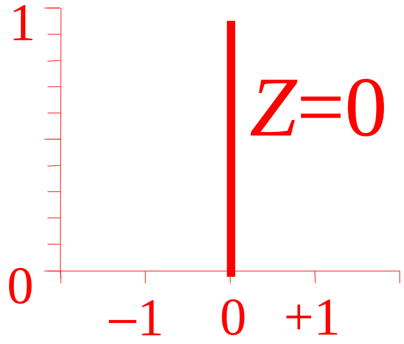
SKIP

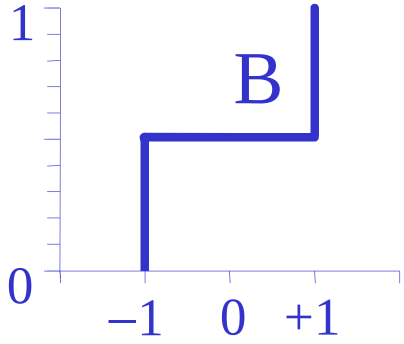
Best possible

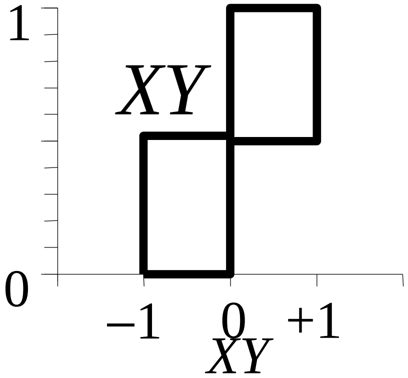
- These bounds are realized by solutions

  If $X = 0$, then $Z=0$

  If $X = Y = B = \{(-1, ½),(+1, ½)\}$, then $Z = B$





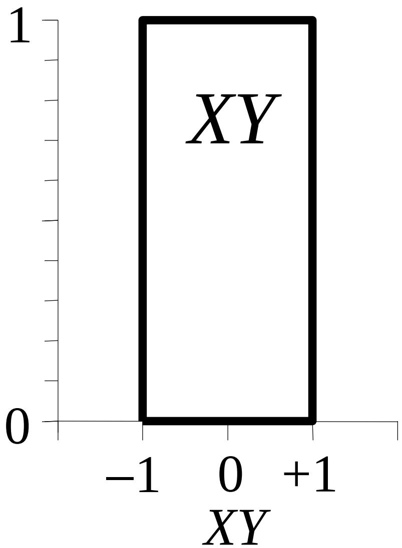


- So these bounds are also best possible

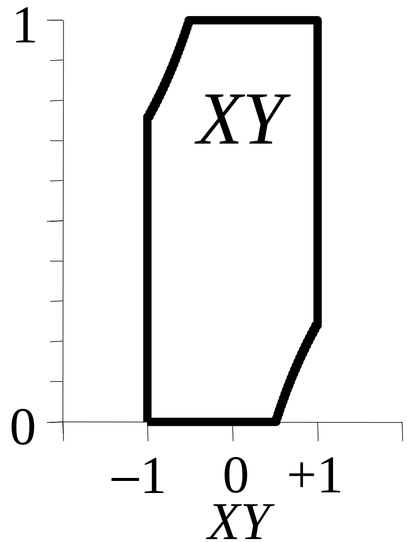
SKIP

So which is correct?

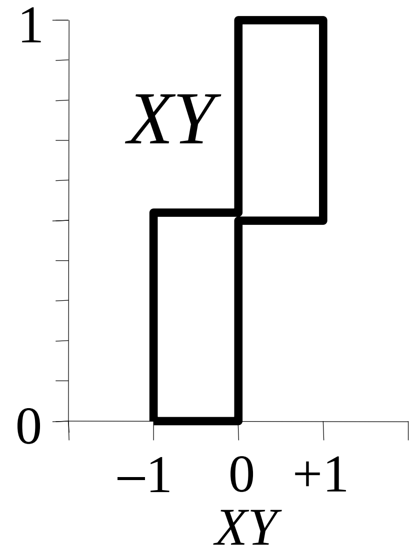
Random-set independence



Moment independence



Strong independence



The answer depends on what one meant by “independent”

SKIP

So what?

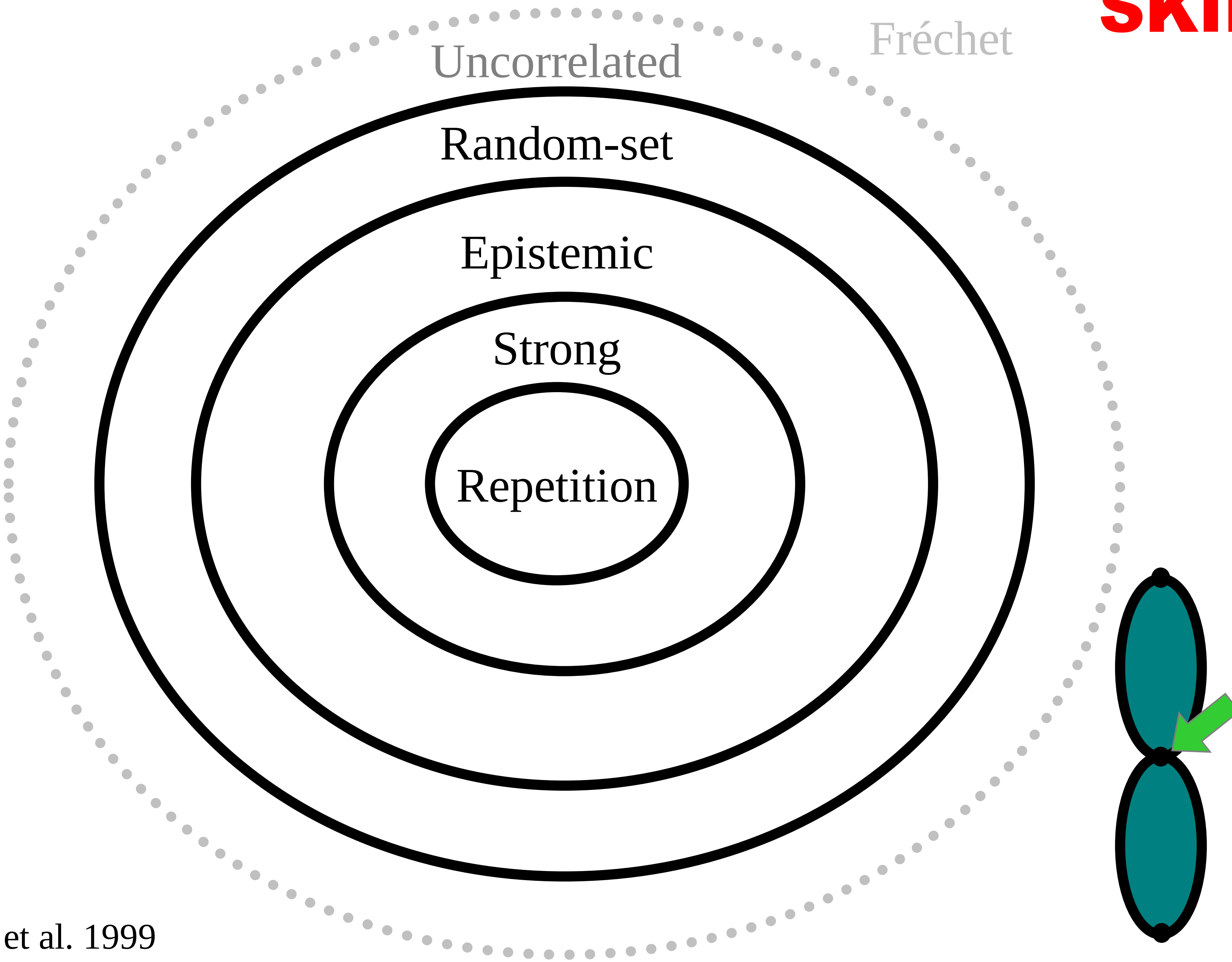
- The example illustrates a practical difference between random-set independence and strong independence
- It disproves the conjecture that the convolution of uncertain numbers is not affected by dependence assumptions if at least one of them is an interval
- It tempers the claim about the best-possible nature of convolutions with probability boxes and Dempster-Shafer structures

SKIP

Strategy for risk analysts

- Random-set independence is conservative
- Using the Cartesian product approach is always rigorous, though may not be optimal
- Convenient methods to obtain tighter bounds under stronger kinds of independence await discovery

SKIP



Couso et al. 1999

Sensitivity analysis

When to break down and get more data

Justifying further empirical effort

- If the incertitude associated with the results of an analysis is too broad to make practical decisions, and the bounds are best possible, more data is needed
  - *Strong argument for collecting more data*

- Planning empirical efforts can be improved by doing sensitivity analysis of the model

Sensitivity analysis <u>of</u> p-boxes

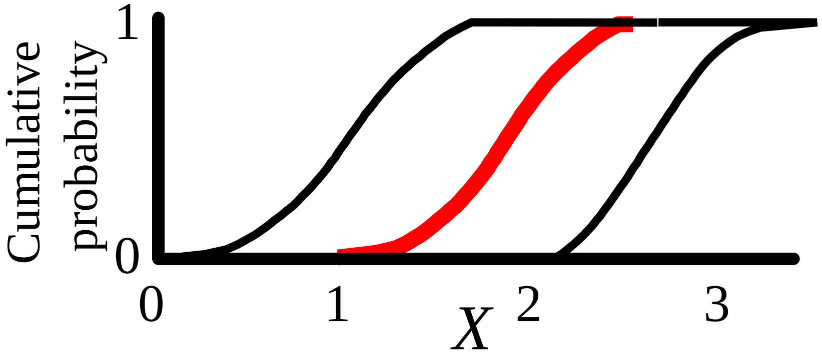
- Quantifies the reduction in uncertainty of a result when an input is pinched

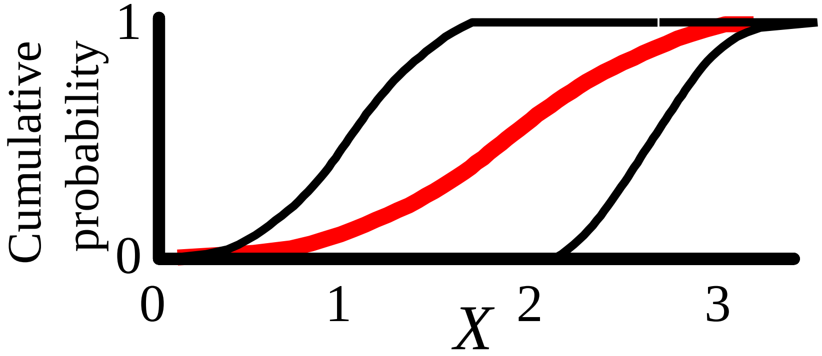
- *Pinching* is hypothetically replacing it by a less uncertain characterization

Pinching sensitivity analyses

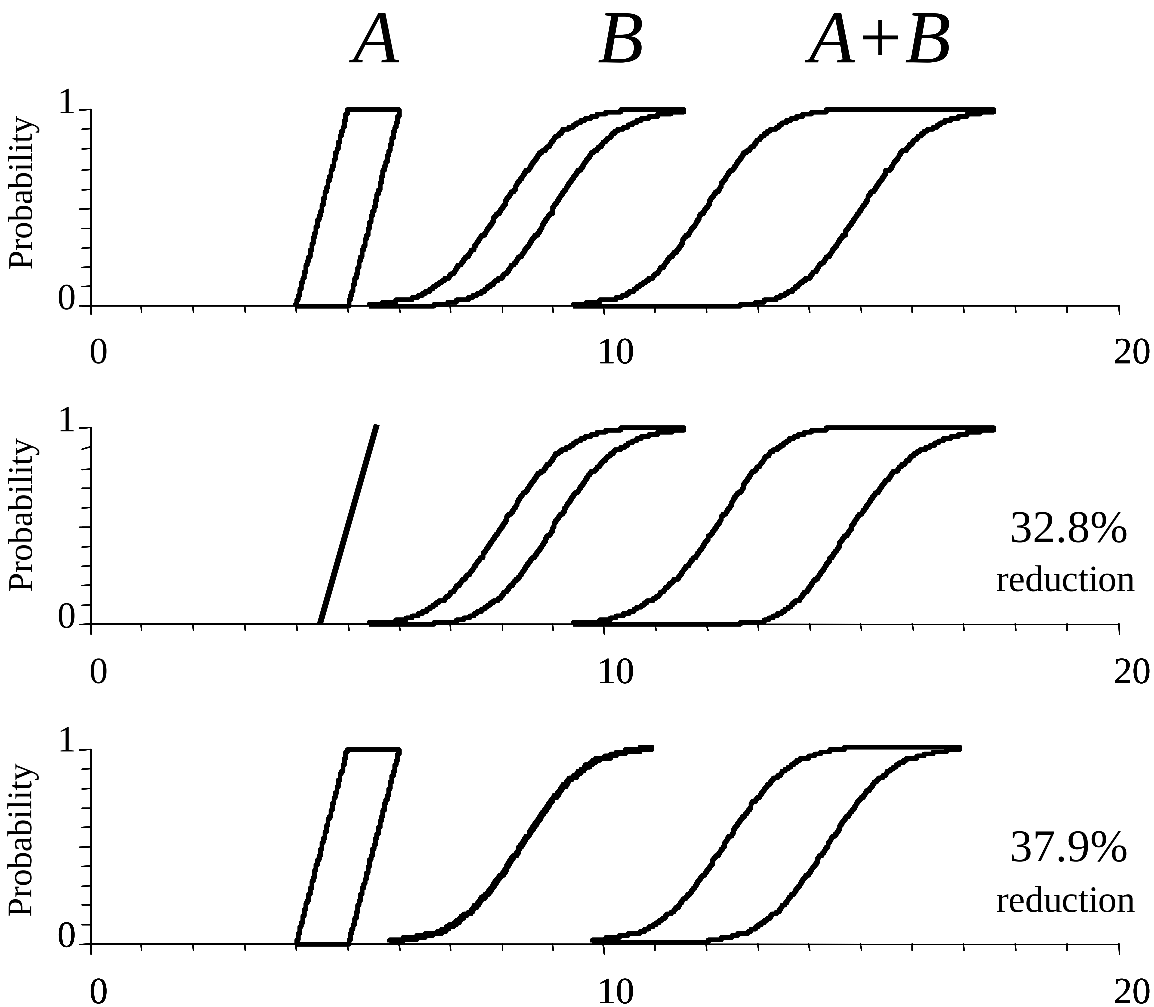
- Model the possible contraction of incertitude in each input p-box from additional data to be collected
- Recompute analysis with this tighter p-box
  - Others inputs held in their original form
  - Or possibly tighten several if appropriate for planned data
- Estimate improvement in results
  - breadth(tighter input) / breadth(base)
- Repeat for all inputs or data collection strategies
  - Don’t omit variables from sensitivity study or “shortlist”
- Allocate empirical effort by the magnitude of the potential improvement in uncertainty

Pinching to a precise distribution



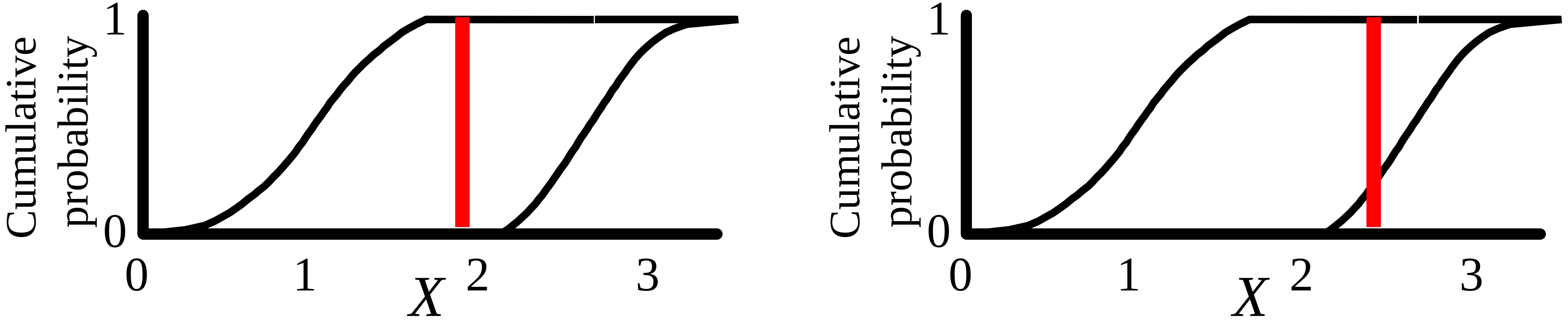


Examples

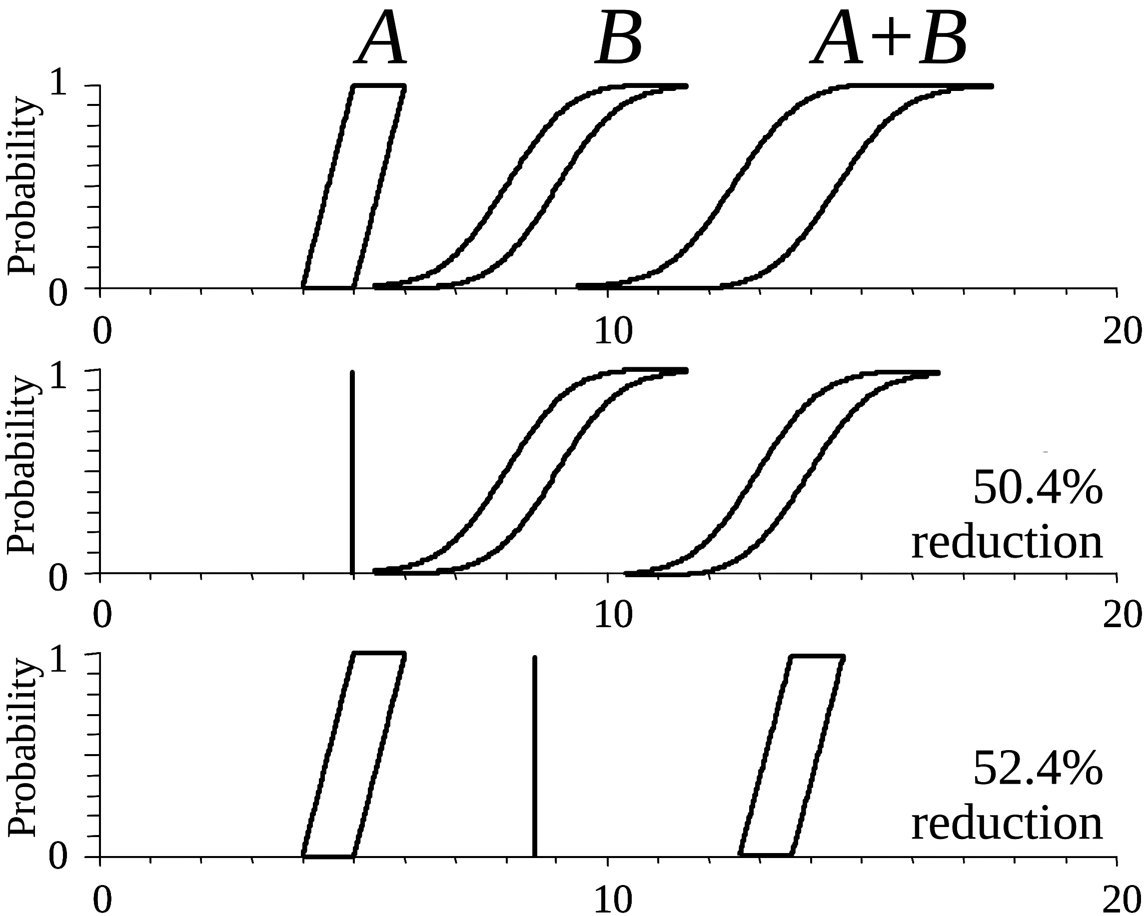


Uses no assumption about dependence between *A* and *B*

Pinching to a point value



Examples


A
B
A+B
1
0
Probability
0
10
20
1
0
Probability
0
10
20
50.4%
reduction
1
0
Probability
0
10
20
52.4%
reduction


Under
independence

Pinching dependence

- An uncertain dependence can also be pinched to a more specific dependence

positive $\rightarrow$ independence

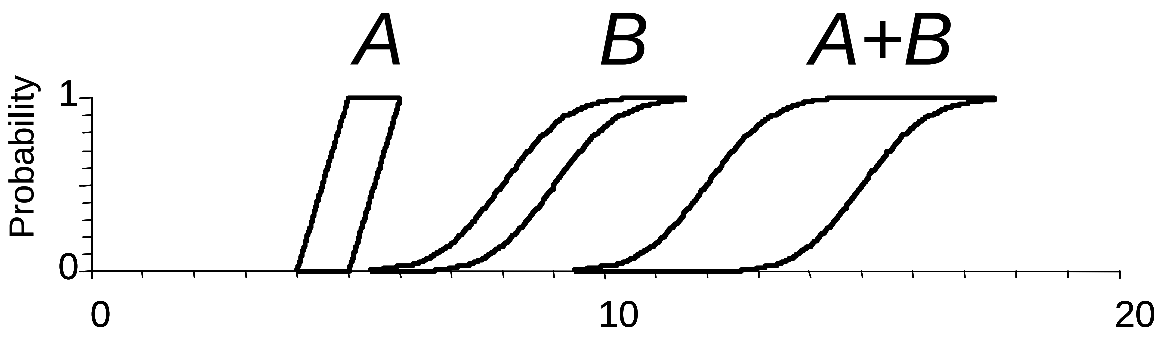
positive $\rightarrow$ perfect

positive $\rightarrow$ normal copula with correlation 0.3

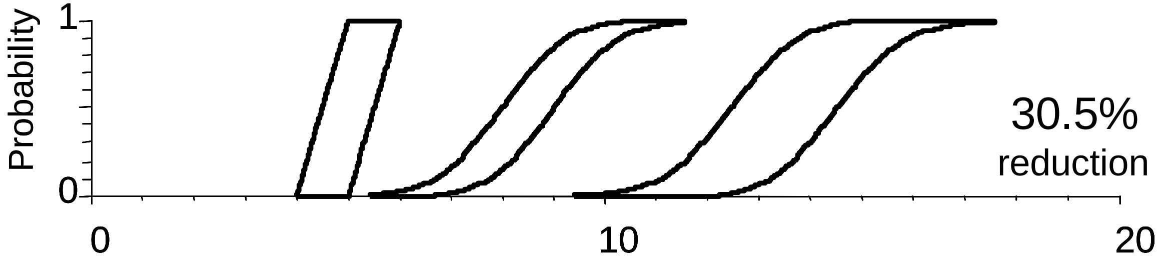
Fréchet $\rightarrow$ independence

etc.

Example

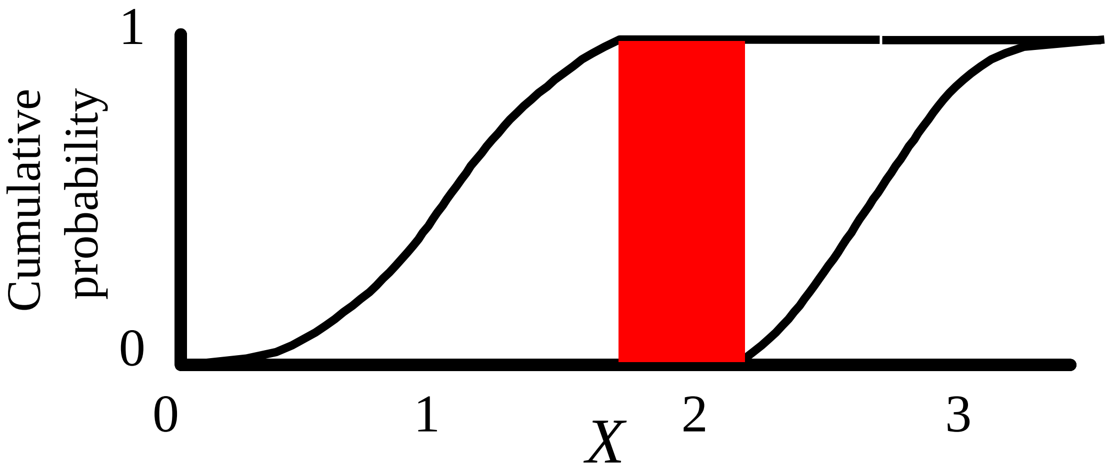


Uses no assumption about dependence between *A* and *B*



Pinch the dependence to *independence*

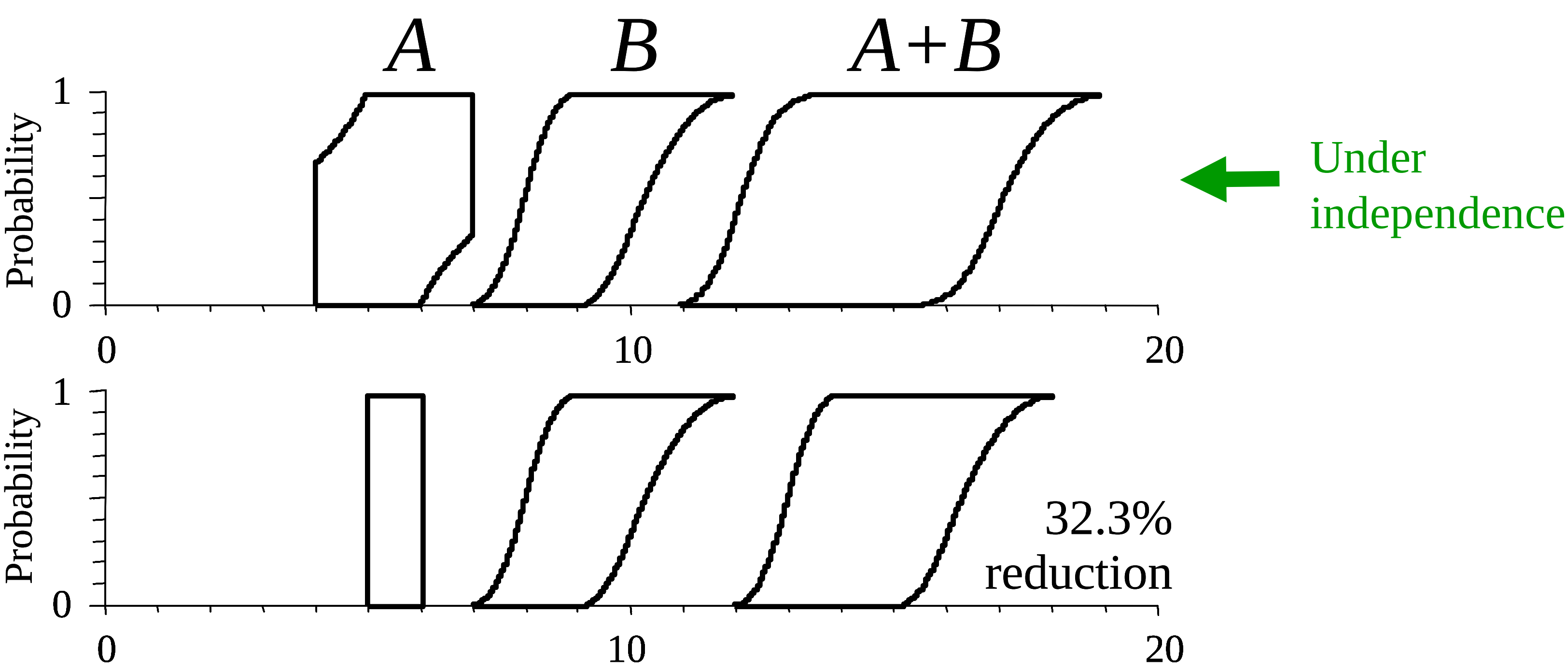
Pinching to a zero-variance interval



Assumes value is constant, but unknown

There's no analog of this in Monte Carlo

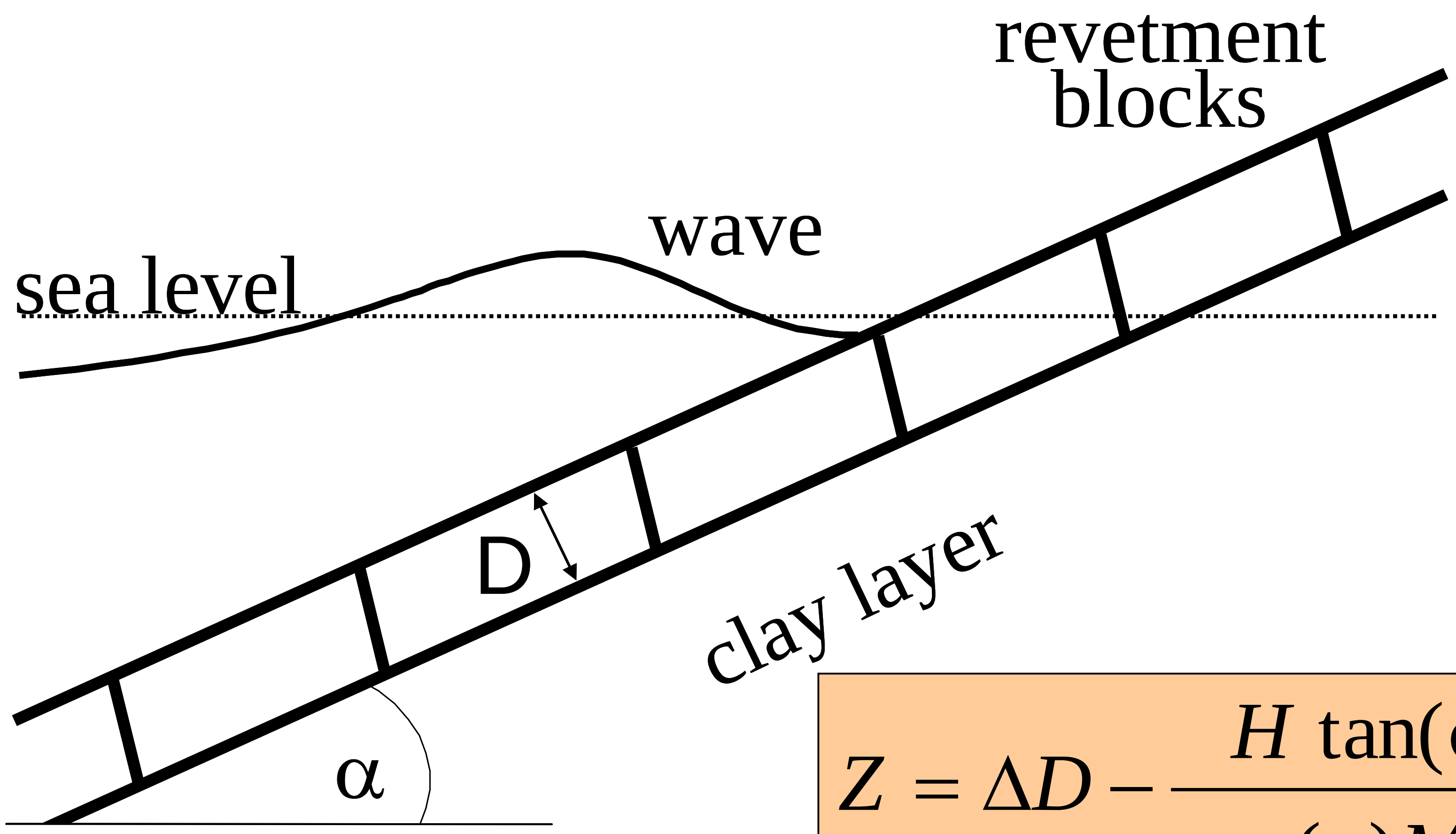
Pinching to a zero-variance interval



Pinching to other targets

- Pinchings need not be to *particular* targets, e.g., a single point or a precise distribution
- Traditionally, the results of pinching to various targets were *averaged* (called “freezing”), but that approach erases the effect of uncertainty
- Instead, we ask: what is the *range* of observed reductions in uncertainty as the pinching target is varied over many different possibilities?

Case study: dike reliability



$$Z = \Delta D - \frac{H \tan(\alpha)}{\cos(\alpha) M \sqrt{s}}$$

The inputs

Relative density of the revetment blocks
Δ = [1.60, 1.65]

Block thickness
D = [0.68, 0.72] meters

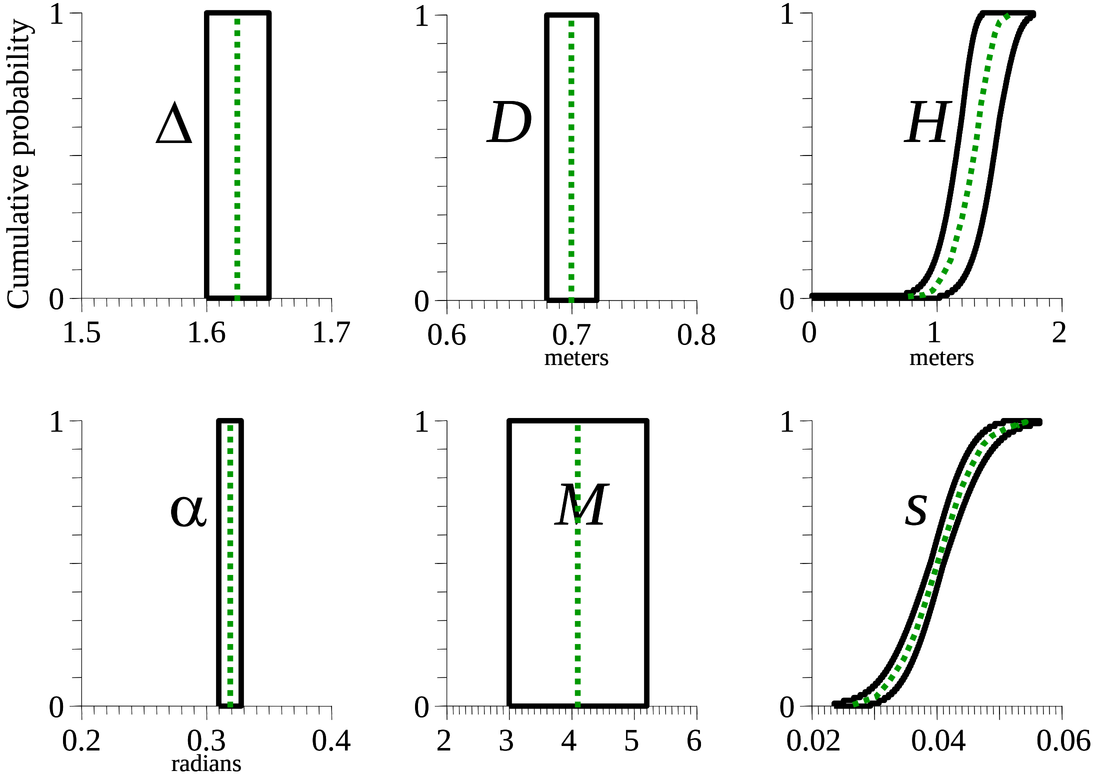
Slope of the revetment
α = atan([0.32, 0.34]) = [0.309, 0.328] radians

Analysts' "model uncertainty" factor
M = [3.0, 5.2]

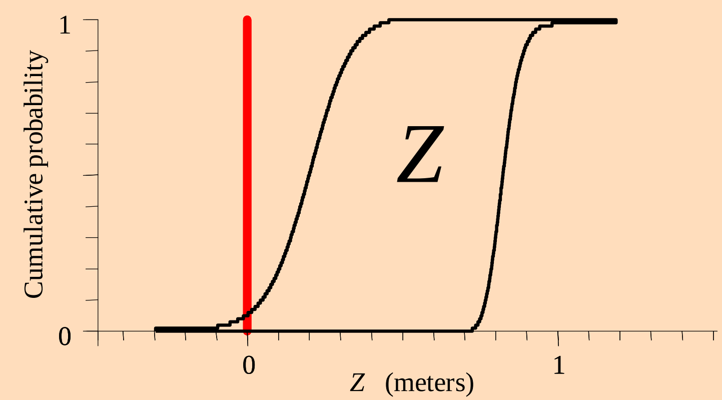
Significant wave height (average of the highest third of waves)
H = weibull([1.2, 1.5] meters, [10, 12])

Offshore peak wave steepness
s = normal([0.039, 0.041], [0.005, 0.006])

Analysis



Nominal pinching targets are dotted



Probability of negative Z (i.e., dike failure) is less than 0.05

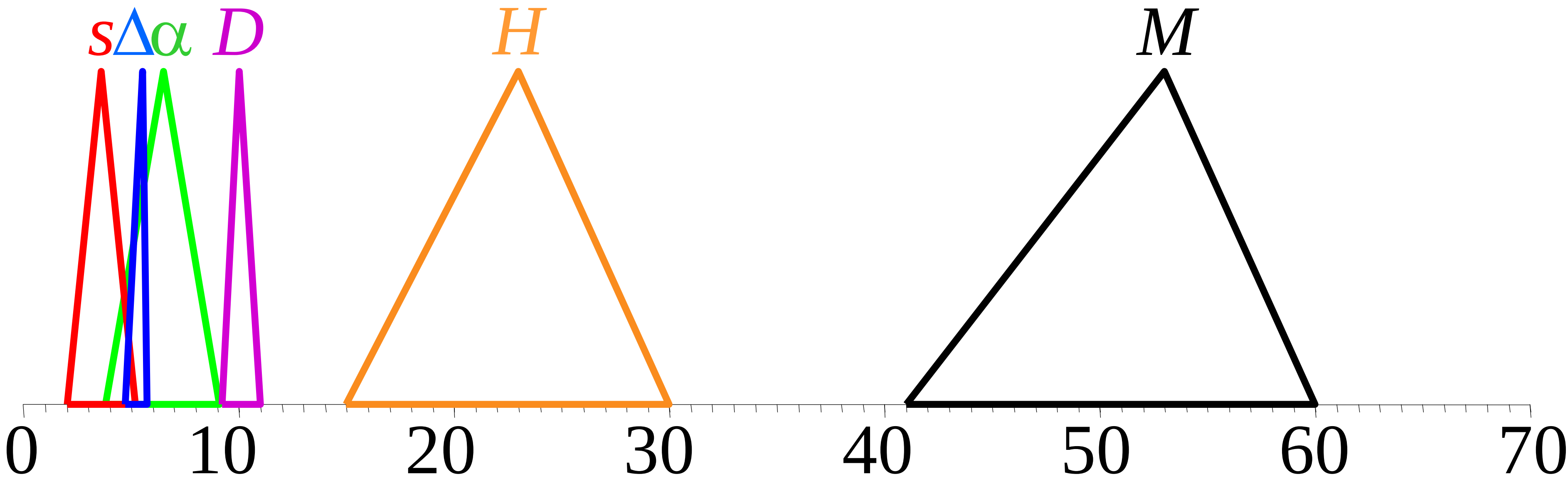
Percent reduction of uncertainty

Input	*Nominal pinching*	*All possible pinchings*
Δ	5.5	[4.7, 5.7]
D	10.0	[9.2, 11.0]
M	53.0	[41.0, 60.0]
α	6.5	[3.8, 9.1]
H	23.0	[15.0, 30.0]
s	3.6	[2.0, 5.2]

Empirical study rank order: $M\ \ H\ \ D\ \ \alpha\ \ \Delta\ \ s$

On the same axis

(intervals depicted as triangles)



Empirical study rank order: $M\ \ H\ \ D\ \ \alpha\ \ \Delta\ \ s$

Caveat

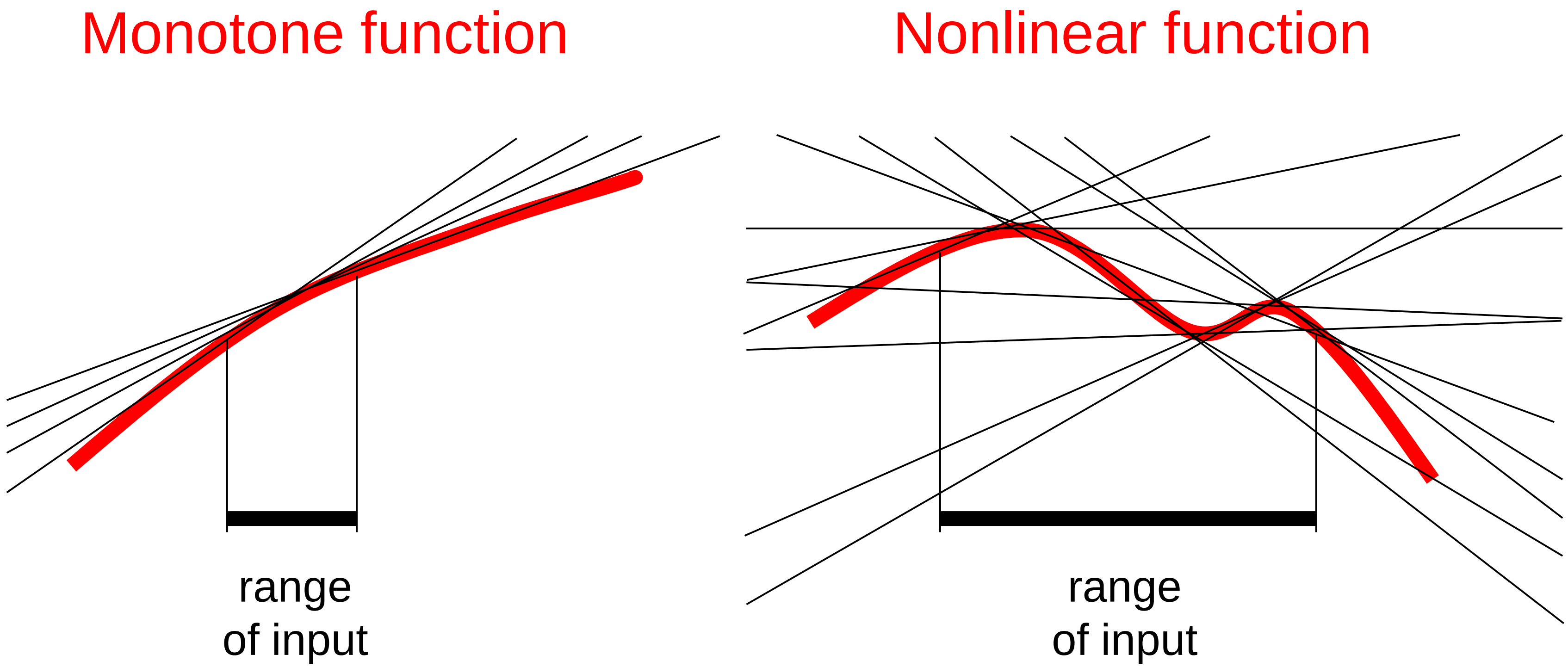
Omitting some variables from a sensitivity analysis ("shortlisting") because their uncertainties are small is a bad idea.

Doing this reduces dimensionality,
but it also *erases uncertainty*.

What about engineering control?

