Computing with confidence

Scott Ferson
Applied Biomathematics

Reliable Engineering Computing, Ruhr-Universität Bochum, Germany, 17 June 2016

We need distributions

- Risk analyses
- Safety assessments
- Reliability analysis
- Environmental models
- Financial forecasts
- Uncertainty modeling

Problem: selecting distributions

- How should we chose a distribution given limited sample or constraint information?

- And what should we do when the available data and tenable assumptions do not specify a single distribution to use?

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?

Two related problems

- Estimating the <u>distribution</u> for *x*-values
  - Observable values
- Estimating <u>parameters</u> for the *x*-distribution
  - Unobservable quantities
- We need solutions for both problems

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

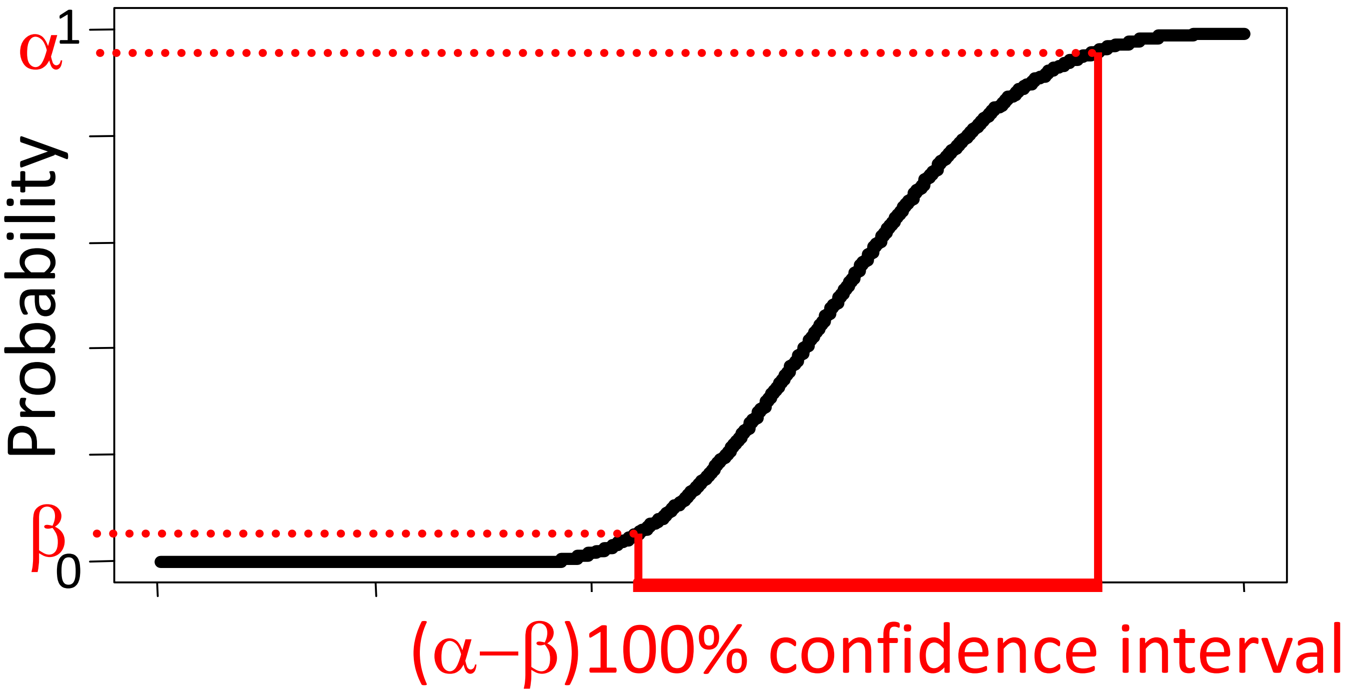
- Permit mathematical calculations
- But lack guarantees ensuring long-run statistical performance
- Many engineers are reluctant to use Bayesian methods

Confidence distributions

- Not widely used in engineering or statistics
- Introduced by Cox in the 1950s
- Closely related to other better-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*

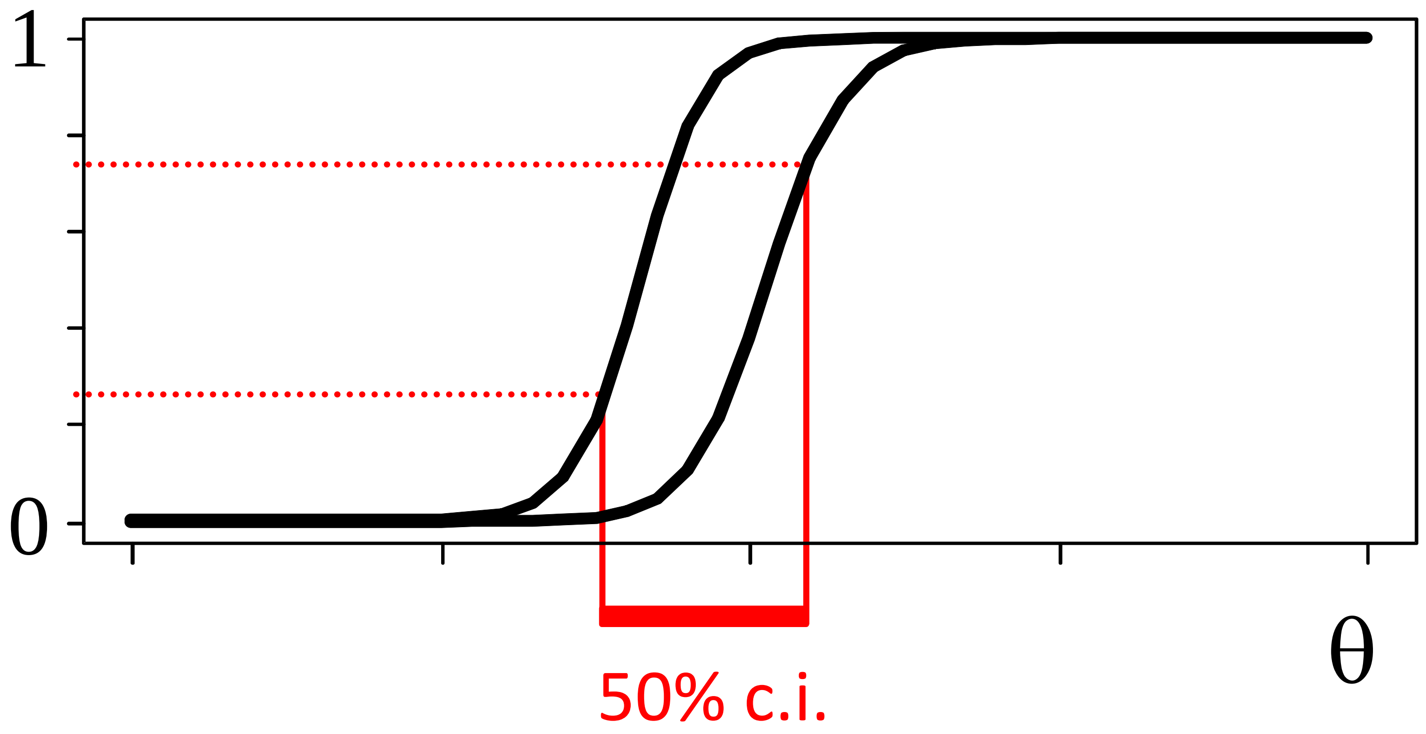
Confidence distributions

- 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)