Sensitivity analysis <u>with</u> p-boxes

- Local sensitivity via derivatives
- Explored macroscopically over the uncertainty in the input
- Describes the ensemble of tangent slopes to the function over the range of uncertainty


Monotone function
Nonlinear function
range
of input
range
of input

Local derivatives

$$\frac{\partial Z}{\partial \Delta} = D$$

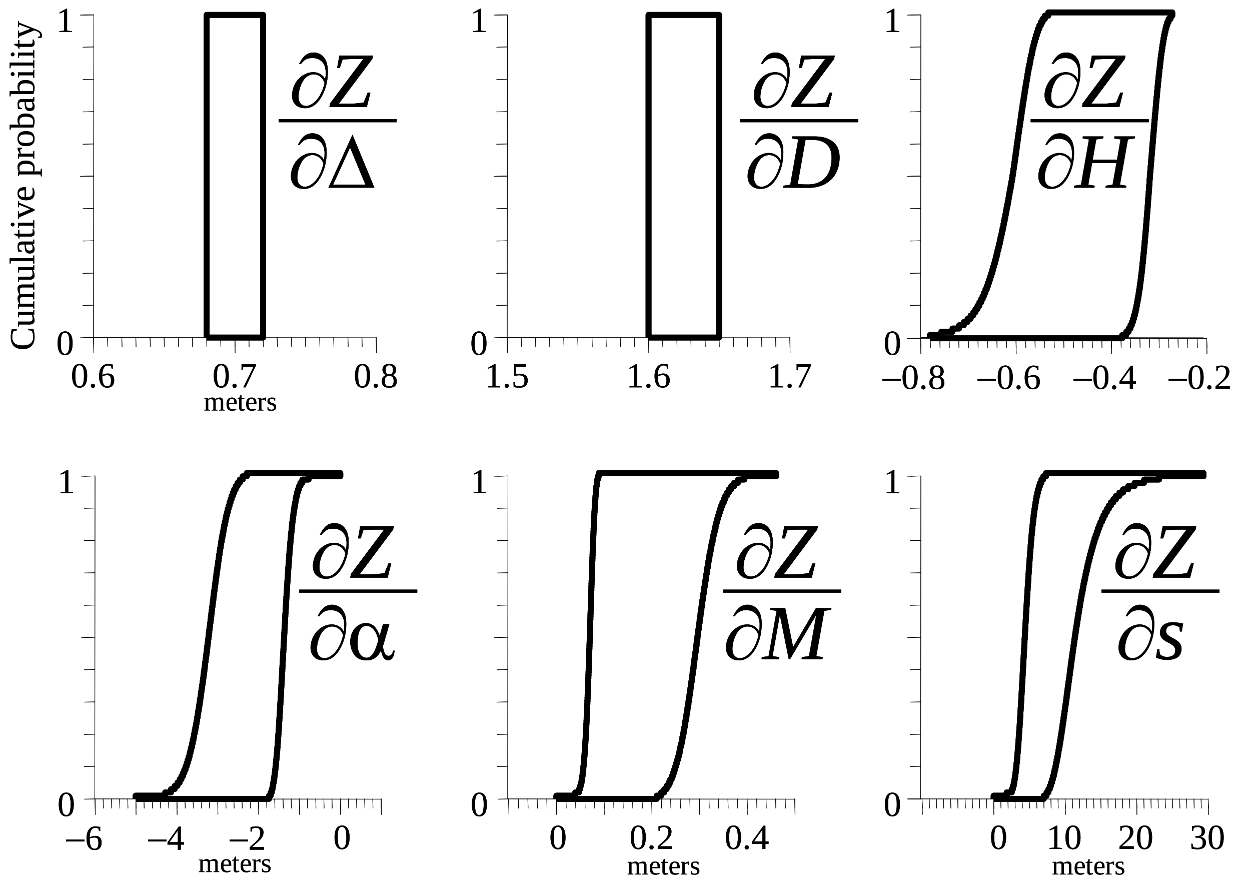
$$\frac{\partial Z}{\partial D} = \Delta$$

$$\frac{\partial Z}{\partial \alpha} = \frac{-H\left(1+\sin^2(\alpha)\right)}{\cos^3(\alpha)M\sqrt{s}}$$

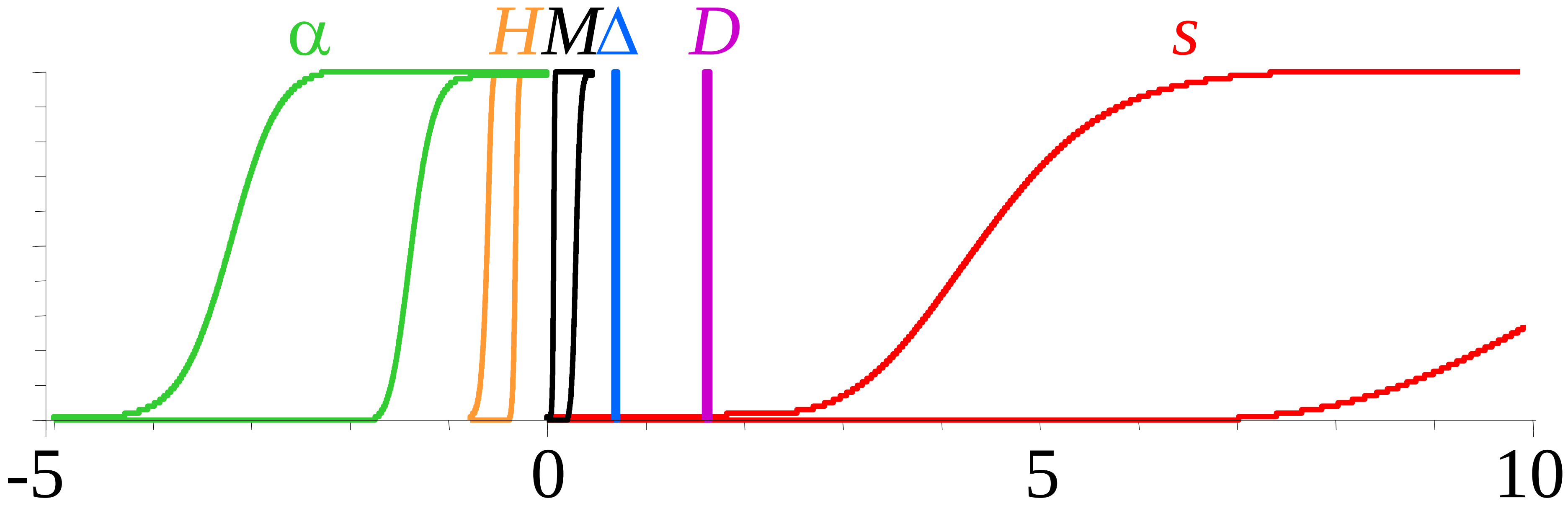
$$\frac{\partial Z}{\partial M} = \frac{H\tan(\alpha)}{\cos(\alpha)M^2\sqrt{s}}$$

$$\frac{\partial Z}{\partial H} = \frac{-\tan(\alpha)}{\cos(\alpha)M\sqrt{s}}$$

$$\frac{\partial Z}{\partial s} = \frac{H\tan(\alpha)}{2\cos(\alpha)Ms^{3/2}}$$



On the same axis



Engineering control rank order: s α D Δ H M

Completely different rankings

- The uncertainty reduction in hypothetical pinching assesses the possible effect of more or better data
- Evaluating local derivates as p-boxes tells how effective control or management can be

Best for study

Empirical study: $M \;\; H \;\; D \;\; \alpha \;\; \Delta \;\; s$

Engineering control: $s \;\; \alpha \;\; D \;\; \Delta \;\; H \;\; M$

Best for control

Different 'sensitivity' questions

- ✓ Empirical planning
  - What variables need study to reduce uncertainty?
- ✓ Engineering control
  - What variables can be modified to change the result?
- ✓ Robustness analysis
  - How robust are the results of the assessment?
- ✗ Tracking analysis
  - Which input combinations yield extreme results?

Validation

Goals

- Objectively measure the conformance of predictions with empirical data
- Use this measure to characterize the reliability of other predictions

Initial setting

- The model is fixed, at least for the time being
  - No changing it on the fly *during* validation
- A prediction is a probability distribution
  - Expressing stochastic uncertainty
- Observations are precise (scalar) numbers
  - Measurement uncertainty is negligible

relaxed later

Validation metric

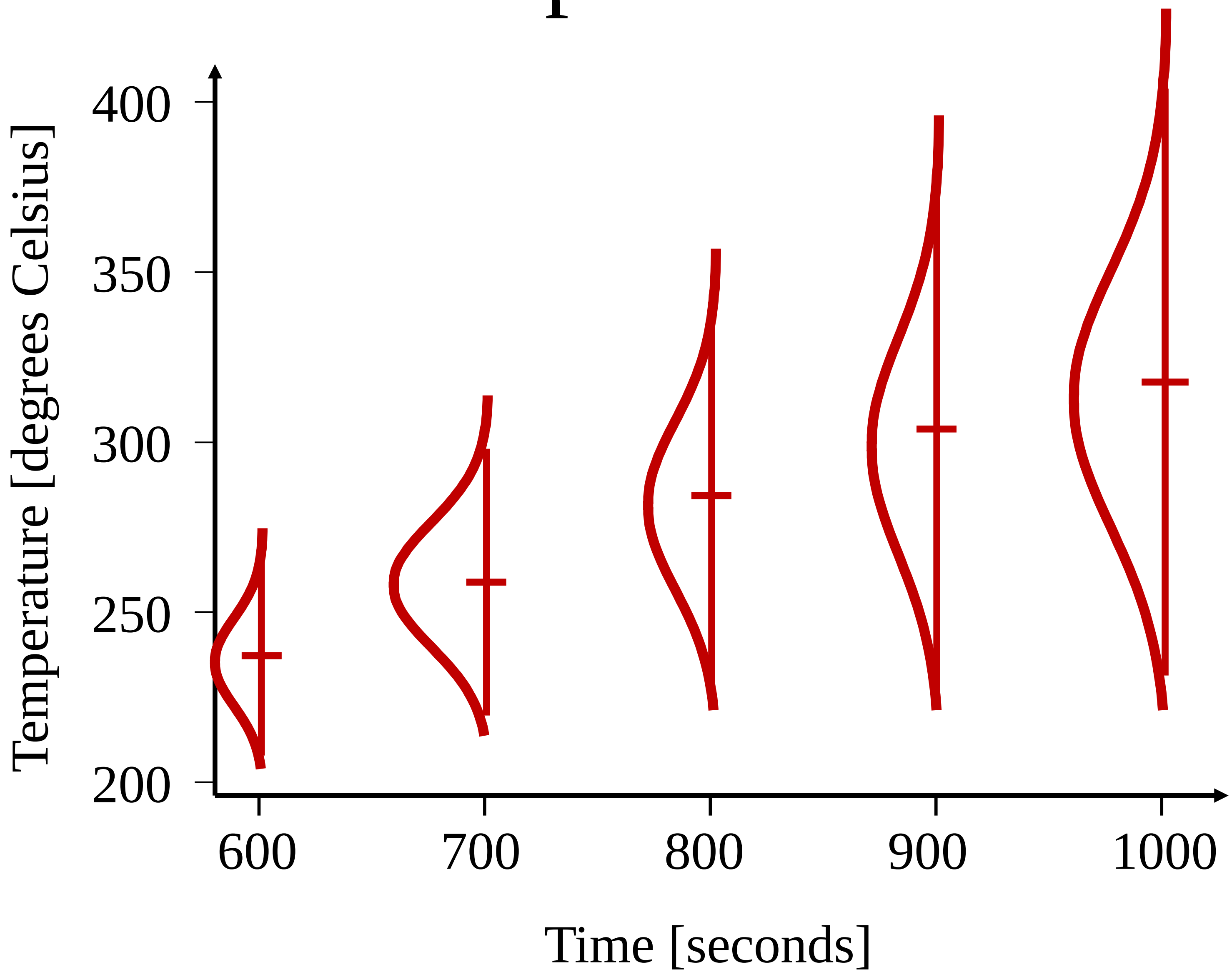
- A measure of the mismatch between the observed data and the model's predictions
  – Low value means a good match
  – High value means they disagree

- *Distance* between prediction and data

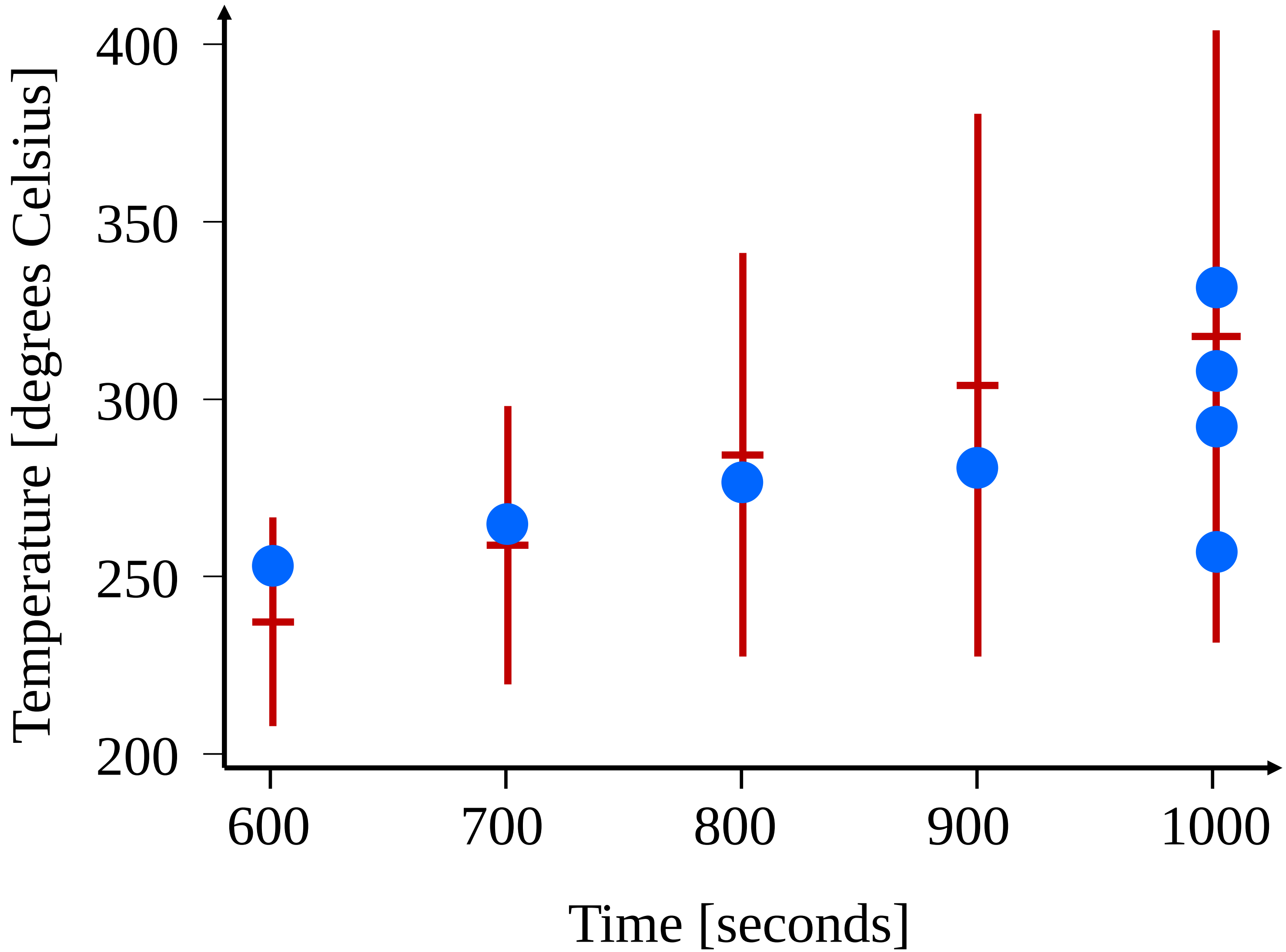
Desirable properties of a metric

- Expressed in physical units
- Generalizes deterministic comparisons
- Reflects full distribution
- Not too sensitive to long tails
- Mathematical metric
- Unbounded (you can be *really* off)

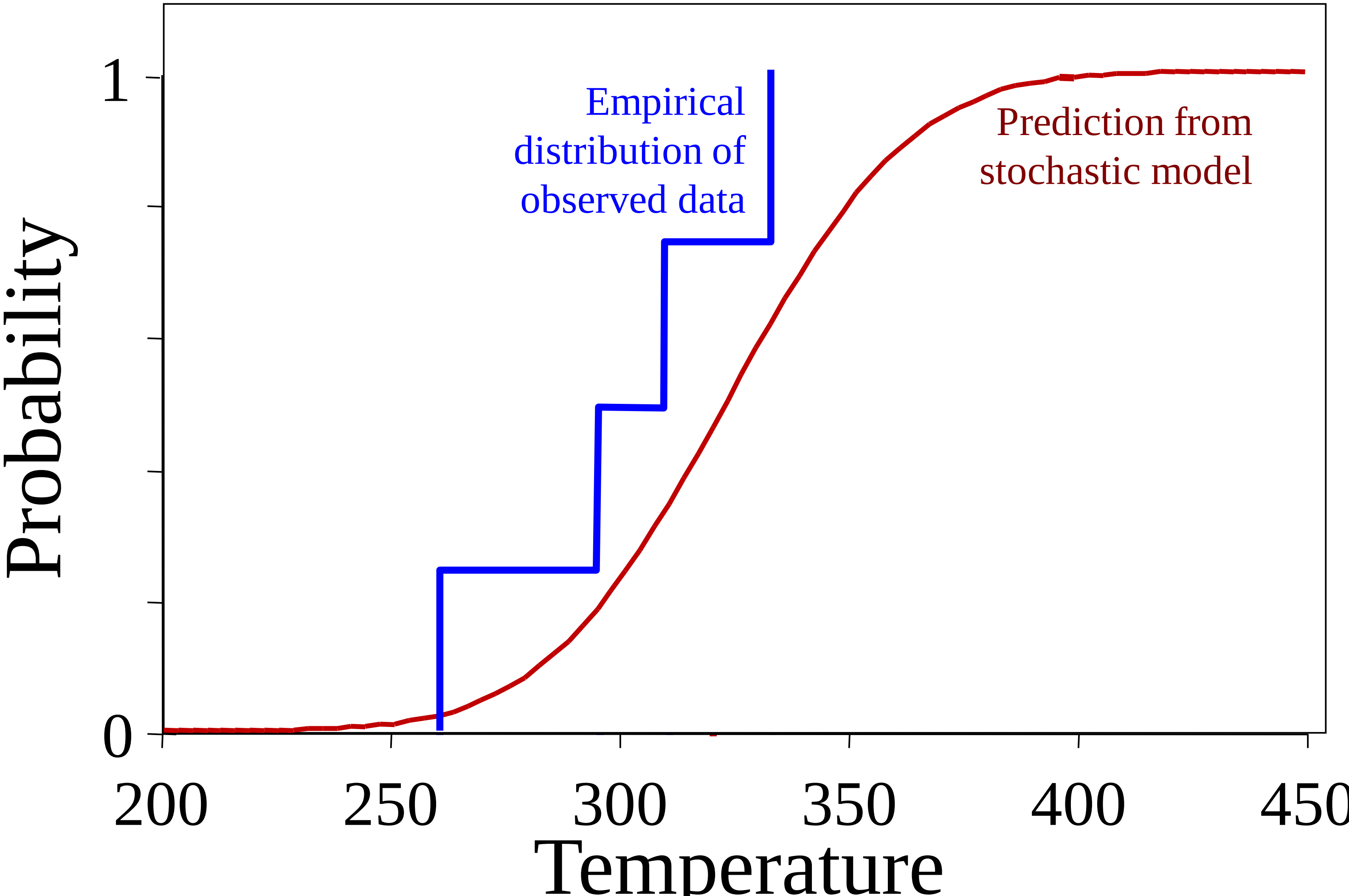
How the predictions come



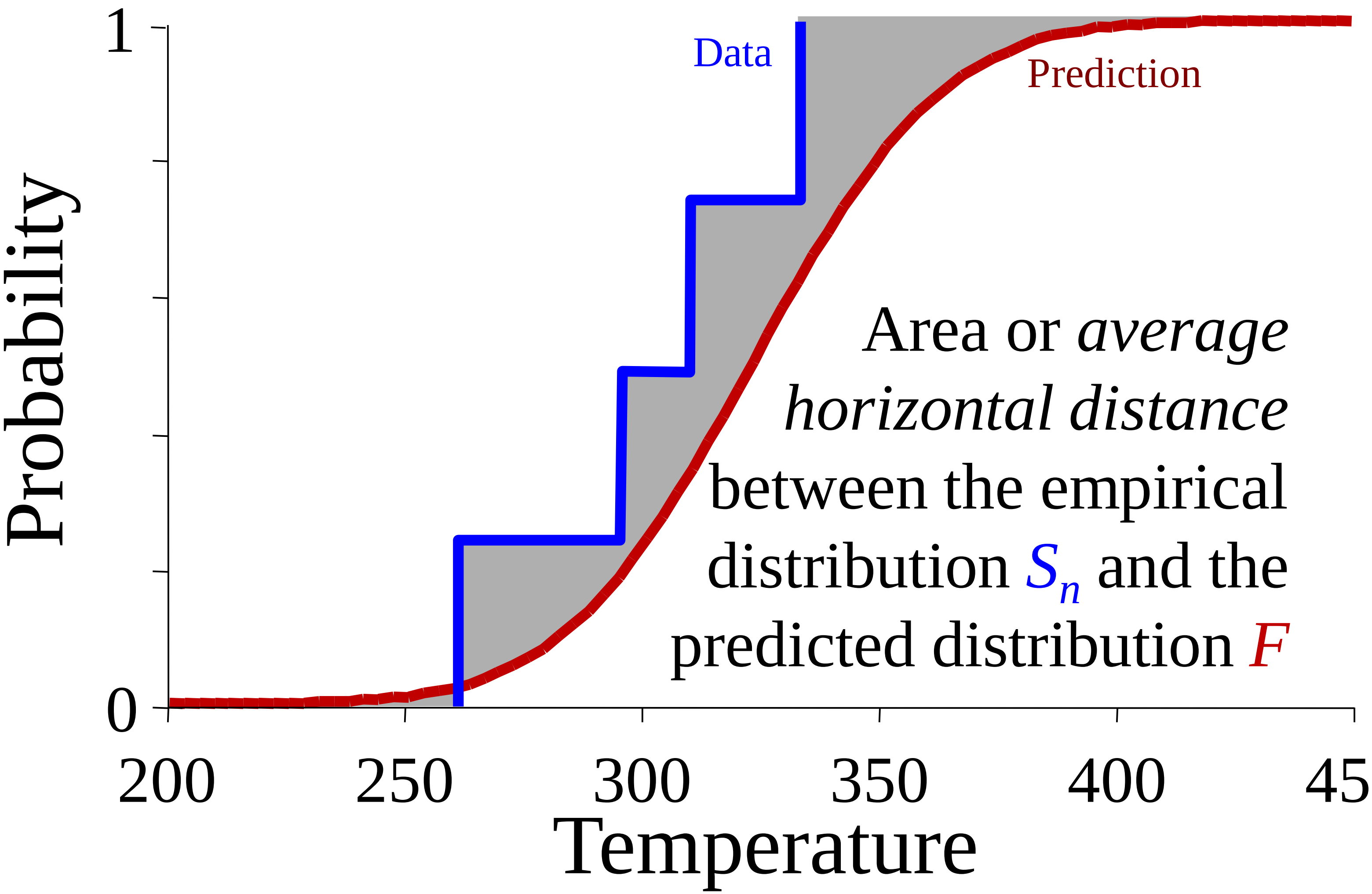
How the data come



How we look at them



One suggestion for a validation metric

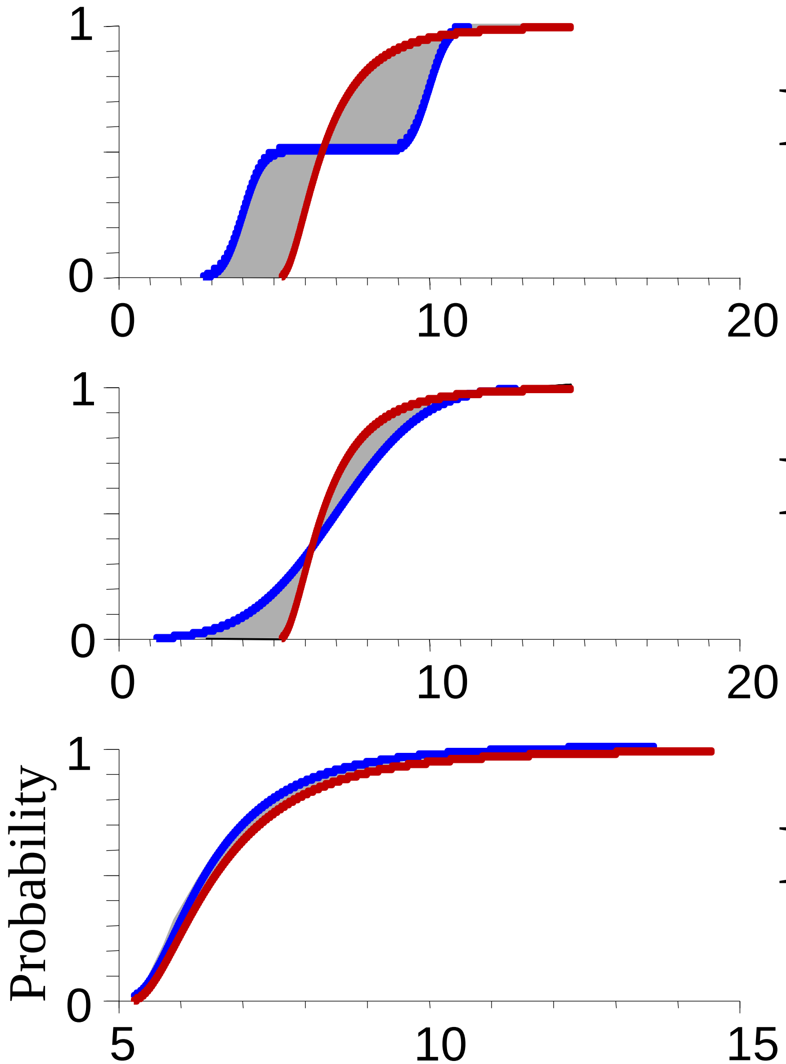


Area metric

- “Distance” between two distributions
- *Smallest mean absolute difference of deviates*

$$\min_{\substack{X \sim F \\ Y \sim S_n}} \mathrm{E}\left| X - Y \right|$$

Reflects full distribution

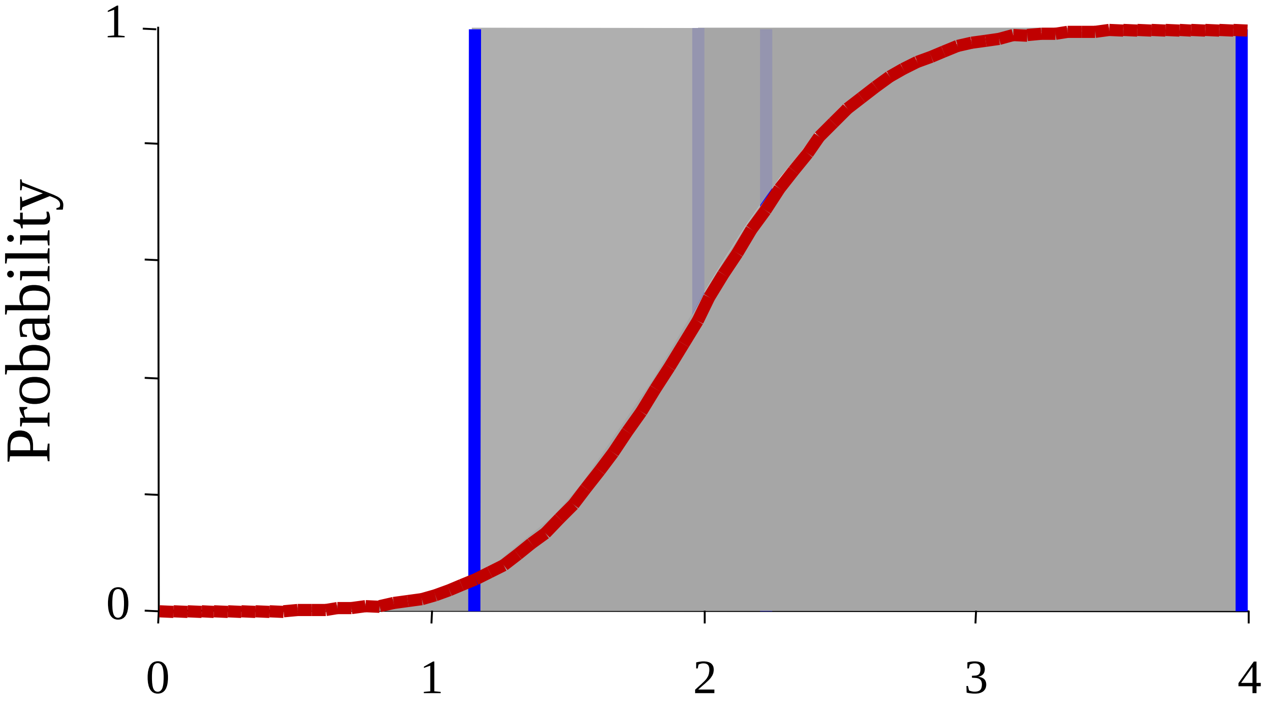


Matches in mean

Both mean and variance

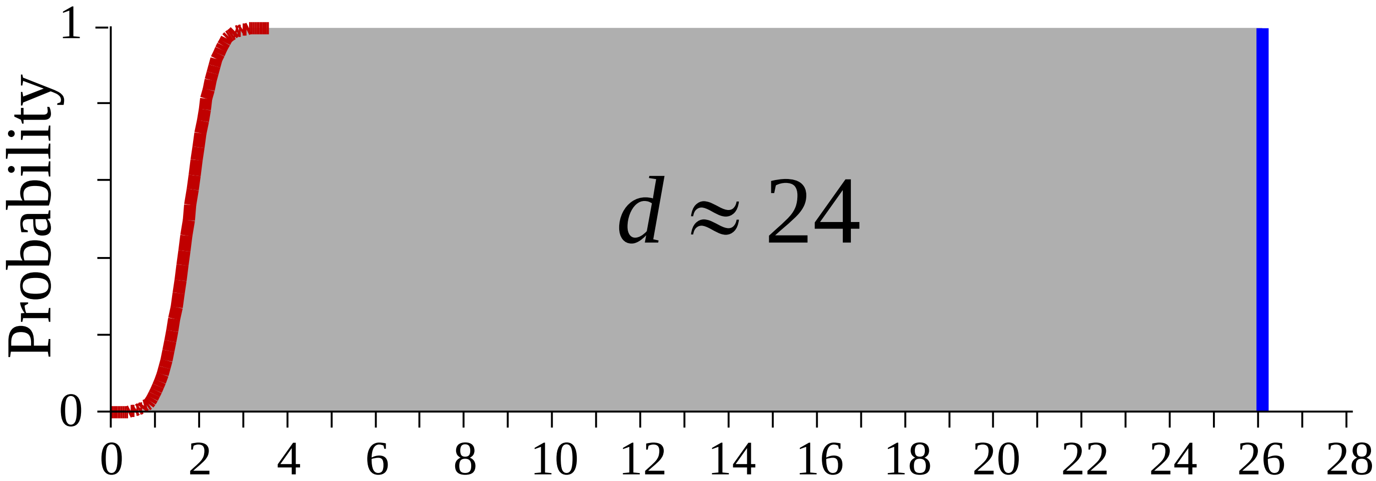
Matches well overall

Single observation



A single datum can't match an entire distribution (unless it's degenerate)

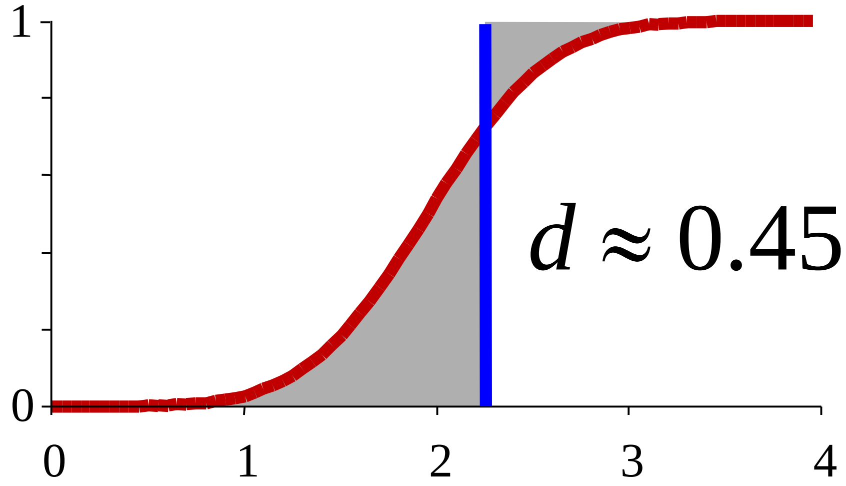
When the prediction is really bad



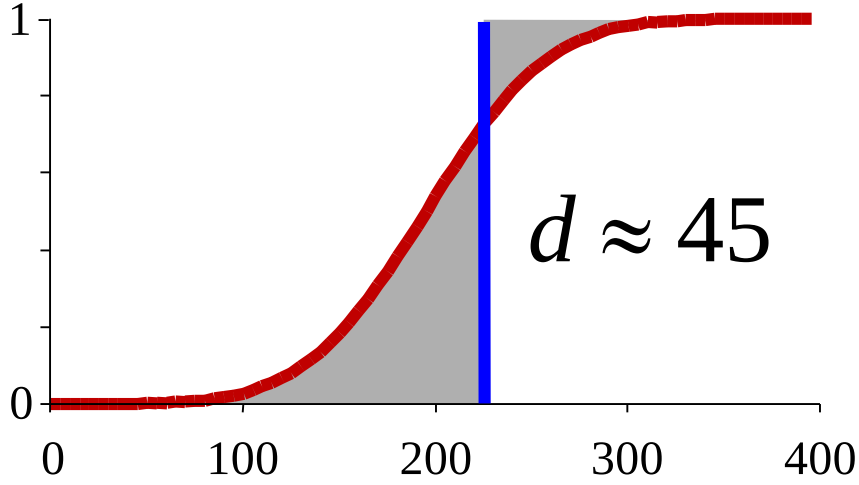
- The metric degenerates to simple distance
- Probability is dimensionless, so units are the same

Depends on the local scale

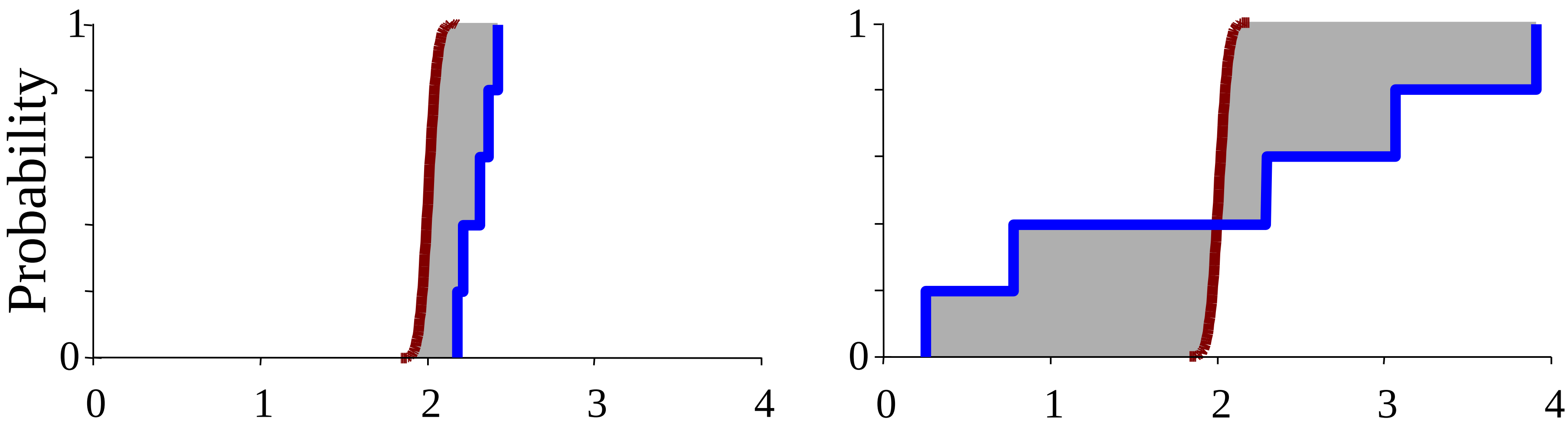
- Depends on the units



- Could standardize, but would no longer be in physical units

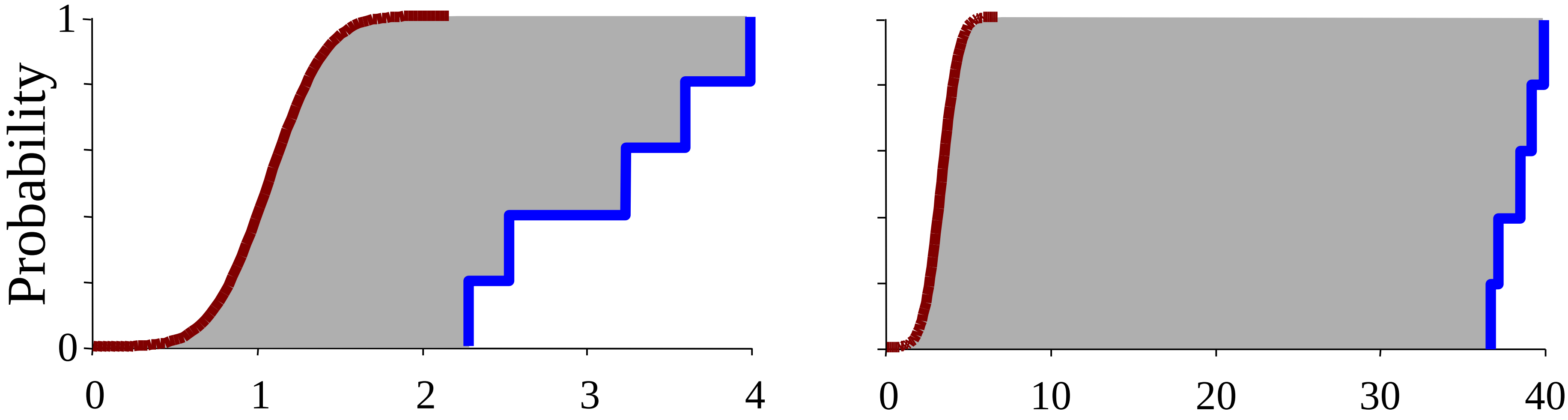


Why physical units?



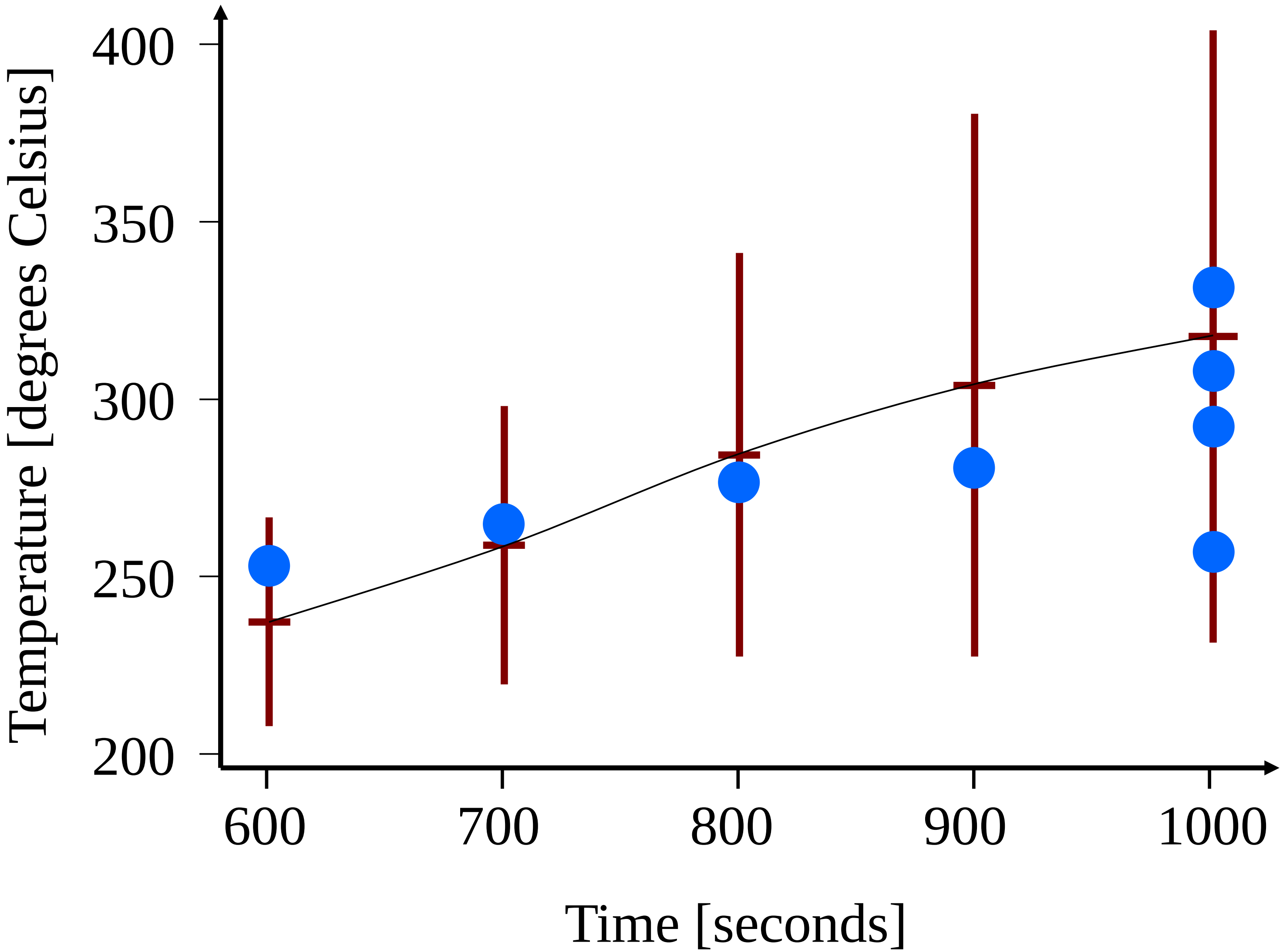
- Distributions in the *left* graph don't even overlap but they seem closer than those on the *right*

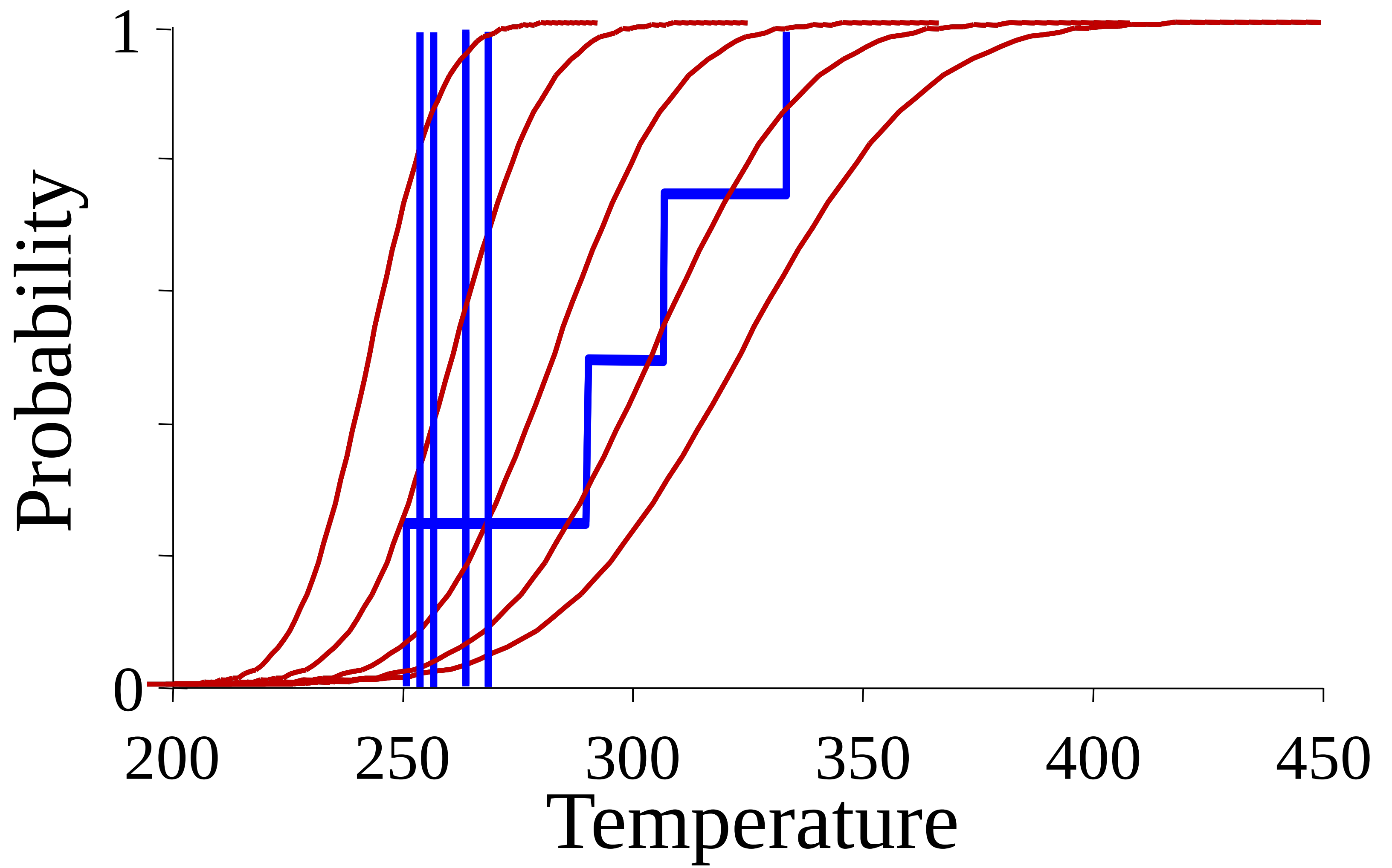
Why an unbounded metric?