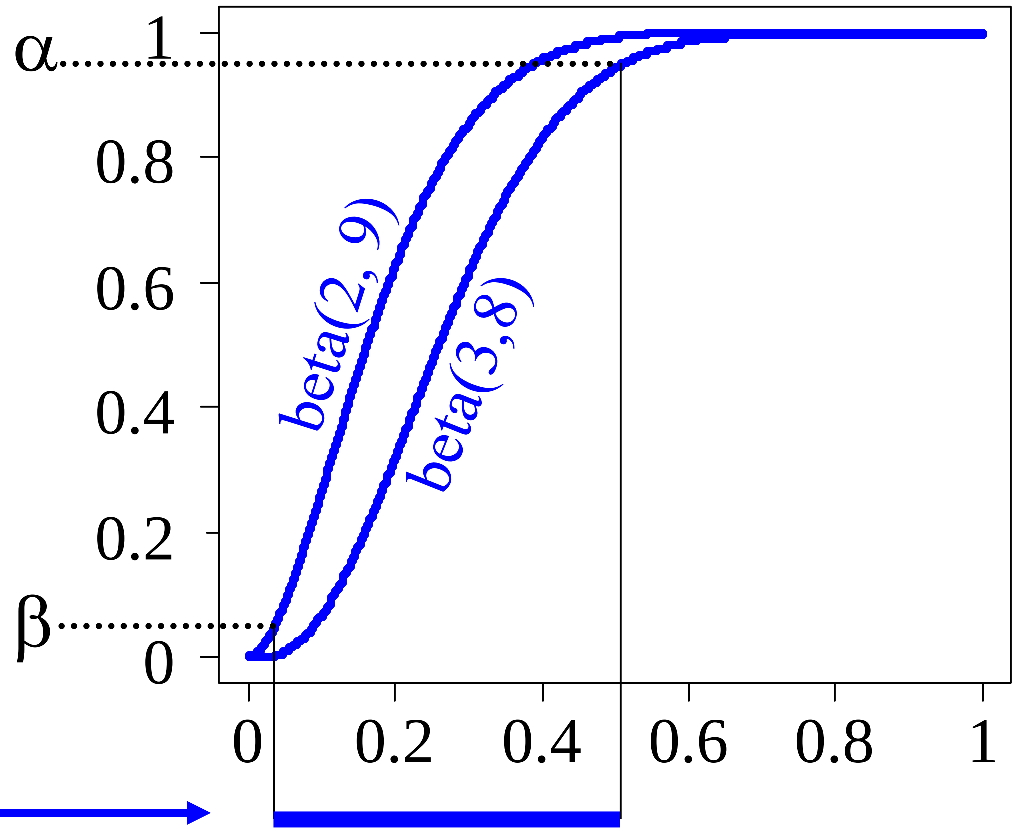
Binomial rate *p* for *k* of *n* trials

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

Data

$k = 2$

$n = 10$



(α–β)100% confidence interval for *p*

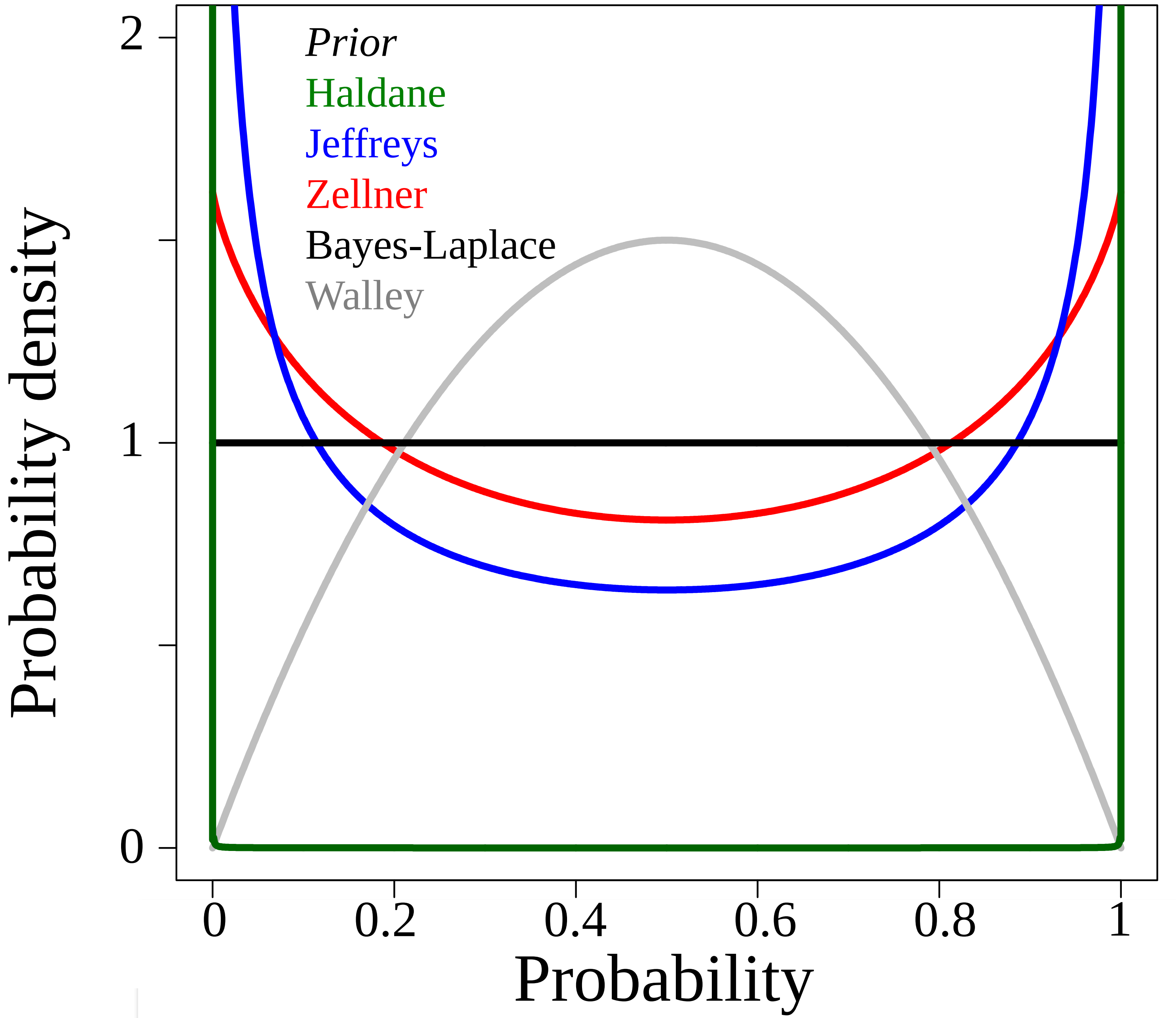
If 1–α = β, result is identical to classical Clopper–Pearson interval

How does the Bayes analysis compare?

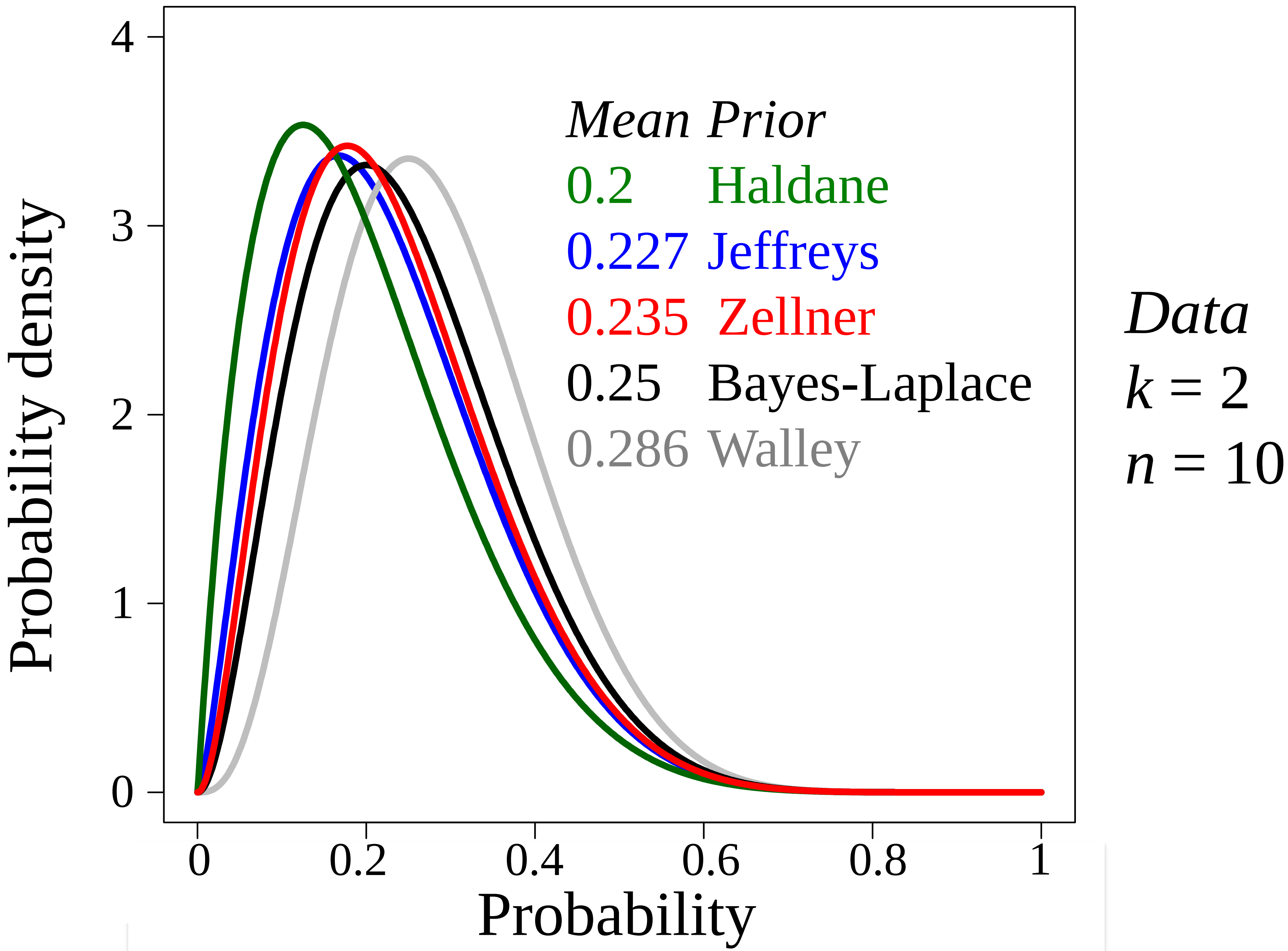
- No such thing as *the* Bayes analysis

- There are always many possible analyses
  - Different priors, which yield different answers
  - When data sets are small, the differences are big