- Neither overlaps, but *left* is better fit than *right*
- Smirnov's metric $D_{\max}$ considers these two cases indistinguishable (they're both just 'far')

The model says different things

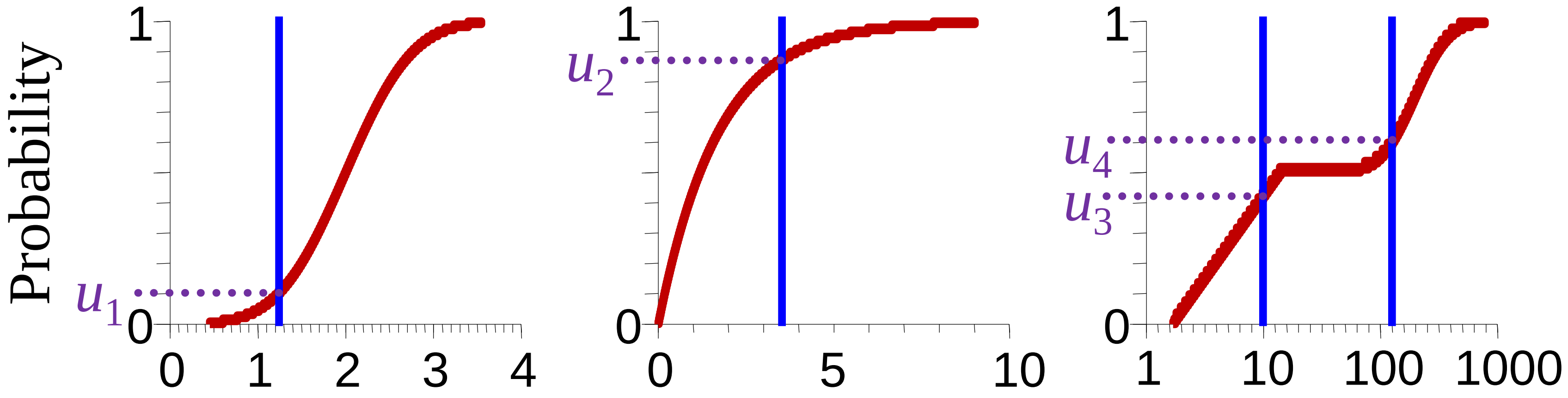



1
0
Probability
200
250
300
350
400
450
Temperature

Pooling data comparisons

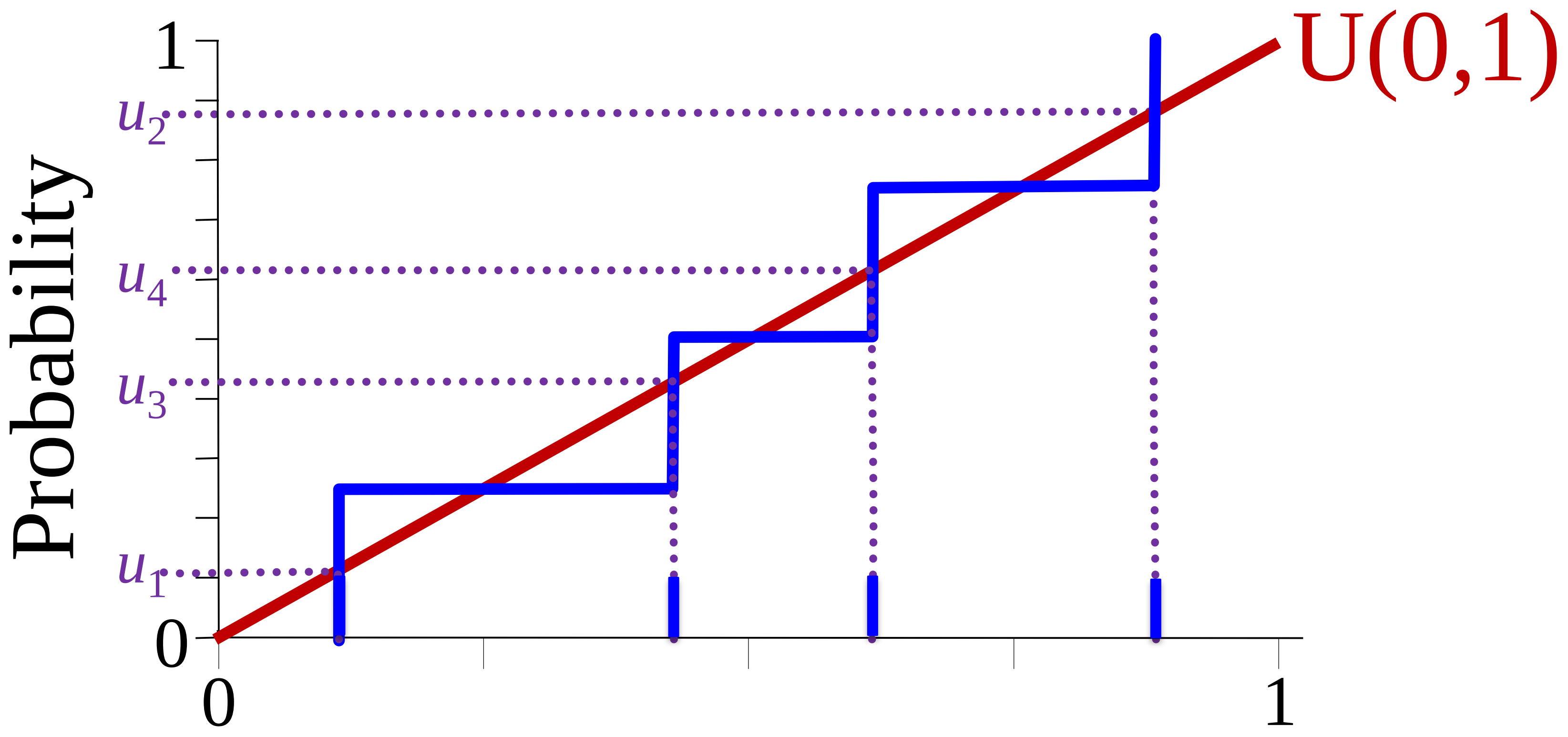
- When data are to be compared against a single distribution, they're pooled into S_n
- When data are compared against *different* distributions, this isn't possible
- Conformance must be expressed on some universal scale

Universal scale



$u_i = F_i\,(x_i)$ where x_i are the data and F_i are their respective predictions

u-pooling



Statistical test for model accuracy

- Kolmogorov-Smirnov test of distribution of u_i's against uniform over [0,1]

- Tests whether the data were drawn from the respective prediction distributions

  Probability integral transform theorem (Angus 1994) says the u's will be distributed as uniform(0,1) if $x_i \sim F_i$

- Assumes the empirical data are independent of each other

Extends to correlated multivariate case

- Li, et al. (2014) say *u*-pooling doesn't generalise to multivariate predictions with correlations
  - But they just misunderstood how to do it

- Two kinds of *u*-pooling:

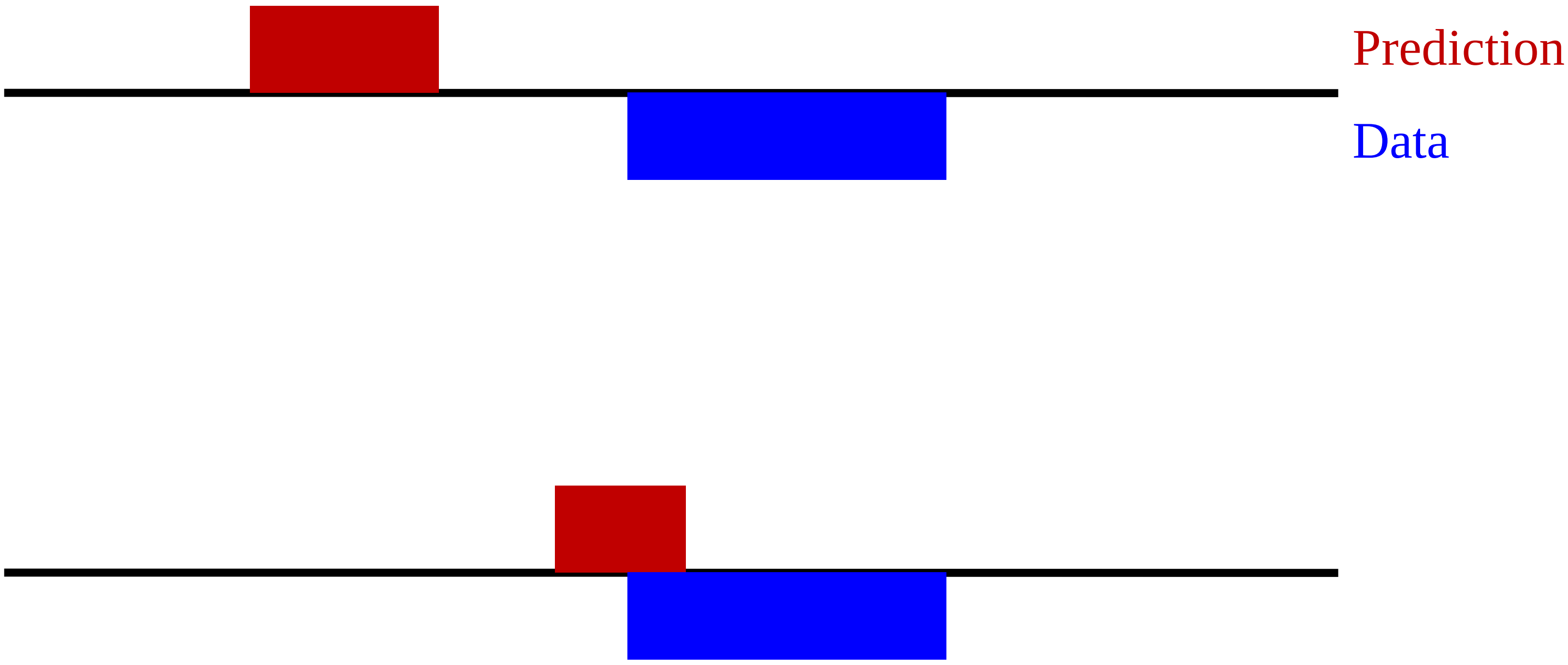
➡ Are not predicting (or don't know) dependencies

➡ Are making predictions about dependencies

Small *d* means match in distributions *and dependencies*

Epistemic uncertainty

How should we compare intervals?



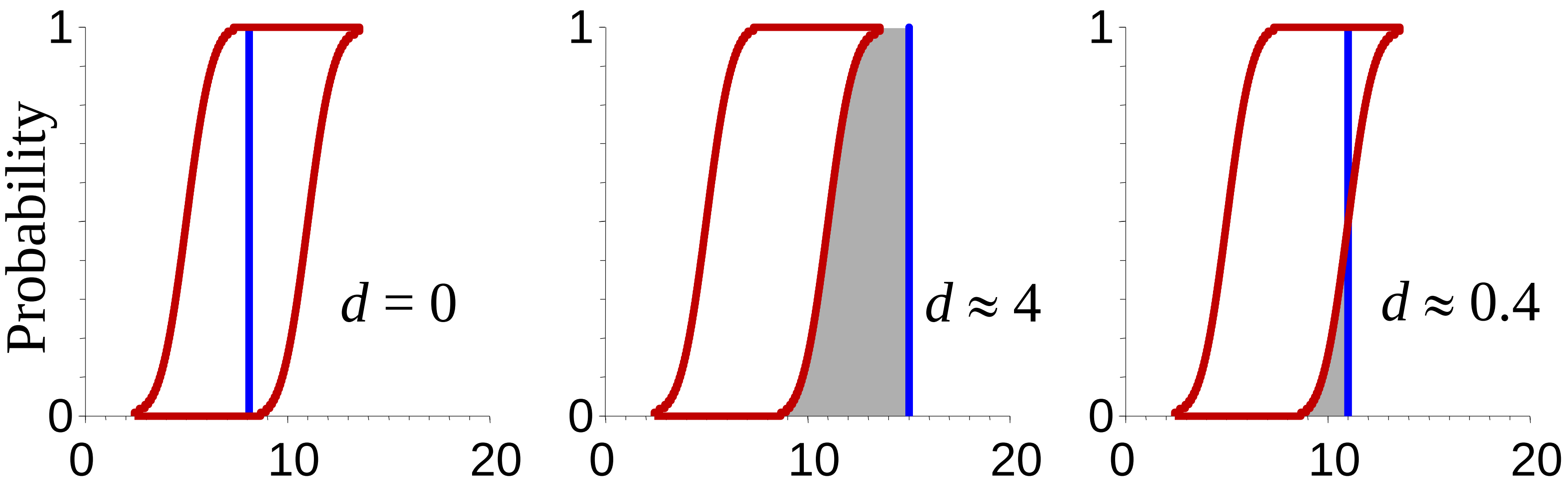
Validation for intervals

- Validation measure is the *smallest* difference
- Overlapping intervals match perfectly
- Validity is distinct from precision
  - Otherwise no value in an uncertainty analysis

277

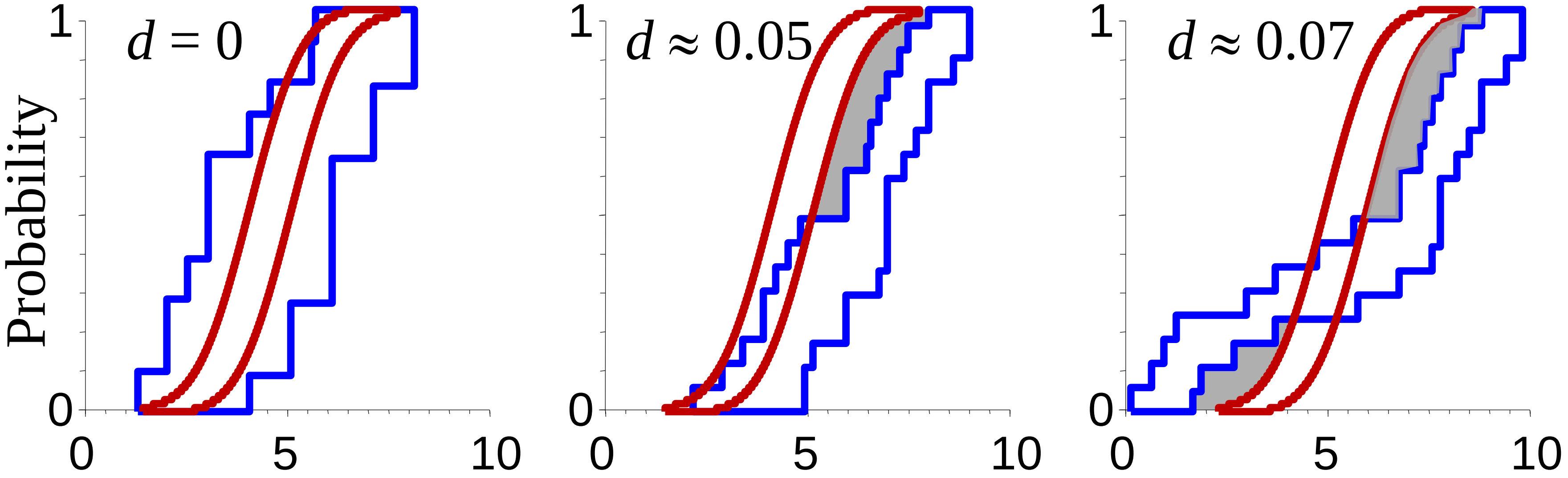
http://encarta.msn.com/map_701512318/English_Channel.html

Epistemic uncertainty in predictions



- In *left*, the datum evidences no discrepancy at all
- In *middle*, the discrepancy is relative to the edge
- In *right*, the discrepancy is even smaller

Epistemic uncertainty in both



Predictions in red

Observations in blue

Predictive capability

Predictive capability

- Not a measure of how good the model is
- A characterization of *how much we should trust its predictions*

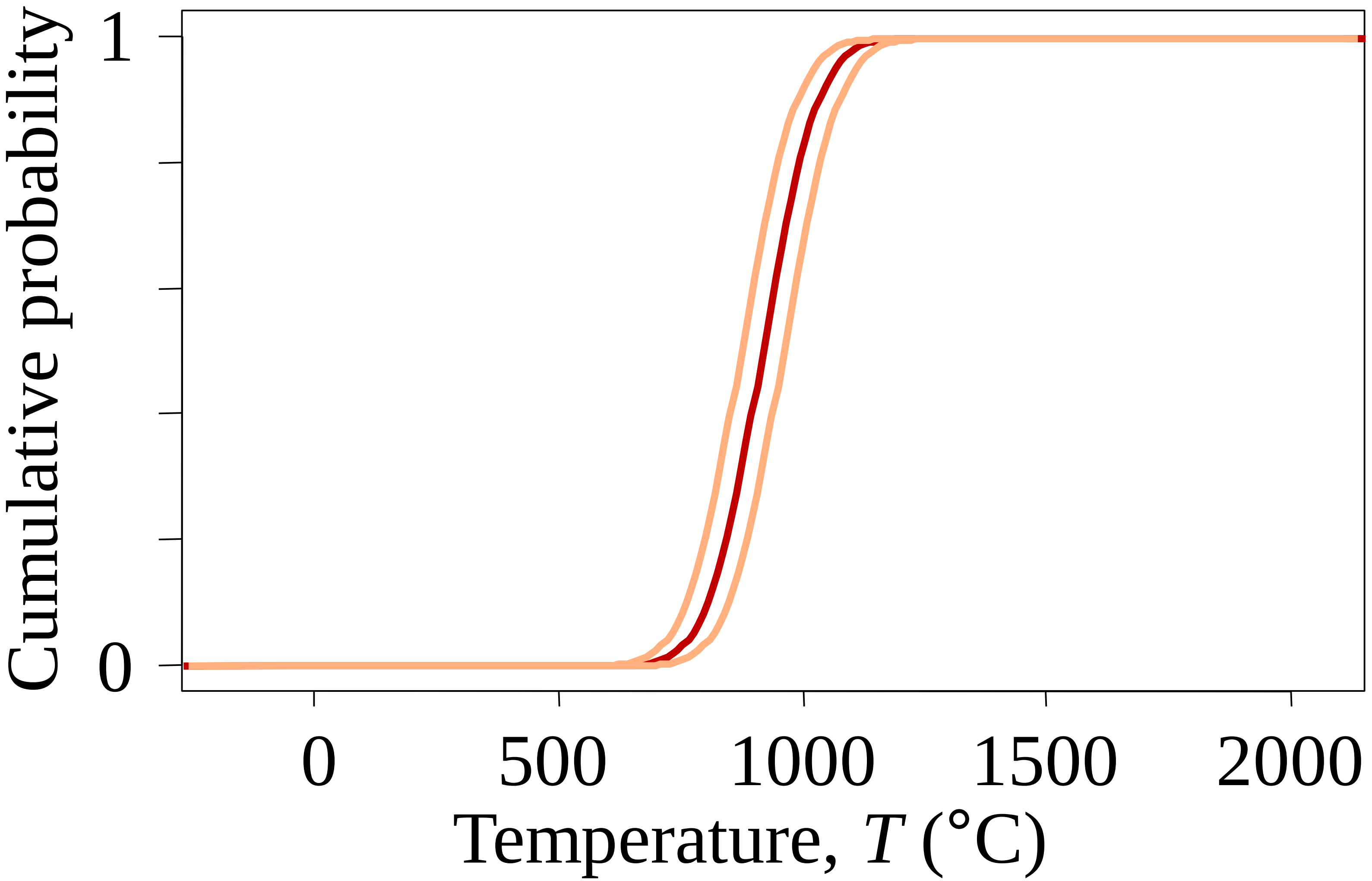
Predictive capability

- Additional uncertainty (estimate of possible error) that is recognized by comparing a model's predictions against available data

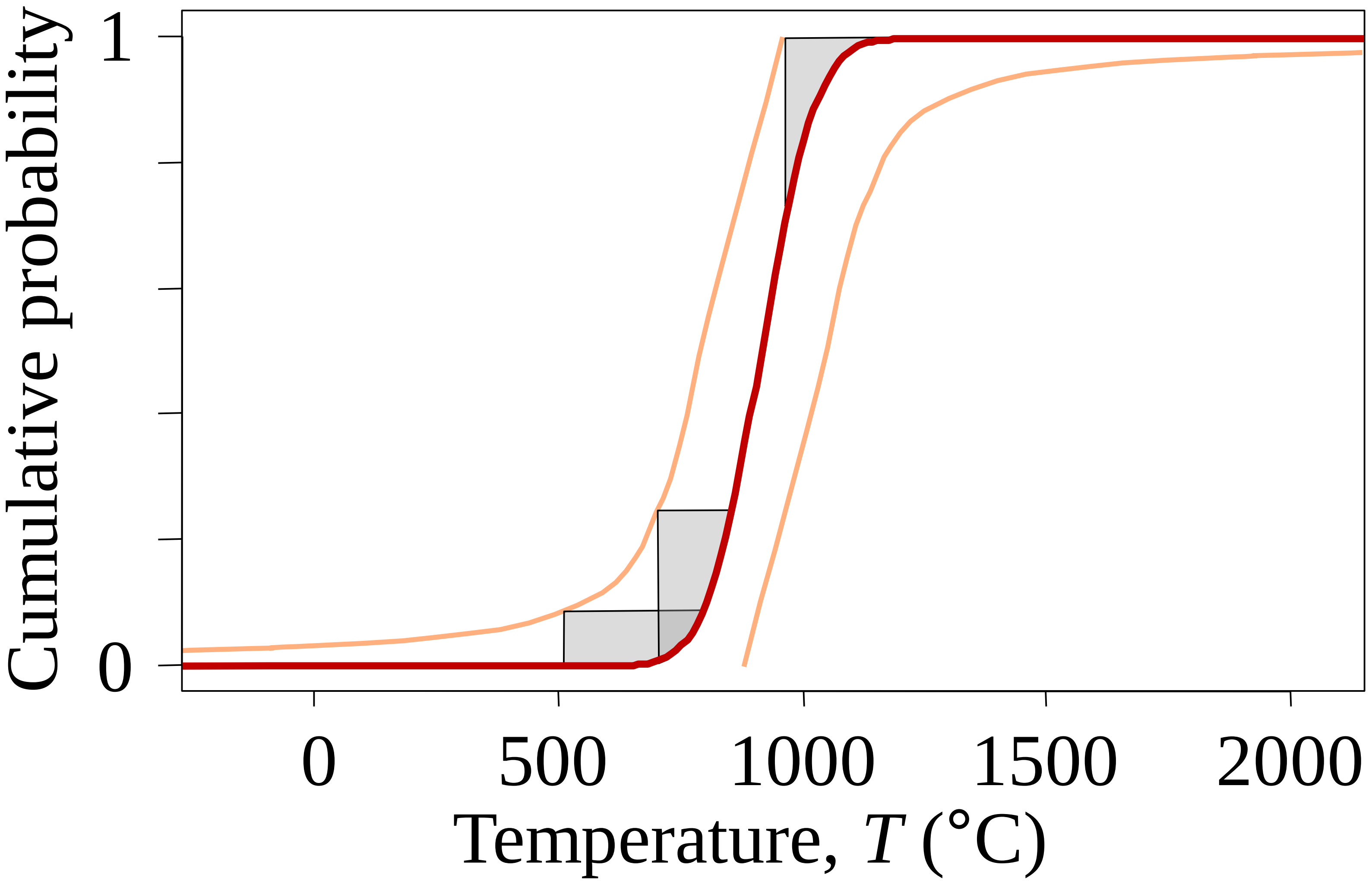
- *On top of* introspective model uncertainty estimates already embodied in the analysis

- How should we characterize this uncertainty?
- Let's consider three possible approaches

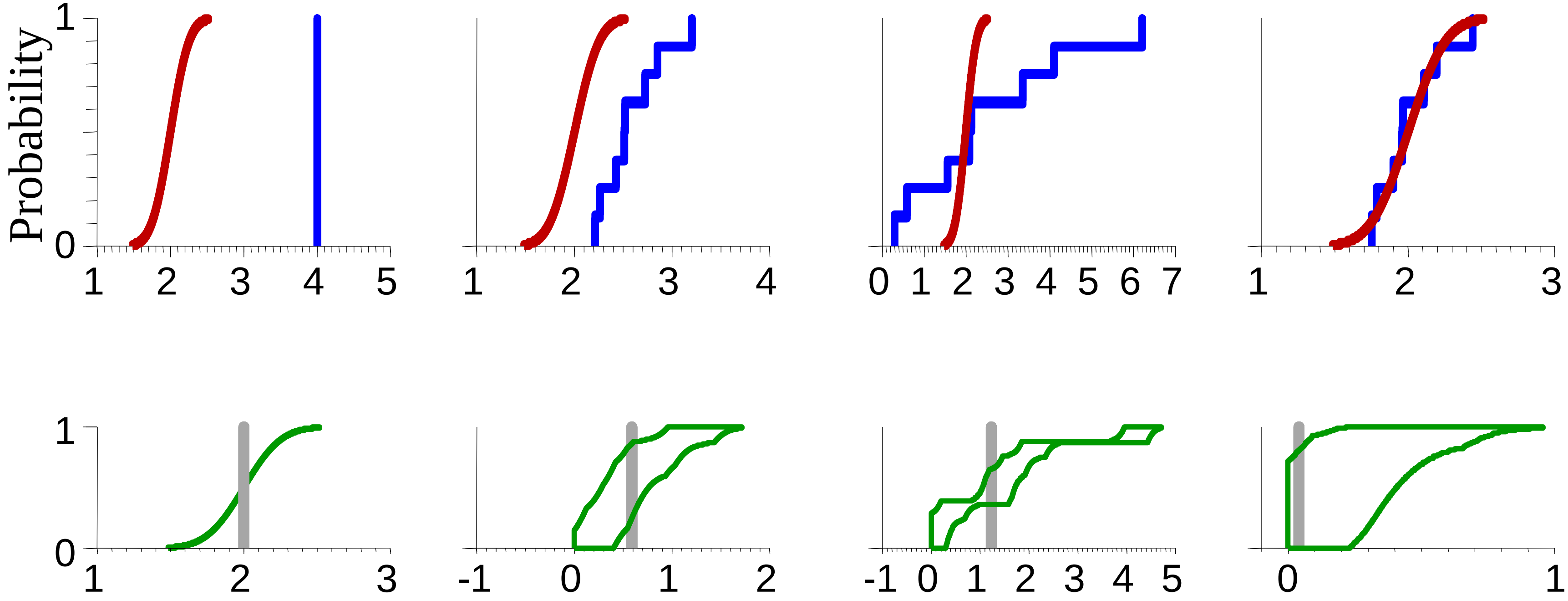
Plus or minus *d* (area metric)



All distributions that are *d*-close

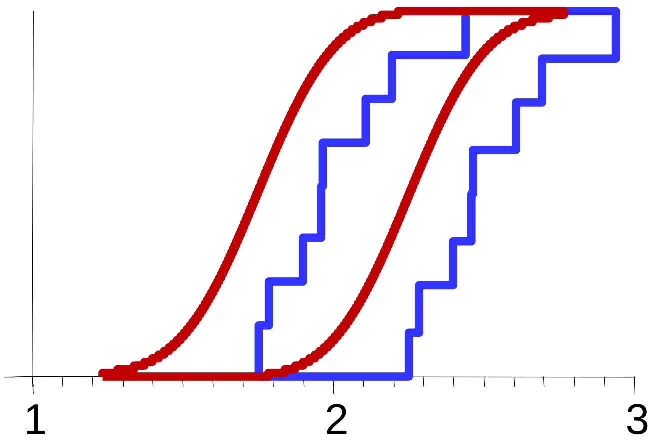


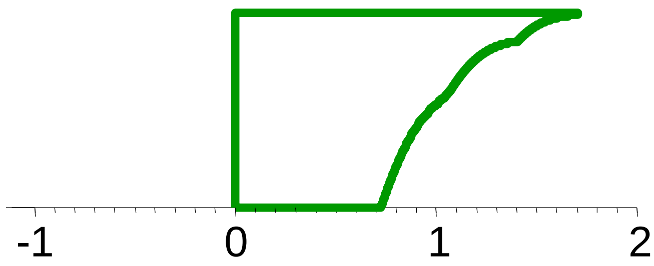
Area versus distribution of differences



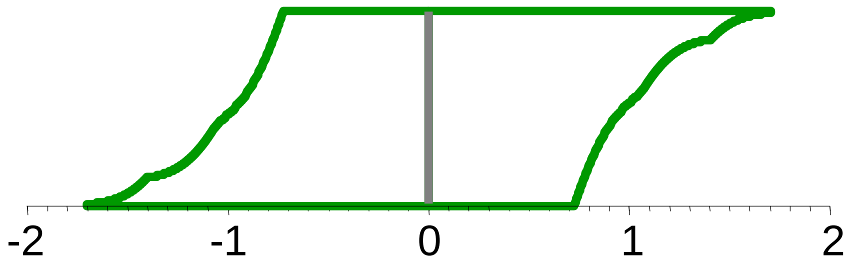
Area between distributions Distribution (or p-box) of differences

Signed difference and bias





Estimate of bias

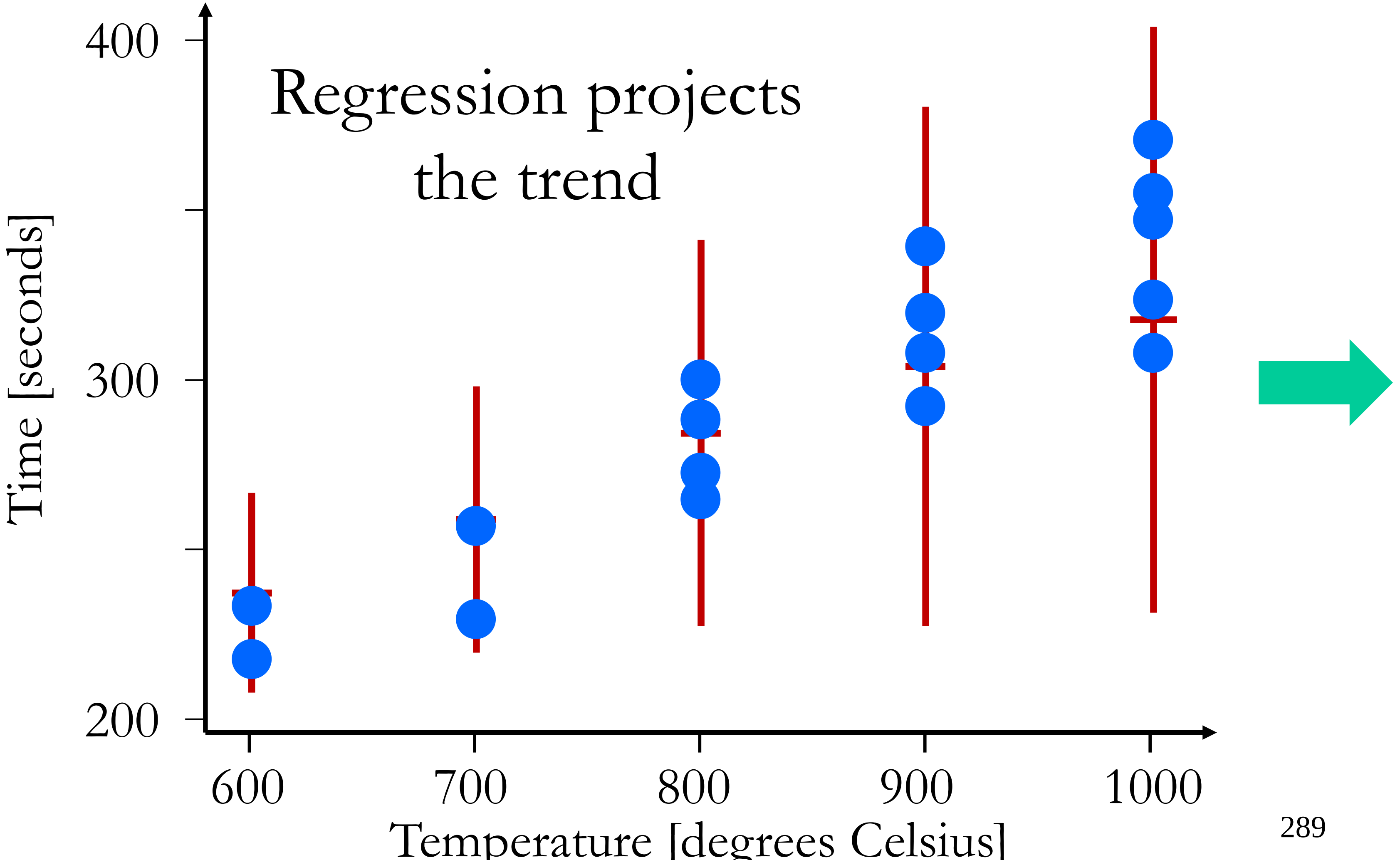


Estimate of uncertainty

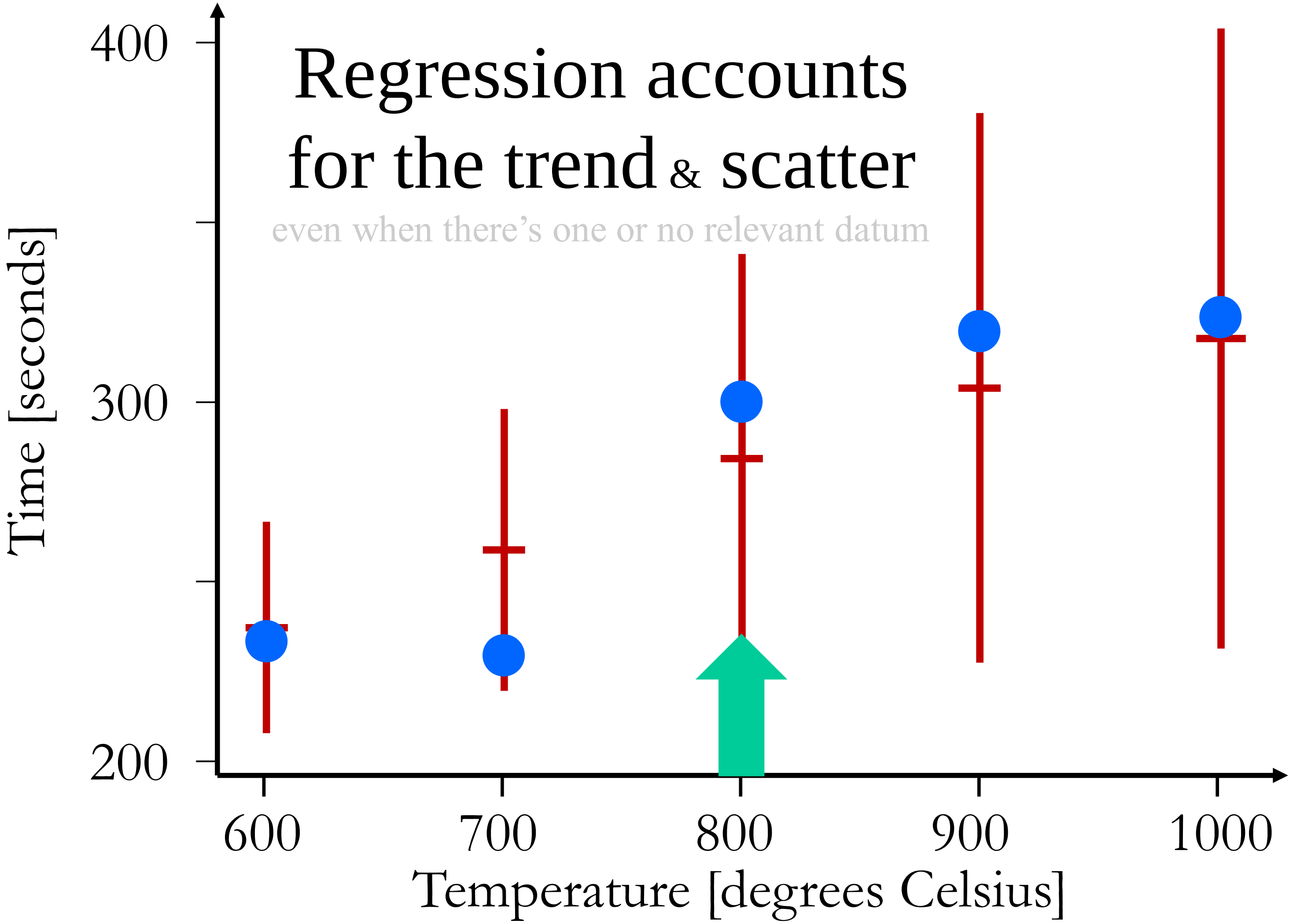
Research topic

- Difference interval, distribution, or p-box
- Sign
  - Signed differences to account for model bias
  - Plus-or-minus absolute differences
- Extrapolation by regression analysis
  - From conditions for which data are available to conditions at which the prediction is to be made
- Interpolation by regression analysis
  - Sampling uncertainty about the differences

Extrapolate to 2000 degrees



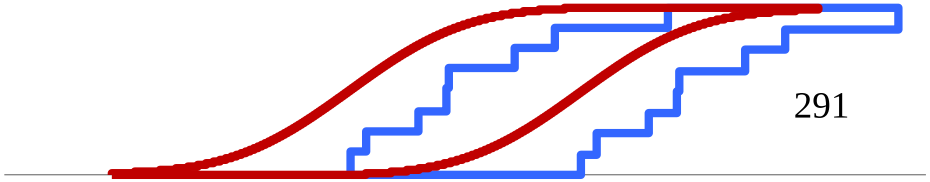
Interpolating at 800 degrees



Sampling uncertainty about data

- Regression methods account for the trends and scatter (variation over some axis) of the data

- But, for predictive capability, the blue data p-box should reflect sampling uncertainty too
  - May be important when sample size is small
  - Can use confidence or tolerance band to form the blue p-box summarizing the data



Predictive capability

- Turns on epistemic and aleatory uncertainty
- Post-hoc measure of model uncertainty
- Distinct from validation
- Clearly crucial to any sort of modeling
- Still the subject of broad discussion and debate

Assessment versus reliability

Validation

- Discrepancy between prediction and data
- Absolute value
- May neglect sampling uncertainty of data
- Summarizes across predictions and even dimensions

Predictive capability

- Distribution of differences between prediction and data
- May have sign
- Must account for sampling uncertainty about data
- Uses regression across predictions, but limited to one dimension

Summary

- Both assessment and reliability of extrapolation
  - How good is the model? Validation
  - Should we trust its pronouncements? Predictive capability
  - Calibration and updating are separate activities
- Validation measure must be bespoke but universal; predictive capability is still part of modeling

- Epistemic uncertainty introduces some wrinkles
  - Full credit for being modest about predictions

Backcalculation

Backcalculation

- Engineering design requires backcalculation

- How can we untangle the expression

$$A+B = C$$

when we know A and C, and need B?