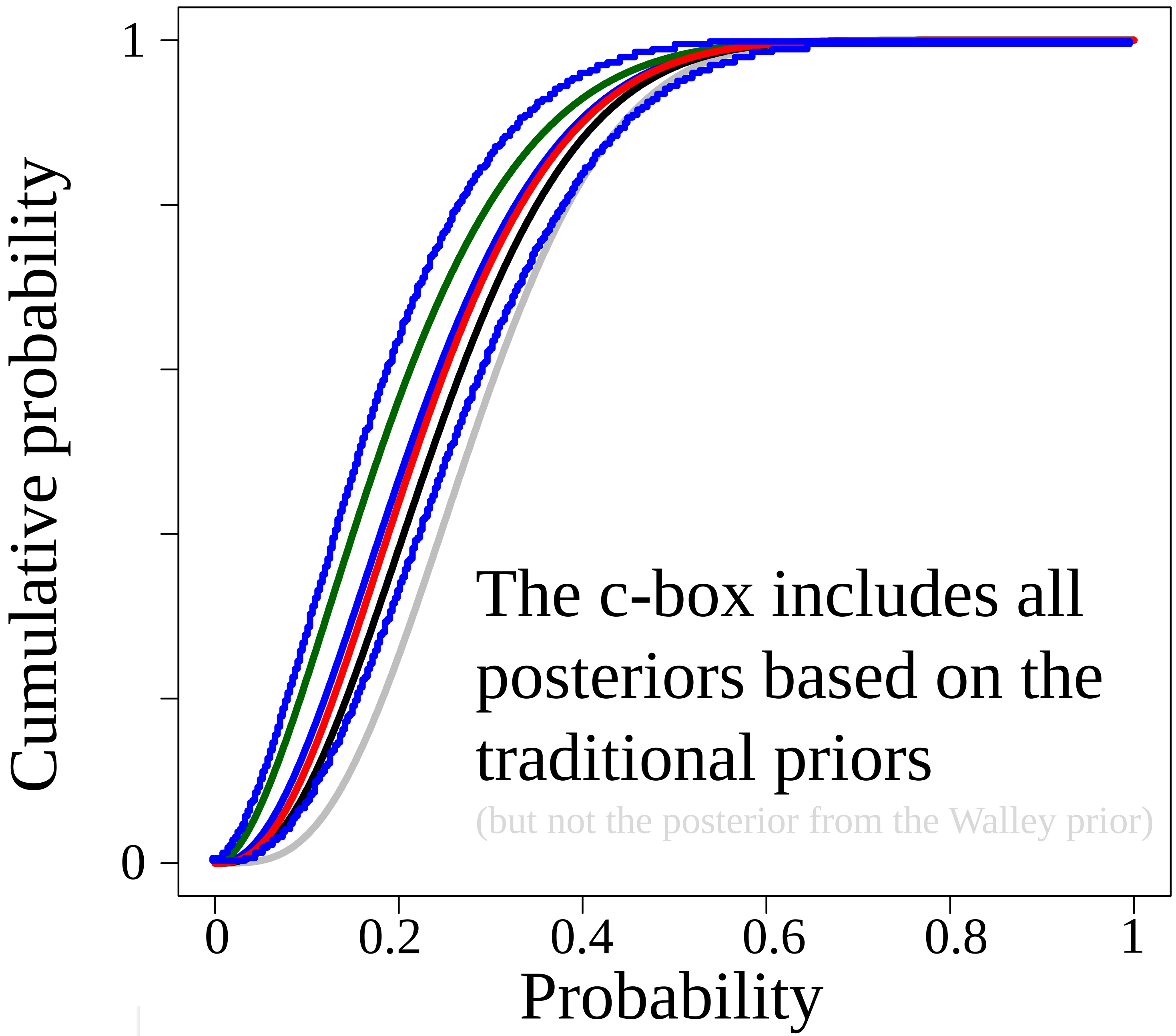
- For binomial rate there are four or five priors Bayesians have not been able to chose among



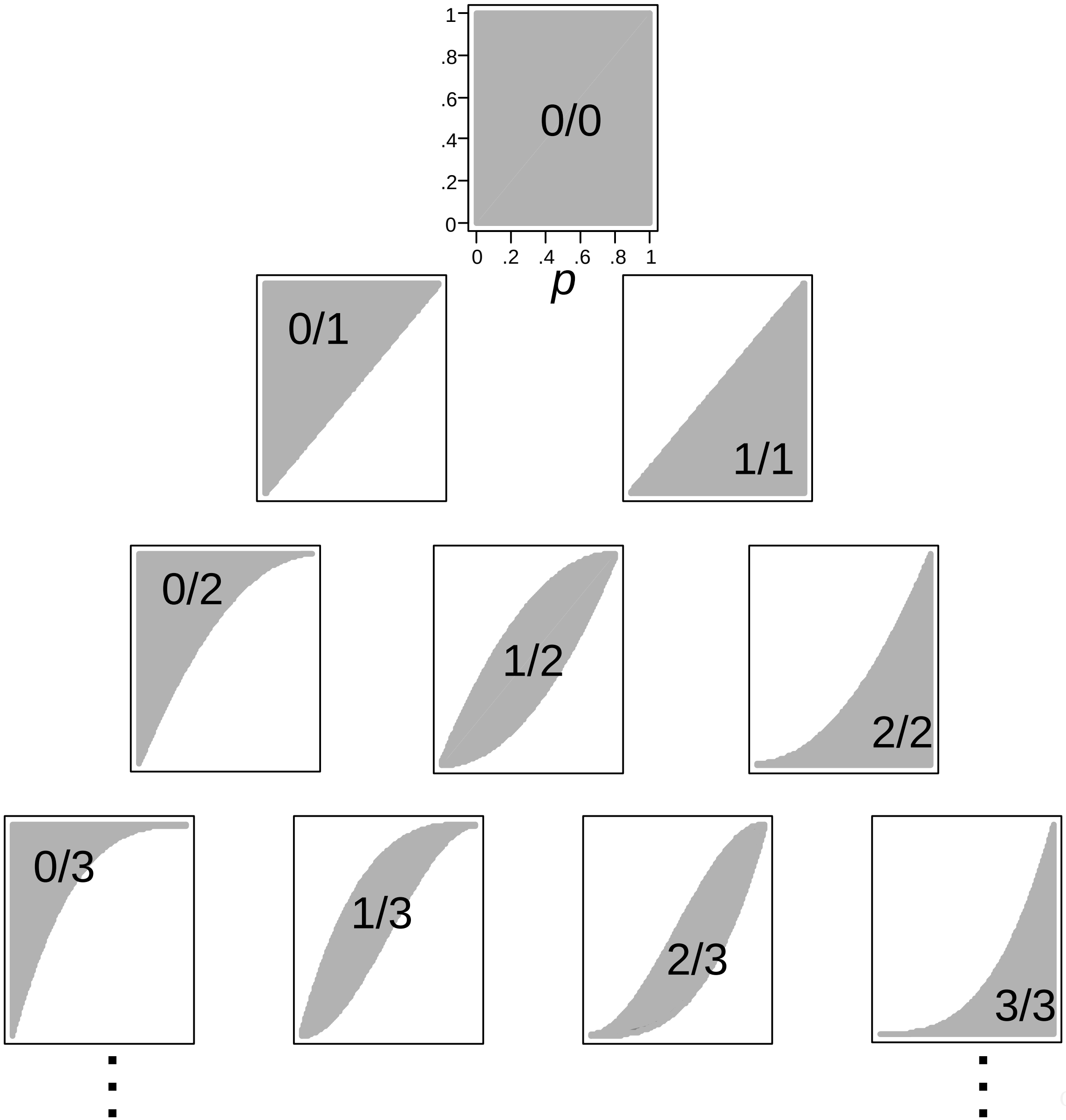
Prior
Haldane
Jeffreys
Zellner
Bayes-Laplace
Walley
Probability density
2
1
0
0
0.2
0.4
0.6
0.8
1
Probability



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
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1
Probability


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

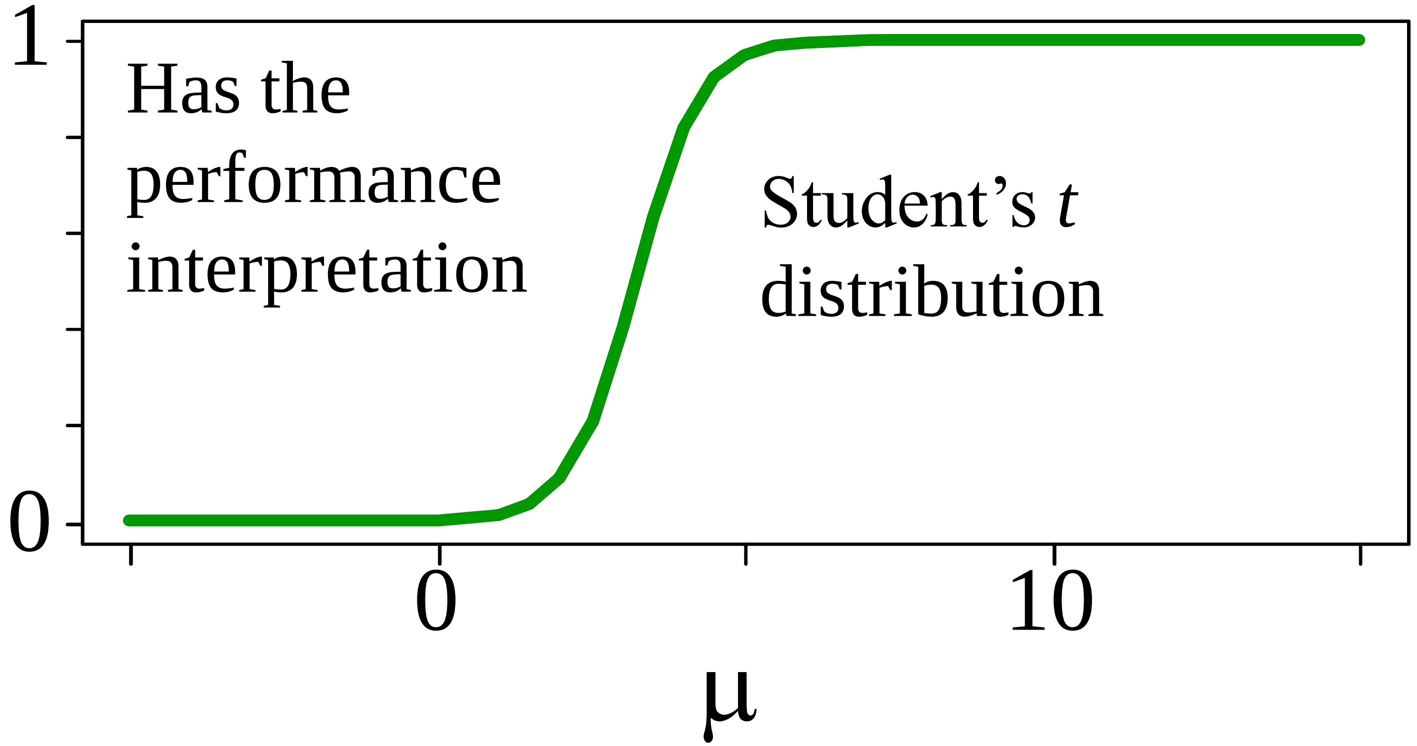


Example: normal mean

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

Data

8
5.5
–1.3
3.5
0.8
2.8
1.8
2.2
3.5
5.3

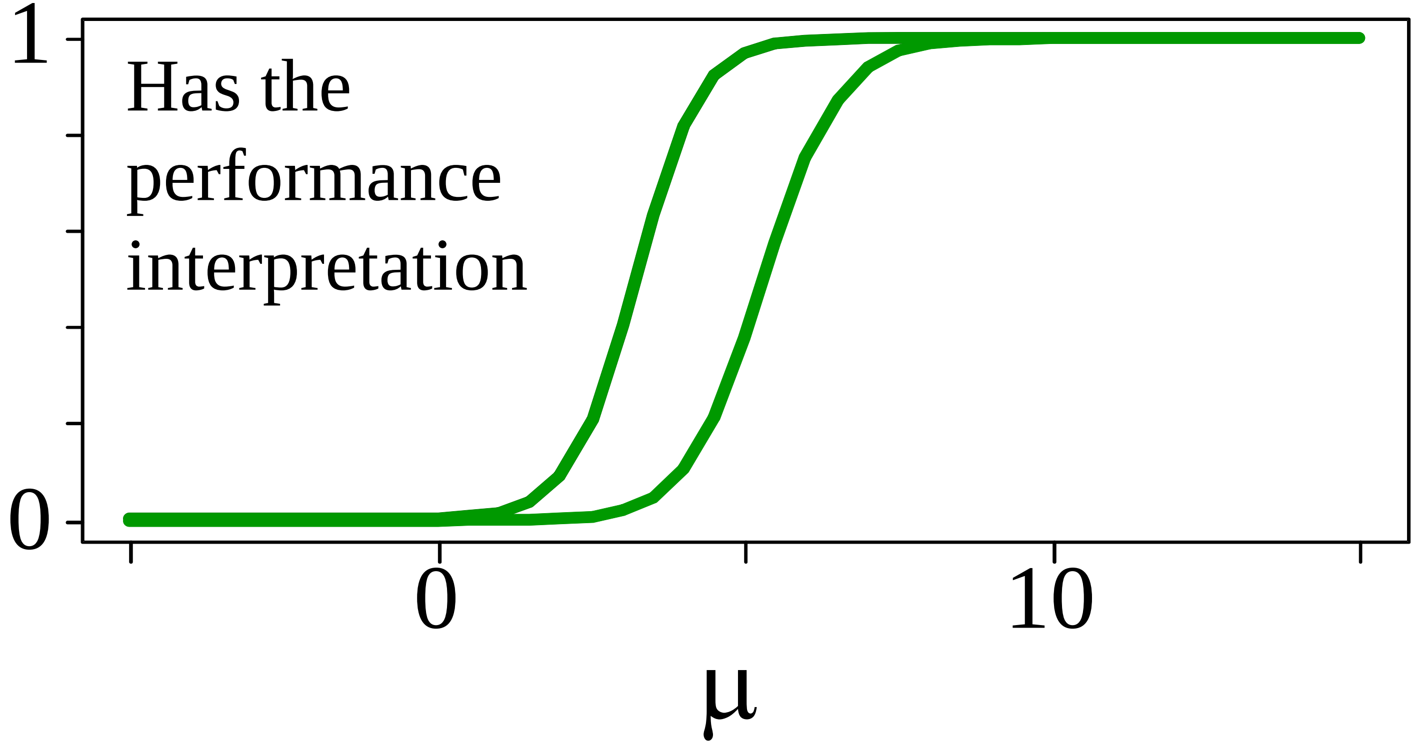


Example: normal mean

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

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 distribution shape
- Traditional approaches based on pivots
- Many solutions have been worked out

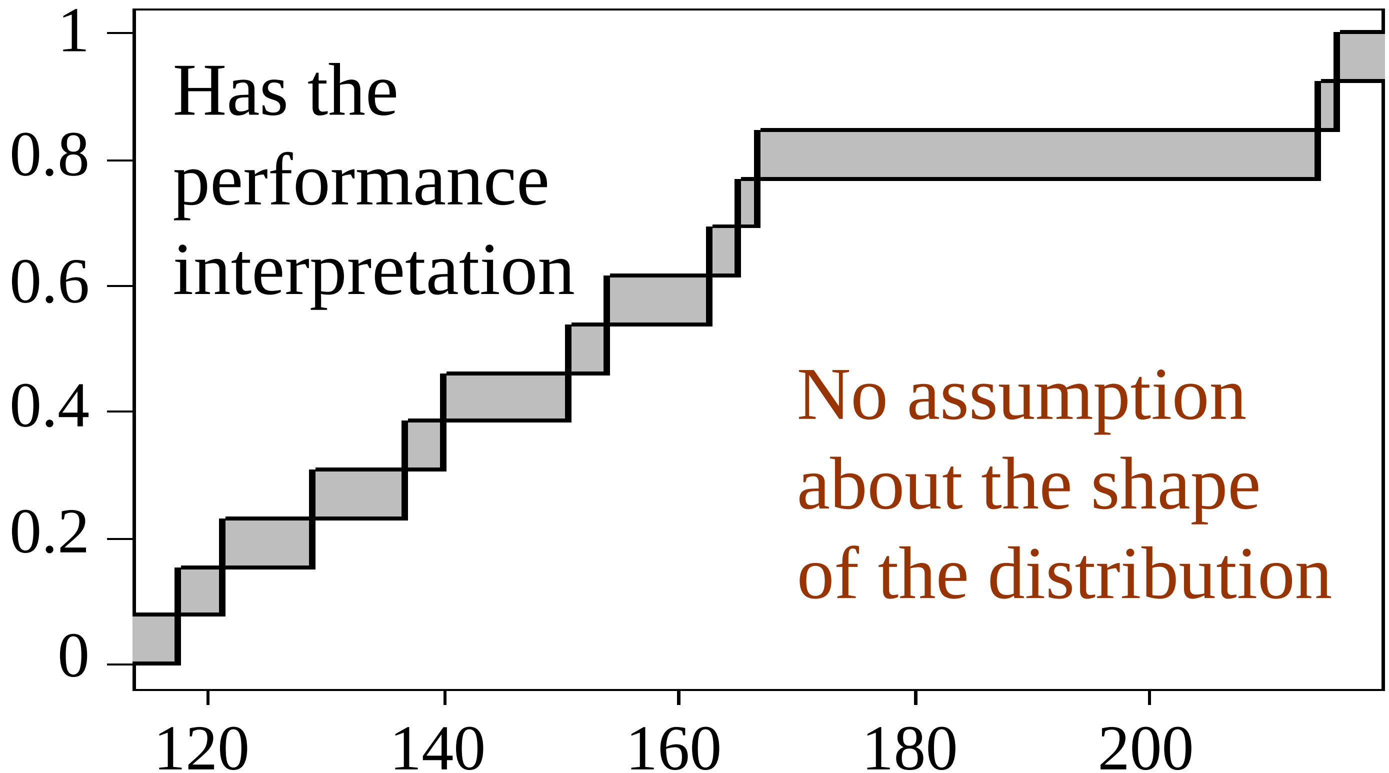
binomial(p, n), given n	normal(μ, σ)
binomial(p, n), given p	lognormal(μ, σ)
binomial(p, n)	gamma(a, b)
Poisson(p)	exponential(λ)
⋮	⋮

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 stochasticity (integrality of data)