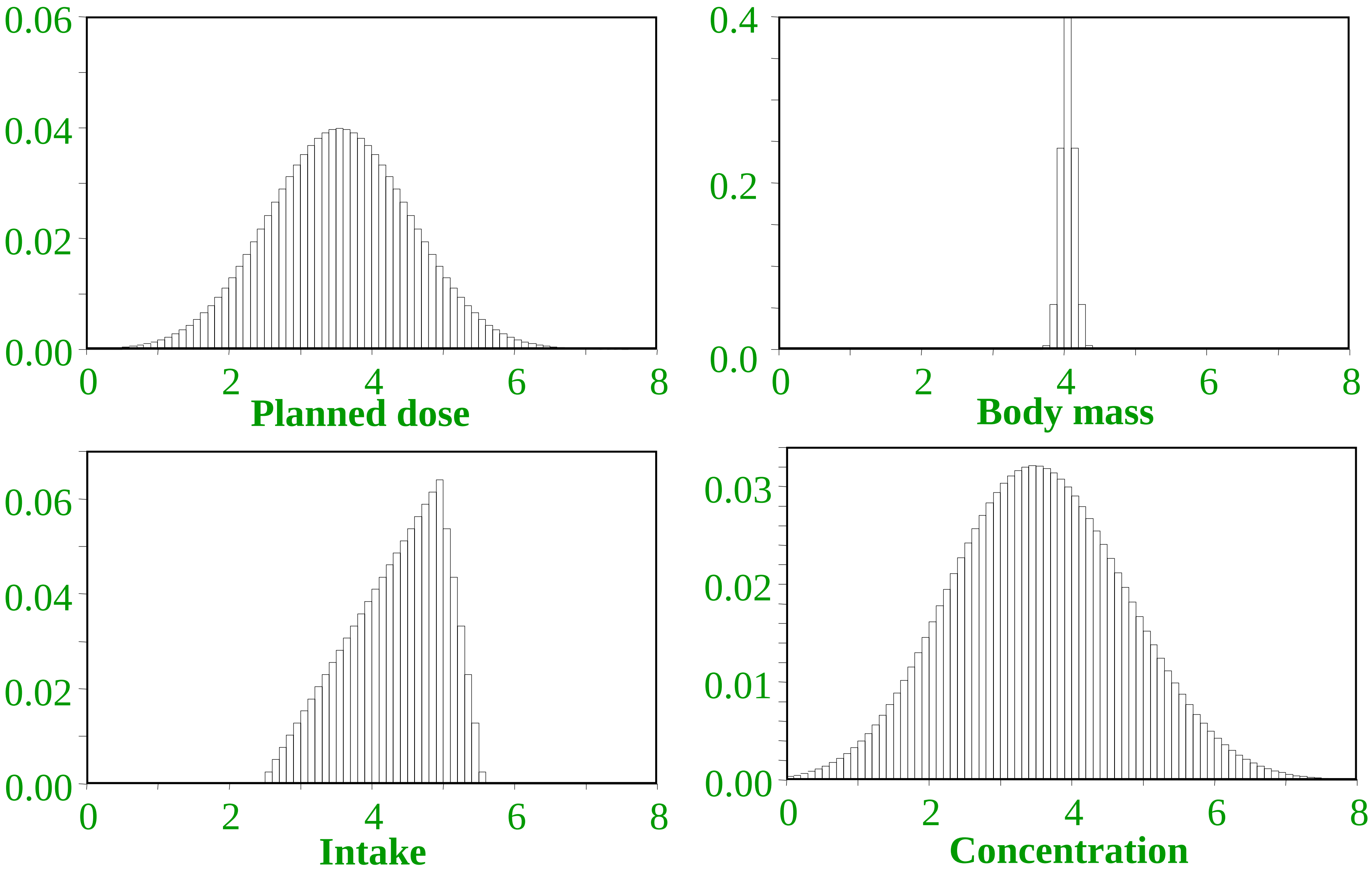
Can’t just invert the equation

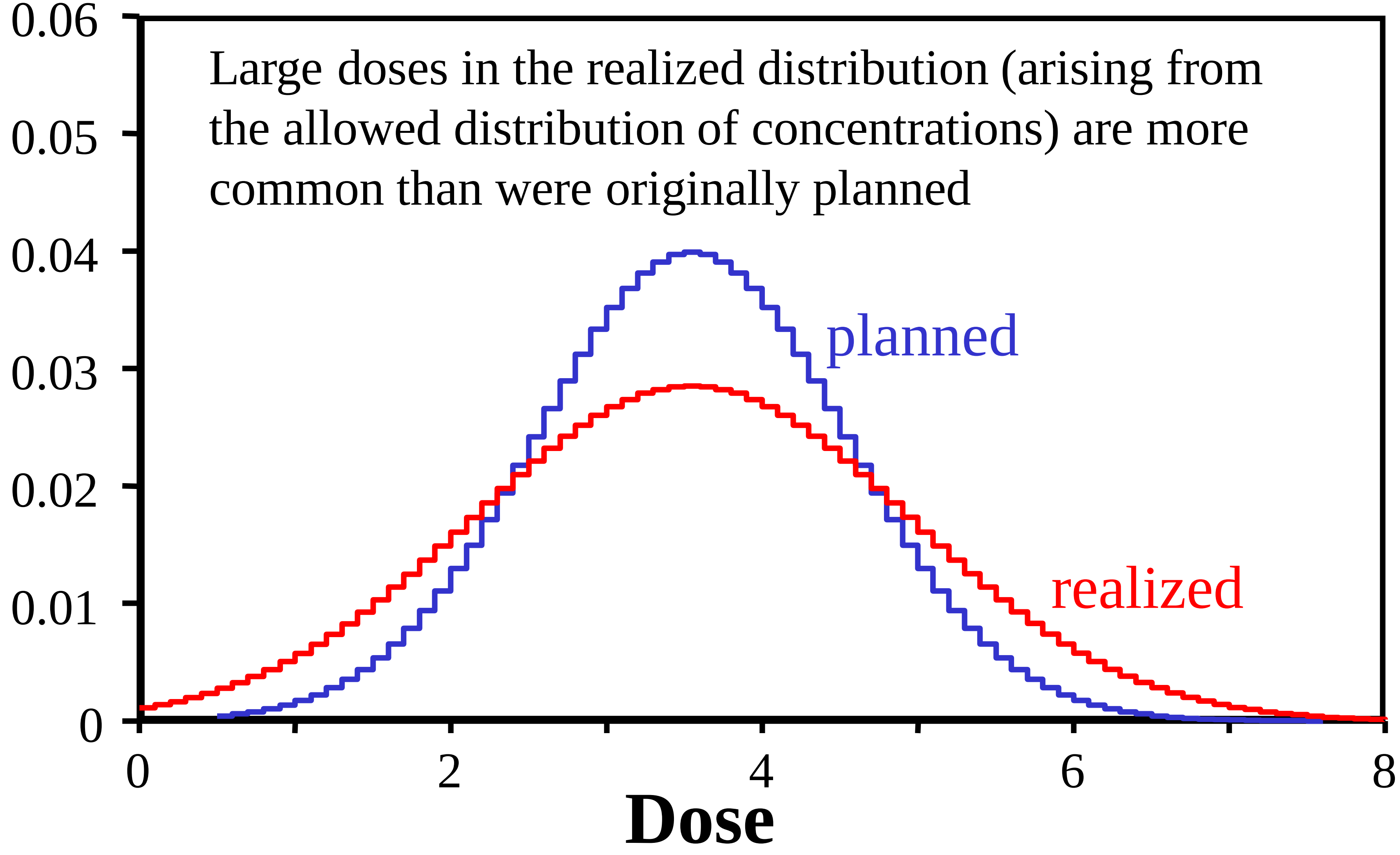
prescribed unknown known

$$\mathit{Dose} = \mathit{Concentration} \times \mathit{Intake} \,/\, \mathit{Bodymass}$$

$$\mathit{Concentration} = \mathit{Dose} \times \mathit{Bodymass} \,/\, \mathit{Intake}$$

When concentration is put back into the forward equation, the resulting dose is wider than planned


0.06
0.04
0.02
0.00
0
2
4
6
8
Planned dose
0.4
0.2
0.0
0
2
4
6
8
Body mass
0.06
0.04
0.02
0.00
0
2
4
6
8
Intake
0.03
0.02
0.01
0.00
0
2
4
6
8
Concentration



0.06
0.05
0.04
0.03
0.02
0.01
0
Large doses in the realized distribution (arising from the allowed distribution of concentrations) are more common than were originally planned
planned
realized
0
2
4
6
8
Dose

Normal approximation

- If $A+B=C$, compute B as $C-A$ under the assumption the correlation between A and C is r = sd(A)/sd(C)

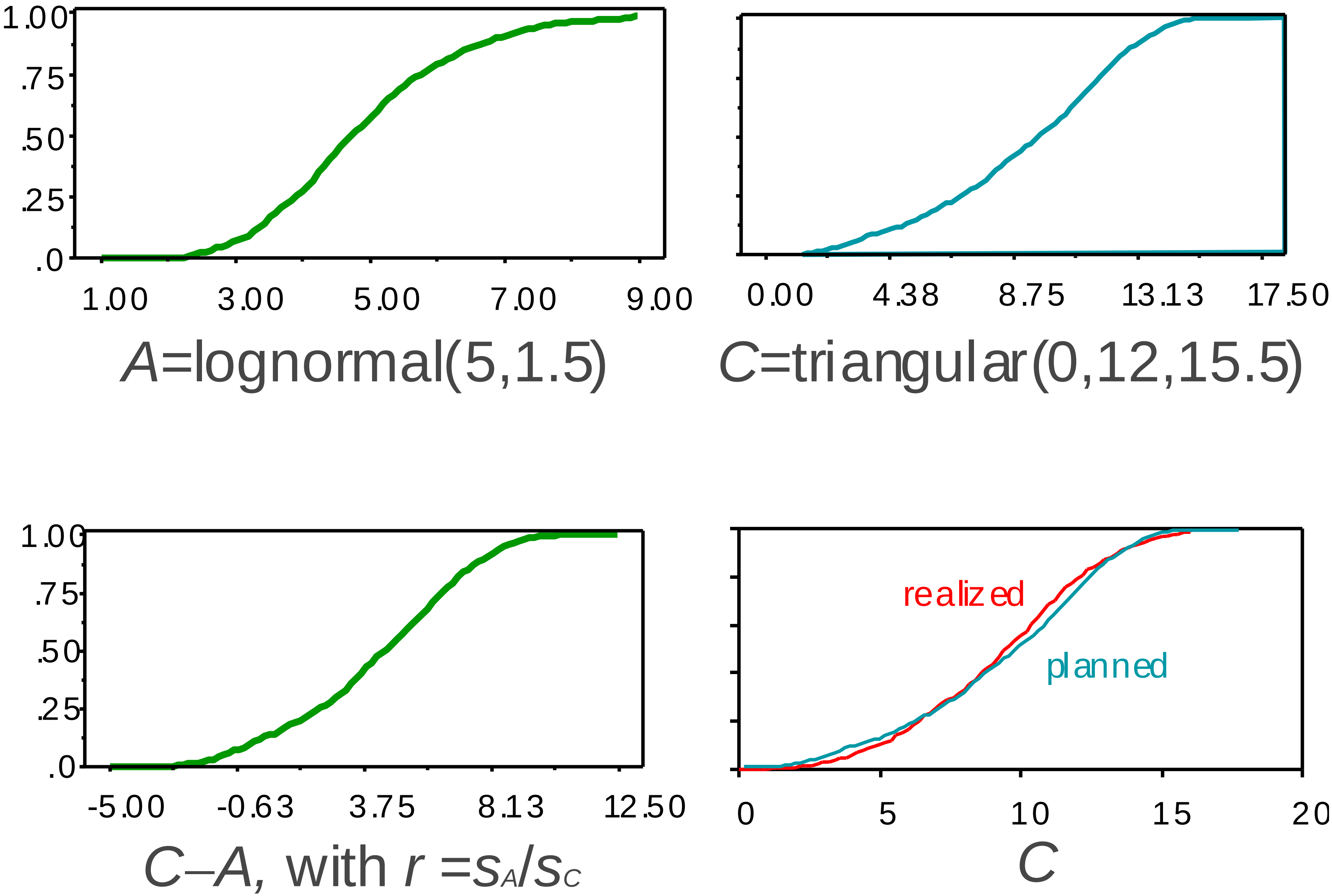
To simulate normal deviates with correlation r, compute

$$Y_1 = Z_1 \times \sigma_1 + \mu_1$$
$$Y_2 = (rZ_1 + Z_2 \sqrt{(1-r^2)}) \times \sigma_2 + \mu_2$$

where Z_1 and Z_2 are independent standard normal deviates, and μ and σ denote the respective desired means and standard deviations

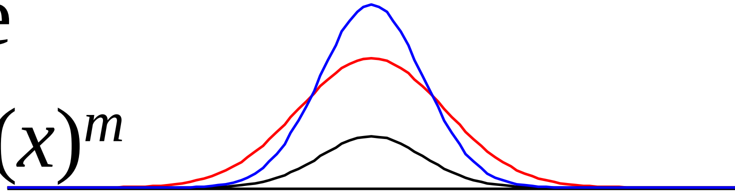
- Uses Pearson correlation (not rank correlation)
- Good for multivariate normal, and maybe more

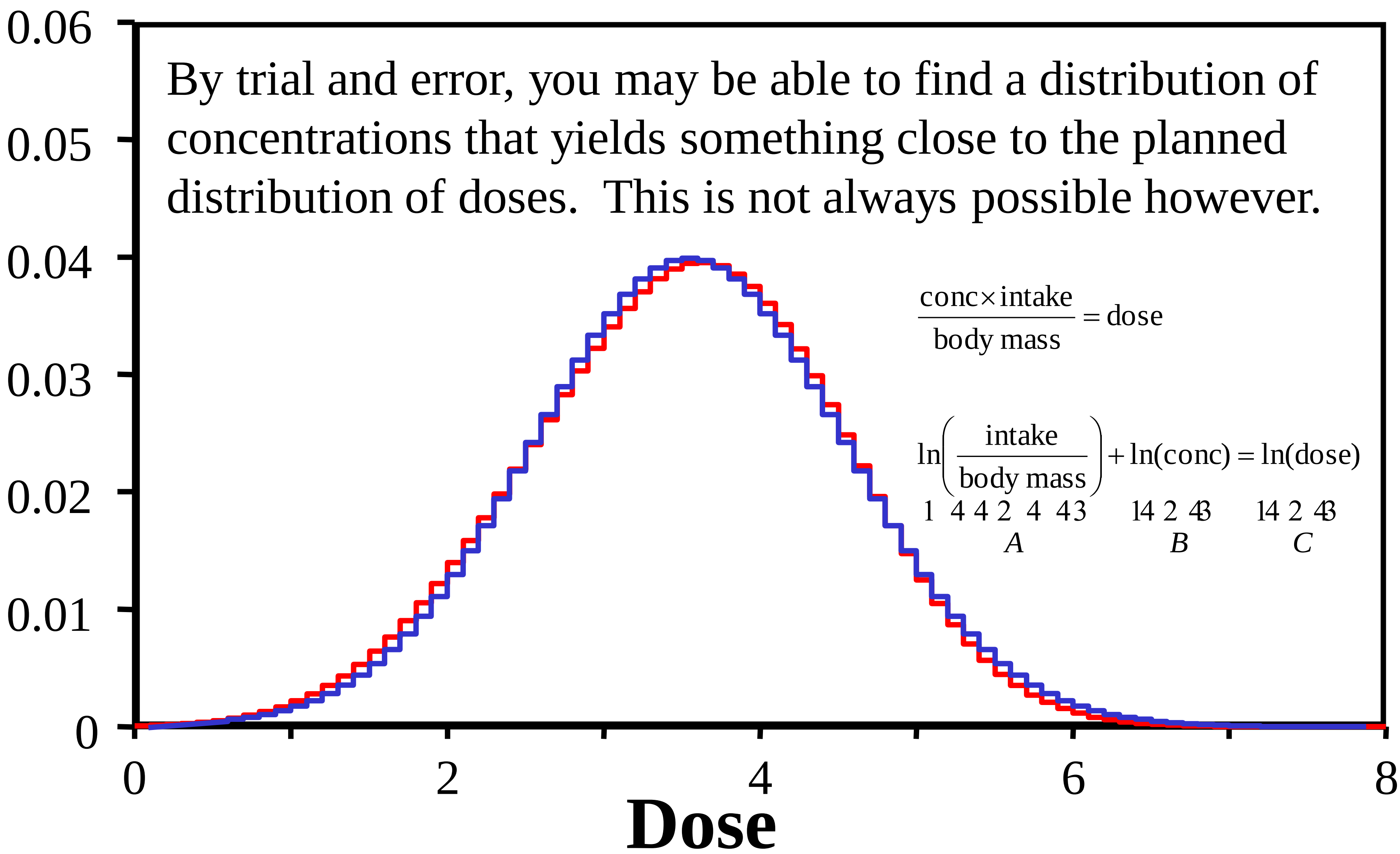
Normal approximation with non-normal distributions



Iterative (trial & error) approach

- Initialize B with C–A
- This distribution is too wide
- Transform density $p(x)$ to $p(x)^m$
- Rescale so that area remains one
- Whenever $m > 1$ dispersion decreases
- Vary m until you get an acceptable fit





By trial and error, you may be able to find a distribution of concentrations that yields something close to the planned distribution of doses. This is not always possible however.
$\frac{\text{conc} \times \text{intake}}{\text{body mass}} = \text{dose}$
$\underbrace{\ln\left(\frac{\text{intake}}{\text{body mass}}\right)}_{A} + \underbrace{\ln(\text{conc})}_{B} = \underbrace{\ln(\text{dose})}_{C}$
0.06
0.05
0.04
0.03
0.02
0.01
0
0
2
4
6
8
Dose

Backcalculation with intervals

- Backcalculation is well understood for intervals

- Each interval operation has multiple modes:

Shell (or "united") solution is forward projection

The shell contains every solution

Kernel (or "tolerance") solution is backward

Every element of the kernel is a solution

Kernel versus shell

$A = [1,2] \qquad C = [2,6] \qquad C = A + B$

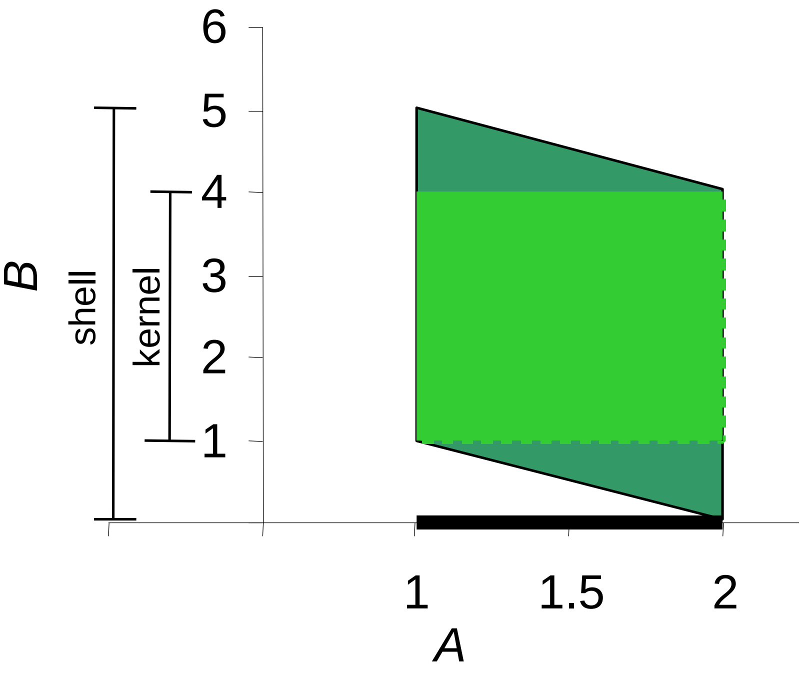
There are different ways to solve for B

Shell

$$B = C - A$$
$$= [C_1 - A_2, C_2 - A_1]$$
$$= [0,5]$$

Kernel

$$B = \text{ backcalc}(A,C)$$
$$= [C_1 - A_1, C_2 - A_2]$$
$$= [1,4]$$



When you need for	And you have estimates for	Use this formula to find the unknown
	A, B	$C = A + B$
$A + B \subseteq C$	A, C	$B = \text{backcalc}(A,C)$
$\subseteq$ means "is a subset of"	B, C	$A = \text{backcalc}\ (B,C)$
	A, B	$C = A - B$
$A - B \subseteq C$	A, C	$B = -\text{backcalc}(A,C)$
	B, C	$A = \text{backcalc}(-B,C)$
	A, B	$C = A * B$
$A \times B \subseteq C$	A, C	$B = \text{factor}(A,C)$
	B, C	$A = \text{factor}(B,C)$
	A, B	$C = A / B$
$A\ / B \subseteq C$	A, C	$B = 1/\text{factor}(A,C)$
	B, C	$A = \text{factor}(1/B,C)$
	A, B	$C = A \wedge B$
$A \wedge B \subseteq C$	A, C	$B = \text{factor}(\log A, \log C)$
	B, C	$A = \exp(\text{factor}(B, \log C))$
$2A \subseteq C$	A	$C = 2 * A$
	C	$A = C / 2$
$A^2 \subseteq C$	A	$C = A \wedge 2$
	C	$A = \text{sqrt}(C)$

$\text{backcalc}(a,c) = [c_1 - a_1, c_2 - a_2]$

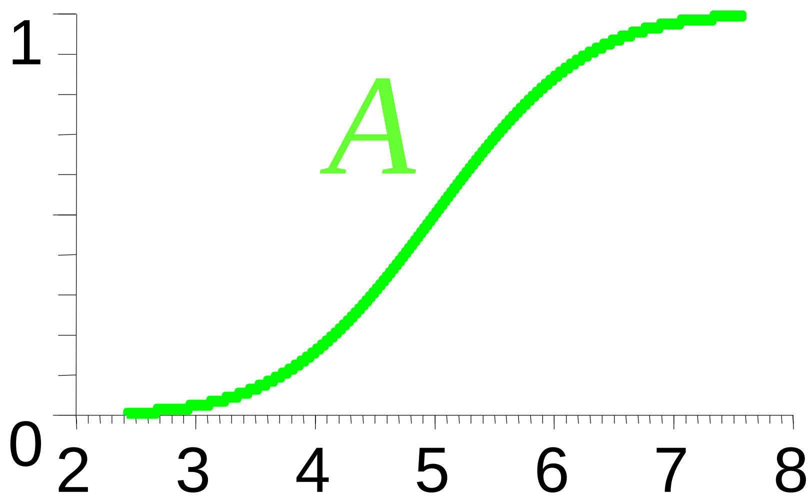
$\text{factor}(a,c) = [c_1/a_1, c_2/a_2]$

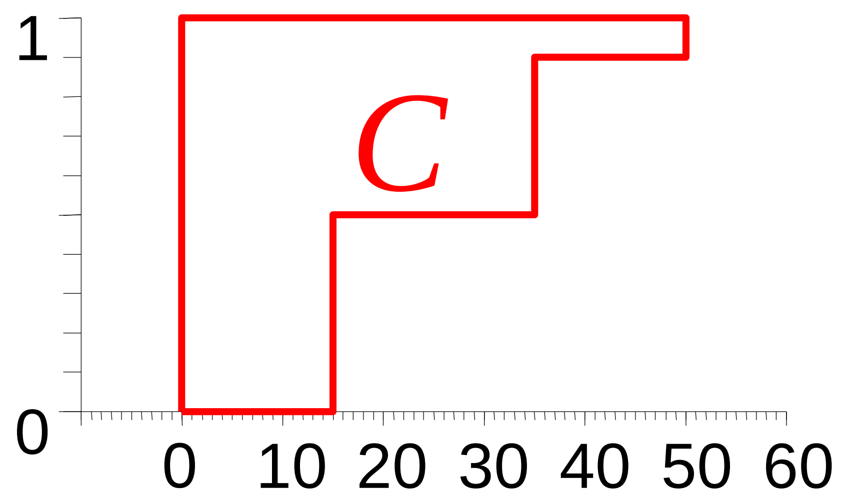
Backcalculation with p-boxes

Suppose $A + B = C$, where

A = normal(5, 1)

$C = \{0 \leq C$, median ≤ 1.5, 90th %ile ≤ 35, max $\leq 50\}$





Backcalc algorithm for p-boxes

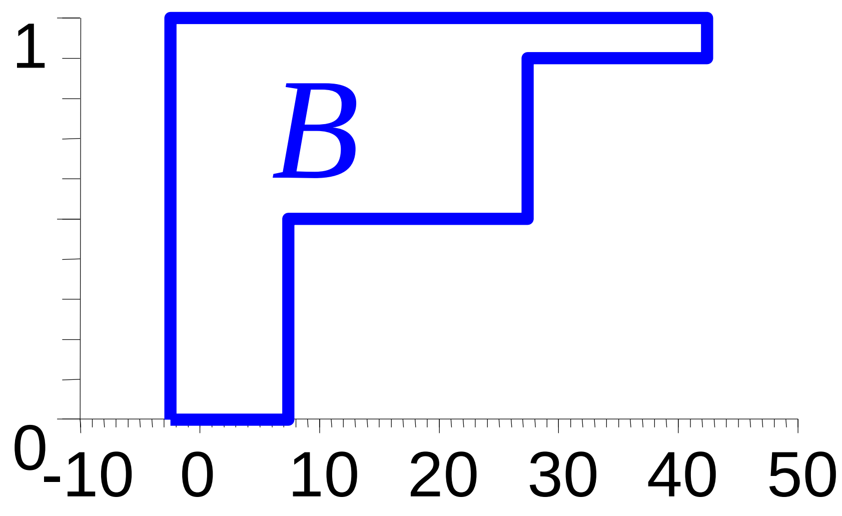
Additive untangling of the kernel for *B* from the equation *C*=*A*+*B*, where p-boxes for *A* and *C* are known. Only the algorithm for the left bound is shown; the right bound is similar.

```
N = number of percentiles (%ile) used in the representation;
left limit on B at lowest %ile = left limit on C at lowest %ile – left limit on A at lowest %ile;
for i = 1 to N do begin
     set flag to "not done";
      for j = 0 to i–1 do begin
         if (left limit on C at ith %ile)≤(left limit on A at [i–j]th %ile)+(left limit on B at the jth %ile)
             then set flag to "done";
         if flag is "done"
             then left limit on B at ith %ile = left limit on B at [i–1]th %ile {the one right below it}
            else left limit on B at ith %ile = left limit on C at ith %ile – left limit on A at 0th %ile;
          end; {for j}
      end; {for i, left bound}
```

Getting the answer

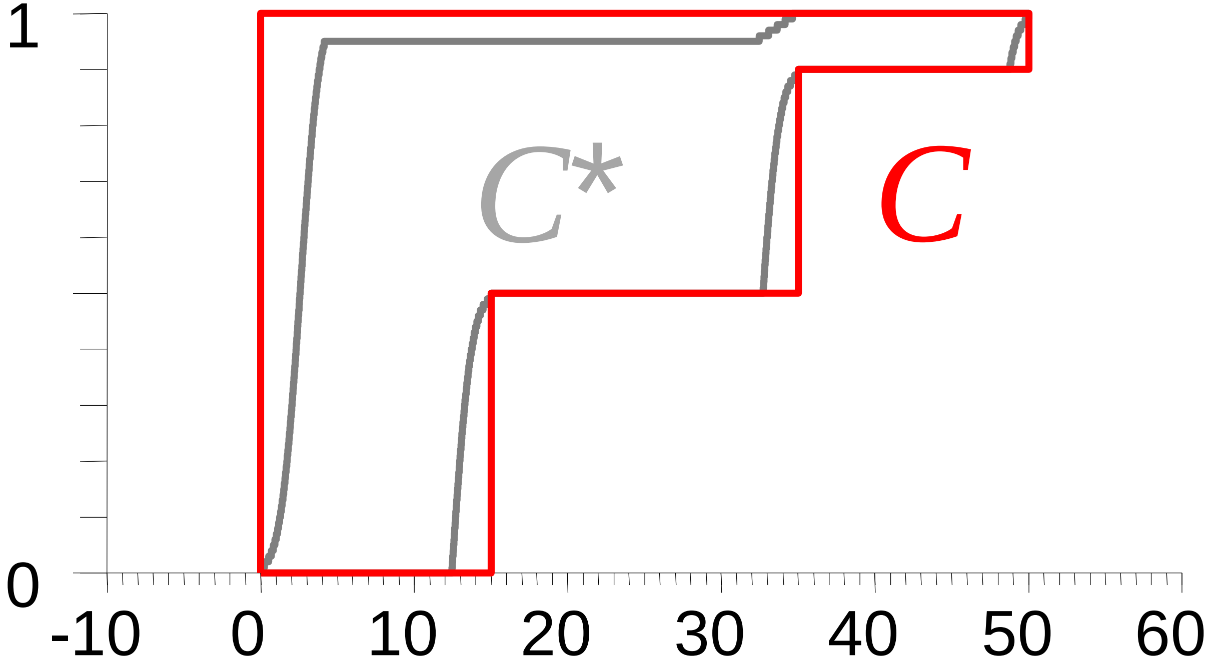
- The backcalculation algorithm basically reverses the forward convolution

- Any distribution totally inside *B* is sure to satisfy the constraint … it’s a “kernel”



Check it by plugging it back in

$A + B = C^* \subseteq C$



Precise distributions don’t work

- Precise distributions can’t express the target
- A concentration distribution giving a prescribed distribution of doses seems to say we *want* some doses to be high
- But any distribution to the left would be better
- A p-box on the dose target expresses this idea

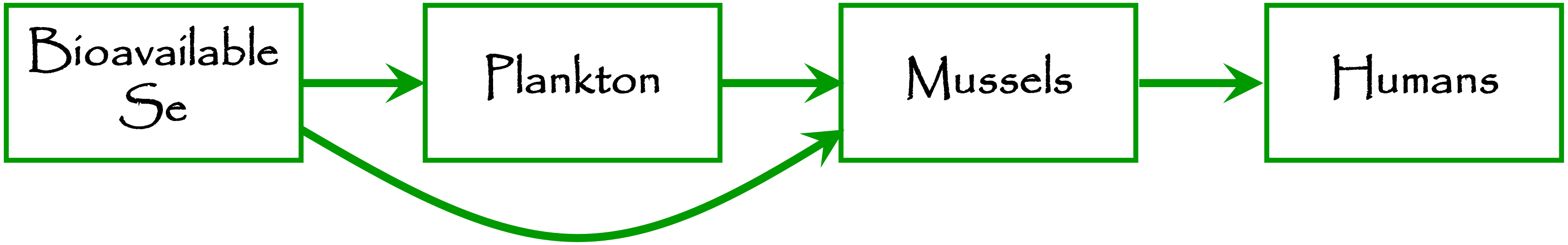
Backcalculation algebra

- Can define untanglings for all basic operations
  e.g., if $A \times B = C$, then $B = \exp(\text{backcalc}(\ln A, \ln C))$
- Can chain them together for big problems
- Assuming independence *widens* the result
- Repeated terms need special strategies

Se concentration in SF Bay mussels

San Francisco Bay blue mussels *Mytilus edulis* are contaminated with the heavy metal selenium (Se)

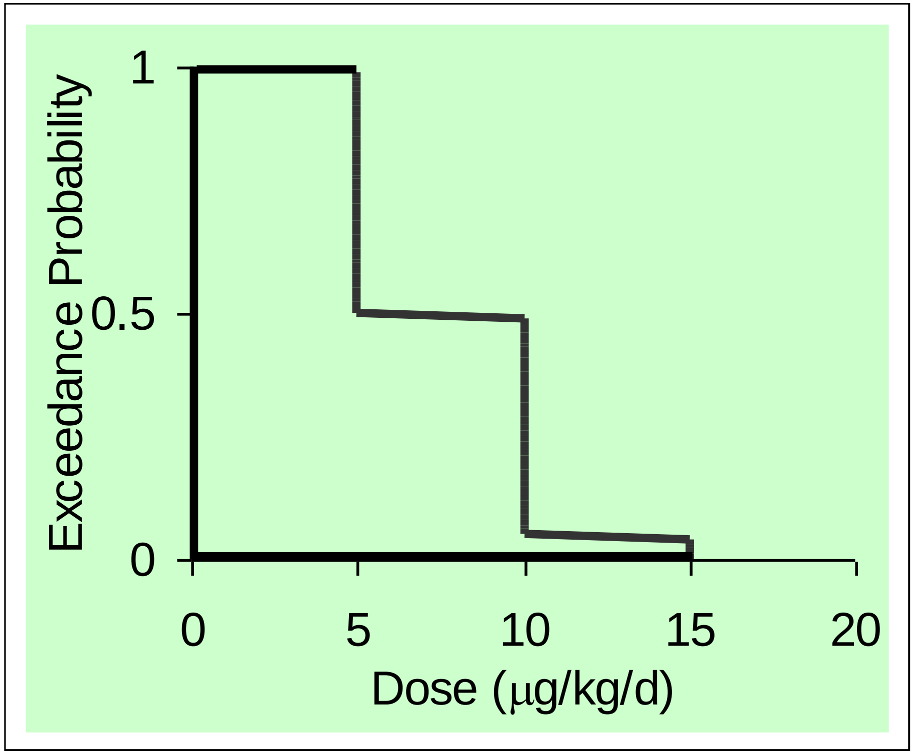
What concentration [Se] in the bay is safe for humans given that it concentrates in food chains



Tolerable Se doses for humans

NOAEL = 15 mg/kg/d (ATSDR 1996)

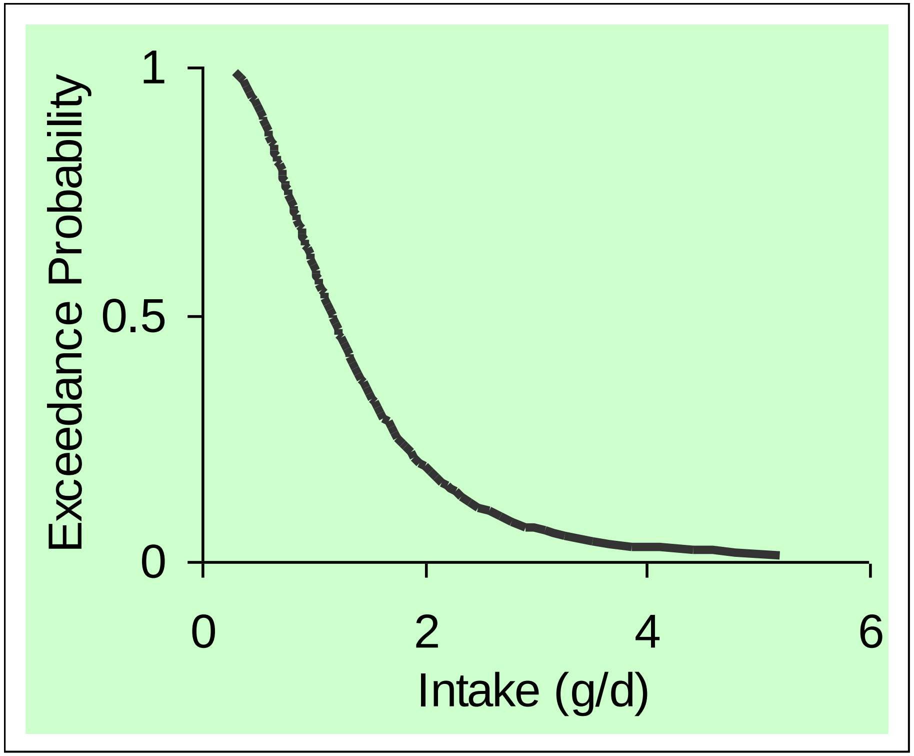
- Median exposures no greater than 5 mg/kg/d
- 95th %tile exposure no greater than 10 mg/kg/d
- Maximum exposure of 15 mg/kg/d



Any dose distribution entirely in this box meets the constraints

Mussel consumption

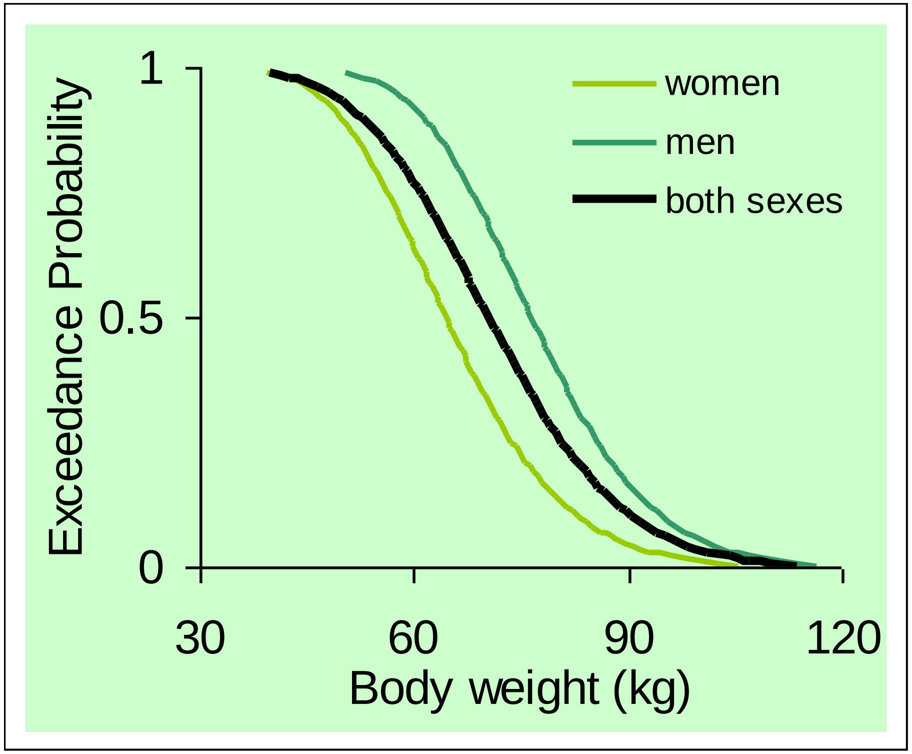
- Data on mussel consumption by Americans were collected by the U.S. National Marine Fisheries Service in a nationwide survey between 1973 and 1974 (Rupp et al. 1980). These data were re-analyzed by Ruffle et al. (1994) and fit to lognormal distributions. Genders are combined for adults aged 19-98, and the consumption data is divided into U.S. regions. We chose the data for the Pacific region (WA, OR, CA, AK, and HI)
- The data are very old, but the best available. Ruffle et al. suggest adding 0.22 to the mean intake rate to account for increased shellfish consumption since 1974, and the intake distribution is then lognormal with mean (1.177+0.22) and standard deviation 0.938 g/d.



Should this have been a p-box too?

Human body mass

- Brainard and Burmaster (1992) provide body weight distributions for men and women aged 18-74 using data in the NHANES II data set (collected between 1976 and 1980).



Use “factor” to backcalculate

$$Dose = WConc \frac{(k + h \times B \times g \times p)}{(l + f \times N \times g \times p)} \times \frac{Intake}{BW}$$

$$WConc = factor\left(\frac{(k + h \times B \times g \times p)}{(l + f \times N \times g \times p)} \times \frac{Intake}{BW}, Dose \right)$$

The factor algorithm untangles the convolutions

- Finds bounds on *WConc* distributions leading to safe doses
- Repeated terms (*g* and *p*) need “region-growing” strategy
- Result has thickest possible tails

Factor algorithm

Multiplicative untangling of the kernel for *B* from the equation *C*=*AB*, where p-boxes for *A* and *C* are known. Only the algorithm for the left bound is shown; the right bound is similar.

N = number of percentiles (%ile) used in the representation;

left limit on B at lowest %ile = left limit on C at lowest %ile / left limit on A at lowest %ile;

for i = 1 to N do begin

    set flag to "not done";

    for j = 0 to i–1 do begin

        if (left limit on C at i^{th} %ile)≤(left limit on A at $[i–j]^{th}$ %ile)×(left limit on B at the j^{th} %ile)

            then set flag to "done";

        if flag is "done"

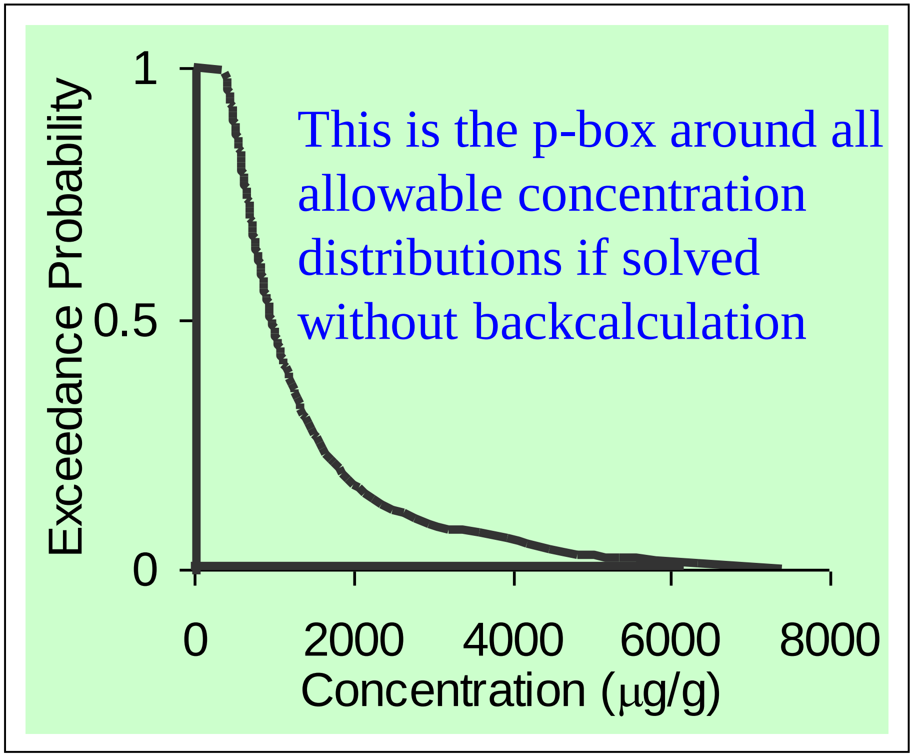
            then left limit on B at i^{th} %ile = left limit on B at $[i–1]^{th}$ %ile {the one right below it}

            else left limit on B at i^{th} %ile = left limit on C at i^{th} %ile / left limit on A at 0^{th} %ile;

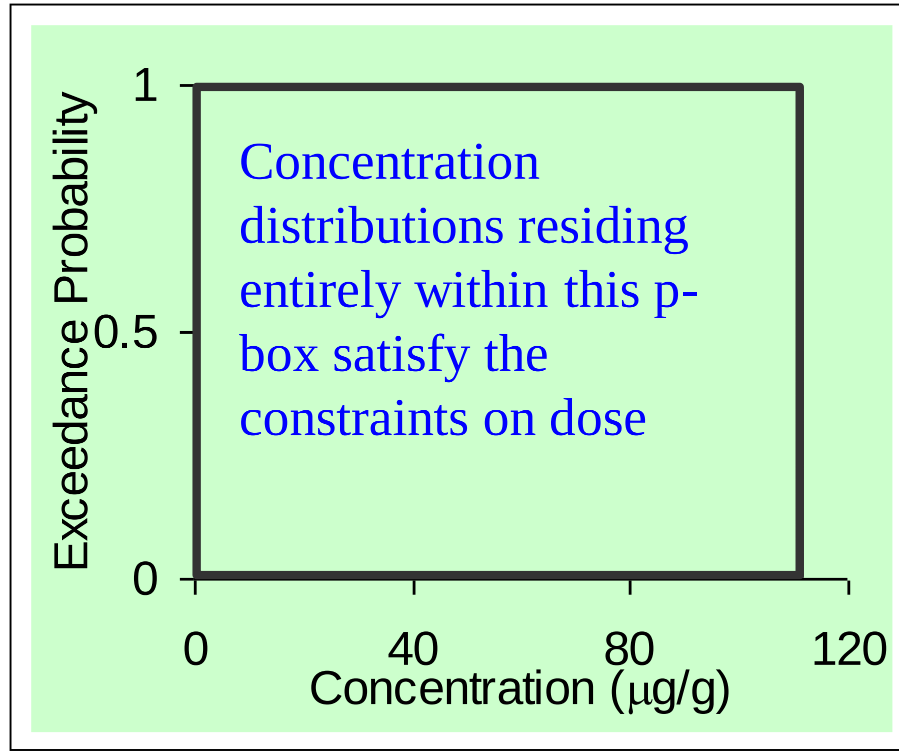
        end; {for j}

    end; {for i, left bound}

Results

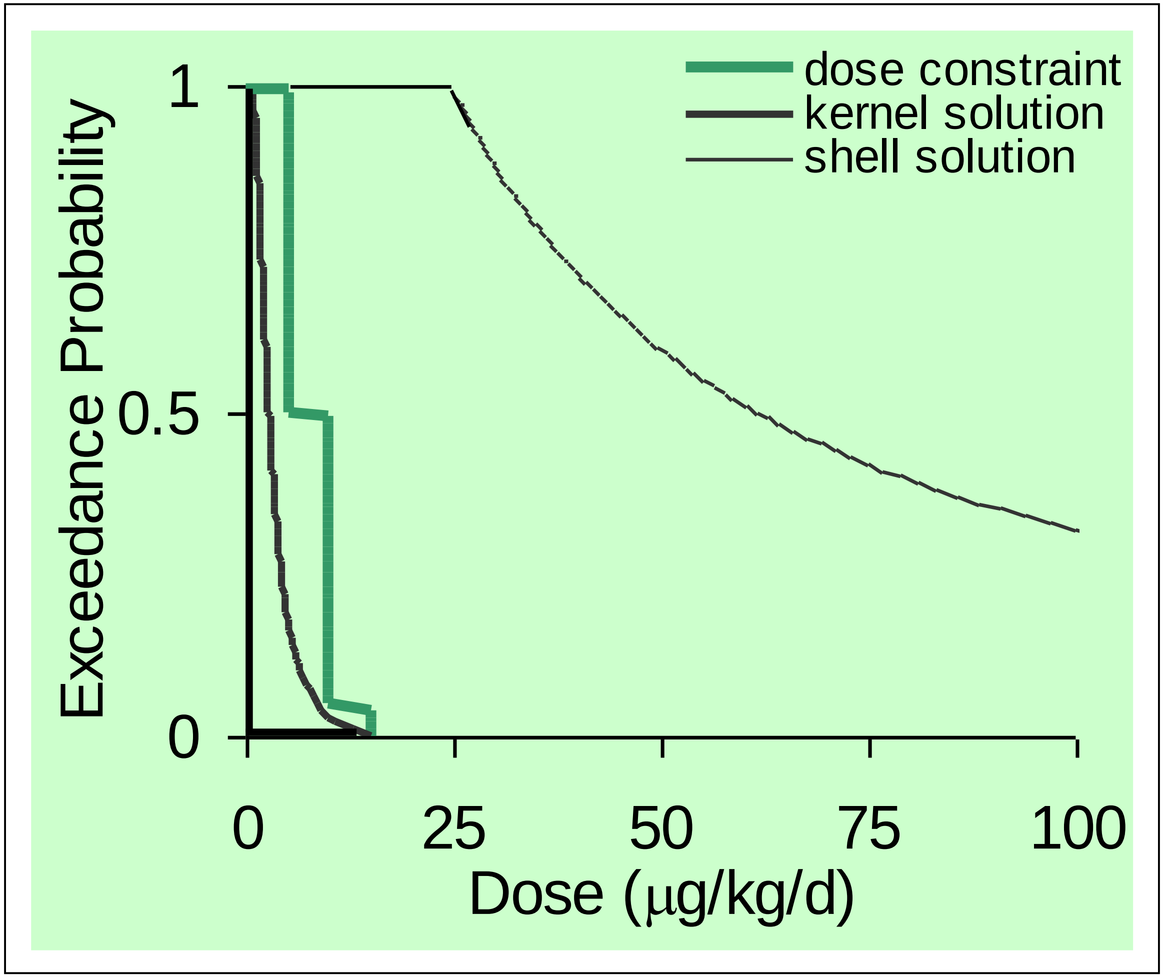


1
0.5
0
Exceedance Probability
0
2000
4000
6000
8000
Concentration (μg/g)
This is the p-box around all allowable concentration distributions if solved without backcalculation



1
0.5
0
Exceedance Probability
0
40
80
120
Concentration (μg/g)
Concentration distributions residing entirely within this p-box satisfy the constraints on dose

Check

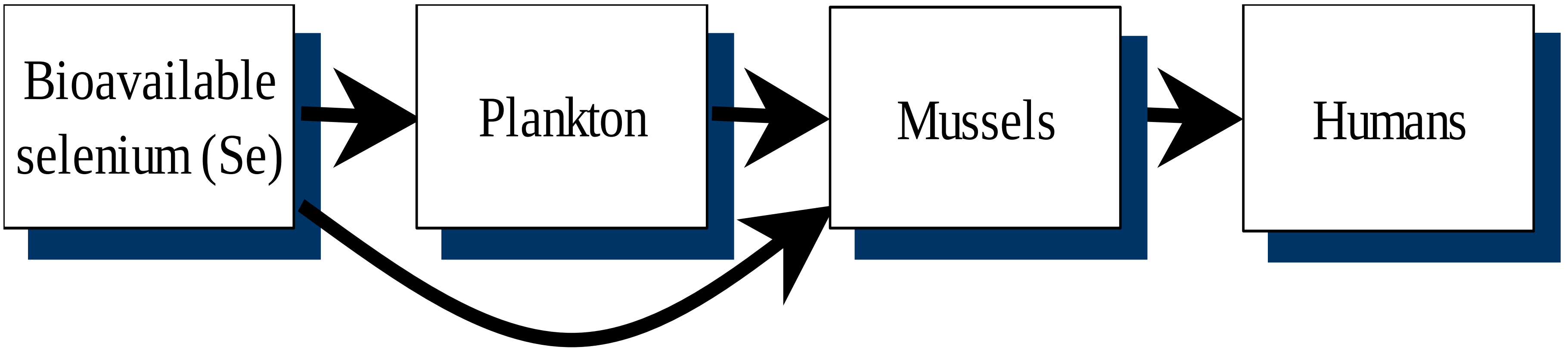


dose constraint
kernel solution
shell solution
Exceedance Probability
1
0.5
0
0
25
50
75
100
Dose (μg/kg/d)

Now what?

- So we know how low the Se concentrations in mussels have to be to be safe
- What does that imply about concentrations for the water in the bay?