Propagated as probability boxes

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

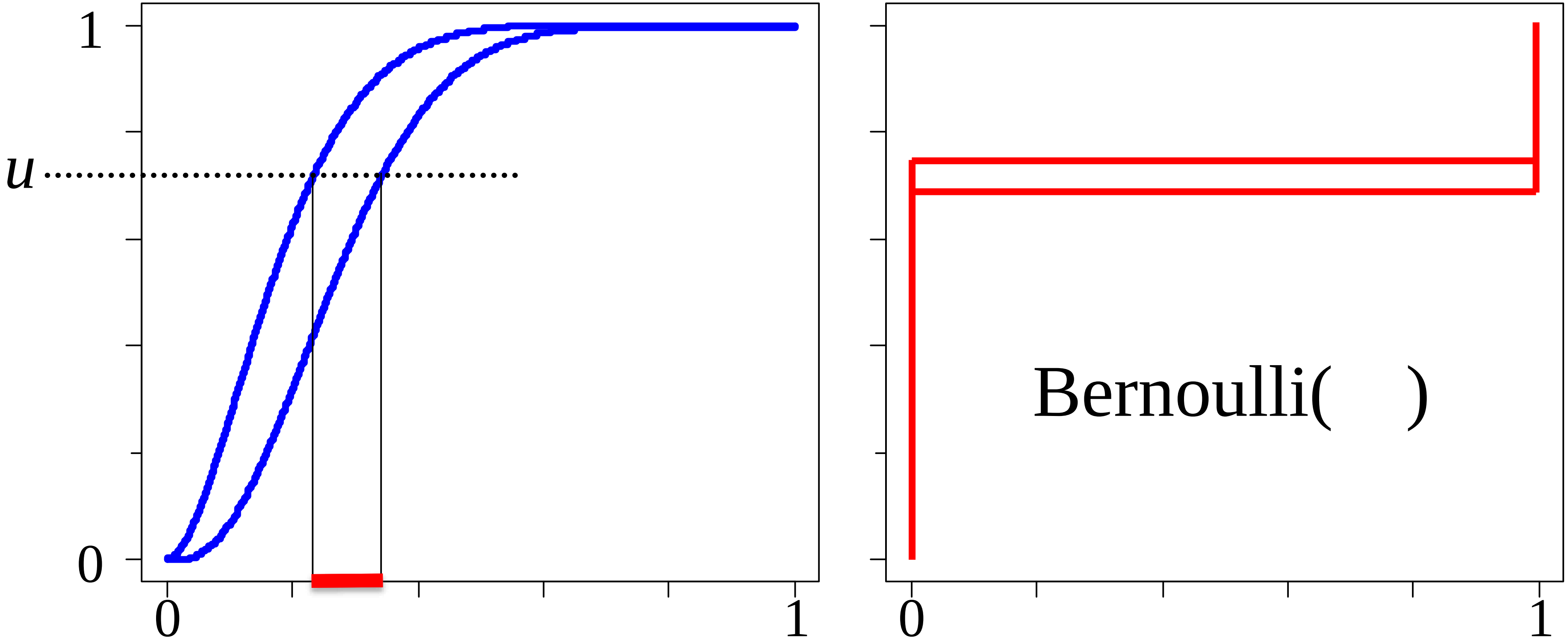
- *Results also have performance interpretations*

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

Prediction structures

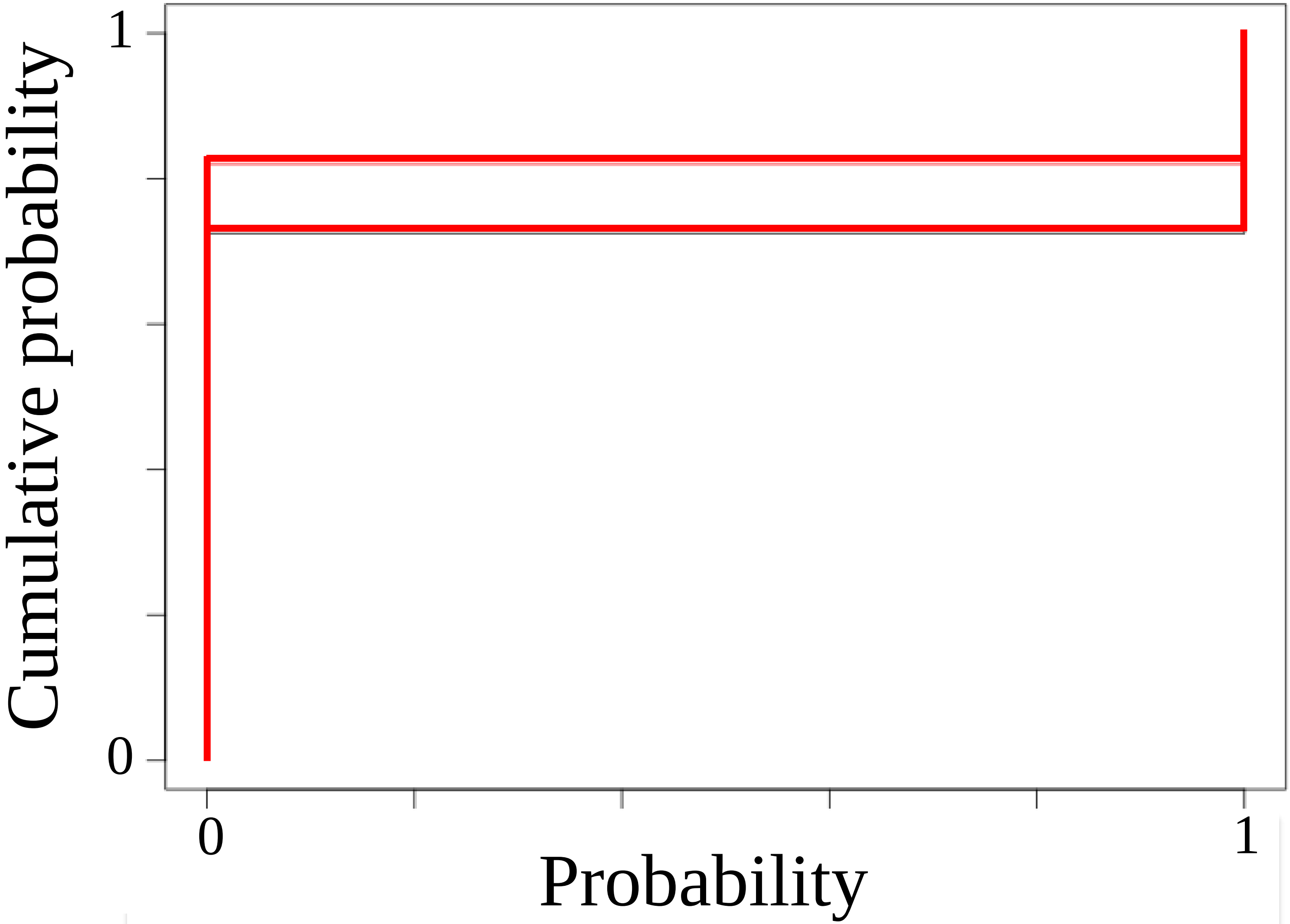
- C-boxes can model the uncertainty about the *underlying distribution* that generated the data
- This is a *composition* of the c-box through the probability model to make a p-box
- Stochastic mixture of p-boxes from interval parameters specified by slices from the c-box

Example: Bernoulli distribution

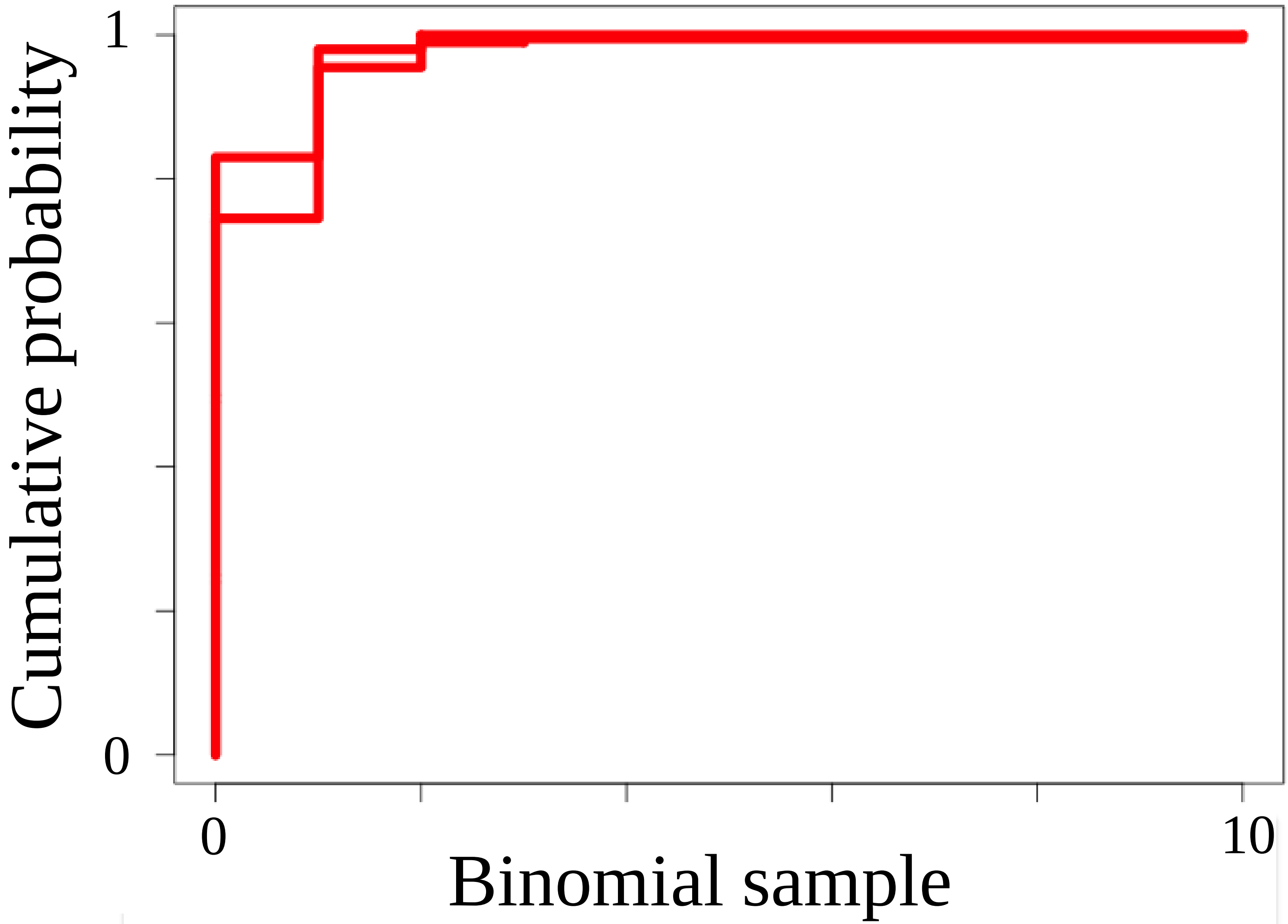


Each interval slice defines a p-box for the underlying distribution (rather than a precise distribution)

Average all such p-boxes



Beta-binomial predictive p-box

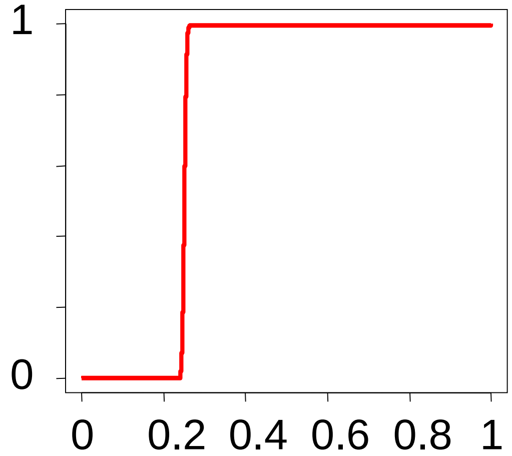


Prediction structures are p-boxes

- Also have confidence interpretation
  - Results are *prediction intervals* enclosing specified percentage of observable values, on average

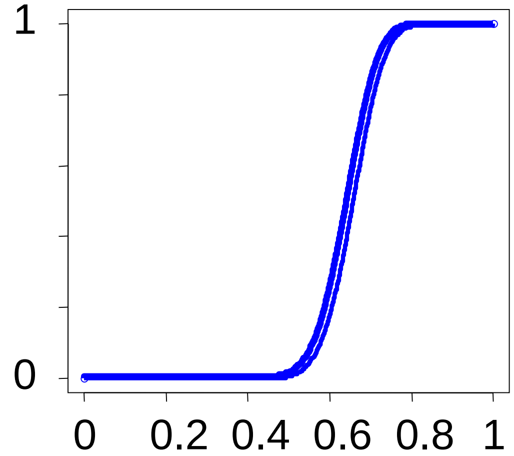
- Can also define analogous tolerance structures
  - *Tolerance intervals* are *X*% sure to enclose *Y*% of the population

Computing with c-boxes directly



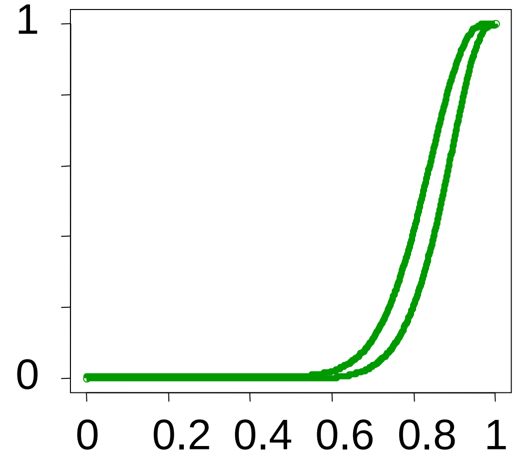
Plan A

25% fail



Plan B

39 out of 60 failed

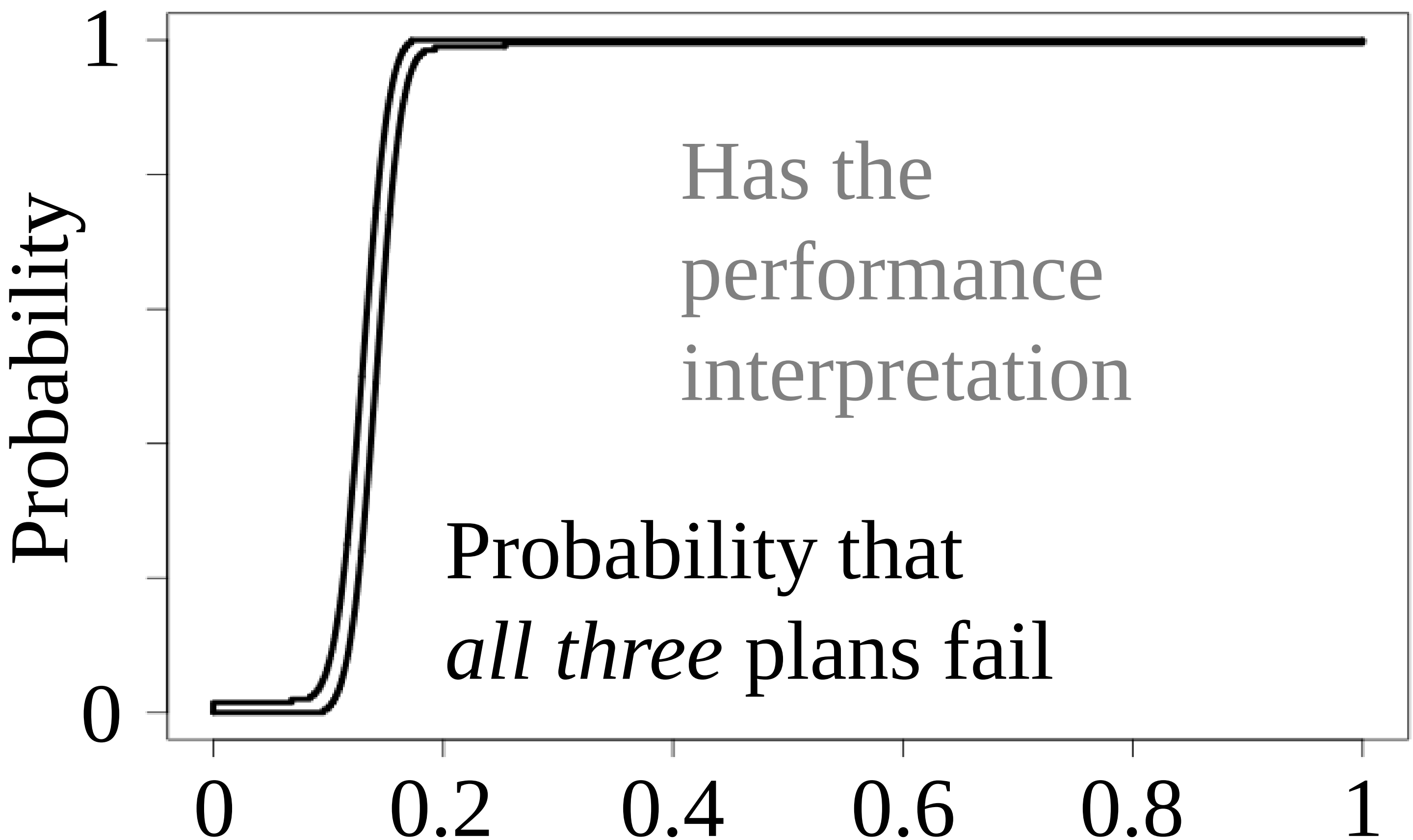


Plan C

17 out of 20 failed

What if we used all three plans independently?