Multiple pathways



Untangling the trophic chain

- Steady-state model for the trophic chain (Spencer et al. 2001)
- Assumes all parameters, including [Se] in water, are at equilibrium
- Needs <u>two backcalculations</u>
  - First determines the p-box around the set of Se concentration distributions in edible mussel tissue that are protective of human health
  - Second derives the p-box around the set of distributions of Se concentration in water that result in concentrations of Se in edible mussel tissue meeting the constraint

Steady-state fate/transport model

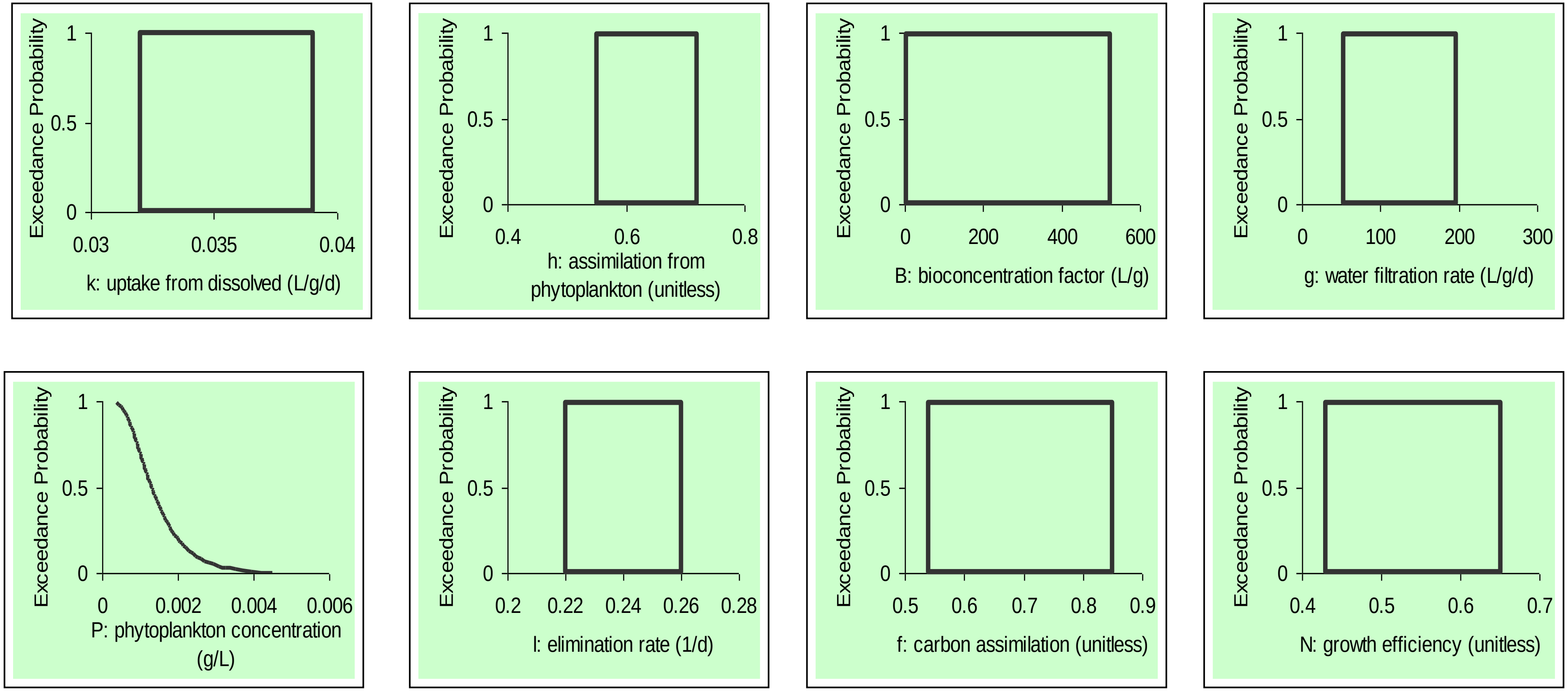
$$Dose = MConc \times \frac{Intake}{BW} \qquad MConc = WConc \frac{(k + h \times B \times g \times p)}{(l + f \times N \times g \times p)}$$

We want to backcalculate

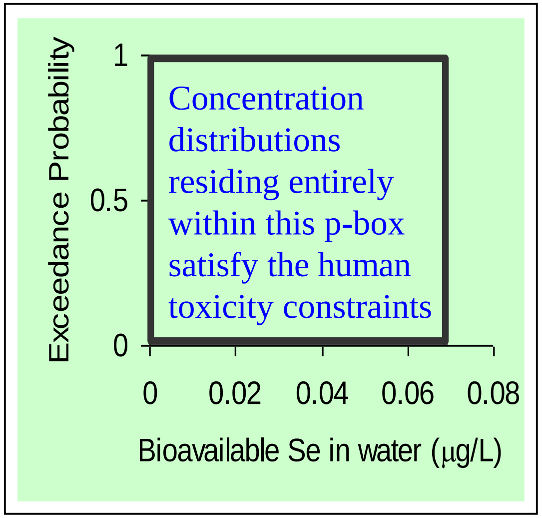
Mconc = Se concentration in mussels (μg/g)
Wconc = Se concentration in bay water (μg/L)
Intake = mussel intake rate (g/d)
BW = adult body weight (kg)
k = mussel Se uptake rate from the dissolved phase (L/g/d)
h = mussel Se assimilation from phytoplankton (unitless)
B = Se bioconcentration factor in phytoplankton (L/g)
g = mussel water filtration rate (L/g/d)
p = phytoplankton concentration (g/L)
l = mussel selenium elimination rate (d^{-1})
f = mussel carbon assimilation efficiency (unitless)
N = mussel net growth efficiency (unitless)

The repeated variables, g and P, were accounted for by subinterval reconstitution

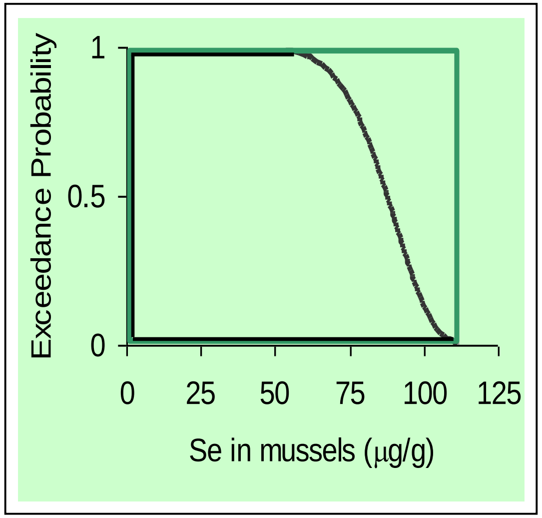
Steady-state parameters



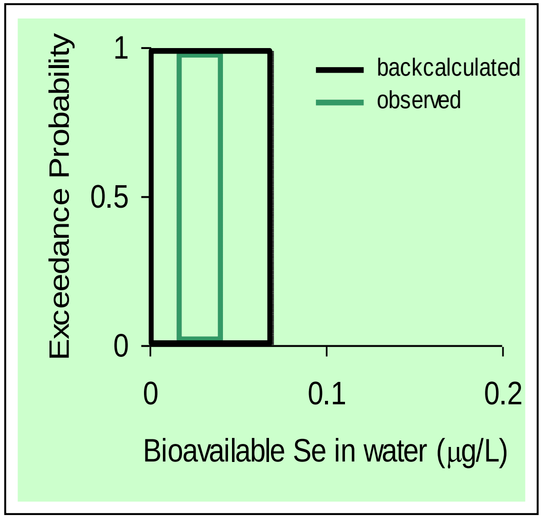
Why is it that *P* is a precise distribution?



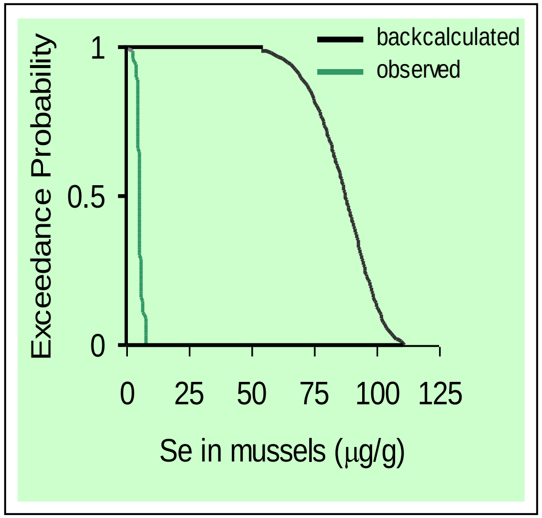
Backcalculated allowable concentration of Se in water.



Concentration of Se in mussels given Se in water (at left). Allowable concentration in green.



Se in water, San Francisco Bay



Se in mussels, San Francisco Bay

So what?

- Many distributions drawn from the shell are not protective of human health
- Algorithms can backcalculate the kernel for both simple and complex exposure models
- The bioavailable Se concentration in San Francisco Bay is well within the kernel of tolerable concentration distributions, as is the Se concentration in edible mussel tissue

Conclusion

- Engineering design requires backcalculation
- Monte Carlo methods don’t generally work except in a trial-and-error approach
- Can express the constraint target as a p-box, although this is awkward in a precise distribution

Not the same as deconvolution

Deconvolution is a related procedure for equations involving uncertainty

Deconvolution is used, for instance, to improve the estimate of the distribution of *X* given a distributions of measurements *Y* and measurement error ε

$$Y = X + \varepsilon$$

Variance of *X* should be smaller than that of *Y*

Deconvolution doesn't satisfy a constraint like backcalculation does

Se/mussel references

- Agency for Toxic Substances and Disease Registry (ATSDR). 1993. Toxicological profile for chromium. U.S. Dept. of Health and Human Services, Public Health Service, Atlanta, GA.
- Brainard, J. and D.E. Burmaster. 1992. Bivariate distributions for height and weight of men and women in the United States. Risk Analysis 12: 267–275.
- Burmaster, D.E. and R.H. Harris. 1993. The magnitude of compounding conservatisms in Superfund risk assessments. Risk Analysis 13: 131.
- Karstadt, M. 1988. Quantitative risk assessment: qualms and questions. Teratogenesis, Carcinogenesis, and Mutagenesis 8:137.
- Longnecker, M.P. P.R. Taylor, O.A. Levander et al. 1991. Selenium in diet, blood and toenails in relation to human health in a seleniferous area. Am. J. Clin. Nutr. 53(5): 1288-1294.
- McKone, T.E. and P.B. Ryan. 1989. Human exposures to chemicals through food chains: an uncertainty analysis. Environmental Science and Technology 23: 1154–1163.
- NS&T Program, Mussel Watch Project. No date. http://ccmaserver.nos.noaa.gov/NSandT/New_NSandT.html.
- Roberts, L. 1990. Risk assessors taken to task. Science 248: 1173.
- Ruffle, B., D.E. Burmaster, P.D. Anderson, and H.D. Gordon. 1994. Lognormal distribution for fish consumption by the general U.S. population. Risk Analysis: 14: 395–404.
- Spencer, M., N.S. Fisher, W. Wang and S. Ferson. 2001. Temporal variability and ignorance in Monte Carlo contaminant bioaccumulation models: a case study with selenium in Mytilus edilus. Risk Analysis 21: 383-394.
- von Stackelberg, K.E., D. Burmistrov, D.J. Vorhees, and T.S. Bridges. 2002. Importance of uncertainty and variability to predicted risks from trophic transfer of PCBs in dredged sediments. Risk Analysis 22: 499–512.
- Yang, G., S. Yin, R. Zhou, et al. 1989. Studies of safe maximal daily dietary Se-intake in a seleniferous area in China. II. Relation between Se-intake and the manifestation of clinical signs and certain biochemical alterations in blood and urine. J. Trace Elem. Electrolytes Health Dis. 3(3): 123-130.

Case studies

Case study:

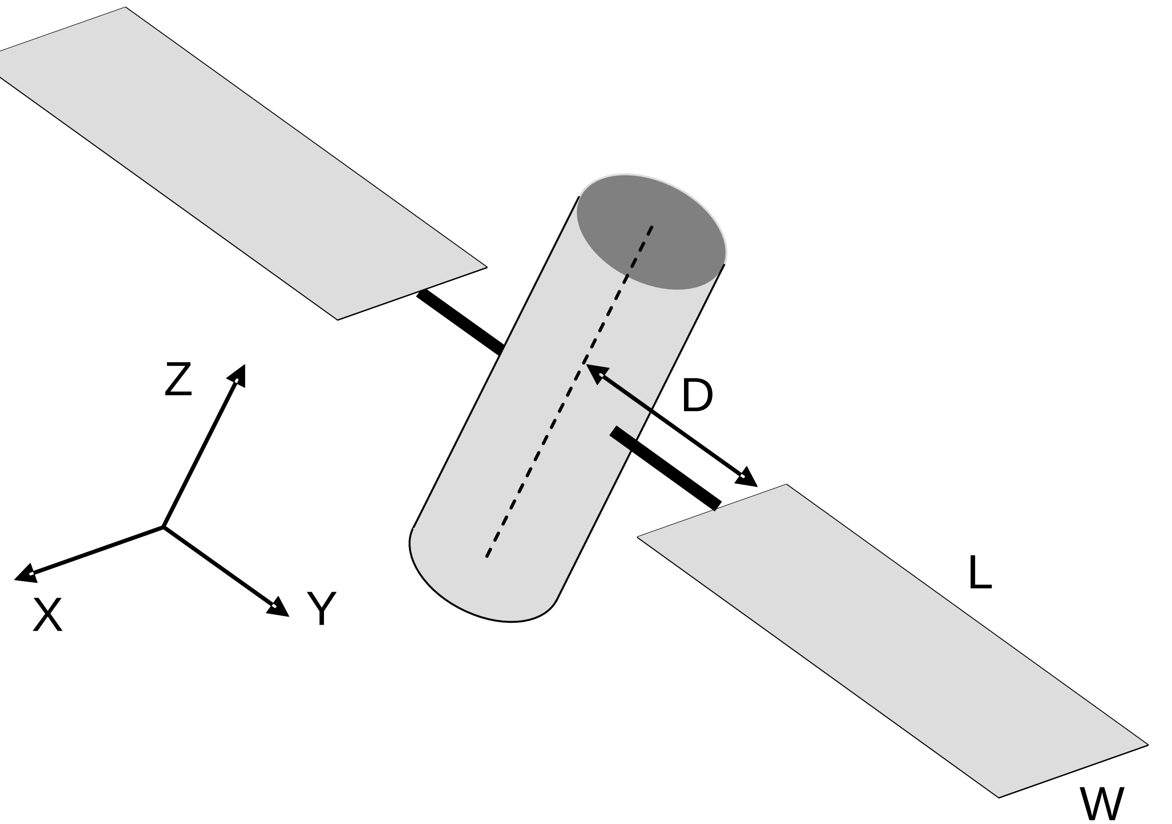
Spacecraft design under mission uncertainty

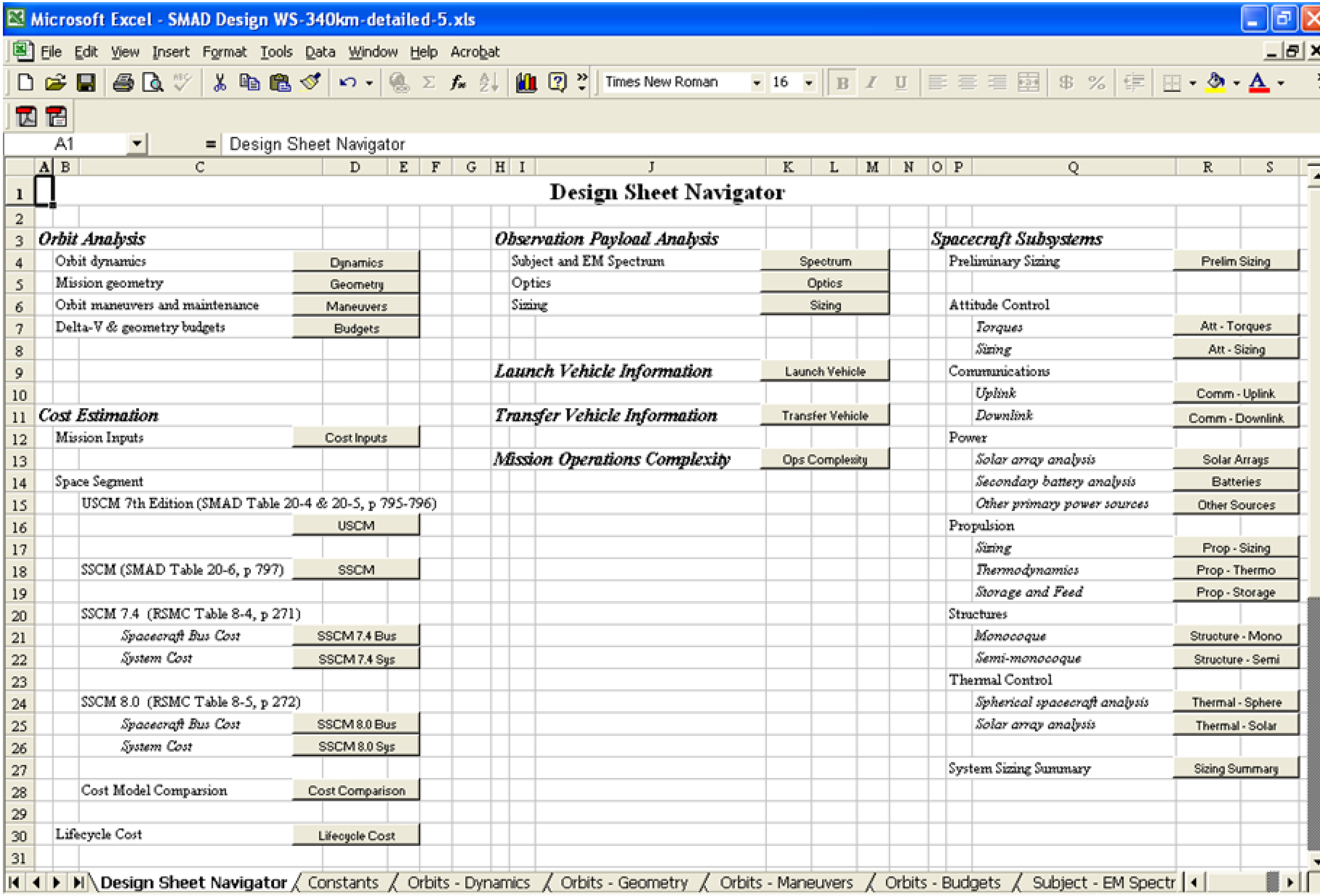
Integrated concurrent engineering

- Real-time collaborative, interactive process
  - Implemented at JPL, LaRC, and others
- Reduces design time by an order of magnitude
  - But quantitative risk assessment is difficult
- Design solutions are iterative
  - So a Monte Carlo approach may not be practical

Mission

Deploy satellite carrying a large optical sensor





Wertz and Larson (1999) *Space Mission Analysis and Design* (***SMAD***). Kluwer.

Microsoft Excel - SMAD Design WS-340km-detailed-5.xls

File Edit View Insert Format Tools Data Window Help Acrobat

Times New Roman 10

M17 =

Attitude Control - Torque Estimates

Return to Navigator

(All information on this sheet is contained in the block from Cell A1 to Cell Q27)

Orbit characteristics					
Altitude		340.000	km		
Satellite velocity		7.703	km/s		
Atmospheric density		1.983E-11	kg/m^3		
Sun incidence angle		0.00	deg		
Maximum deviation from local vertical		10.00	deg		
Spacecraft characteristics					
Largest moment of inertia	*7315.000*	7315.000	kg-m^2		
Smallest moment of inertia	*4655.000*	4655.000	kg-m^2		
Projected surface area	*14.063*	14.063	m^2		
Moment arm for solar radiation torques		0.250	m		
Moment arm for aerodynamic torques		0.250	m		
Drag coefficient		3.13			
Surface reflectivity		0.60			
Residual dipole		1.00	A-m^2		
Slew characteristics					
Maximum slewing angle	*38.00*	38.00	deg		
Minimum maneuver time	*760.00*	760.00	sec		

Environmental torques		
Gravity gradient	1.794E-03	N-m
Solar radiation	2.565E-05	N-m
Magnetic	5.250E-05	N-m
Aerodynamic	6.473E-03	N-m
Total (RSS)	6.717E-03	N-m
Slewing torque	3.360E-02	N-m

Obs Payload - Optics / Obs Payload - Sizing / SC Bus - Preliminary Sizing / **Attitude Control - Torques** / Attitude Control - Sizing / Comm

Typical subsystems

Attitude control

Command data systems

Configuration

Cost

Ground systems

Instruments

Mission design

Power

Program management

Propulsion

Science

Solar array

Systems engineering

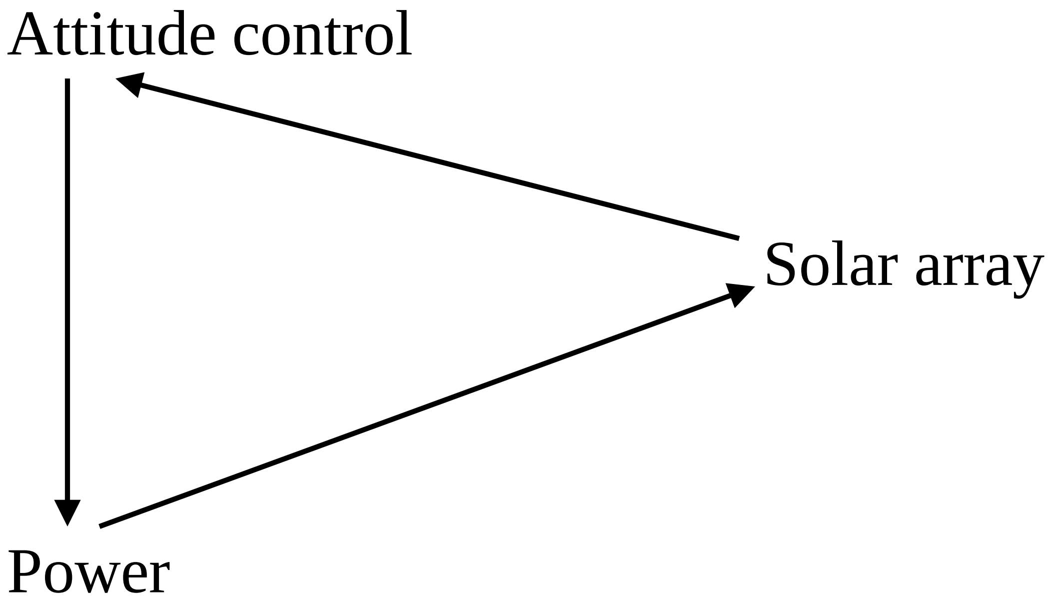
Telecommunications – System

Telecommunications – Hardware

Thermal control

Moments of inertia determine
Torque needed determines
Power requirements determines
Solar panel size determines
Moments of inertia

Demonstration system



- Calculations within a single subsystem (ACS)
- Calculations within linked subsystems

Attitude control subsystem (ACS)

- 3 reaction wheels
- Design problem: solve for *h*
  - Required angular momentum
  - Needed to choose reaction wheels
- Mission constraints
  - Δt_{orbit} = 1/4 orbit time
  - θ_{slew} = max slew angle
  - Δt_{slew} = min maneuver time
- Inputs from other subsystems
  - I, I_{max}, I_{min} = inertial moment
  - Depend on solar panel size, which depends on power needed, so on *h*

$$h = \tau_{\text{tot}} \times \Delta t_{\text{orbit}}$$

$$\tau_{tot} = \tau_{slew} + \tau_{dist}$$

$$\tau_{\text{slew}} = \frac{4\theta_{\text{slew}}}{\Delta t_{\text{slew}}^{2}} I$$

$$\tau_{dist} = \tau_{g} + \tau_{sp} + \tau_{m} + \tau_{a}$$

$$\tau_{\text{g}} = \frac{3\mu}{2(R_{\text{E}} + H)^{3}} |I_{\max} - I_{\min}| \sin(2\theta)$$

$$\tau_{\text{sp}} = L_{\text{sp}} \frac{F_{\text{s}}}{c} A_{\text{s}} (1 + q) \cos(i)$$

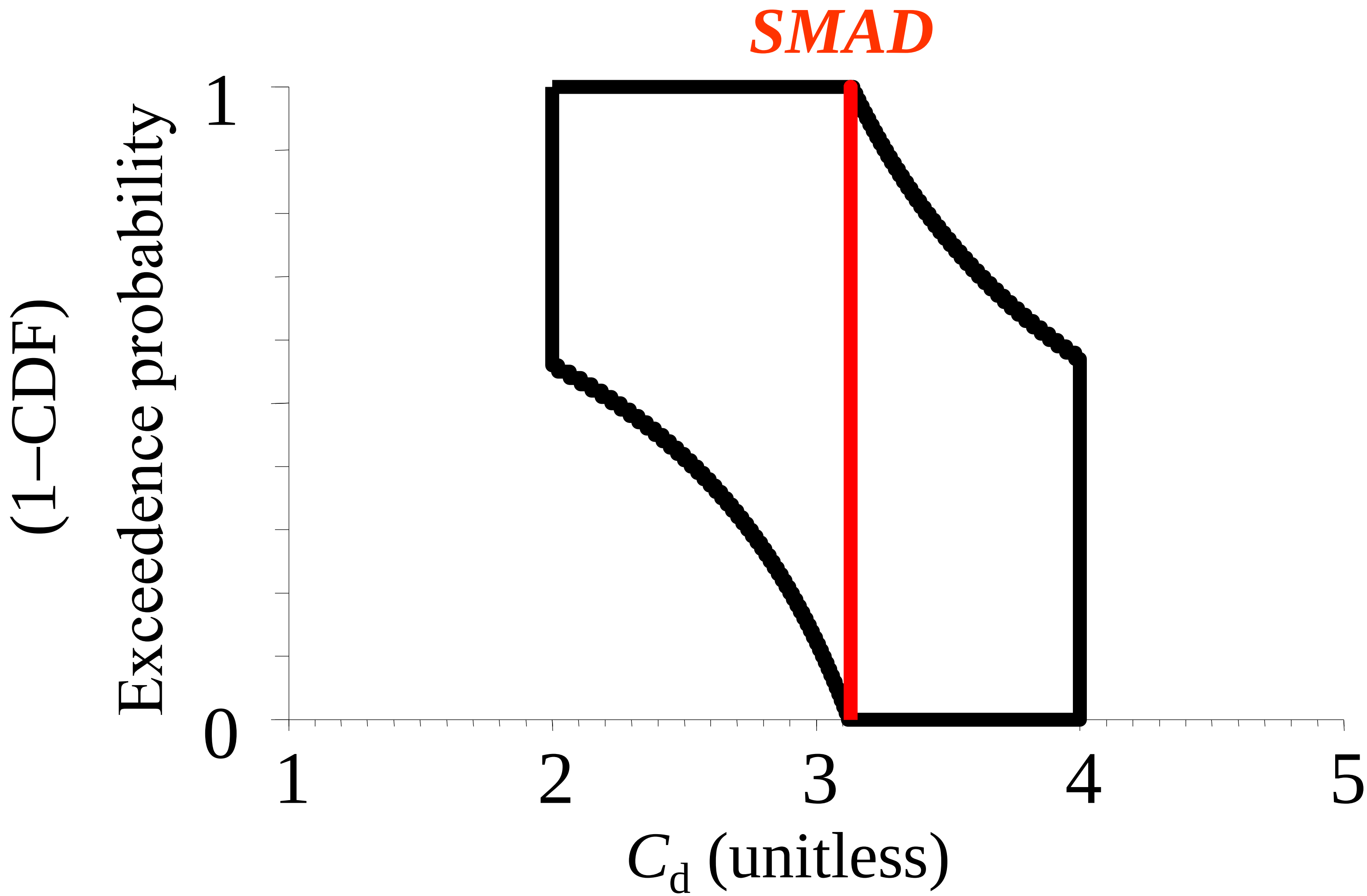
$$\tau_{\text{m}} = \frac{2MD}{(R_{\text{E}} + H)^{3}}$$

$$\tau_{\text{a}} = \frac{1}{2} L_{\text{a}} \rho C_{\text{d}} A V^{2}$$

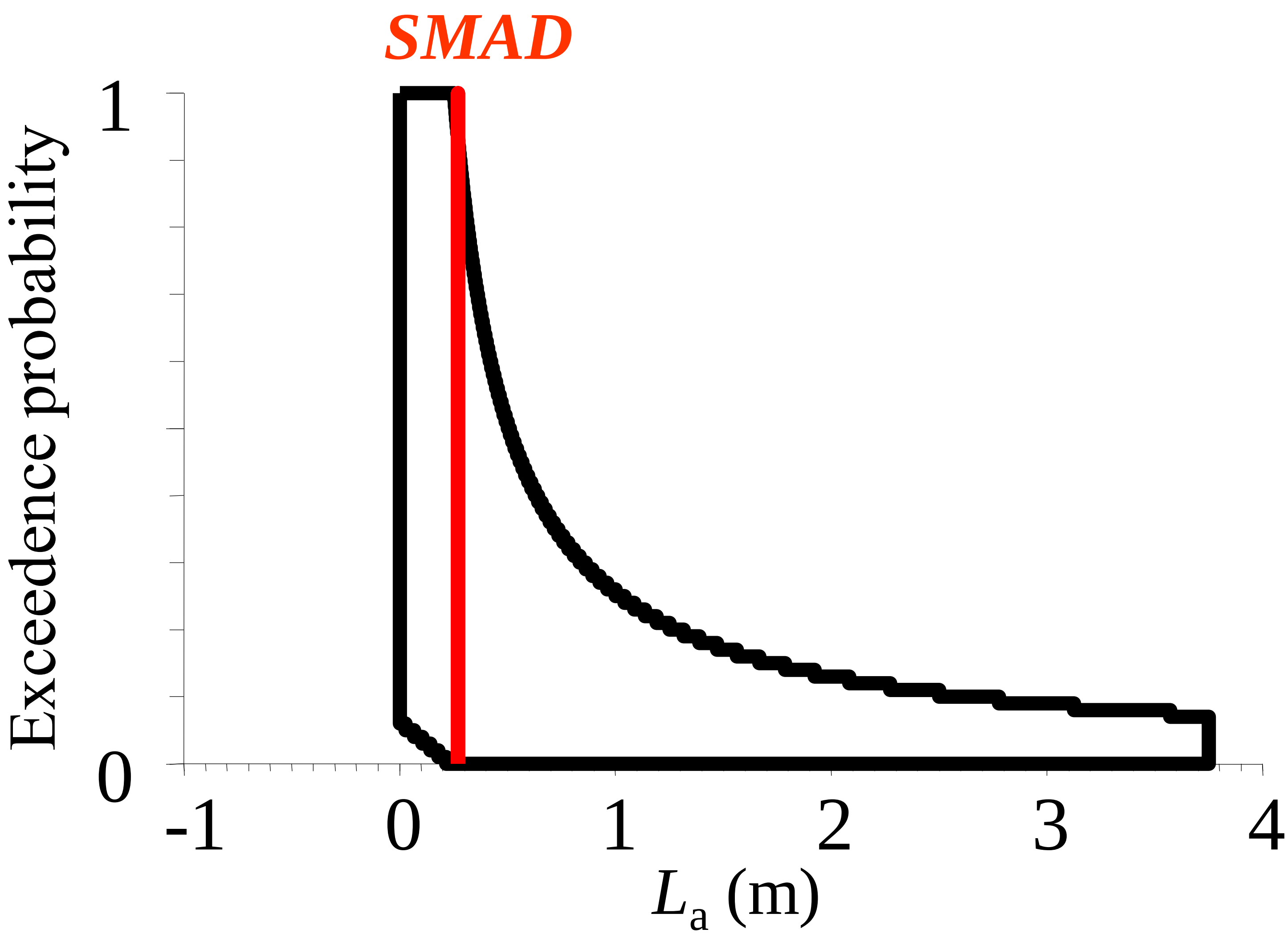
Attitude control input variables

Symbol	Unit	Variable	Type	Value	*SMAD*
C_d	unitless	Drag coefficient	p-box	range=[2,4] mean=3.13	3.13
L_a	m	Aerodynamic drag torque moment	p-box	range=[0,3.75] mean=0.25	0.25
L_{sp}	m	Solar radiation torque moment	p-box	range=[0,3.75] mean=[0.25]	0.25
D	A m^2	Residual dipole	interval	[0,1]	1
i	degrees	Sun incidence angle	interval	[0,90]	0
ρ	kg m^3	Atmospheric density	interval	[3.96e-12, 9.9e-11]	1.98e-11
θ	degrees	Major moment axis deviation from nadir	interval	[10,19]	10
q	unitless	Surface reflectivity	interval	[0.1,0.99]	0.6
I_{min}	kg m^2	Minimum moment of inertia	interval	[4655]	4655
I_{max}	kg m^2	Maximum moment of inertia	interval	[7315]	7315
μ	m^3 s^{-2}	Earth gravity constant	point	3.98e14	3.98e14
A	m^2	Area in the direction of flight	point	3.75^2	3.75^2
RE	km	Earth radius	point	6378.14	6378.14
H	km	Orbit altitude	point	340	340
F_s	W m^{-2}	Average solar flux	point	1367	1367
θ_{slew}	degrees	Maximum slewing angle	point	38	38
c	m s^{-1}	Light speed	point	2.9979e8	2.9979e8
M	A m^2	Earth magnetic moment	point	7.96e22	7.96e22
Δt_{slew}	s	Minimum maneuver time	point	760	760
A_s	m^2	Area reflecting solar radiation	point	3.75^2	3.75^2
Δt_{orbit}	s	Quarter orbit period	point	1370	1370

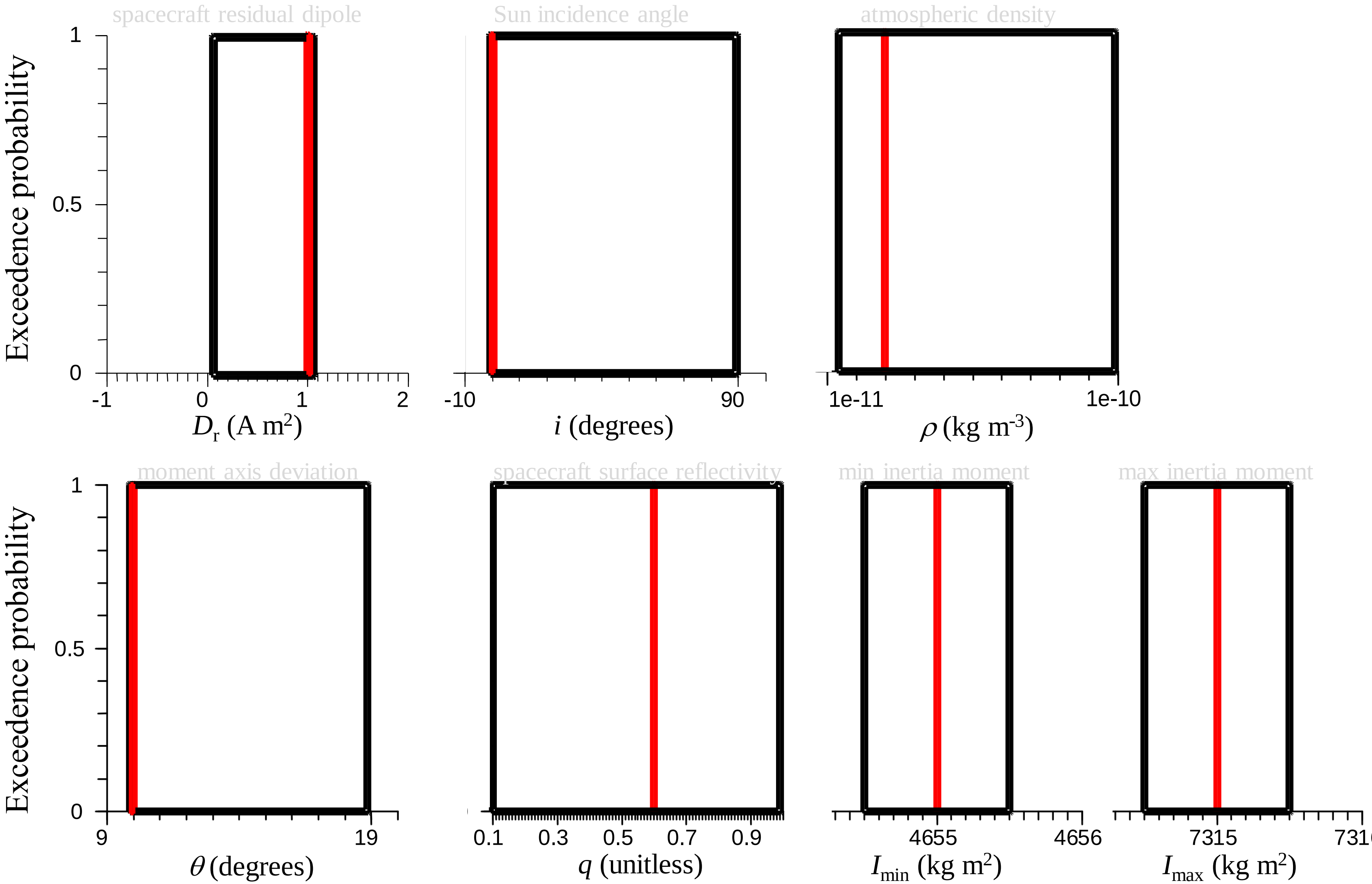
Coefficient of drag, C_d



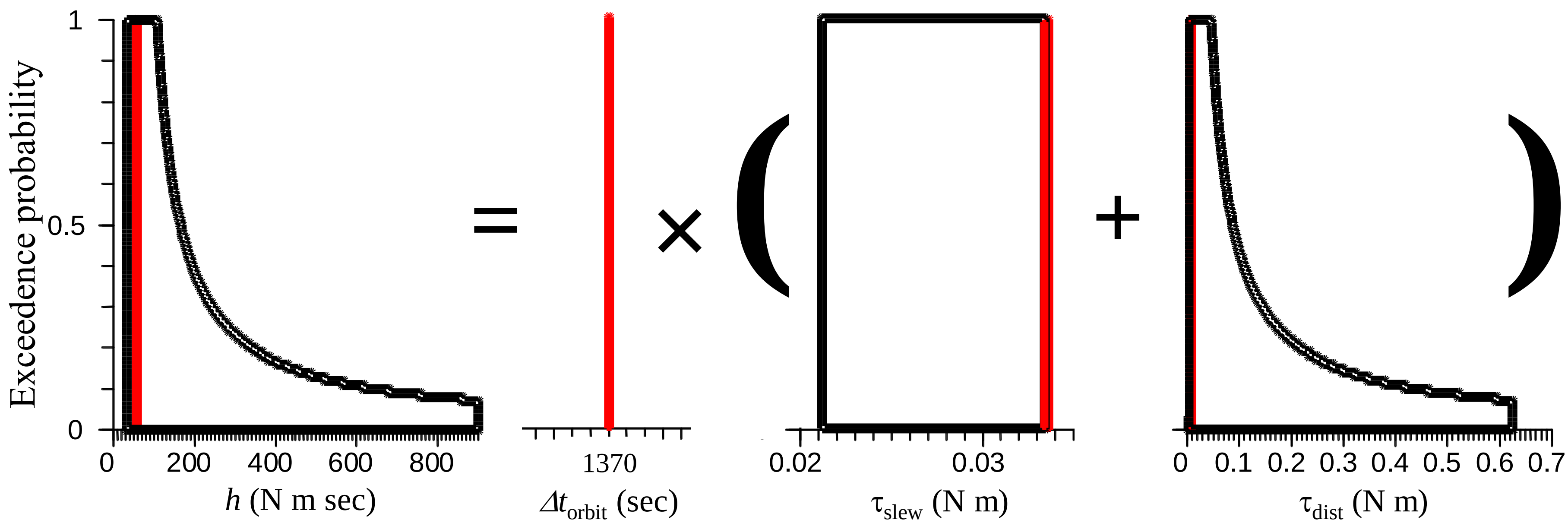
Aerodynamic drag torque moment, L_a



Interval inputs and *SMAD* points

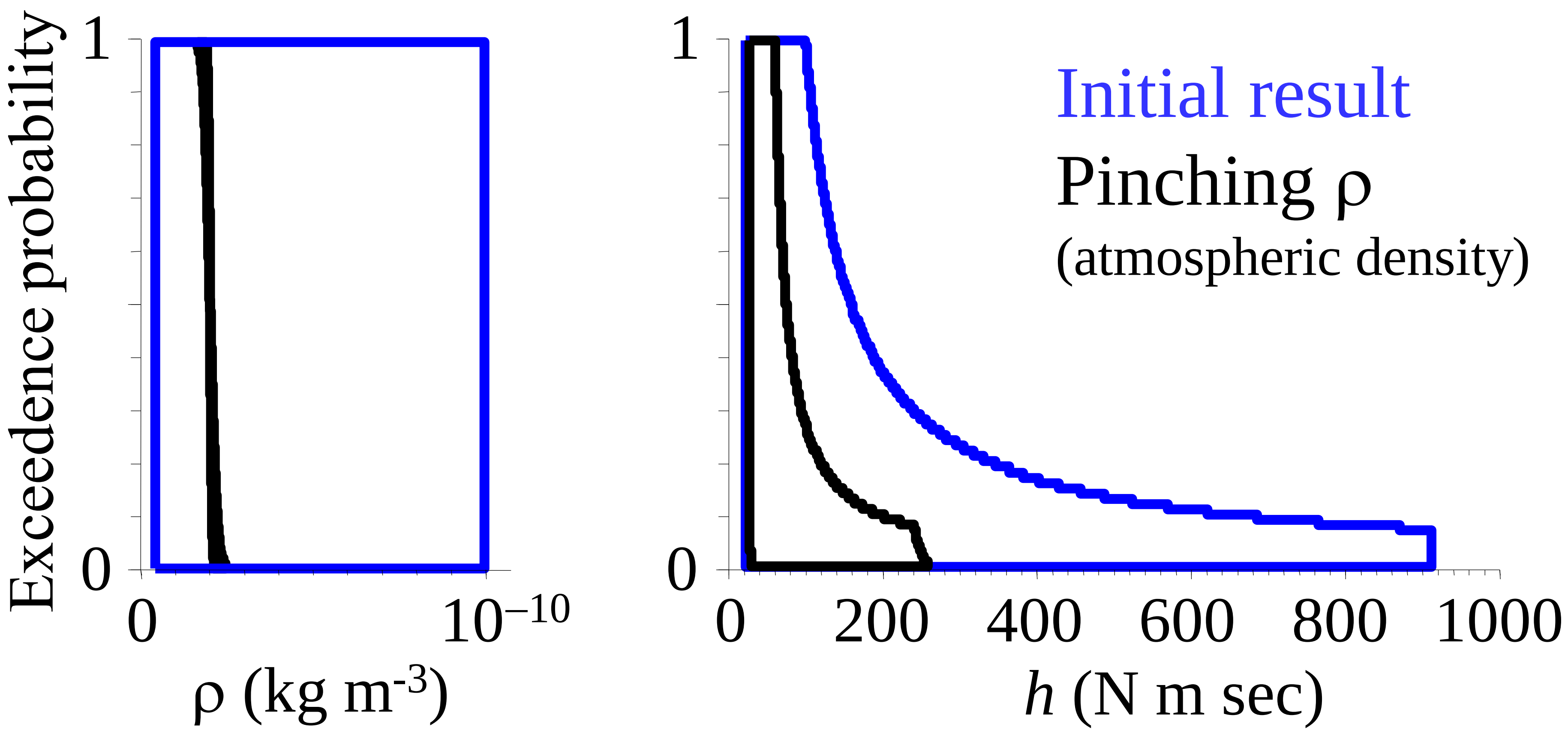


Required angular momentum, *h*



$$h = \Delta t_{orbit} \times \left(\tau_{slew} + \tau_{dist} \right)$$

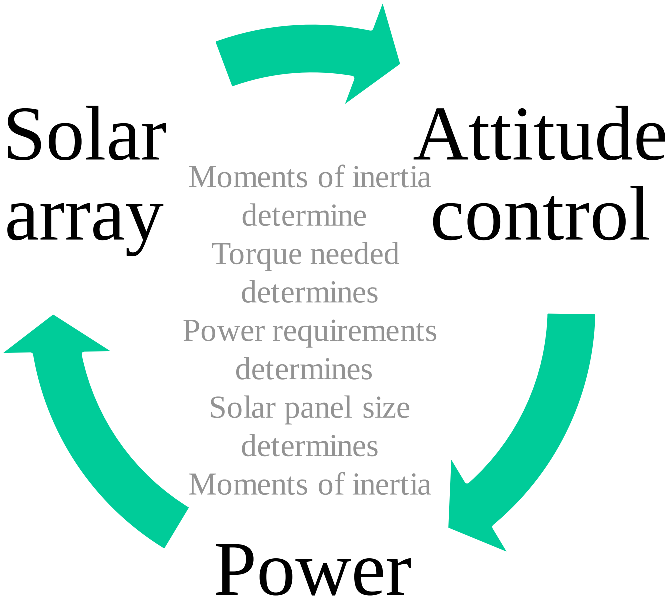
Value of information: pinching ρ



Linked subsystems

- Minimum moment of inertia I_{min}
- Maximum moment of inertia I_{max}
- Total torque τ_{tot}
- Total power P_{tot}
- Solar panel area A_{sa}