Conjunction (AND)



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

Applies even with zero data

- There may be no *sample* data at all
- 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

Maximum likelihood

Maximum entropy

Bayesian inference

Estimation

Expert elicitation

“the great frontier of making things up”

Method of moments

PERT

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

C-boxes are Bayesian

Bayesian sensitivity analysis

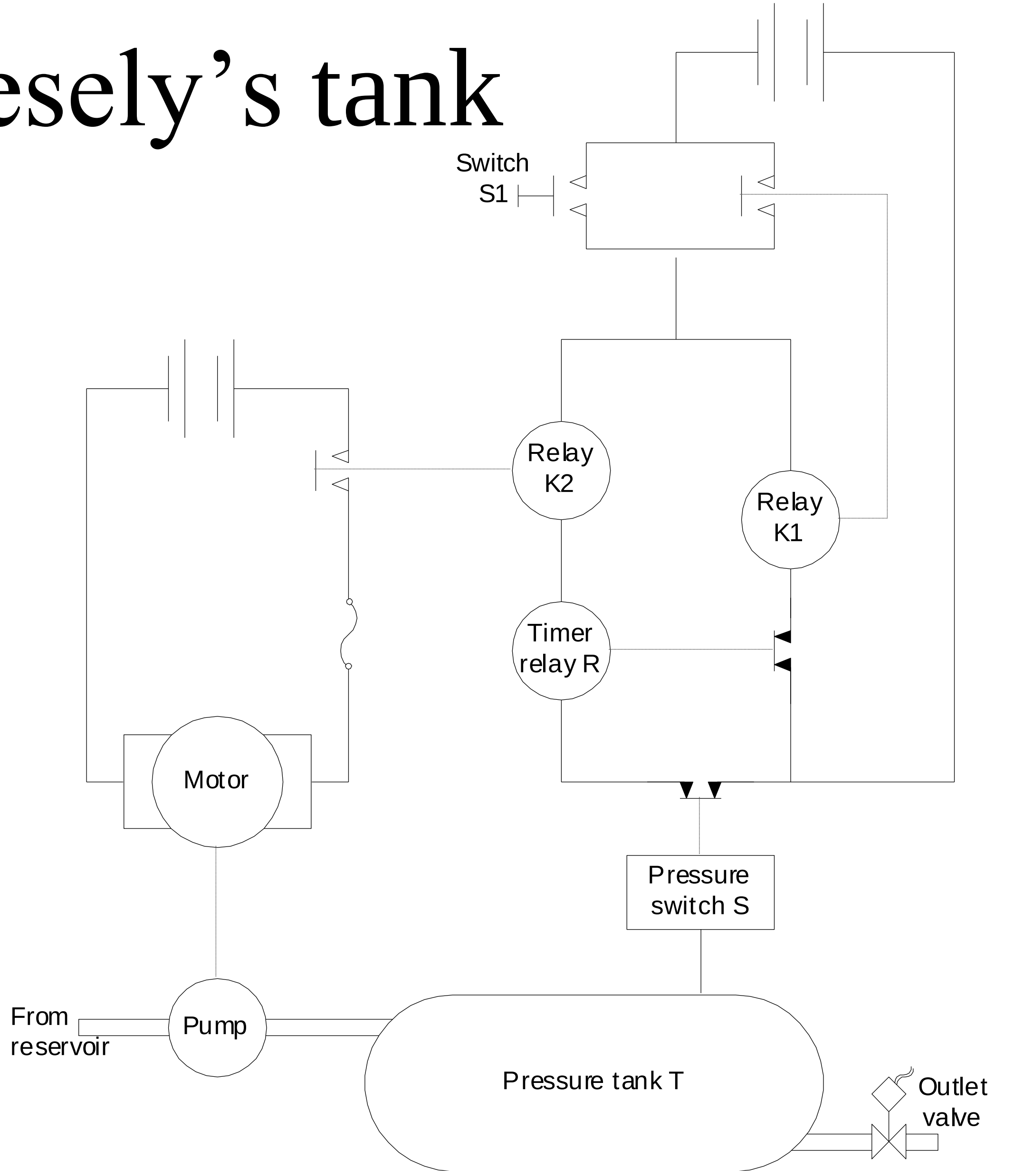
- Under robust Bayes approach, c-boxes can be thought of as Bayesian posteriors
  - Don't require specification of a unique prior
  - Have added feature of statistical performance
  - Imply posterior predictive distributions
  - Compatible with specifying a robust or precise prior when that's desirable

A single c-box or prediction box

- Expresses confidence (prediction) intervals at *all* possible α levels
- Including central, high-density, two-sided and left- or right-sided intervals at any desired level
- So you don't have to decide in advance which probability level or which kind you want

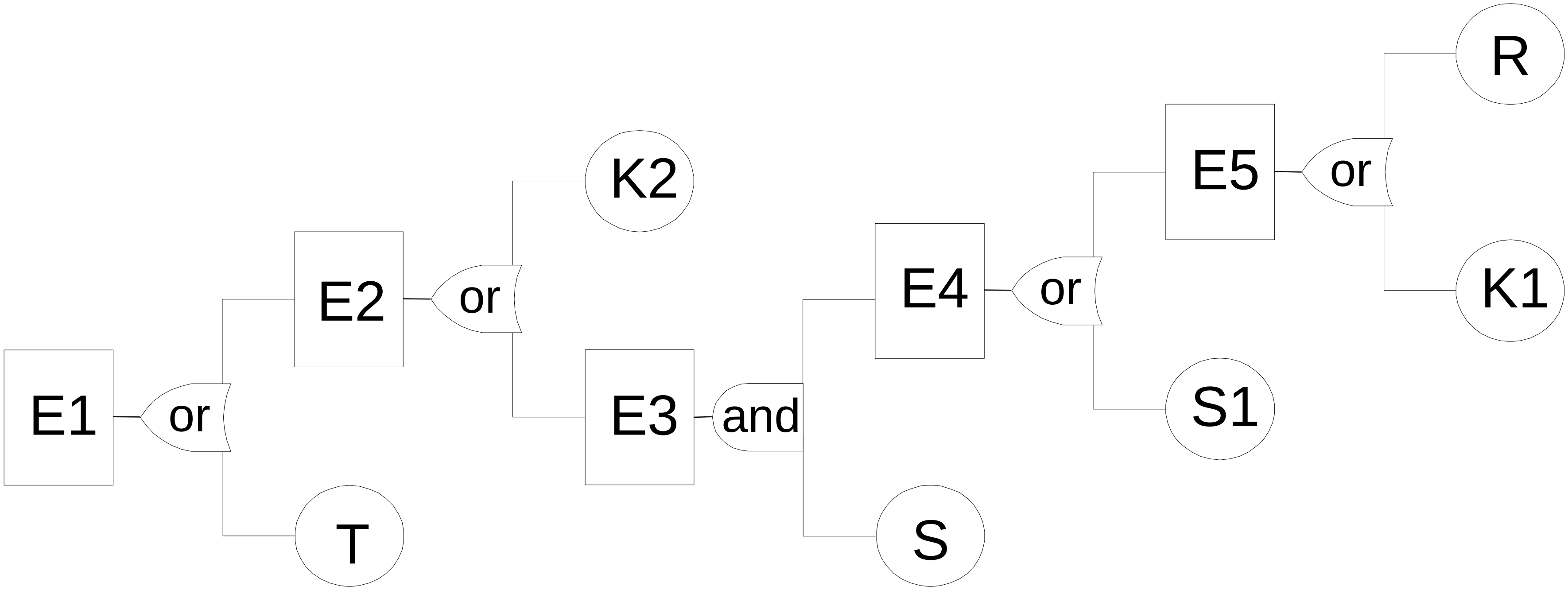
Vesely's tank

What's the chance the tank ruptures under pumping?