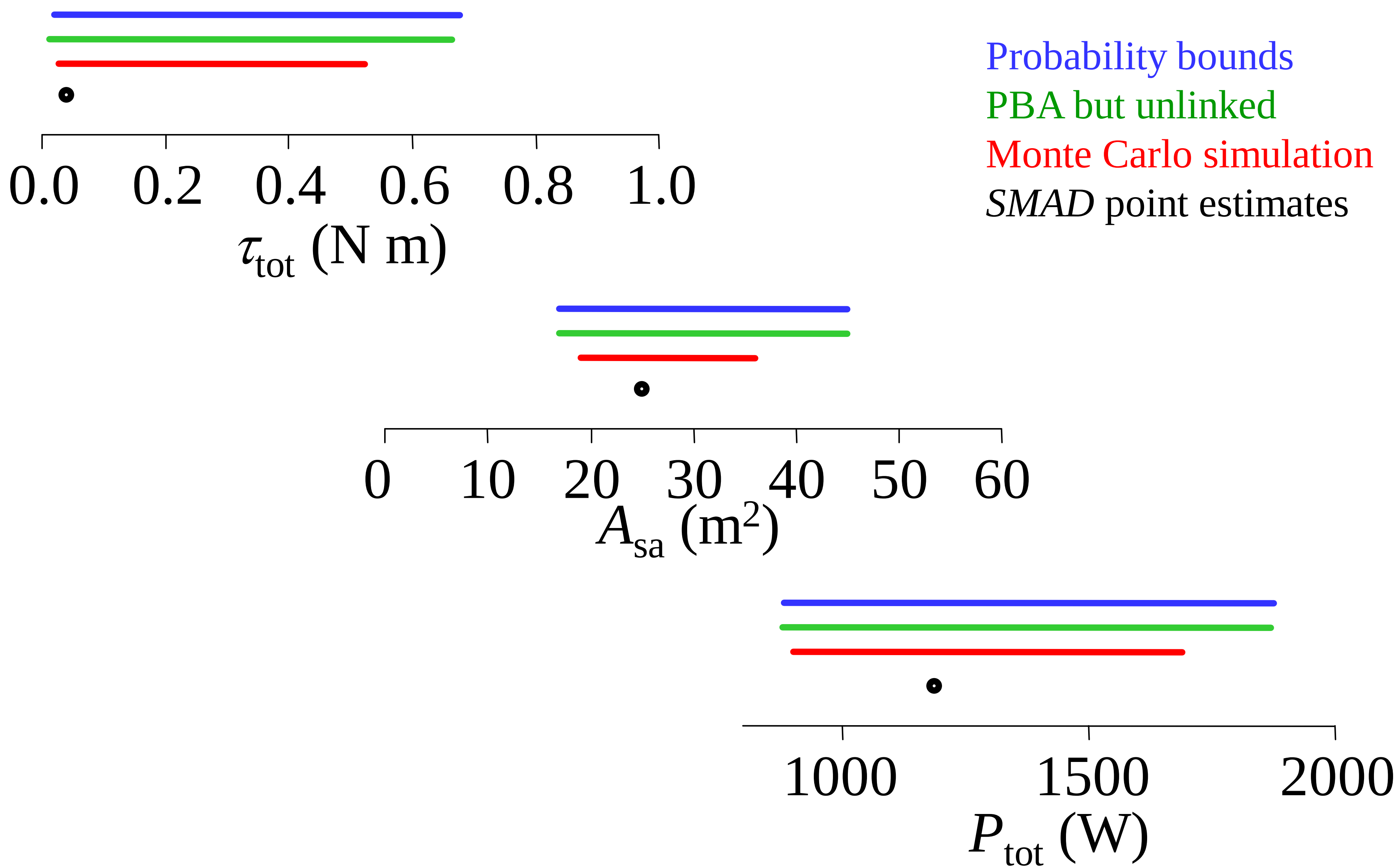
Iteratively calculated



Analysis of calculations

- Do p-boxes enclose Monte Carlo and *SMAD*?
- Does iteration through links cause runaway uncertainty growth (or contraction)?
- Four parallel analyses
  - *SMAD*'s point estimates
  - Monte Carlo simulation
  - P-boxes but without linkage among subsystems
  - P-boxes with fully linked subsystems

Ranges of results



Findings

- Calculations workable
  - No runaway inflation (or loss) of uncertainty
  - Comprehensive bounds easier than via Monte Carlo
- Practical and useful results
  - Uncertainty influences engineering decisions
  - Reducing uncertainty about ρ (by picking a launch date) strongly reduces design uncertainty

Attitude control subsystem

Symbol	Unit	Type	Value	SMAD
C_d	unitless	p-box	range=[2,4] mean=3.13	3.13
L_a	m	p-box	range=[0,3.75] mean=0.25	0.25
L_{sp}	m	p-box	range=[0,3.75] mean=[0.25]	0.25
D	A m^2	interval	[0,1]	1
i	degrees	interval	[0,90]	0
ρ	kg m^3	interval	[3.96e-12, 9.9e-11]	1.98e-11
θ	degrees	interval	[10,19]	10
q	unitless	interval	[0.1,0.99]	0.6
I_{min}	kg m^2	interval	[4655]	4655
I_{max}	kg m^2	interval	[7315]	7315
μ	m^3 s^{-2}	point	3.98e14	3.98e14
A	m^2	point	3.75^2	3.75^2
RE	km	point	6378.14	6378.14
H	km	point	340	340
F_s	W m^{-2}	point	1367	1367
θ_{slew}	degrees	point	38	38
c	m s^{-1}	point	2.9979e8	2.9979e8
M	A m^2	point	7.96e22	7.96e22
Δt_{slew}	s	point	760	760
A_s	m^2	point	3.75^2	3.75^2
Δt_{orbit}	s	point	1370	1370

$$h = \tau_{\text{tot}} \times \Delta t_{\text{orbit}}$$

$$\tau_{tot} = \tau_{slew} + \tau_{dist}$$

$$\tau_{\text{slew}} = \frac{4\theta_{\text{slew}}}{\Delta t_{\text{slew}}^{\;2}} I$$

$$\tau_{dist} = \tau_g + \tau_{sp} + \tau_m + \tau_a$$

$$\tau_{\text{g}} = \frac{3\mu}{2(R_{\text{E}} + H)^3} \left| I_{\max} - I_{\min} \right| \sin(2\theta)$$

$$\tau_{\text{sp}} = L_{\text{sp}} \frac{F_{\text{s}}}{c} A_{\text{s}} (1 + q) \cos(i)$$

$$\tau_{\text{m}} = \frac{2MD}{(R_{\text{E}} + H)^3}$$

$$\tau_{\text{a}} = \frac{1}{2} L_{\text{a}} \rho C_{\text{d}} A V^2$$

Be careful with units!

Neuroscience of risk

Decade of the Brain ('90s)

Psychometry

- Probability and decision theory are rife with paradoxes that no other areas in math have
- Cottage industry in documenting ways in which humans mess up probabilities

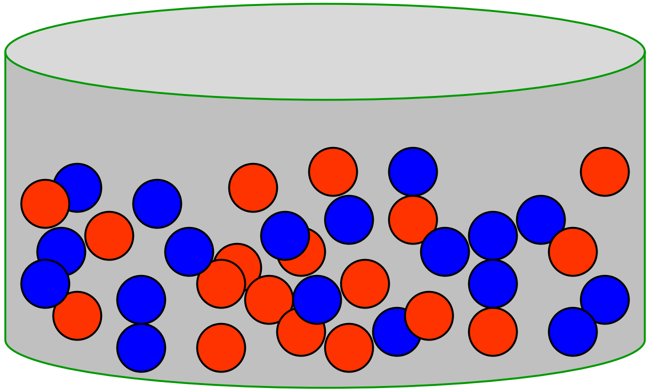
Paradoxes & biases

- Ellsberg paradox
- St. Petersburg paradox
- Two-envelopes problem
- Monty Hall problem
- Simpson’s paradox
- Risk aversion
- Loss aversion

Risk aversion

- Suppose you can get $100 if a randomly drawn ball is red from an urn with half red and half blue balls…or you can just get $50
- Which prize do you want?

$50

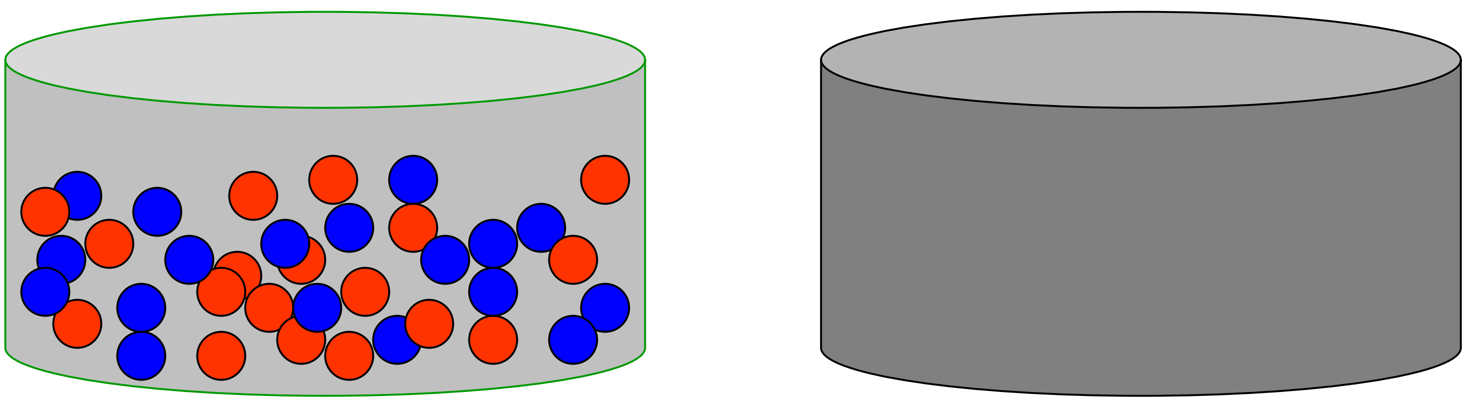


EU is the same, but most people take the sure $50

Ambiguity aversion

Keynes; Dempster

- Balls can be either red or blue
- Two urns, both with 36 balls
- Get $100 if a randomly drawn ball is red
- Which urn do you wanna draw from?



Ellsberg Paradox

- Balls can be red, blue or yellow (probs are R, B, Y)
- A well-mixed urn has 30 red balls and 60 ***other*** balls
- Don't know how many are blue or how many are yellow

Gamble A $R > B$

Get \$100 if draw red

Gamble B

Get \$100 if draw blue

Gamble C

Get \$100 if red or yellow

Gamble D $R < B$

Get \$100 if blue or yellow

Persistent paradox

- People always prefer unambiguous outcomes
  - Doesn’t depend on your utility function or payoff
  - Not related to risk aversion
  - We simply don’t like ambiguity

- Not explained by probability theory, or by prospect theory

Ambiguity (incertitude)

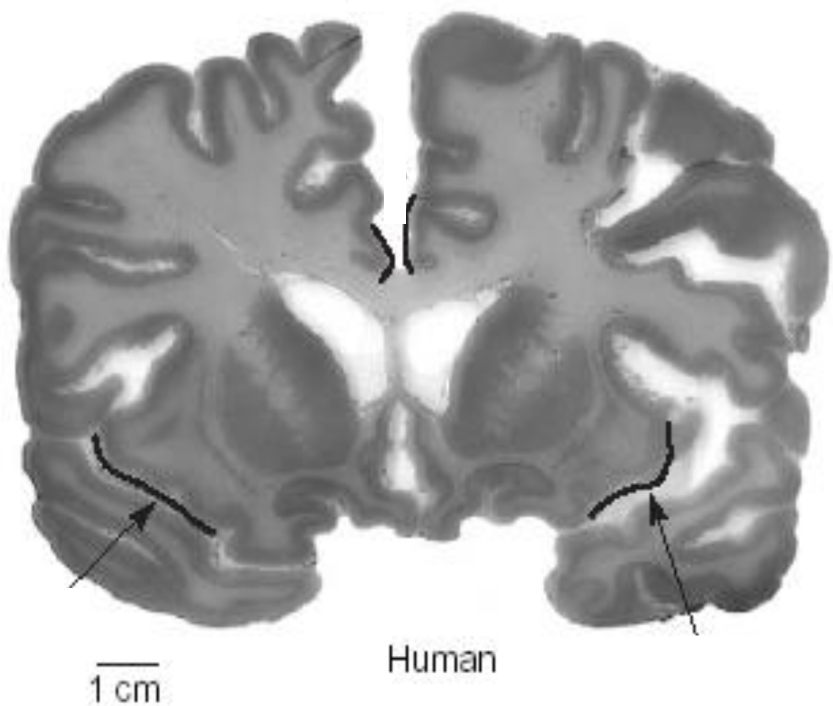
- Ambiguity aversion is ubiquitous in human decision making, and is utterly incompatible with Bayesian norms

- Humans are *wired* to process incertitude separately and differently from variability

Neuroscience of risk perception

(Marr 1982; Barkow et al. 1992; Pinker 1997, 2002)

Instead of being divided into rational and emotional sides, the human brain has *many* special-purpose calculators



Partial list of mental calculators

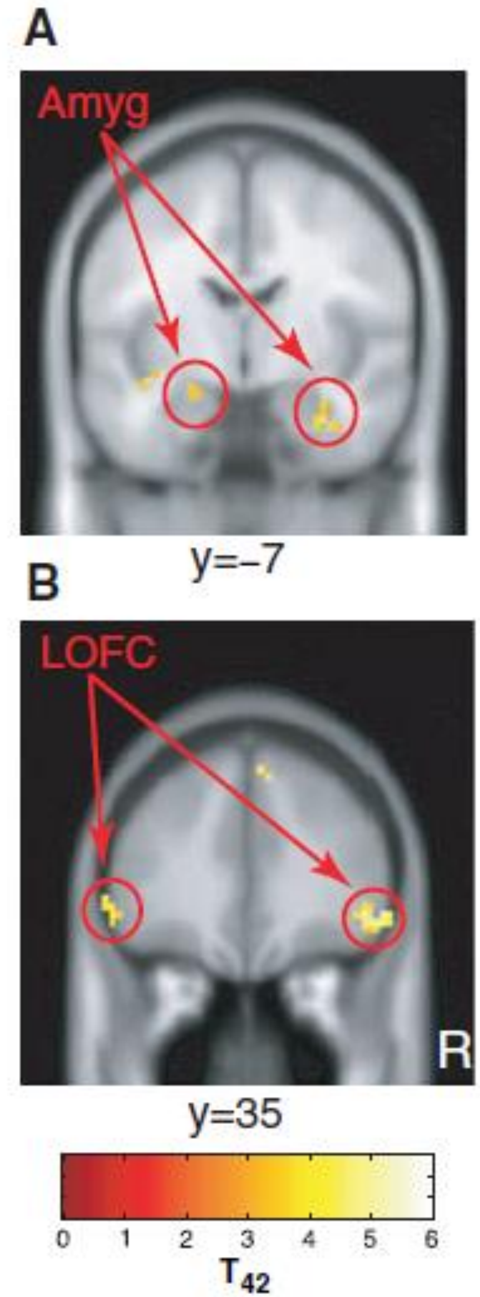
(after Pinker 2002; 1997; Marr 1982; Barkow et al. 1992)

- Language (grammar and memorized dictionary)
- Practical physics (pre-Newtonian)
- Intuitive biology (animate differs from inanimate)
- Intuitive engineering (tools designed for a purpose)
- Spatial sense (dead reckoner and mental maps)
- Number sense (1, 2, 3, many)
- Probability sense (frequentist Bayes)
- Uncertainty detector (procrastination)
- Intuitive economics (reciprocity, trust, equity, fairness)
- Intuitive psychology (theory of mind, deception)

What’s the evidence for an ‘uncertainty detector’ in humans?

fMRI

- Hsu et al. (2005) found localized regions of activity in the brain under situations of ambiguity (incertitude)
- Amygdala associated with processing fear and threat



Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680-1683.

Ambiguity/incertitude detector

- Humans have an incertitude processor
  - Triggered by situations with ambiguity
  - Especially focused on the worst case
  - Common response is procrastination
- Functional organ
  - Normal feature of the human brain, also apes, rats
  - Not a product of learning
  - Visible in fMRI
- Brain lesions can make people insensitive to incertitude…so they behave as Bayesians

Other species

- Chimpanzees and bonobos preferred peanuts (which they like less than bananas) when they don't know the probability of getting bananas

A pair of bonobos enjoy banana slices at the Lola ya Bonobo sanctuary Vanessa Woods

Rosati, A., and B. Hare 2010. Chimpanzees and bonobos distinguish between risk and ambiguity. *Proceedings of Royal Society: Biology Letters*. See also "Apes unwilling to game when odds are uncertain" http://www.physorg.com/print209830622.html

What’s the evidence for a ‘probability sense’ in humans?

Probability sense

- One of the ways we know there is a probability calculator is that we can watch it turn on
- Platt and Glimcher (NYU) found particular neurons in the lateral intraparietal cortex in rhesus monkeys encode both the probability of an outcome and its magnitude
- We can also see it in reasoning behaviors

Bayesian reasoning (poor)

If a test to detect a disease whose prevalence is 0.1% has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? ___%

12-18% of medical students get this correct

Casscells et al. 1978 replicated in Cosmides and Tooby 1996

Bayesian reasoning (good)

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 50/1000, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? __**1**__ out of __**51**__.

76-92% of medical students get this correct.
That it's so easy to solve suggests hardwiring.

Casscells et al. 1978 replicated in Cosmides and Tooby 1996

Calculators must be triggered

Cosmides & Tooby 1996; Gigerenzer 1991

- Humans have an innate probability sense, but it is triggered by *natural frequencies*

  Format of sensory data matters, not the meaning

- This calculator kicked in for the medical students who got the question in terms of natural frequencies, and they mostly solved it

- The mere presence of the percent signs in the question hobbled the other group

Multiple calculators may fire

- There are distinct calculators associated with
  - Probabilities and risk (variability) medical students
  - Ambiguity and uncertainty (incertitude) Hsu et al.
  - Trust and fairness Ultimatum Game
- Brain processes them differently
  - Different brain regions
  - Different chemical systems
- They can give conflicting answers
-

(e.g. Glimcher & Rustichini 2004 and references therein)

Conflict explains Ellsberg paradox

- Ambiguity detector countermands any risk estimate that might be produced by the probability sense

- What matters most, and perhaps exclusively, is *how bad it could be*…not how likely or unlikely that outcome is

Biological basis for Ellsberg

Hsu et al. 2005

- Probability sense and the ambiguity detector interfere with each other
- Humans do not make decisions based purely on probability in such cases
- Probabilists use equiprobability to model incertitude which confounds it with variability

Conflict may explain other biases too

Kahneman and Tversky

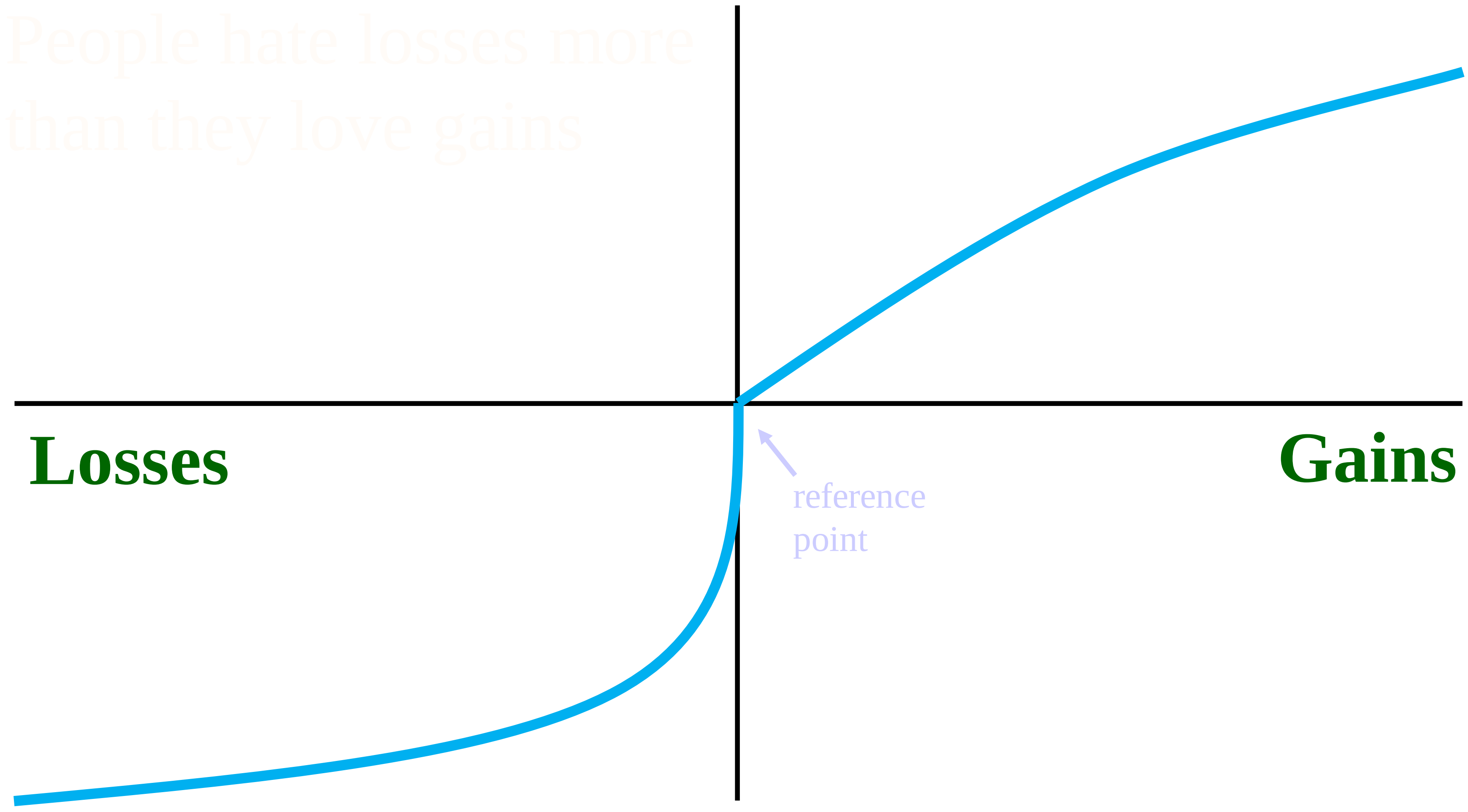
- Ambiguity aversion
  Avoiding options when probabilities seem unknown
- Loss aversion
  Disliking a loss more fervently than liking a gain of the same magnitude
- Hyperbolic discounting
  Preferring immediate payoffs over later (more intense the closer to present payoffs are)
- Base rate fallacy
  Neglecting available statistical data in favor of particulars
- Neglect of probability
  Disregarding probability in decision making under uncertainty
- Pseudocertainty
  Making risk-averse choices for positive outcomes, but risk-seeking for negative

People just seem downright *stupid* about risks and uncertainty

Loss aversion

(asymmetry in perceptions about losses and gains)

Loss aversion

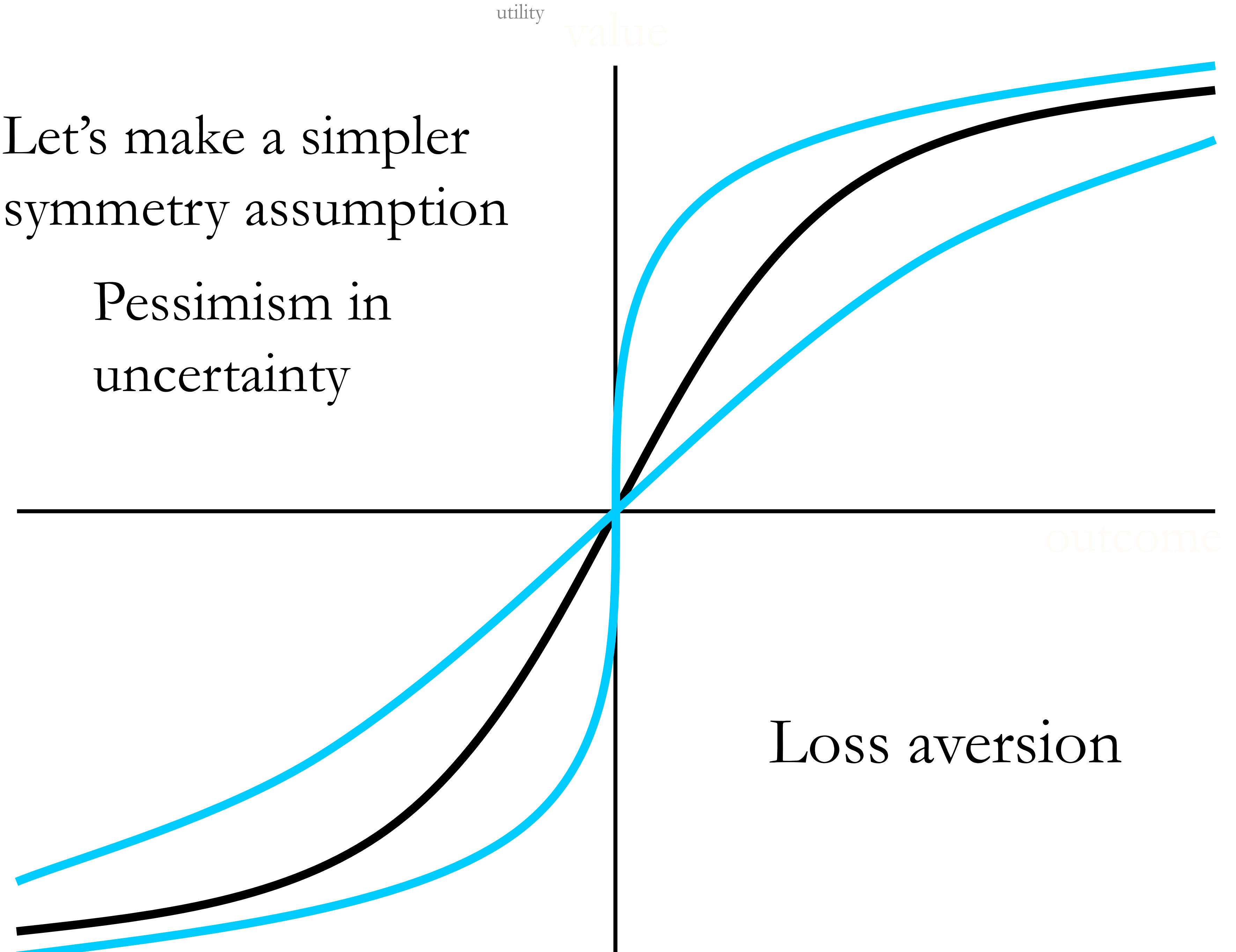


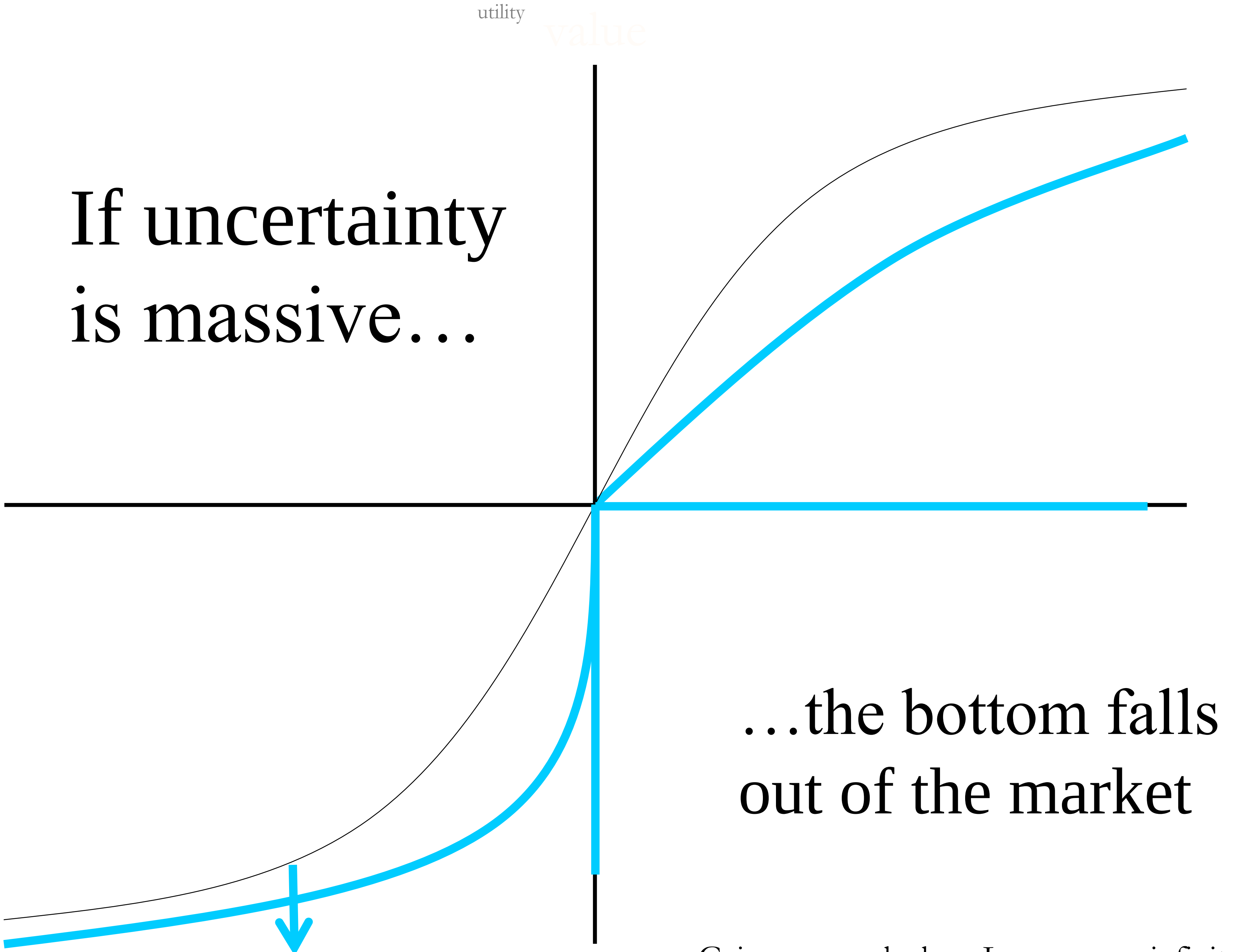
Prospect theory adopts (but does not explain) loss aversion

But *why*?

- Prospect theory is the state of the art
- Purely descriptive
- Doesn’t say *why* loss aversion should exist

- What is the biological basis for loss aversion?
- How could it have arisen in human evolution?


utility
Let's make a simpler symmetry assumption
Pessimism in uncertainty
Loss aversion



Gains seem valueless; Losses seem infinite

Loss aversion disappears with certainty

- Loss aversion disappears
  - with a person you trust, or
  - *after* the gamble has been realized
    - Gilbert et al. 2004
    - Kermer et al. 2006
    - Yechiam & Ert 2007
    - Erev, Ert, & Yechiam 2008
    - Ert & Erev 2008
- When losses and gains are surely exchangeable, the uncertainty contracts to the symmetric utility

Direct experimental evidence

- Ellsberg made the probabilities ambiguous

- Psychologist Christian Luhmann (Stony Brook) made *rewards* ambiguous
  - Visually obscured the promised payoffs
  - “I’ll pay you between 1 and 10 bucks”

- Loss aversion varies with the size of uncertainty
- Disappears with certainty

Clinical evidence

- Amygdala damage eliminates loss aversion

- But doesn’t affect a person’s ability to gamble and respond to changing values and risk (*n* = 2)

Still normal in risk aversion

- Amygdalectomied rhesus monkeys approach stimuli that healthy monkeys avoid

De Martino, B., C.F. Cramerer and R. Adolphs (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences of the United States of America* 107(8): 3788–3792.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2840433/pdf/pnas.200910230.pdf
Mason et al. (2006). *Emotion* 6: 73-81.

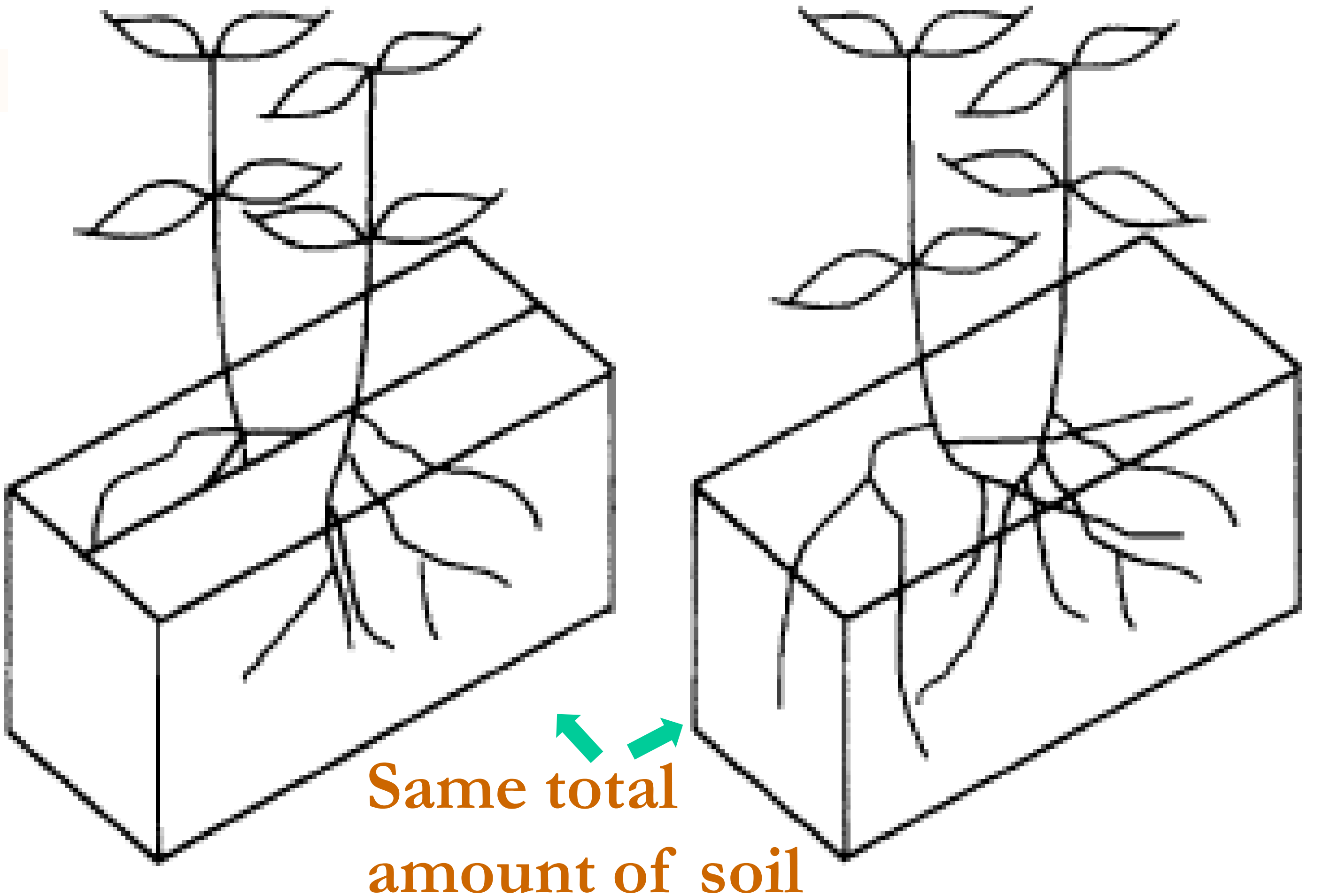
But why pessimism?

- Pessimism is often advantageous evolutionarily

- Natural selection can favor pessimism
  - Death is ‘hard selection’
  - Animal foraging strategies
  - Programmed plant behaviors

- Being wrong often has asymmetric consequences
  - Foraging: Finding dinner versus being dinner
  - Competition: Preemption versus being preempted

And even in plants!



Gersani, M., Brown, J. S., Brien, E. E., Maina, G. M., & Abramsky, Z. (2001). Tragedy of the commons as a result of root competition. *Journal of Ecology* 89: 660-669.

Pessimism is not inevitable

- Pessimism is not the only reaction to uncertainty
  - Normal people in stressful situations
  - Pathological gamblers
  - Maniacs

- Ambiguity aversion decreases with optimism (Pulford 2009)

Pulford, B. D. (2009). Is luck on my side? Optimism, pessimism, and ambiguity aversion. *Quarterly Journal of Experimental Psychology*, 62: 1079-1087.

Third calculator: intuitive economics

- Computes fairness of situations
- Detects cheaters who are getting more than their share, or not shouldering their responsibility

Ultimatum Game: a scientist offers money to two players

- One player proposes how the money should be divided between the two players
- The other player can *accept* (and both get their share) or *reject* the division (and neither gets anything)

How should you play?

- Economists say the rational behavior is
  - Responder: always accept any deal (it's free money!)
  - Proposer: always offer the smallest possible amount

- Actual plays are much closer to fair
  - Proposer offers a split much closer to 50:50
  - Responder accepts only if the split is closer to 50:50

Universal in humans

- Ultimatum Game is only surprising to economists
- The pattern may be universal in human behavior
  - Over 100 papers in 25 western societies
  - 15 non-western societies
- Exceptions
  - Sociopaths
  - Chimpanzees
  - Kids under 5
  - People playing against machines
- Chimps are *more rational* that humans

Why do humans do this?

- Adaptation for reciprocal altruism
  - Cares about fairness and reciprocity
  - Alters outcomes of others at a personal cost
  - Rewards those who act in a prosocial manner
  - Punishes those who act selfishly, even when punishment is costly

- Mediated by the fairness calculator
  - Pattern absent when fairness is not an issue

“Irrationality”

- Irrationality is a hallmark of human decisions
- Why are humans biased, irrational, stupid?
  - Using the wrong mental calculator *(optical illusion)*
  - Disagreement among mental calculators
  - Concerned with issues outside the risk analysis
    - Fairness, justice
    - Outcomes not treated in the analysis
    - Chance the risk analyst is lying
    - Chance the risk analyst is inept

Different calculators

Currently, confusion is guaranteed

- Neuroimagery and clinical psychology show humans distinguish incertitude and variability
- Probabilists traditionally use equiprobability to model incertitude, which confounds the two
- Risk analysts report their findings in ways that we know will create misunderstandings

Import for risk assessment

- Risk analyses woefully incomplete
  - Neglect or misunderstand incertitude
  - Omit important issues and thus understate risks
- Presentations use very misleading formatting
  - Percentages, relative frequencies, conditionals, etc.
- Both problems can be fixed
  - By changing *analysts'* behavior (not the public's)

Risk communication

- Informed consent in medicine
- Effective communication to decision makers

- If you want people to understand your risk calculations, you have to *speak their language*

Take-home messages

- Evolution has wired humans to see incertitude distinctly, and differently, from variability
- Conflict between the two seems to explain many probability and decision paradoxes
- There is a proper calculus that handles both although it is incompatible with current norms

Statistics for the next century

Small sample size

- Student's *t* statistic introduced in 1908
- Statistics has spent a century developing analyses in which sample size is limiting
- But sample sizes are not so small anymore
  - Financial data
  - Continuous mechanized sampling
  - Satellite imagery and other mass collections
  - Commercial data
  - Social media
  - Internet of Things (30 trillion sensors feeding the web)

Other uncertainties

- Sample size will always be an issue
- But it may not always be the *only* issue anymore
- Other issues become important as sample sizes grow
  - Measurement imprecision (mensurational uncertainty)
  - Missingness and censoring
  - Model uncertainty, non-stationarity, etc.
- Many believe it’s always better to collect more samples than to improve the precision of samples
  - This is *not* true
  - Believing this creates suboptimal experimental designs

Measurements aren't reals

- Real line is a poor model for measurements
  - "Measure" by comparing readings against a scale
  - Almost all real values cannot be measurements
- All real measurements have uncertainties
  - A real value cannot express this
- Real values have infinitely many zeros after the last decimal place
  - Such precision is never achievable in the real world
- The real line is totally ordered
  - But a measurement may not be only larger, smaller or the same as another

Missing data

- Traditional methods assume MCAR or MAR
- These assumptions are not always reasonable
- Assuming them anyway leads to wrong answers
- Correctly accounting for missingness can yield *dilation*, in which your uncertainty increases even when you increase sample size
  - If a temperature sensor fails from extremes in either direction, a missing value may mean the temperature is much higher or much lower than you thought

Censoring

- Traditional methods are decidedly bad and can be grossly misleading
- Likelihood strategies make assumptions that may not be tenable
  - Can produce unreasonable results

Interval data

- Calculating variances, *t*-statistics, etc. for data sets that contain intervals are NP-hard problems
- But various special cases are quite easy
  - Censored
  - Binned
  - Same precision
  - Nested
- These let us compute variance in $O(n)$ time
  - Can compute variance for 500,000 samples in less than 0.5 second on a laptop

Interval uncertainty
Incertitude

- Intermittent observations
- Plus-minus intervals
- Non-detects and data censoring
- Missing values
- Blurring for privacy or security reasons
- Bounding studies

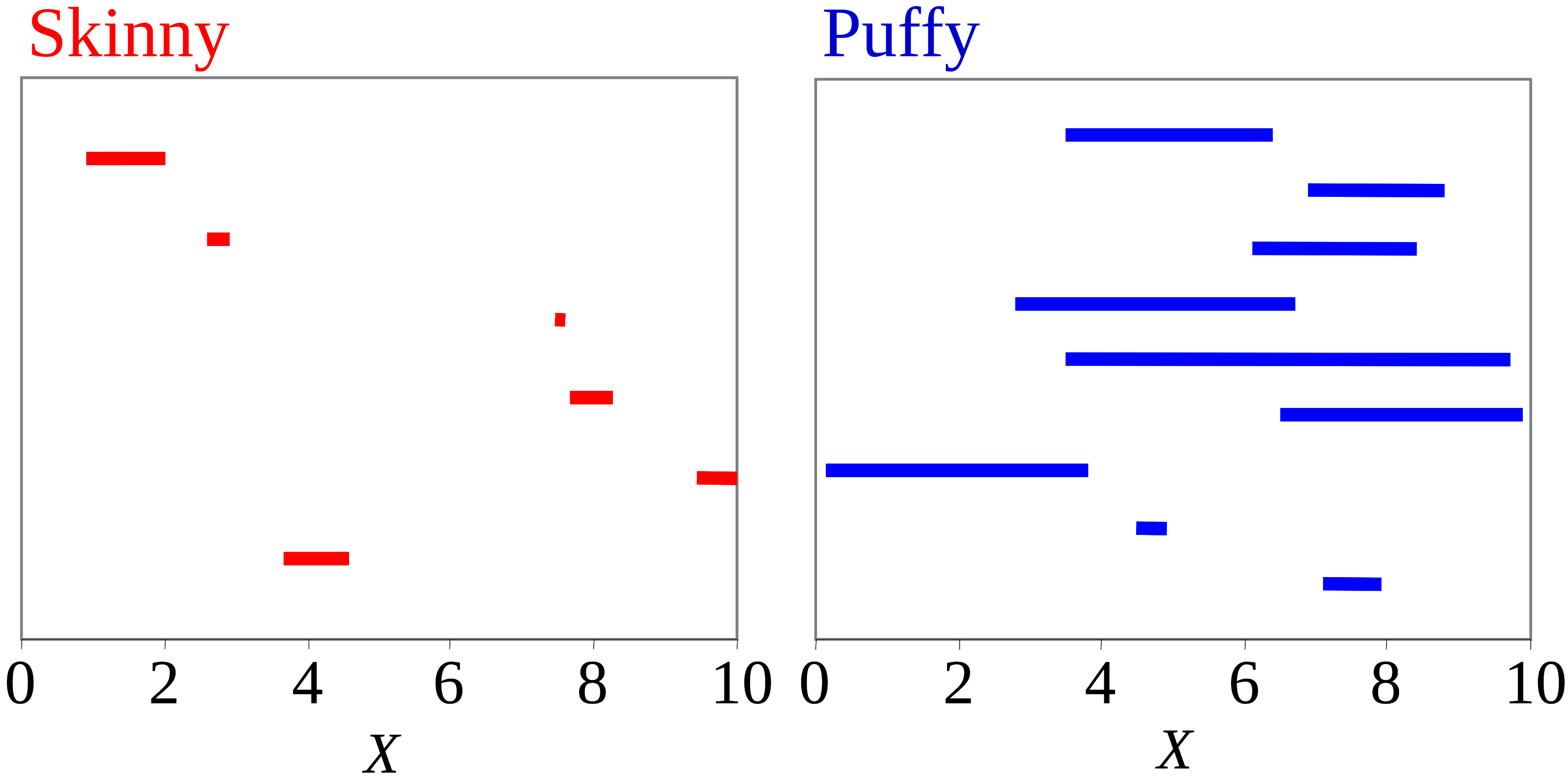
Two approaches

- Model each interval as uniform distribution
  - Presumes different values are equally likely
  - Laplace's principle of insufficient reason
  - Calculations relatively easy, but interpretation subtle

- Model each interval as a *set* of possible values
  - Specifies no single distribution within the range
  - Theory of imprecise probabilities
  - Calculations often NP-hard, but interpretation easy

A tale of two data sets

Skinny data	**Puffy data**
[1.00, 2.00]	[3.5, 6.4]
[2.68, 2.98]	[6.9, 8.8]
[7.52, 7.67]	[6.1, 8.4]
[7.73, 8.35]	[2.8, 6.7]
[9.44, 9.99]	[3.5, 9.7]
[3.66, 4.58]	[6.5, 9.9]
	[0.15, 3.8]
	[4.5, 4.9]
	[7.1, 7.9]

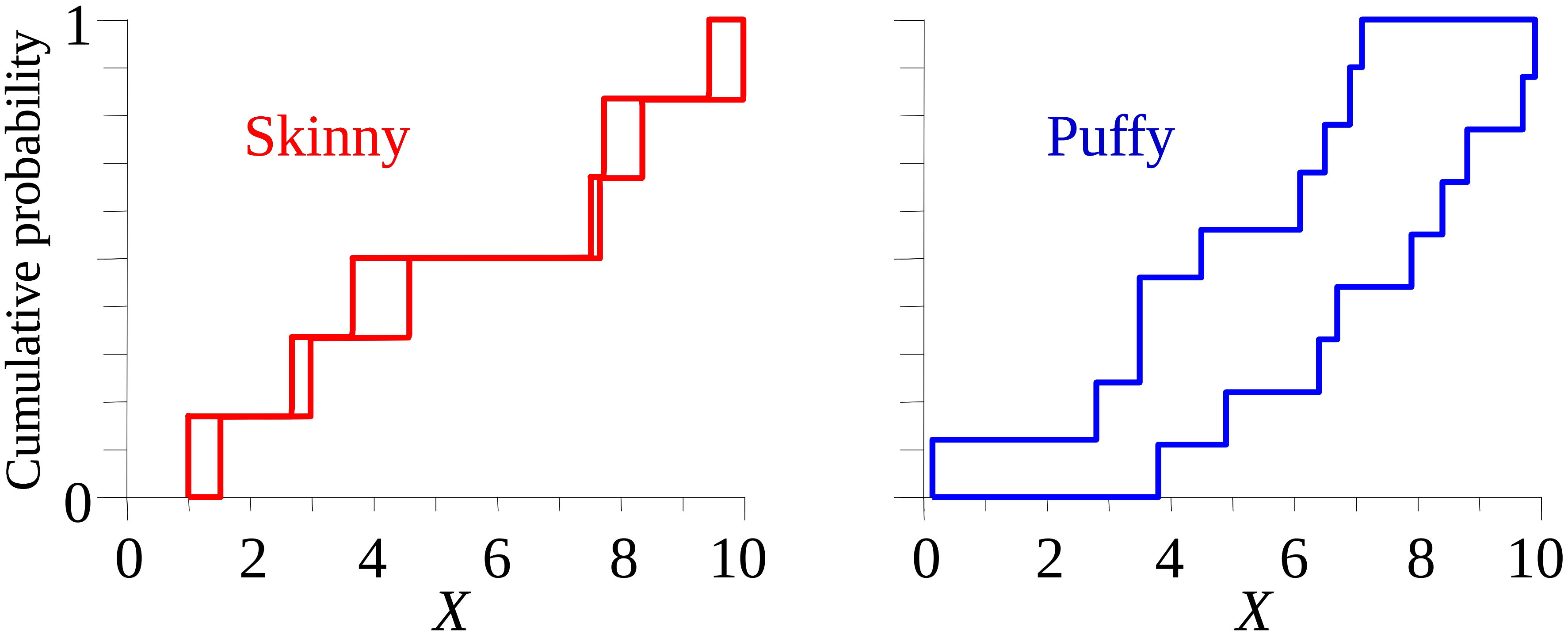


Skinny
Puffy
0
2
4
6
8
10
X
0
2
4
6
8
10
X

Empirical distribution

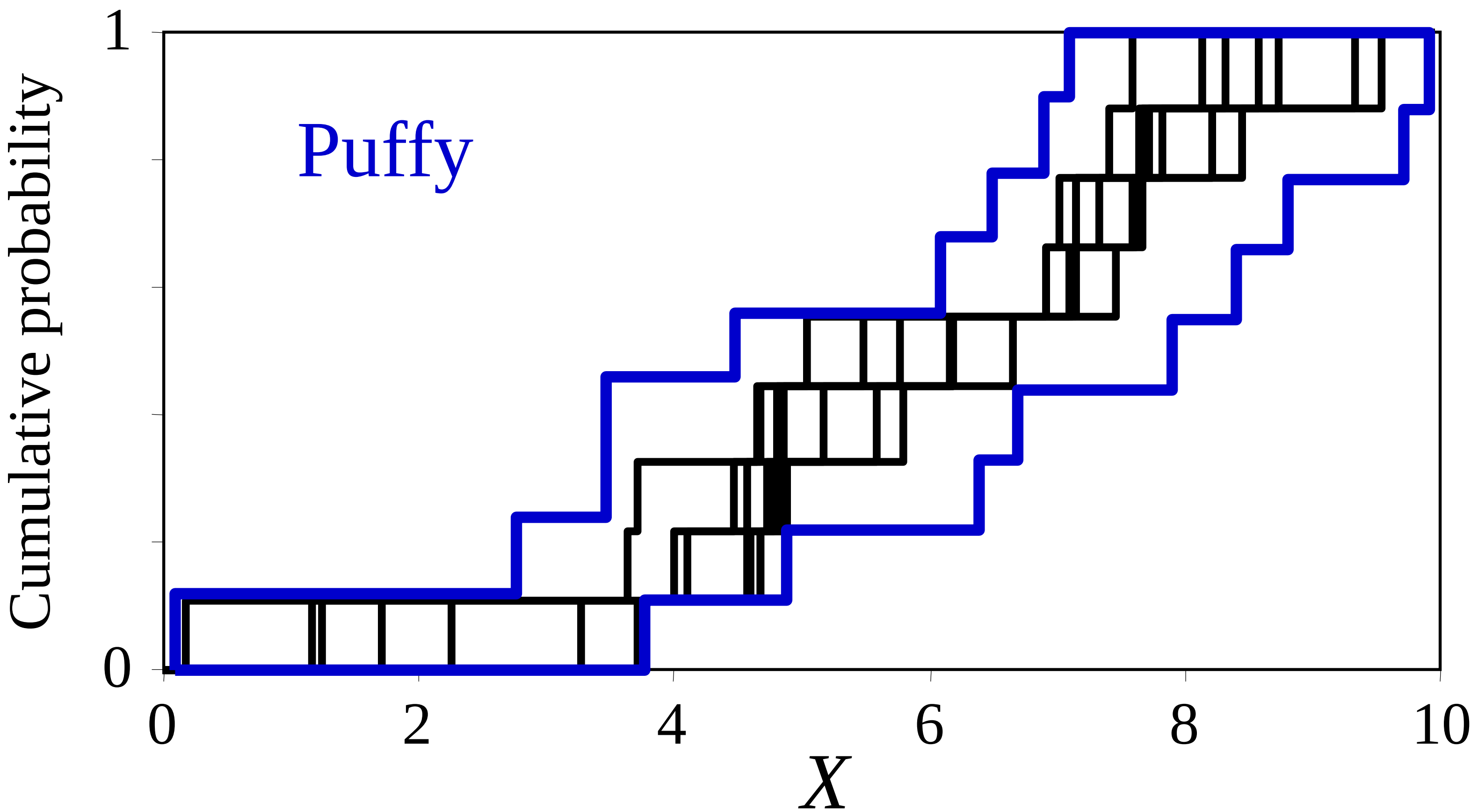
- Summary of the data themselves
- No distributional assumptions
- Uniforms approach yields a single distribution
- Intervals approach yields a probability box (i.e., a class of distributions)

Intervals approach

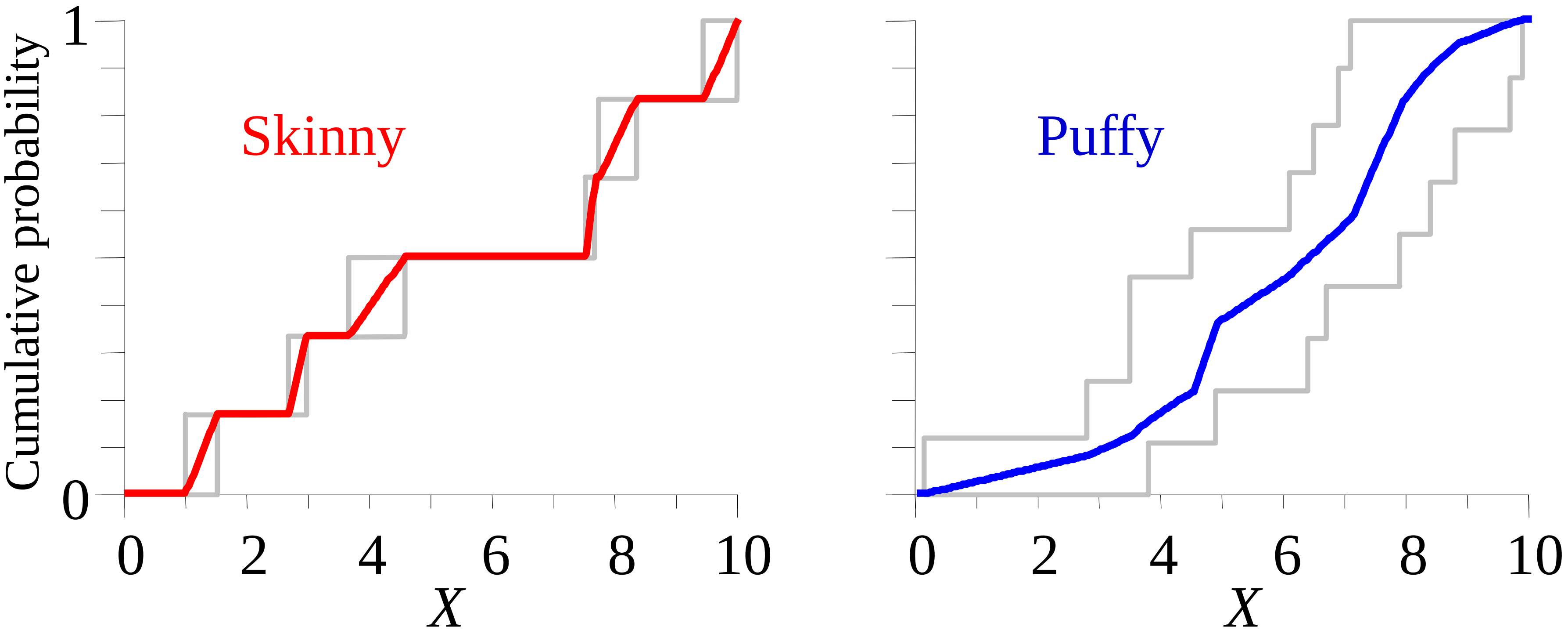


- Each side is cumulation of respective endpoints
- Represents both incertitude and variability

Uncertainty about the EDF



Uniforms approach

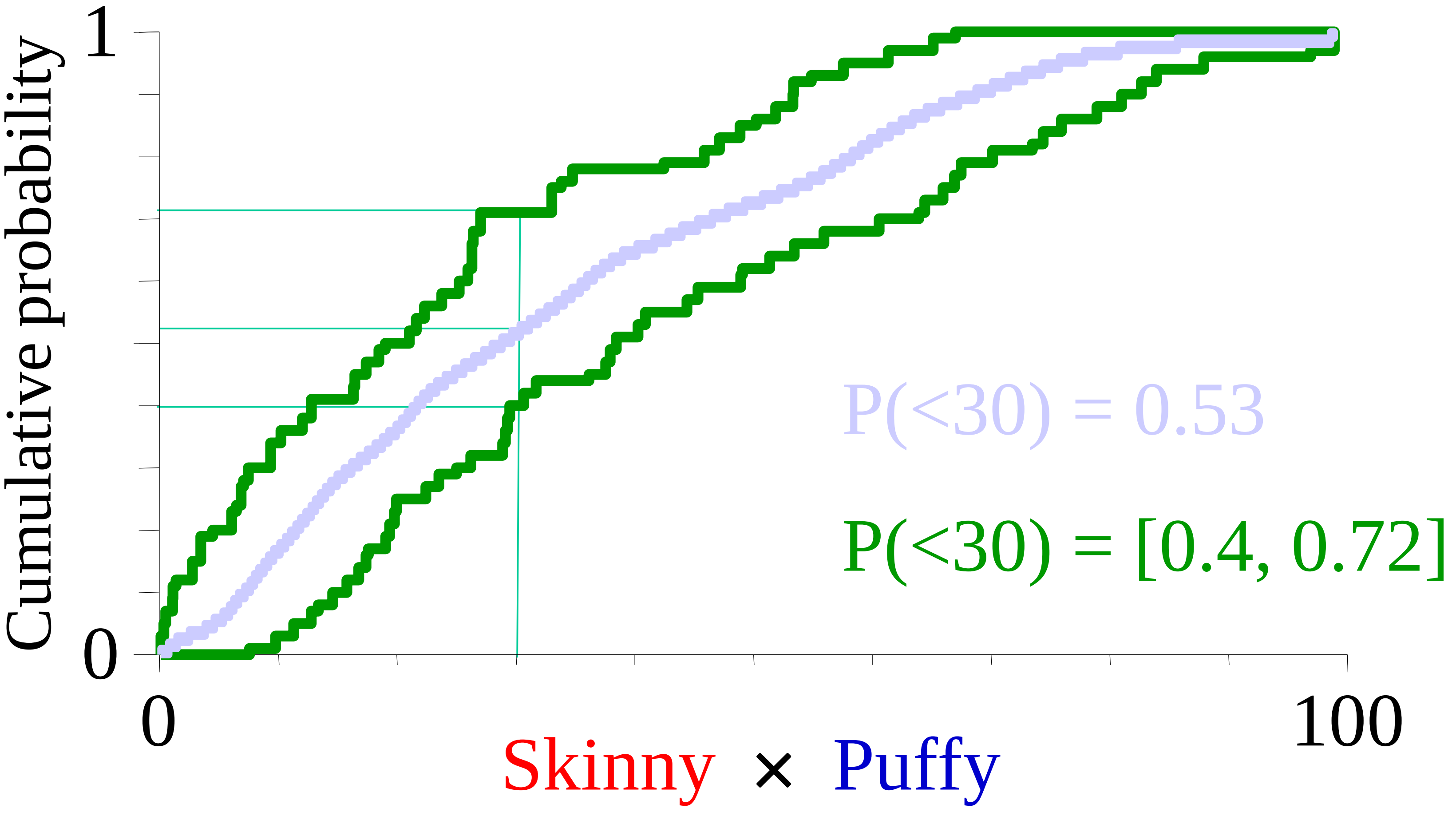


- Mixture (vertical average) of uniforms
- Conflates incertitude and variability

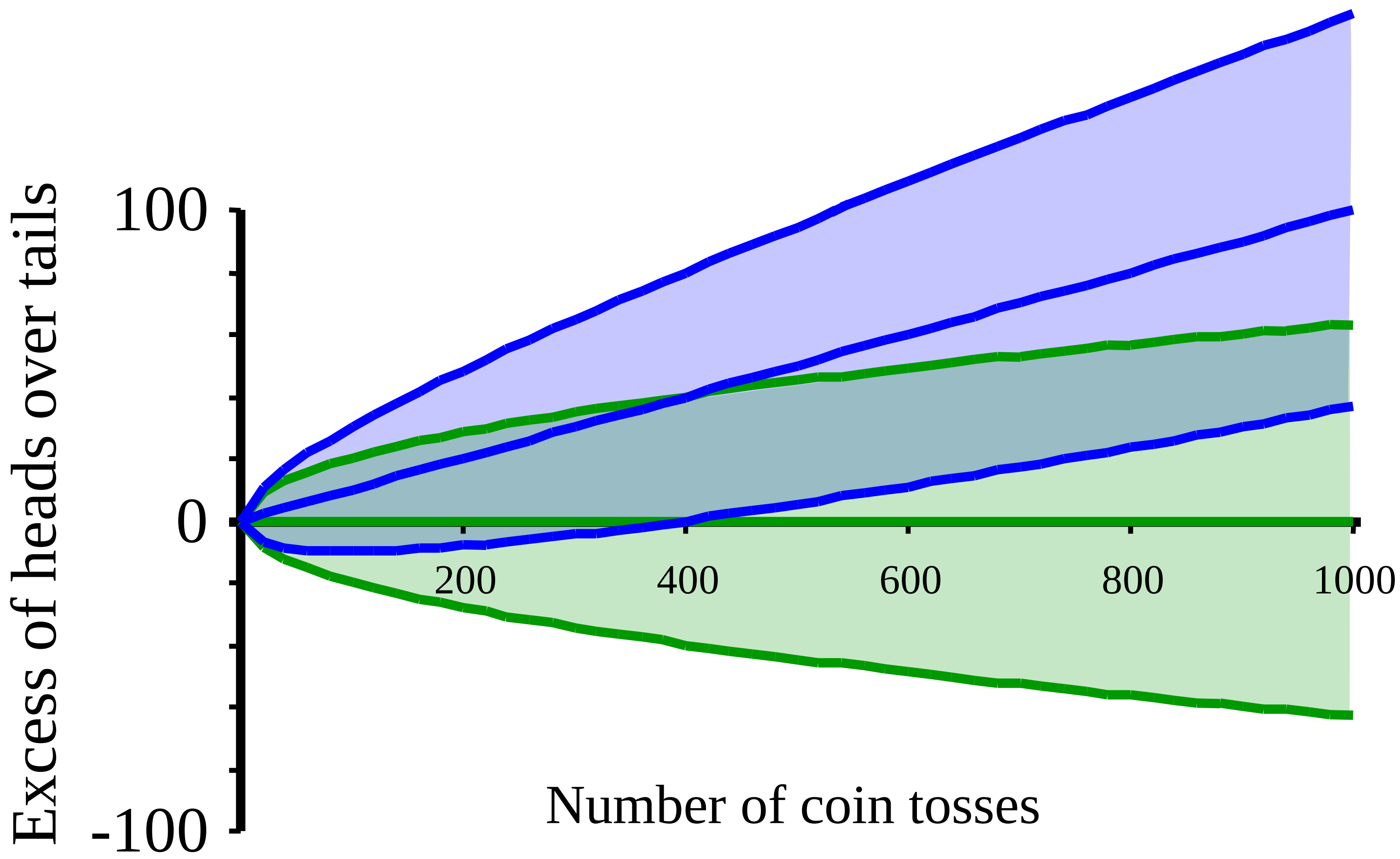
What’s the difference?

- Might be advisable to propagate the two kinds of uncertainty separately and differently
- *Example:* suppose we’re interested in the product of Skinny and Puffy…

Uniforms versus intervals approach



Randomness self-cancels



Fitted distribution

- Assumes some *shape* for the distributions
- Uniforms approach yields a single distribution
- Intervals approach yields a probability box

Interval approach

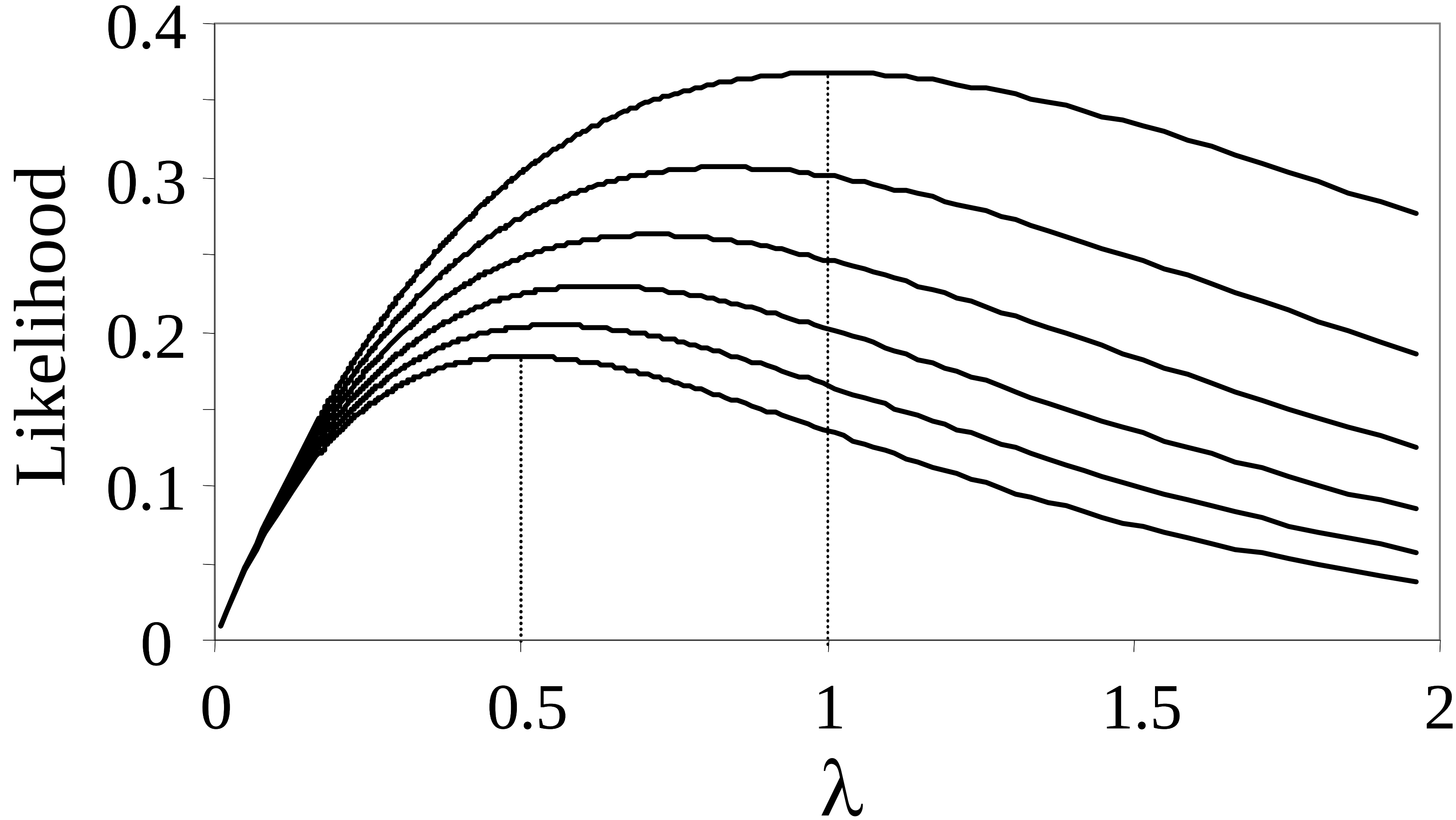
- Creates a class of maximum likelihood solutions

- Every one solves a ML problem for a set of scalar values within the respective intervals

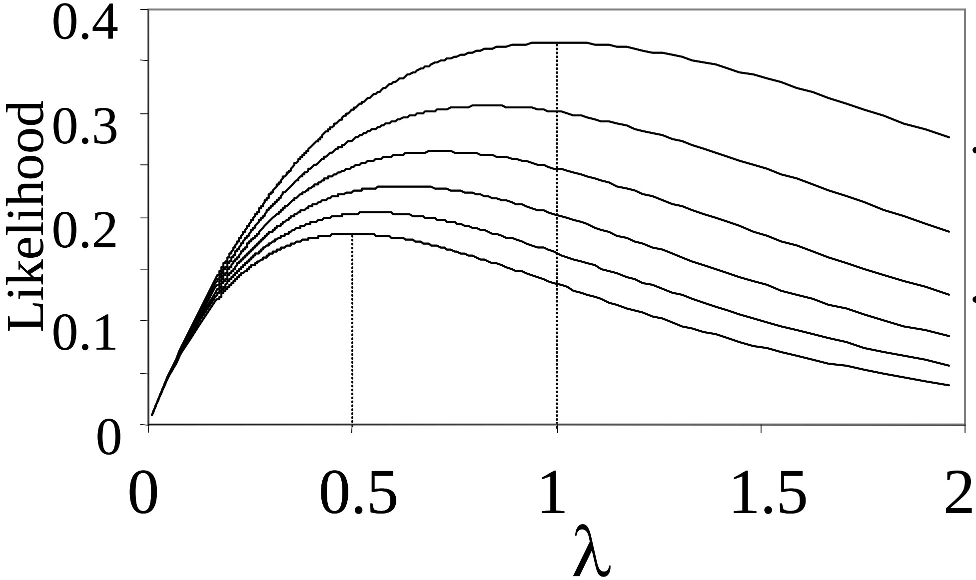
- *Example:* let's fit exponential distributions

$$L(x;\lambda) = \lambda \exp(-\lambda x) \qquad x \in [1,2]$$

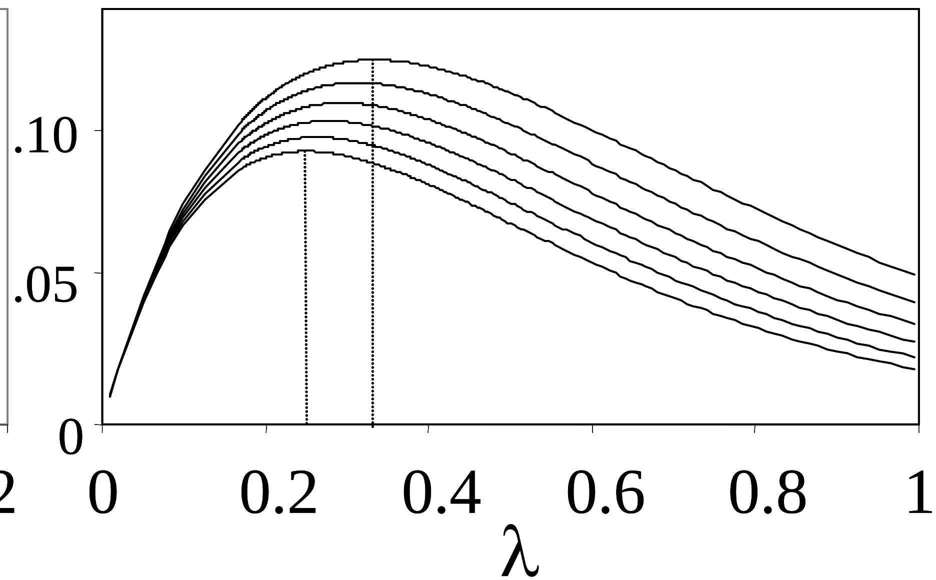
densities

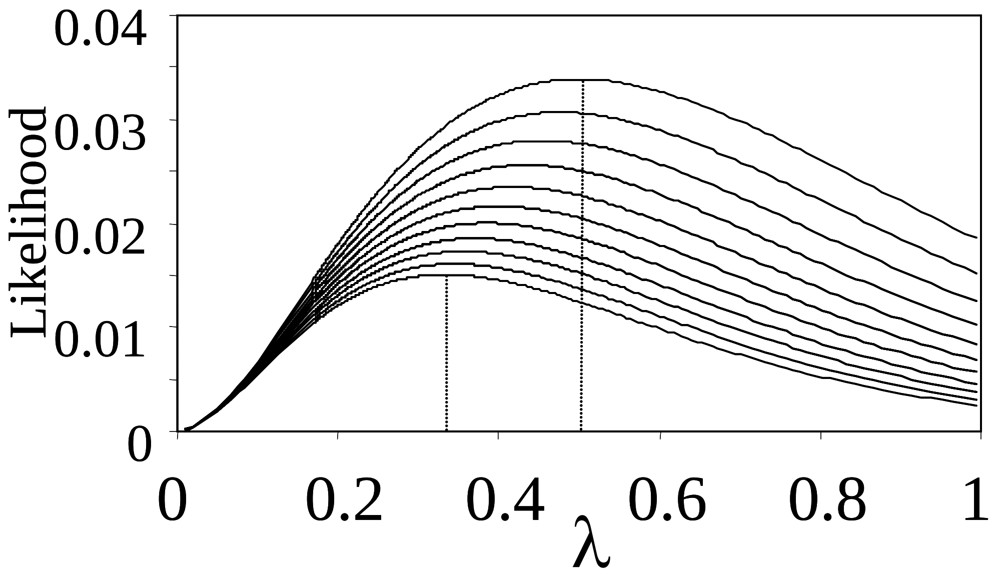


$x_1 \in [1,2]$



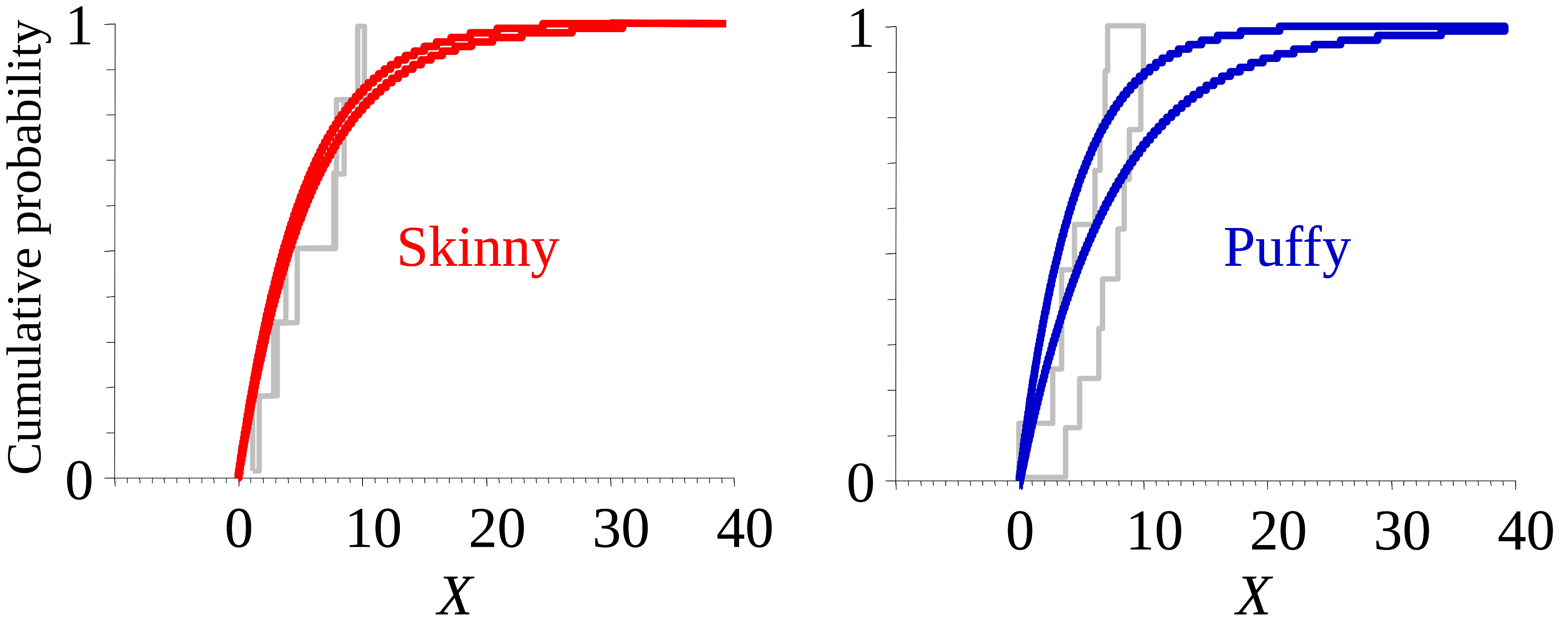
$x_2 \in [3,4]$





Pairwise multiplied if data are independent

Results from intervals approach



Maximum likelihood for censored data

Given a datum $\boldsymbol{x} = [\underline{x}, \overline{x}]$,

$$
\begin{aligned}
L(\boldsymbol{x}) &= \Pr(\underline{x} \le X \le \overline{x}) \\
&= \Pr(\overline{x} \le X) - \Pr(\underline{x} \le X) \\
&= F(\overline{x};\, \lambda) - F(\underline{x};\, \lambda)
\end{aligned}
$$

cumulatives

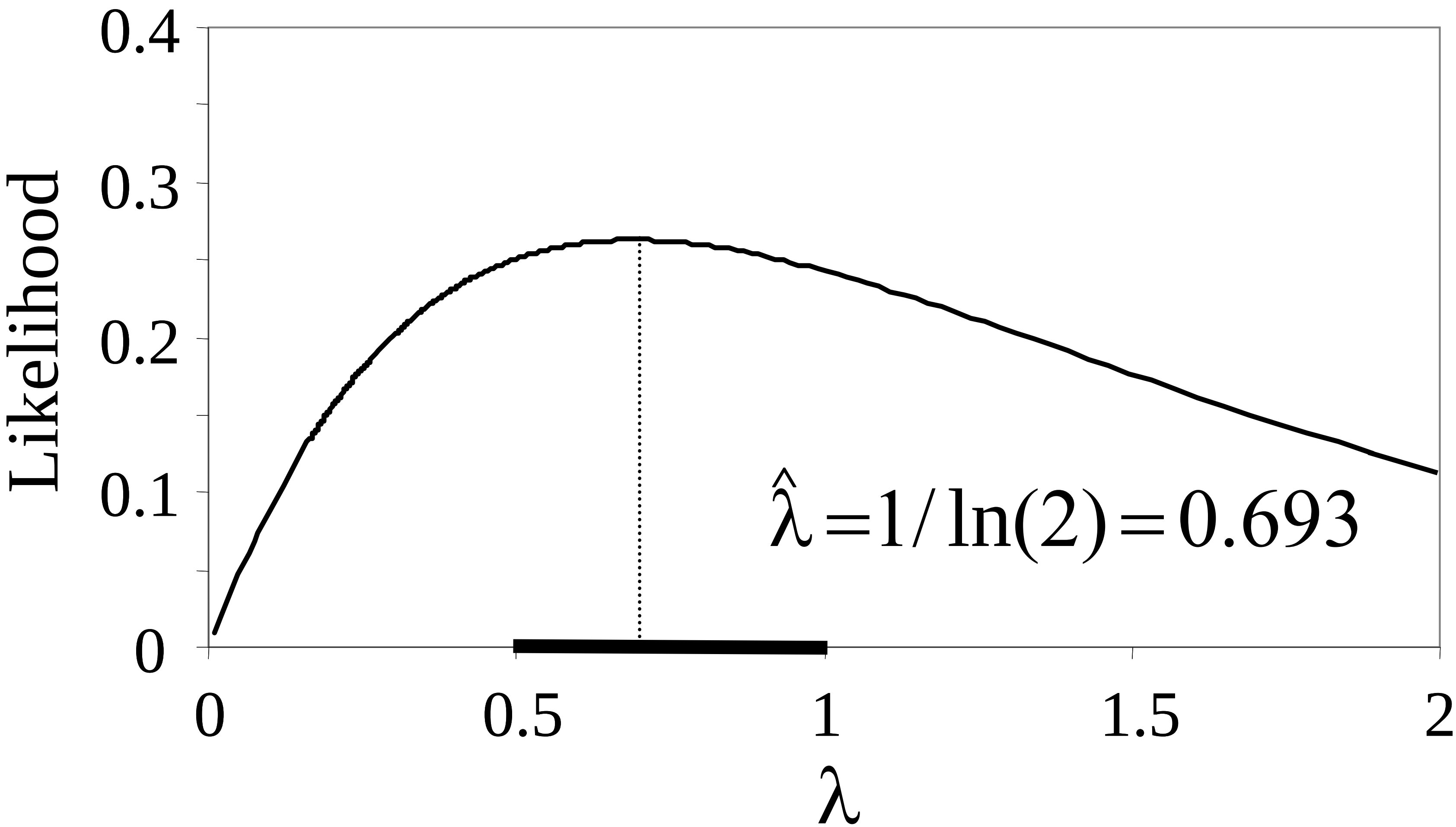
where, for an exponential distribution,

$$F(x;\lambda) = 1 - \exp(-\lambda x)$$

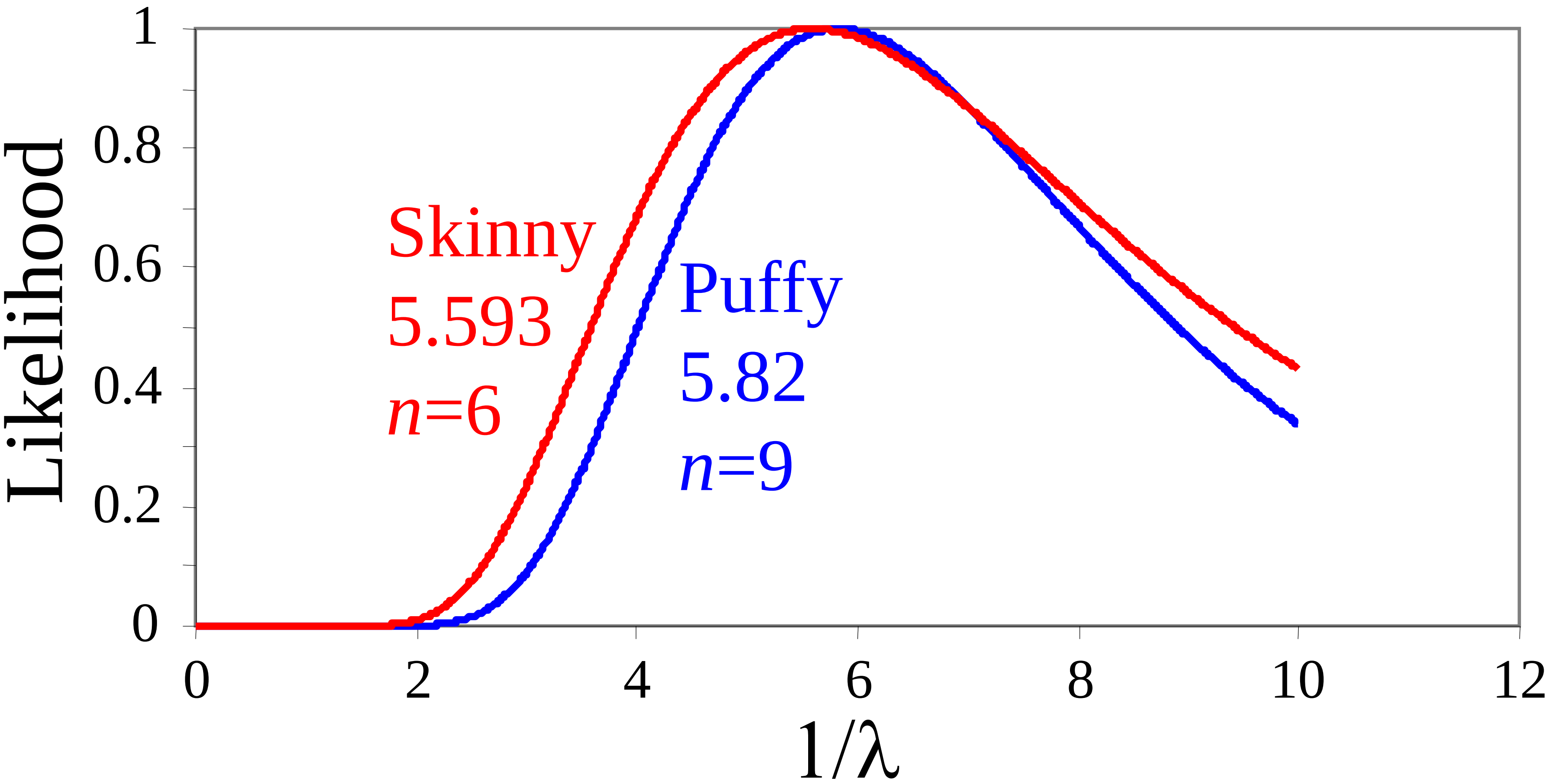
and $1/\lambda$ is the mean

Single datum $x \in [1,2]$

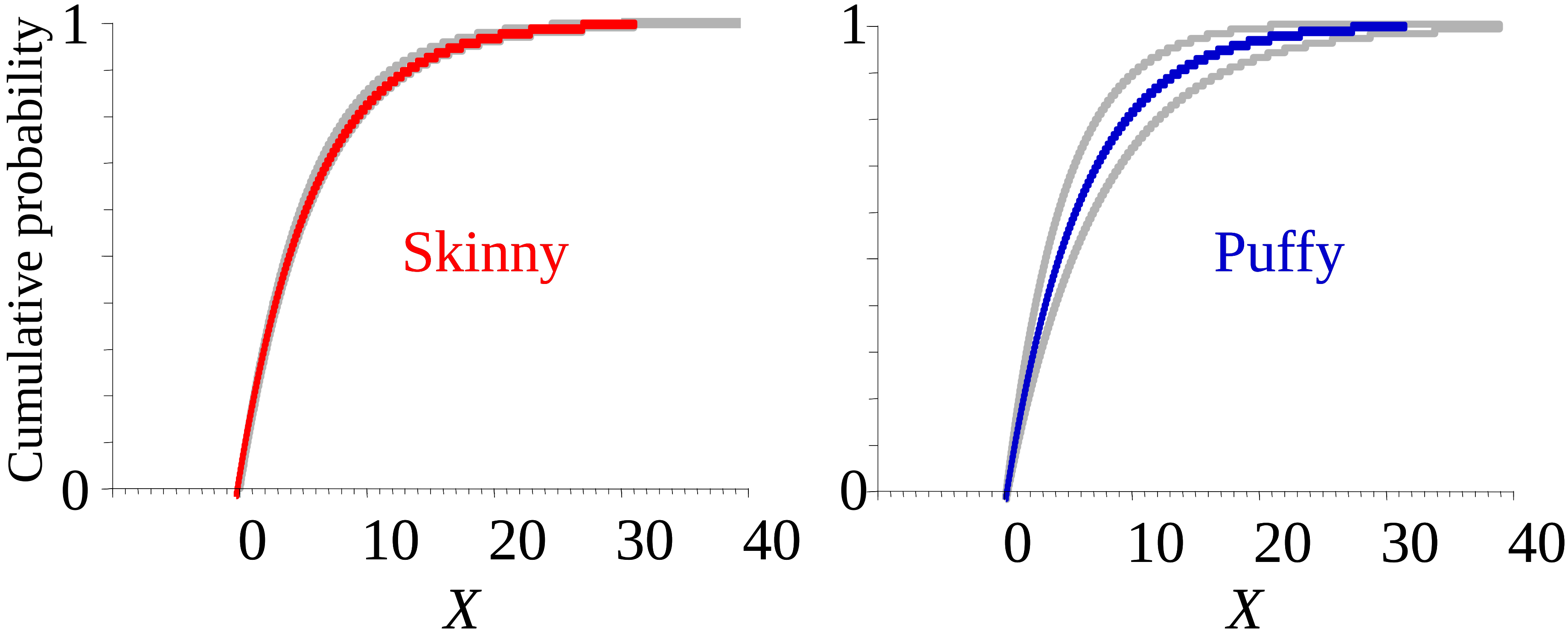
$$L(\lambda) = \exp(-\lambda) - \exp(-2\lambda)$$



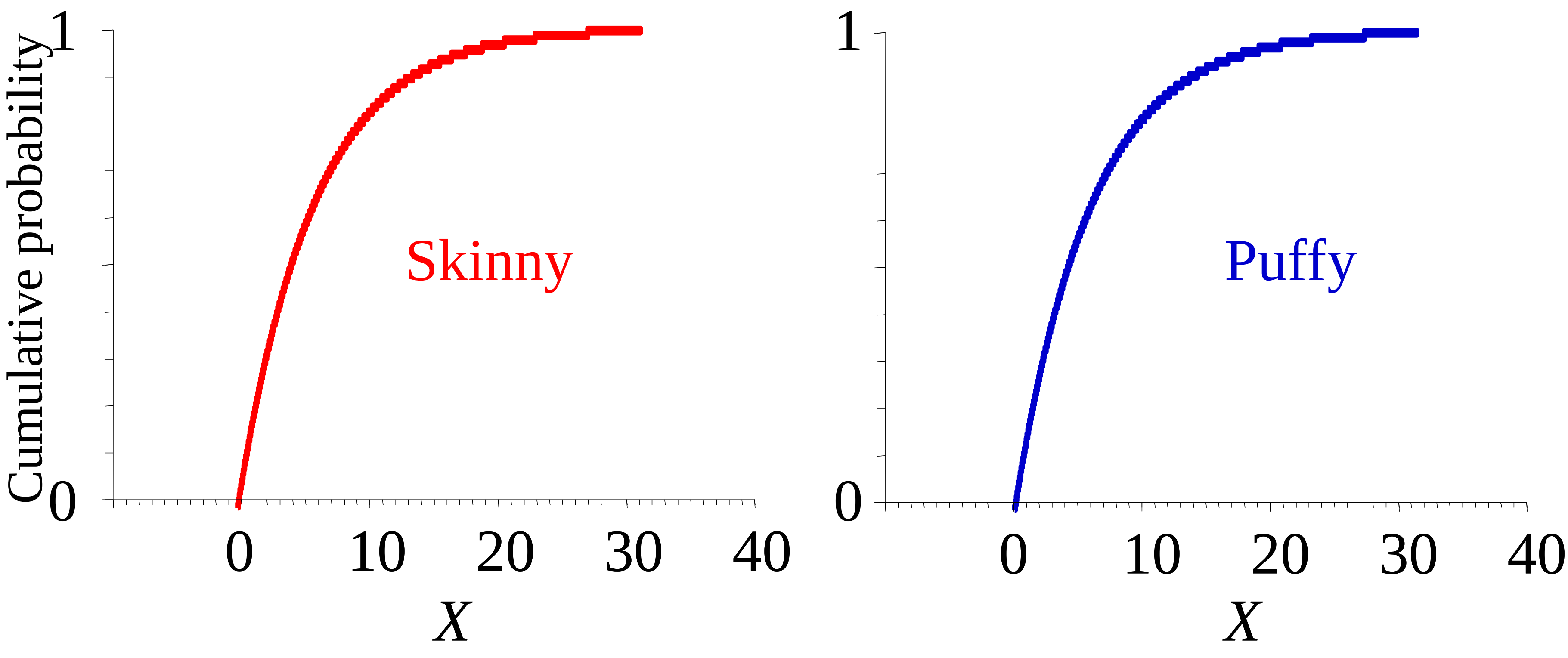
Maximum likelihood for multiple data



Results from uniforms approach



Negligible difference



Are these answers reasonable?

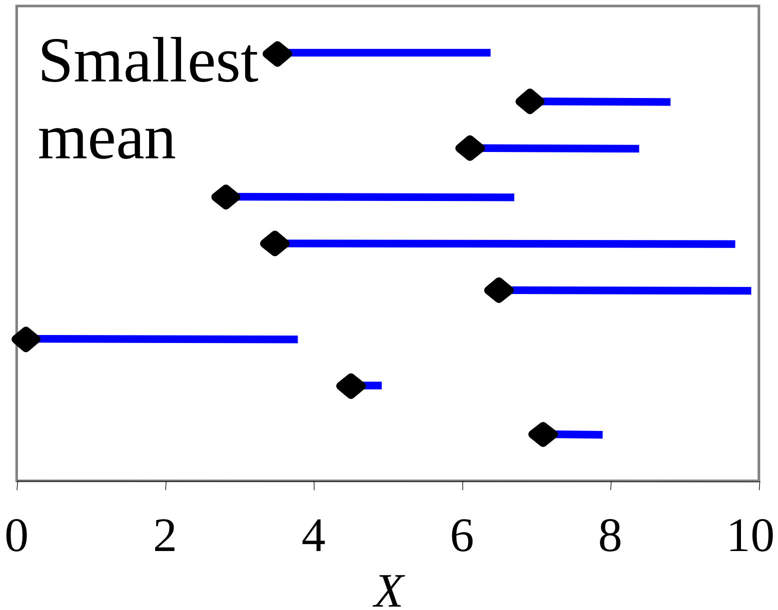
- Picking a single exponential to fit interval data
- Essentially no difference between Skinny and Puffy data sets, despite disparity of uncertainties
- Confidence bands actually smaller for Puffy, even though its uncertainty is much larger
- No guarantee that the answer approaches the true distribution *even if asymptotically many data are collected*

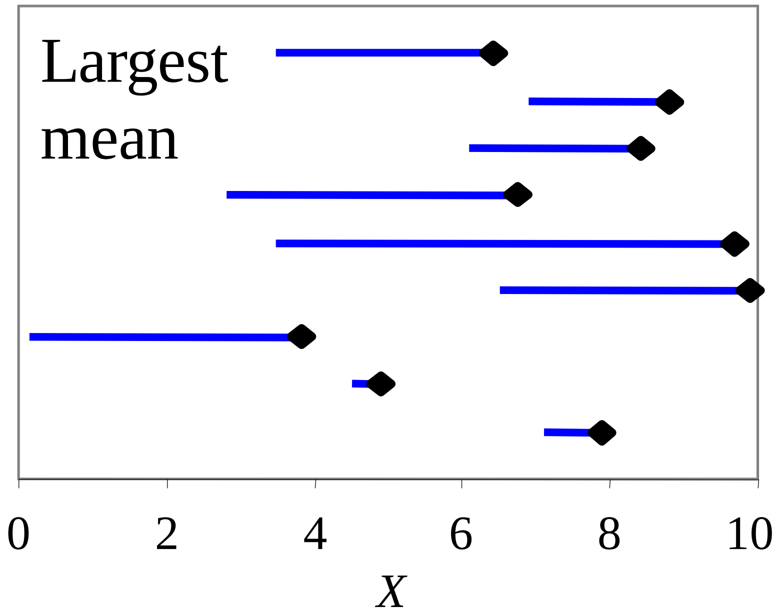
Descriptive statistics

- Practical algorithms for all common statistics
  - Mean, variance, skewness, etc., but not mode
  - Empirical and fitted distributions
  - Confidence limits, outlier statistics, etc.
- Some are simple to compute
- Many are NP-hard to compute, e.g., variance
- Feasible algorithms exist for many special cases
  - Variance can be computed in linear time with *n*
  - In practice, interval data is computationally easy

Central tendency

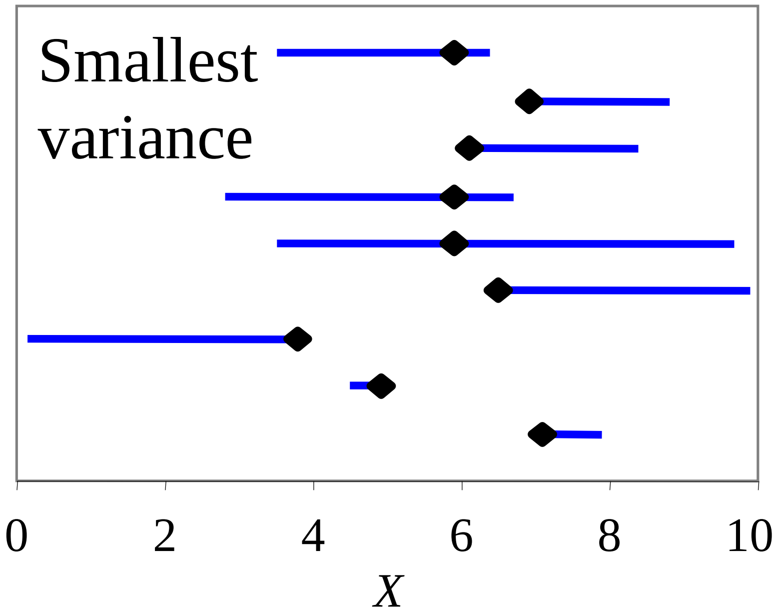
	Skinny	Puffy
Mean	[5.338, 5.849]	[4.56, 7.39]
Geometric mean	[4.186, 4.866]	[3.28, 7.09]
Median	[5.59, 6.125]	[4.5, 7.9]

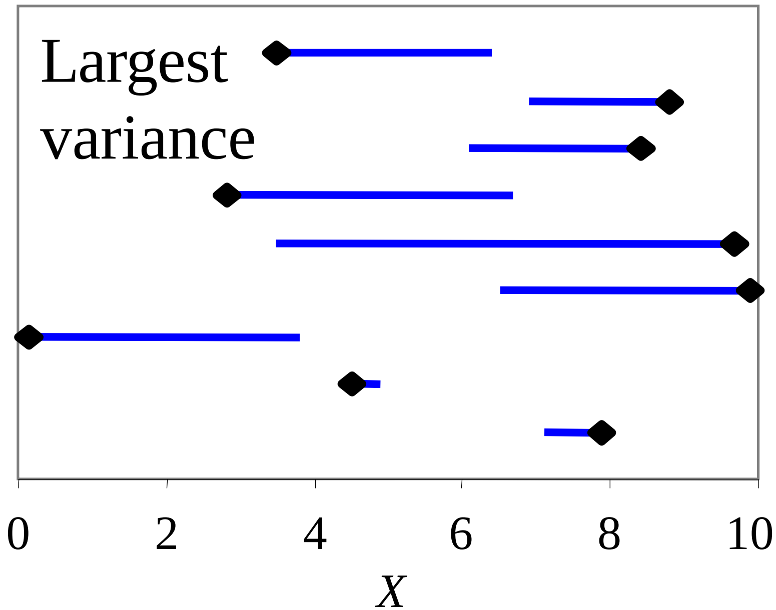




Dispersion

	Skinny	Puffy
Variance	[7.91, 10.77]	[0.91, 10.98]
Sample variance	[9.50, 12.92]	[1.03, 12.35]
Standard deviation	[2.81, 3.29]	[0.95, 3.32]
Interquartile range	[2.68, 8.35]	[3.50, 8.80]





Take-home messages

- Intervals generalise real-valued measurements
- Handle ± measurement precision, data censoring and missingness as special cases
- We *can* mix good and bad data in a consistent way
- Interval approach yields more robust results than methods that depend on subtle assumptions such as missing at random, or the ‘uniforms’ model

Model uncertainty

Model uncertainty

- Doubt about the structural form of the model

model uncertainty $\left\{ \begin{array}{l} \text{Parameters} \\ \text{Distribution shape} \\ \text{Intervariable dependence} \\ \text{Arithmetic equation} \\ \text{Level of abstraction} \end{array} \right.$

- Usually incertitude rather than variability
- Usually considerable in ecosystems models
- Often the elephant in the middle of the room

Uncertainty in probabilistic analyses

- Parameters
- Data surrogacy
- Distribution shape
- Intervariable dependence
- Arithmetic expression
- Level of abstraction

already

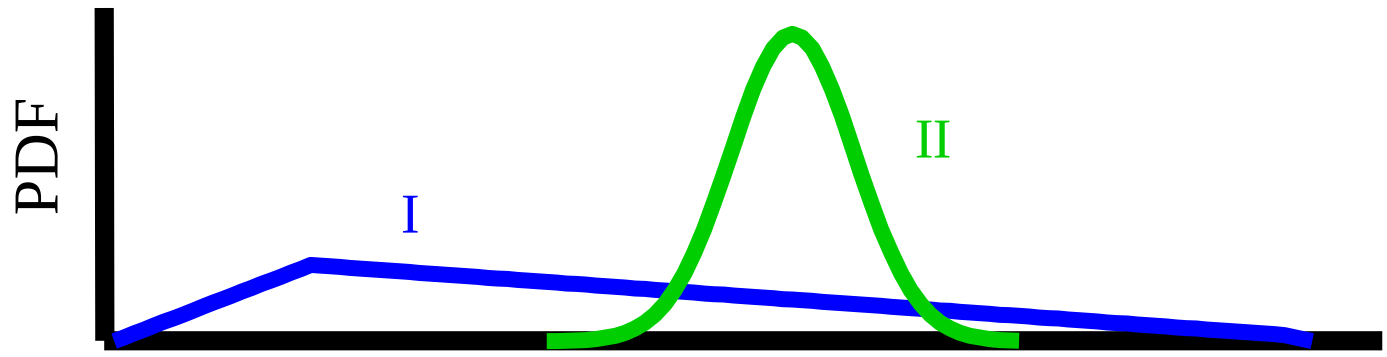
model uncertainty

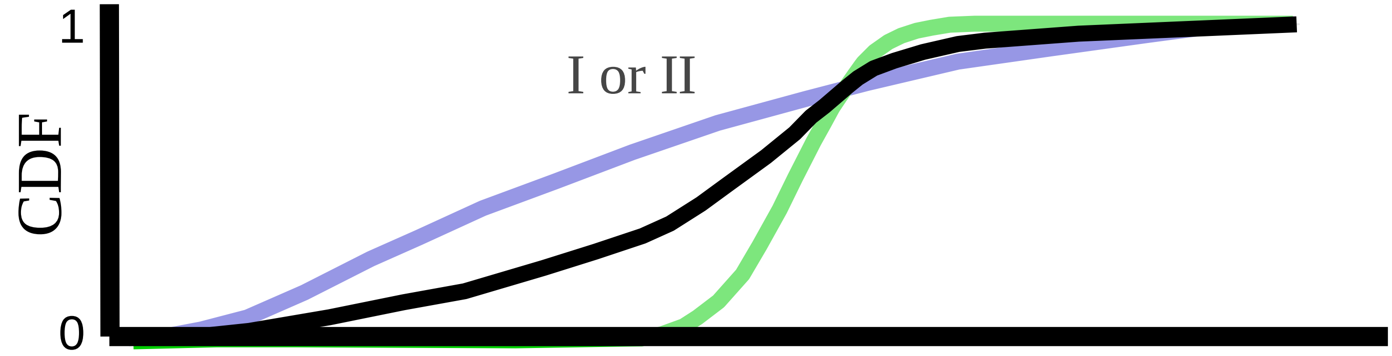
Monte Carlo strategy

- Introduce a new discrete variable
- Let the value of the variable dictate which model will be used in each rep
- Wiggle the value from rep to rep
- Only works for short, explicit list of models (you have to list the models)
- Many theorists object to this strategy

Model uncertainty as a mixture

If u>0.5 then model=I else model=II





General strategies

- Sensitivity (what-if) studies
- Probabilistic mixture
- Bayesian model averaging
- Enveloping and bounding analyses

Sensitivity (what-if) studies

- Simply re-compute the analysis with alternative assumptions
  - Intergovernmental Panel on Climate Change
- No theory required to use or understand

Drawbacks of what-if

- Consider a long-term model of the economy under global warming stress

  3 baseline weather trends
  3 emission scenarios
  3 population models
  3 mitigation plans

  81 analyses to compute, *and to document*

- Combinatorially complex as more model components are considered

- Cumbersome to summarize results

Probabilistic mixture

NRC uses

- Identify all possible models
- Introduce a new discrete random variable whose value says which model to use; let it vary in MC
- This *averages* probability distributions
- Use weights to account for different credibility (or assume equiprobability)

Drawbacks of mixture

- If you cannot enumerate the possible models, you can't use this approach
- Averages together incompatible theories and yields an answer that neither theory supports
- Can underestimate tail risks

Bayesian model averaging

- Similar to the probabilistic mixture
- Updates prior probabilities to get weights
- Takes account of available data

Drawbacks of Bayesian averaging

- Requires priors and can be computationally challenging

- Must be able to enumerate the possible models

- Averages together incompatible theories and yields an answer that neither theory supports

- Can underestimate tail risks

Bounding probabilities

- Translate model uncertainties to a choice among distributions
- Envelope the cumulative distributions
- Treat resulting p-box as single object

Drawbacks of bounding

- Cannot account for different model credibilities
- Can’t make use of data
- Doesn’t account for ‘holes’

Numerical example

The function f is one of two possibilities. Either

$$f(A,B) = f_{\mathrm{Plus}}(A,B) = \mathrm{A} + \mathrm{B}$$

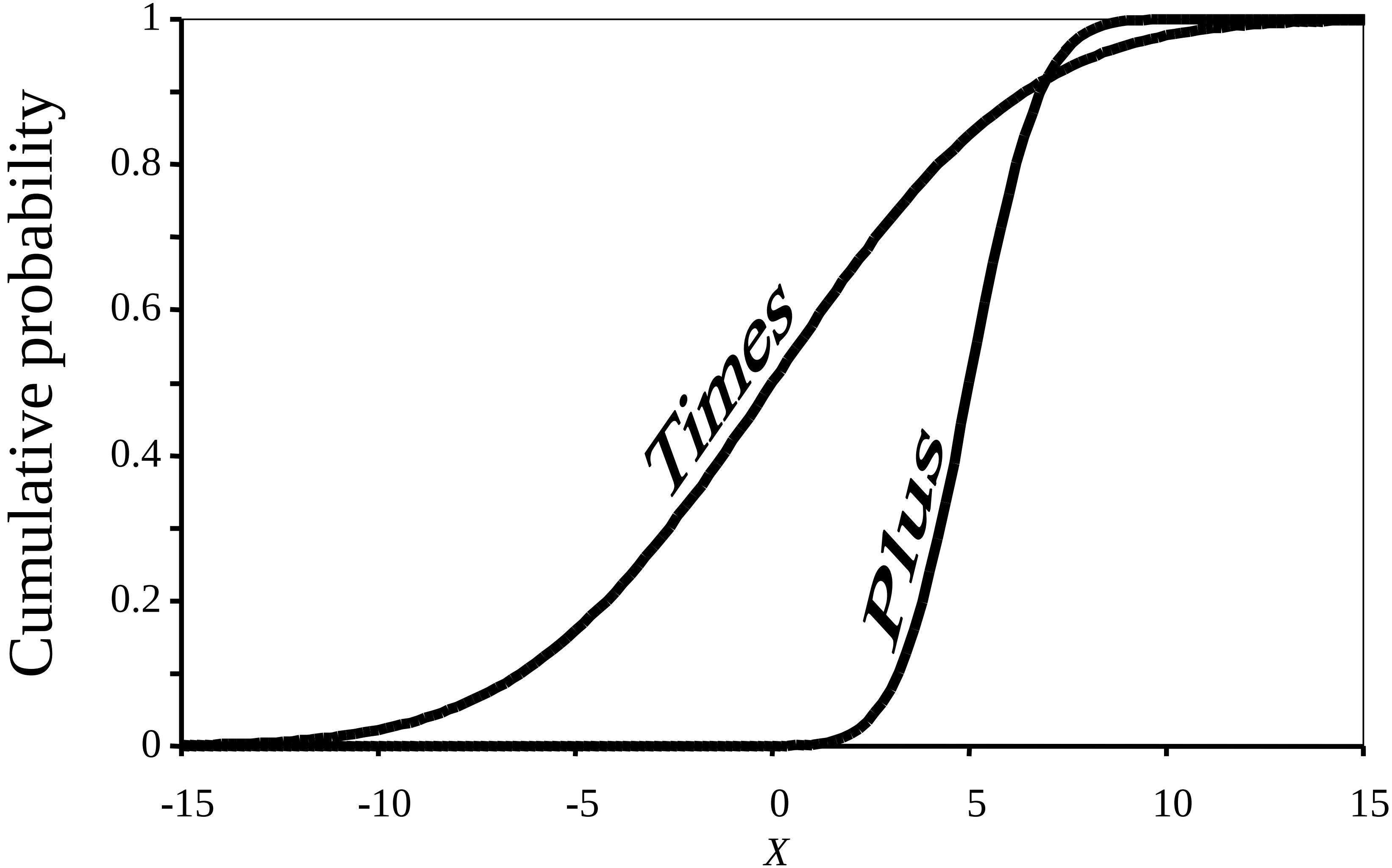
or

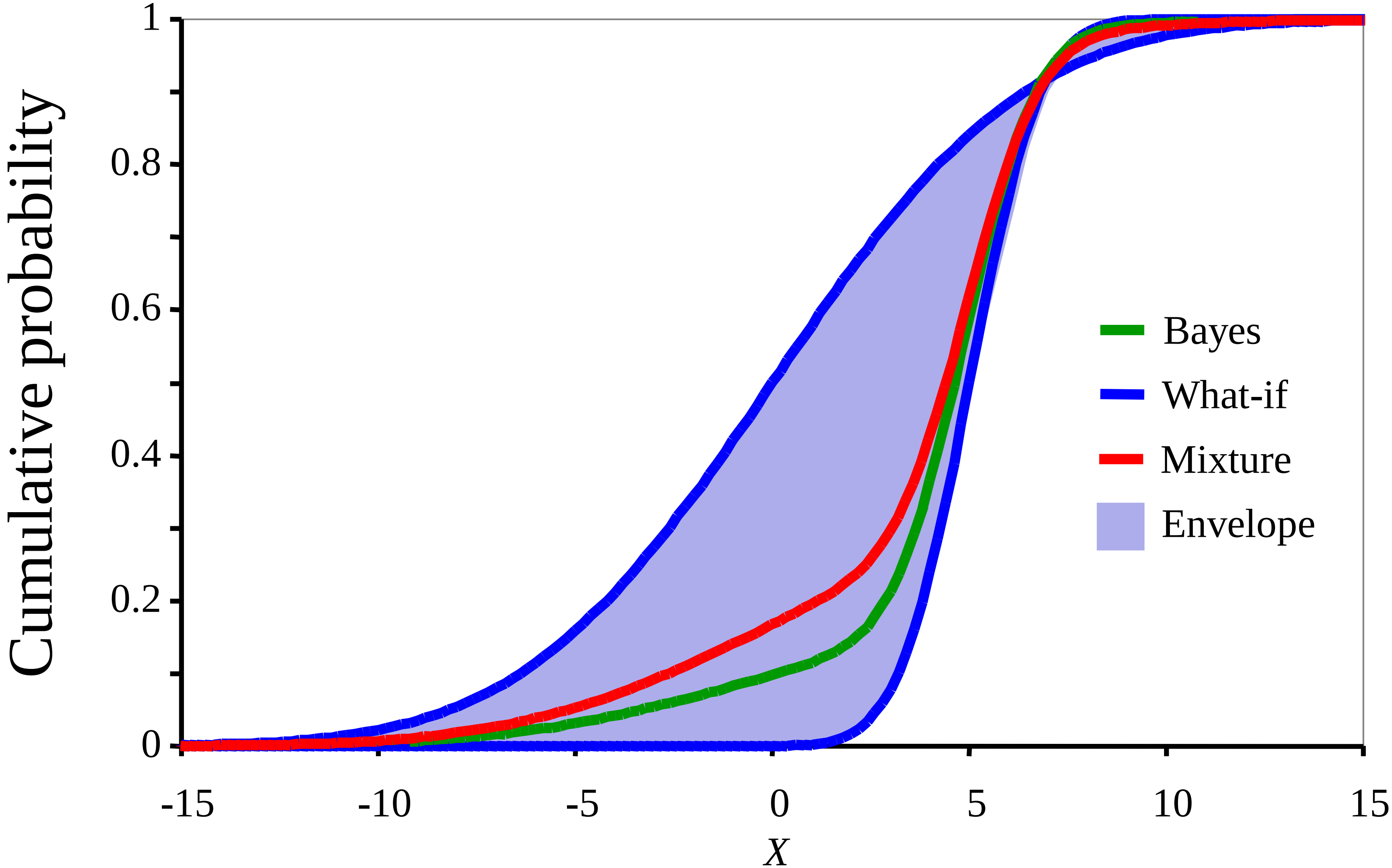
$$f(A,B) = f_{\mathrm{Times}}(A,B) = \mathrm{A} \times \mathrm{B}$$

is the correct model, but the analyst does not know which. Suppose that

$$A \sim \text{triangular}(-2.6, 0, 2.6)$$
$$B \sim \text{triangular}(2.4, 5, 7.6).$$

f_{Plus} is twice as likely as f_{Times}; datum: $f(A,B) = 7.59$


Cumulative probability
1
0.8
0.6
0.4
0.2
0
-15
-10
-5
0
5
10
15
X
Times
Plus



Cumulative probability
1
0.8
0.6
0.4
0.2
0
-15
-10
-5
0
5
10
15
X
Bayes
What-if
Mixture
Envelope

When you can enumerate the models

- What-if analysis isn’t feasible in big problems
- Probabilistic mixture is, at best, *ad hoc*
- For abundant data, Bayesian approach is best
- Otherwise, it’s probably just wishful thinking
- Bounding is reliable, but may be too wide

When you can’t list the models

- If you cannot enumerate all the models, bounding is often the only tenable strategy

- Shape of input distributions
- Dependence
- Functional form
  - Laminar versus turbulent flow
  - Linear or nonlinear low-dose extrapolation
  - Ricker versus Beverton-Holt density dependence

Synopsis of the four approaches

- What-if
  - Straightforward, doesn’t conflate uncertainties
  - Must enumerate, combinatorial
- Probabilistic mixture, Bayesian model averaging
  - Single distribution, accounts for data (and priors)
  - Must enumerate, averages incompatible theories
  - Can underestimate tail risks
- Bounding
  - Yields one object; doesn’t conflate or understate risk
  - Cannot account for data or differential credibility

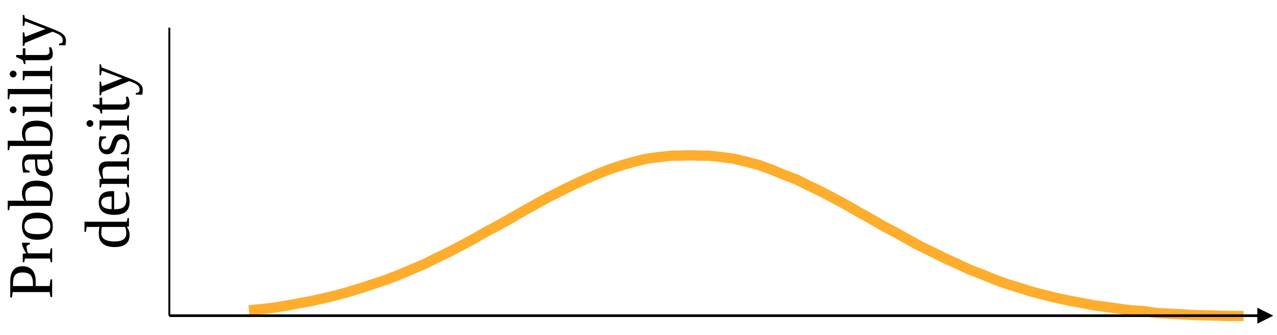
Conclusions

Probability theory isn’t good enough

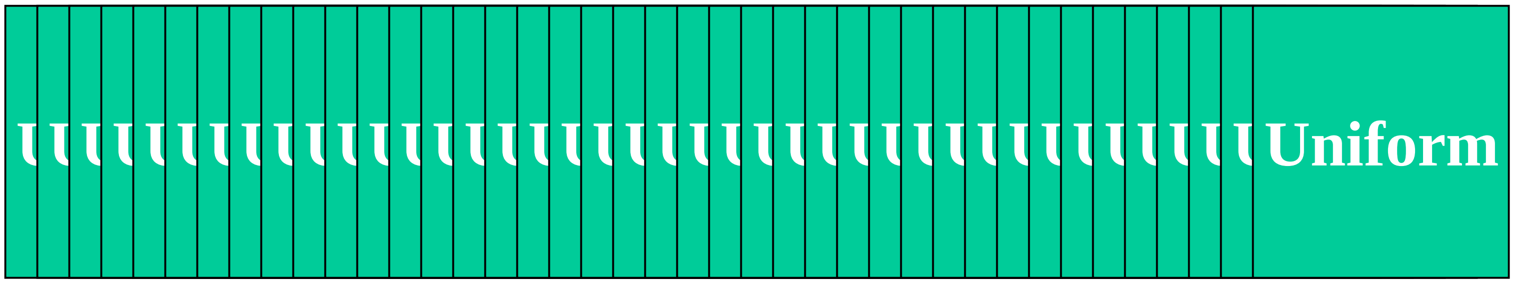
- Probability theory, as its commonly used, doesn’t cumulate gross uncertainty correctly
- Precision of the answer (measured as cv) depends strongly on the number inputs and not so strongly on their distribution shapes, even if they are uniforms or flat priors
- The more inputs, the tighter the answer

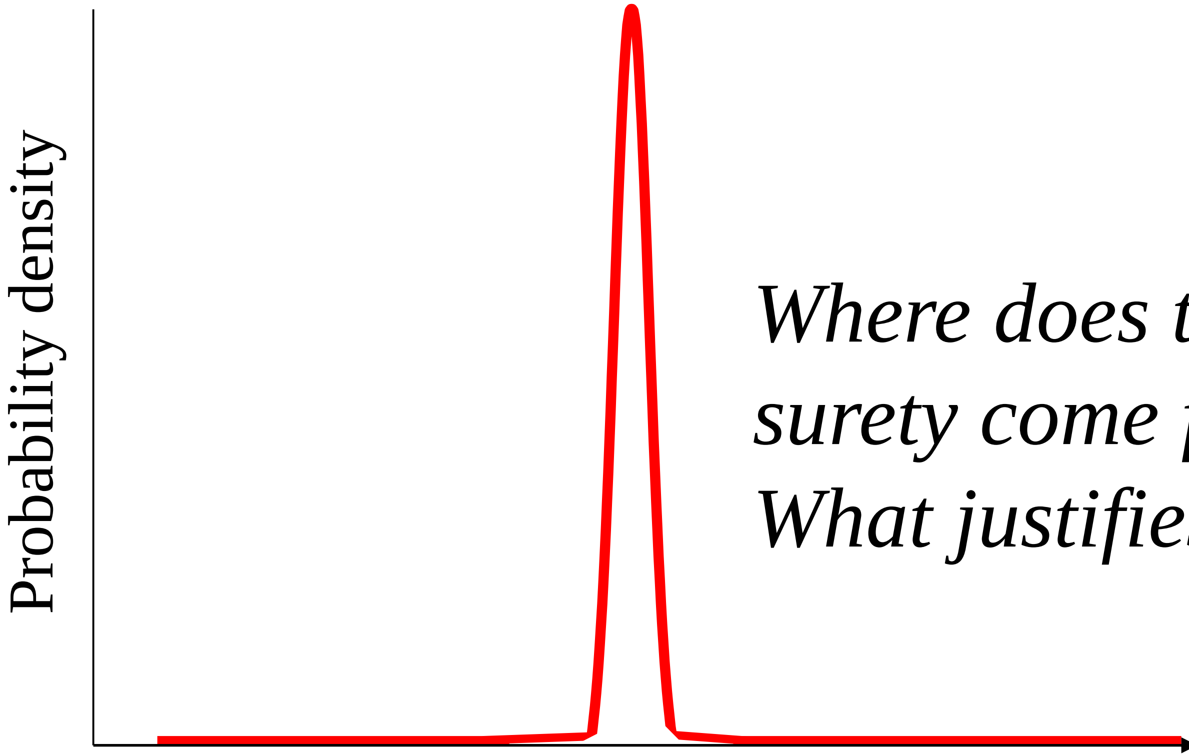
A few grossly uncertain inputs

Uniform + **Uniform** + **Uniform** + **Uniform**



A lot of grossly uncertain inputs...





Where does this surety come from? What justifies it?

Smoke and mirrors certainty

- Conventional probability theory, at least as it’s naively applied, seems to manufacture certainty out of nothing
- This is why some critics say probabilistic analyses are “smoke and mirrors”
- P-boxes give a vacuous answer if all you give them are vacuous inputs

The problem is wishful thinking

In practice, probabilists often use conventions and make assumptions that may be convenient but are not really justified:

1. Variables are independent of one another
2. Uniform distributions capture incertitude
3. Distributions are stationary (unchanging)
4. Specifications are perfectly precise

Untenable assumptions

- Distributions normal
- Uncertainties are small
- Sources of variation are independent
- Uncertainties can cancel each other out
- Linearized models good enough
- Most of the physics is known and modeled

Need ways to relax assumptions

- Possibly large, non-normal uncertainties
- Non-independent, or *unknown* dependencies
- Uncertainties that may not cancel
- Arbitrary mathematical operations
- Model uncertainty

Everyone makes assumptions

- But not all sets of assumptions are equal!

Point value	Linear function
Interval range	Monotonic function
Entire real line	Any function
Normal distribution	Independence
Unimodal distribution	Known correlation
Any distribution	Any dependence

- Like to discharge unwarranted assumptions

"Certainties lead to doubt; doubts lead to certainty"

Take-home messages

- Monte Carlo will always be useful, like Euclidean geometry
- Using bounding, you don't have to pretend you know a lot to get *quantitative* answers
- Approximation and bounding are often complementary
- Paying attention to measurement imprecision, censoring, and missingness requires new approaches
- Integrating them with small- and no-data problems, with model uncertainty, yields a non-Laplacian theory

What is known empirically	**Monte Carlo analysis**	**Probability bounds analysis**
Know only range of variable	Assume uniform distribution	Assume interval
Know some constraints about random variable	Select largest entropy distribution from all thus constrained distributions	Form envelope around class of distributions matching constraints
Uncertainty about distribution family or shape	Repeat analysis for other plausible distribution shapes	Form distribution-free p-box as envelope of all plausible distributions
Sample data	Form empirical distribution function (EDF)	Form nonparametric c-box or KS confidence limits around EDF
Variable follows known marginal distribution	Sample from particular distribution	Use particular distribution
Measurement uncertainty	Ignore it (usually), or perform sensitivity analysis	Express it in intervals and incorporate it into analysis

What is known empirically	Monte Carlo analysis	Probability bounds analysis
Non-detects	Replace non-detect with ½DL (detection limit)	Replace non-detect with interval [0, DL]
Know variables are independent	Assume independence	Assume (random-set) independence
Know magnitude of correlation	Simulate correlation from particular (but usually arbitrary) copula	Bound result from possible copulas with correlation, or use known copula
Know only the general sign (+ or –) of dependence	Assume some correlation of appropriate sign, or repeat analysis for different correlations	Bound result assuming only the sign of the dependence and specific or all possible copulas
Do not know the nature of the dependence	Assume independence (usually), or repeat analysis for different correlations	Bound result for all possible dependencies (Fréchet case)
Model uncertainty	Form stochastic mixture (vertical average) of distribution functions	Form envelope of distribution functions

Cheat sheet for the pba.r library

- plot, lines, show, summary
- mean, sd, var, median, quantile, left, right, prob, cut, percentile, iqr, random, range
- exp, log, sqrt, abs, round, trunc, ceiling, floor, sign, sin, cos, tan, asin, acos, atan, atan2, reciprocate, negate, +, -, *, /, pmin, pmax, ^, and, or, not, mixture, smin, smax

normal, etc.

histogram

quantiles

pointlist

MM<tab><tab>

ME<tab><tab>

ML<tab><tab>

CB<tab><tab>

NV<tab><tab>

Supported named distributions

bernoulli, beta, binomial, cauchy, chi, chisquared, delta, dirac, discreteuniform, exponential, exponentialpower, extremevalue, f, fishersnedecor, fishertippett, fisk, frechet, gamma, gaussian, geometric, generalizedextremevalue, generalizedpareto, gumbel, histogram, inversegamma, laplace, logistic, loglogistic, lognormal, logtriangular, loguniform, negativebinomial, normal, pareto, pascal, powerfunction, poisson, quantiles, rayleigh, reciprocal, shiftedloglogistic, skewnormal, student, trapezoidal, triangular, uniform, Weibull

Nonparametric p-boxes

maxmean, minmax, minmaxmean, minmean, meanstd, meanvar, minmaxmode, minmaxmedian, minmaxmedianismode, minmaxpercentile, minmaxmeanismedian, minmaxmeanismode, mmms, mmmv, posmeanstd, symmeanstd, uniminmax, unimmmv, unimmms

Synopsis: probability bounds analysis

How?

- specify what you are sure about
- establish bounds on probability distributions
- pick dependencies (no assumption, indep., perfect, etc.)

Why?

- account for uncertainty better than maximum entropy, etc.
- puts bounds on Monte Carlo results
- bounds get narrower with better empirical information

Why not?

- does not yield second-order probabilities
- best-possible results can sometimes be expensive to compute

Synopsis: imprecise probabilities

How?

- avoid sure loss, $\underline{P}(A) \leq \overline{P}(A)$
- be coherent, $\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B)$
- use natural extension (mathematical programming) to find consequences

Why?

- most expressive language for uncertainty of *all kinds*
- can provide expectations and conditional probabilities
- provides best possible results that do not lose information

Why not?

- requires mathematical programming
- can strain mathematical ability of the analyst

Software for probability bounds

- R libraries pbox.r and pba.r
- Add-in for Excel (NASA, beta version)
- RAMAS Risk Calc 4.0 (NIH, commercial)
- StatTool (Dan Berleant, freeware)
- PBDemo (NIH, freeware)
- Constructor (Sandia and NIH, freeware)
- Williamson and Downs (1990)

Web presentations and documents

Introduction to probability bounds analysis written for Monte Carlo users
http://www.ramas.com/pbawhite.pdf

Introduction of probability bounds analysis to interval researchers
http://www2.imm.dtu.dk/~km/int-05/Slides/ferson.ppt

Gert de Cooman's gentle introduction to imprecise probabilities
http://maths.dur.ac.uk/~dma31jm/durham-intro.pdf

Fabio's Cozman's introduction to imprecise probabilities
http://www.cs.cmu.edu/~qbayes/Tutorial/quasi-bayesian.html

Notes from a week-long summer school on imprecise probabilities
http://idsia.ch/~zaffalon/events/school2004/school.htm

Introduction to p-boxes and related structures
http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf

Handling dependencies in uncertainty modeling
http://www.ramas.com/depend.zip

Introduction to Bayesian and robust Bayesian methods in risk analysis
http://www.ramas.com/bayes.pdf

Statistics for data that may contain interval uncertainty
http://www.ramas.com/intstats.pdf

Topical websites

- Intervals and Probability Distributions
  http://class.ee.iastate.edu/berleant/home/ServeInfo/Interval/intprob.html
- Imprecise Probabilities Project
  http://ippserv.rug.ac.be/home/ipp.html
- Sandia National Laboratory's Epistemic Uncertainty Project
  http://www.sandia.gov/epistemic/
- R software for confidence boxes
  https://sites.google.com/site/confidenceboxes/
- Applied Biomathematics' Risk Calc website
  http://www.ramas.com/riskcalc.htm
- Society for Imprecise Probabilities Theory and Applications
  http://www.sipta.org/

References

Barmish, B.R. and C. M. Lagoa. 1995. The uniform distribution: a rigorous justification for its use in robustness analysis. Technical Report ECE-95-12, Electrical and Computer Engineering, University of Wisconsin, Madison.

Barmish, B.R. and C. M. Lagoa. 1996. The uniformity principle: a new tool for probabilistic robustness analysis. Proceedings of the 30th Annual Conference on Information Science and Systems. Princeton, New Jersey.

Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Ben-Haim, Y. 2001. *Information-gap Decision Theory*, Academic Press, San Diego.

Bukowski, J., L. Korn and D. Wartenberg. 1995. Correlated inputs in quantitative risk assessment: the effects of distributional shape. *Risk Analysis* 15:215-219.

de Cooman, G., T. L. Fine, et al. 2001. *ISIPTA '01. Second International Symposium on Imprecise Probabilities and Their Applications*, Robert Purcell Community Center of Cornell University, Ithaca, NY, USA, Shaker Publishing BV.

Copeland, T.L., A.M. Holbrow, J.M Otani, K.T. Conner and D.J. Paustenbach 1994. Use of probabilistic methods to understand the conservativism in California's approach to assessing health risks posed by air contaminant. *Journal of the Air and Waste Management Association* 44: 1399-1413.

Couso, I, S. Moral and P. Walley. 1999. Examples of independence for imprecise probabilities. Pages 121-130 in *Proceedings of the First International Symposium on Imprecise Probability and Their Applications*. G. de Cooman, F.F. Cozman, S. Moral and P. Walley (eds.), Imprecise Probabilities Project, Universiteit Gent.

Cox, L.A. Jr., D. Babayev, and W. Huber. 2005. Some limitations of qualitative risk rating systems. *Risk Analysis* 25(3):651-662.

Dempster, A.P. 1967. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics* 38: 325–339.

Feller, W. 1948. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics* 19: 177–189.

Ferson, S., V. Kreinovich, L. Ginzburg, K. Sentz and D.S. Myers. 2002. *Constructing probability boxes and Dempster-Shafer structures*. Sandia National Laboratories, Technical Report SAND2002-4015, Albuquerque, New Mexico. Available at http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf.

Ferson, S. 2002. *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*. Lewis Publishers, Boca Raton.

Fetz, T. 2001. Sets of joint probability measures generated by weighted marginal focal sets. Pages 171-178 in *Proceedings of the Second International Symposium on Imprecise Probability and Their Applications*. G. de Cooman, T.L. Fine and T. Seidenfeld (eds.), Shaker Publishing, Maastricht.

Frank, M. 1979. On the simultaneous associativity of and $F(x, y)$ and $x+y-F(x, y)$. *Aequationes Mathematicae* 19: 194-226.

Frank, M.J., Nelsen, R.B., and Schweizer, B. 1987. Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* 74, 199-211.

Fréchet, M. 1935. Généralisations du théorème des probabilités totales. *Fundamenta Mathematica* 25: 379-387.

Fréchet, M. 1951. *Sur les tableaux de corrélation dont les marges sont données*. Ann. Univ. Lyon, Sect. A 9: 53-77.

Grosof, B.N. 1986. An inequality paradigm for probabilistic knowledge: the logic of conditional probability intervals. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.

Hailperin, T. 1986. *Boole's Logic and Probability*. North-Holland, Amsterdam.

Helsel, D.R. 1990. Less than obvious: statistical treatment of data below the detection limit. *Environmental Science and Technology* 24: 1766–1774.

Insua, D.R. and F. Ruggeri (eds.) 2000. *Robust Bayesian Analysis*. Lecture Notes in Statistics, Volume 152. Springer-Verlag, New York.

Kolmogorov [Kolmogoroff], A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics* 12: 461–463.

Kyburg, H.E., Jr. 1998. "Interval Valued Probabilities," http://ippserv.rug.ac.be/documentation/interval_prob/interval_prob.html, The Imprecise Probabilities Project, edited by G. de Cooman and P. Walley, http://ippserv.rug.ac.be/home/ipp.html.

Lagoa, C.M., and B.R. Barmish. [n.d.] Monte Carlo simulation without probability distributions: a tutorial. [unpublished manuscript].

Lagoa, C.M., P.S. Shcherbakov and B.R. Barmish. 1998. Probabilistic enhancement of classical robustness margins: the unirectangularity concept. *Systems and Control Letters* 35: 31-43.

Loui, R.P. 1986. Interval based decisions for reasoning systems. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.

Mishra, S. 2000. Uncertainty propagation using the point estimate method. *Calibrated Reliability in Groundwater Modeling*, IAHS.

Moore, R.E. 1966. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Moore, R. 1979. *Methods and Applications of Interval Analysi*s. SIAM, Philadelphia.

Nelsen, R.B. 1999. *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer-Verlag, New York.

Rosenblueth, E. 1975. Point estimates for probability moments. *Proceedings of the National Academy of Sciences USA* 72: 3812-3814.

Rowe, N.C. 1988. Absolute bounds on the mean and standard deviation of transformed data for constant-sign-derivative transformations. *SIAM Journal of Scientific Statistical Computing* 9: 1098–1113.

Rustichini, A. 2005. *Science* 310: 1624.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

Shlyakhter A. 1994. Improved framework for uncertainty analysis: accounting for unsuspected errors. *Risk Analysis* 14(4):441-447.

Smith, A.E., P.B. Ryan and J.S. Evans. 1992. The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Analysis* 12:467-474.

Tessem, B. 1992. Interval probability propagation. *International Journal of Approximate Reasoning* 7: 95-120.

Vesely, W.E., F.F. Goldberg, N.H. Roberts, D.F. Haasl. 1981. *Fault Tree Handbook*. Nuclear Regulatory Commission, Washington, DC.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Walley, P. 1996. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society* 58: 3-37.

Walley, P. 2000. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning* 24: 125-148.

Walley, P. and T.L. Fine 1982. Towards a frequentist theory of upper and lower probability. *Annals of Statistics* 10: 741-761.

Whitt, W. 1976. Bivariate distributions with given marginals. *The Annals of Statistics* 4:1280-1289.

Williamson, R.C. and Downs, T. 1990. Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4 89-158.

Wise. B.P. and M. Henrion 1986. A framework for comparing uncertain inference systems to probability. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.

Yager, R.R. 1986. Arithmetic and other operations on Dempster-Shafer structures. *International Journal of Man-machine Studies* 25: 357-366.