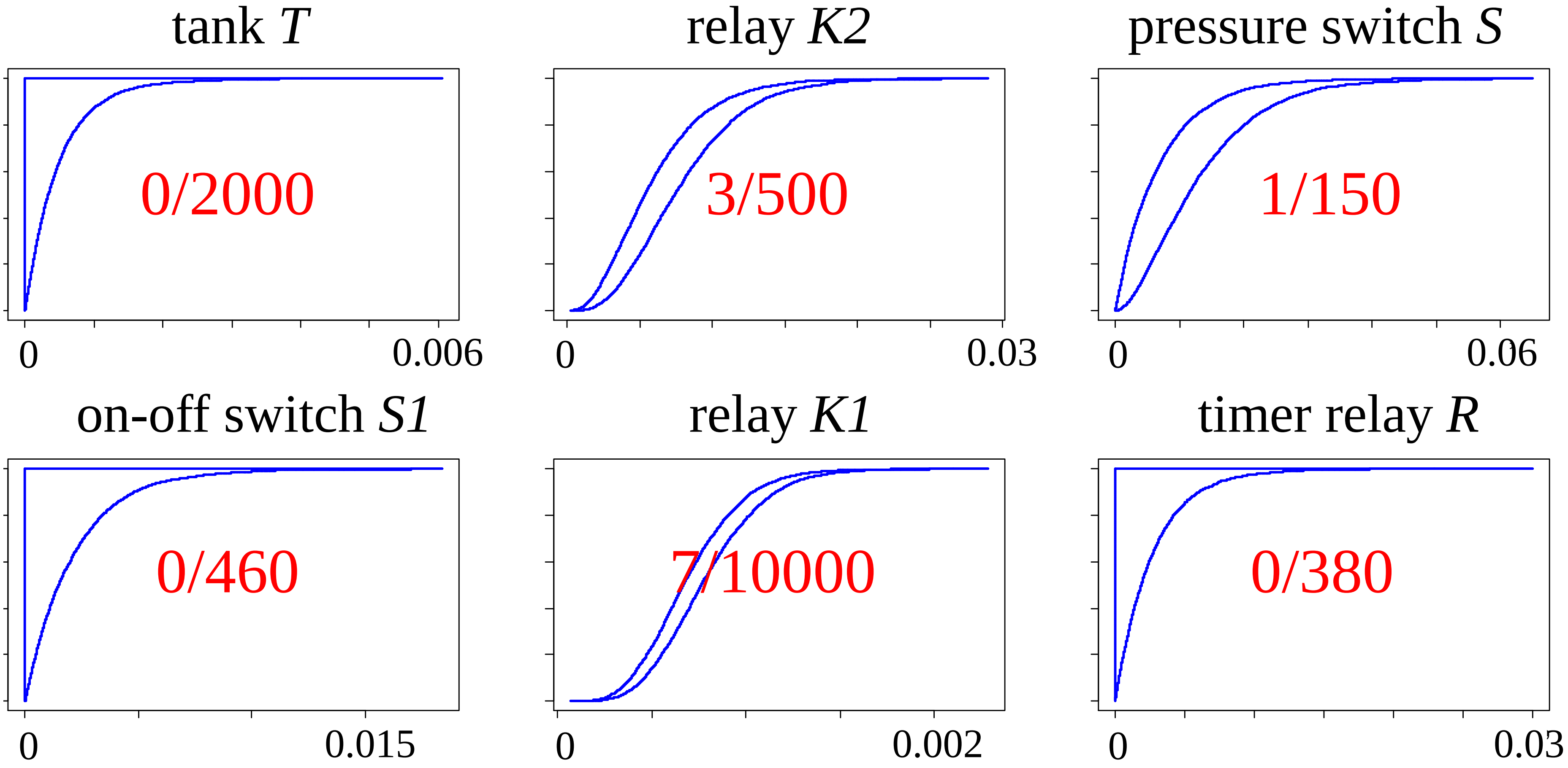
Vesely et al. 1981

Fault tree

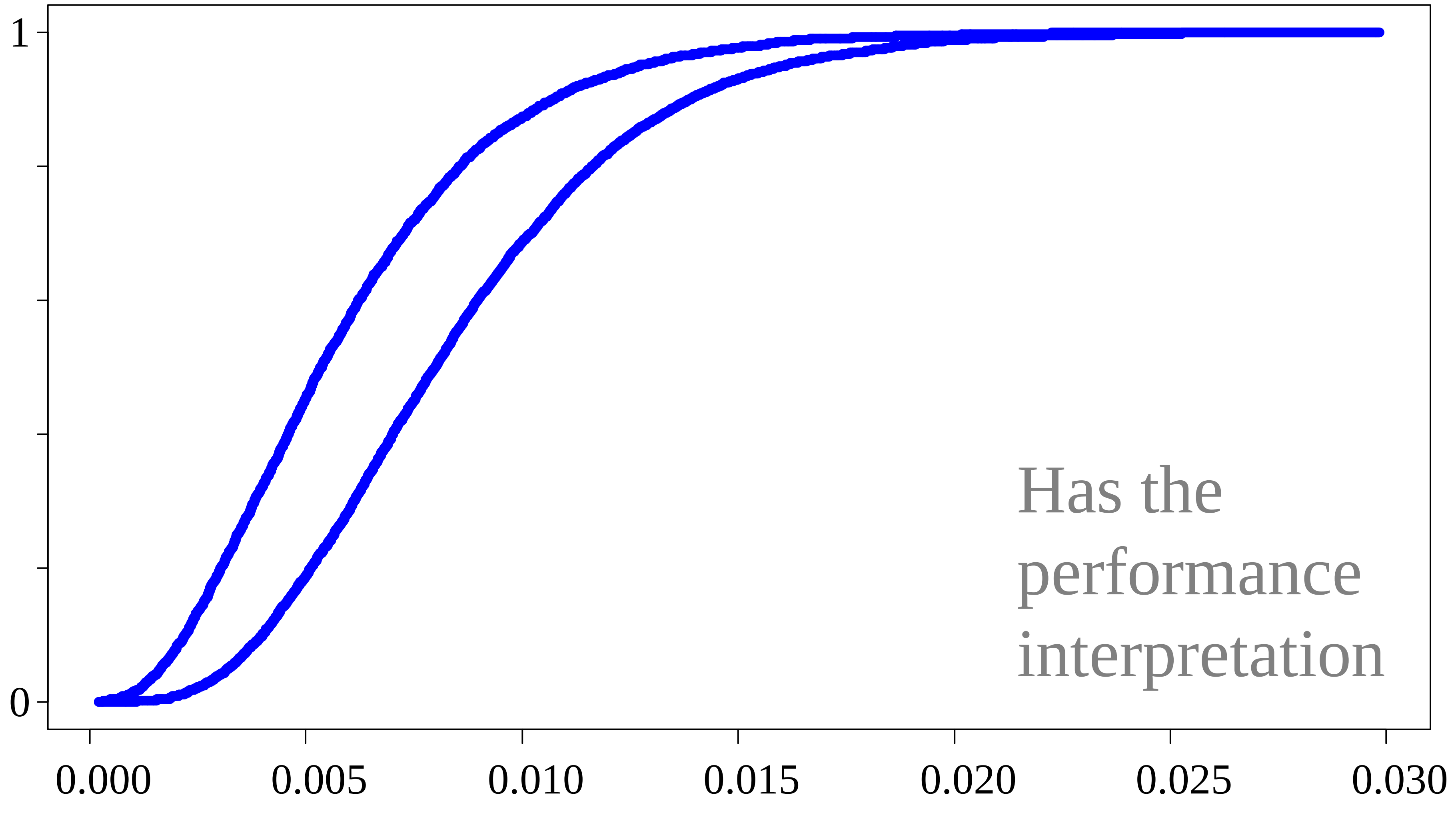


$$E1 = T \vee (K2 \vee (S \,\&\, (S1 \vee (K1 \vee R))))$$

Vesely's pressurized tank



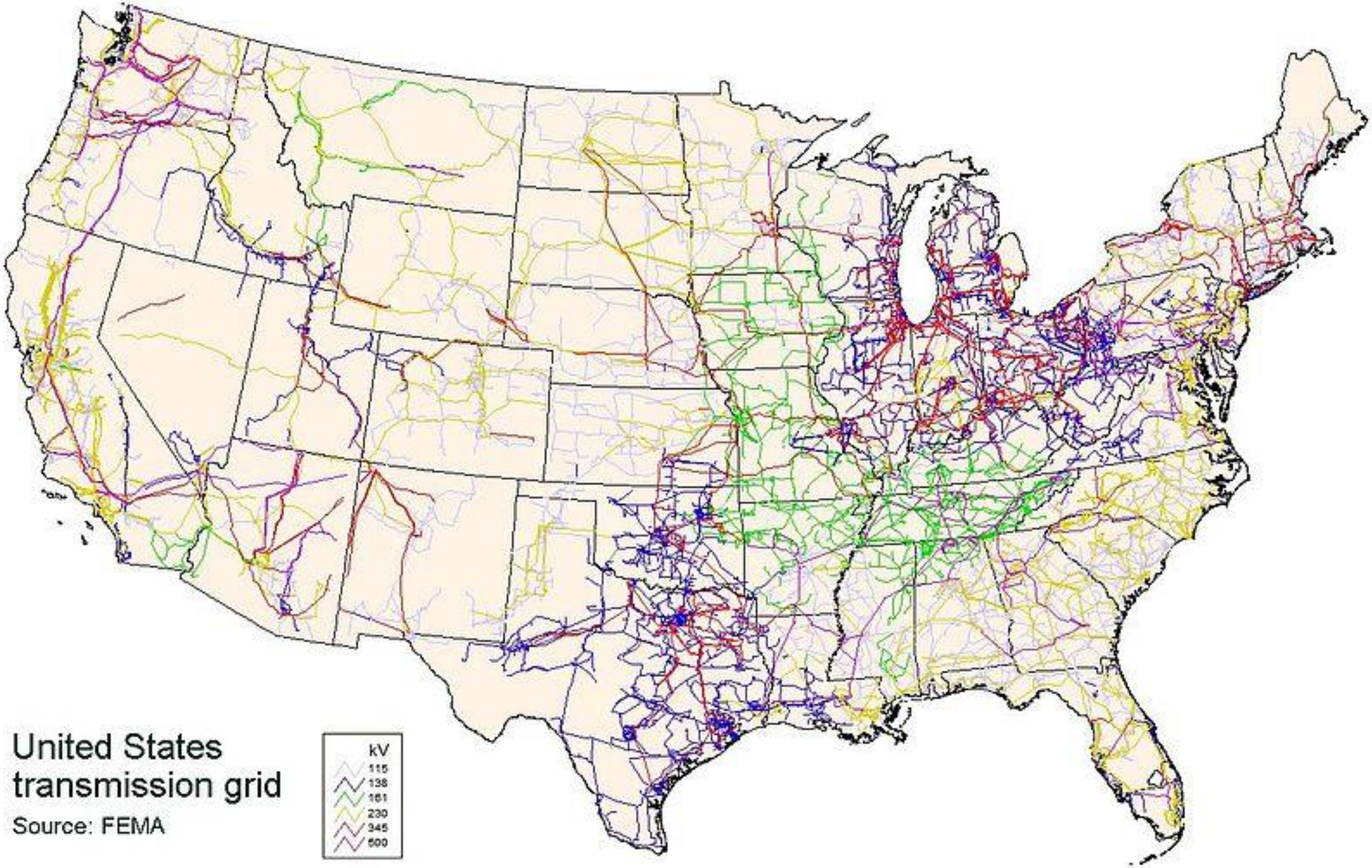
Top event tank rupture under pressurization *E1*



```
# Vesely's pressurized tank system from sparse sample data assuming independence
many = 10000
constant <- function(b) if (length(b)==1) TRUE else FALSE
precise <- function(b) if (length(b)==many) TRUE else FALSE
leftside <- function(b) if (precise(b)) return(b) else return(b[1:many])
rightside <- function(b) if (precise(b)) return(b) else return(b[(many+1):(2*many)])
sampleindices = function() round(runif(many)*(many-1) + 1)
pairsides <- function(b) {i=sampleindices(); return(env(leftside(b)[i],rightside(b)[i]))}
env <- function(x,y) if ((precise(x) && precise(y))) c(x,y) else stop('env error')
beta <- function(v,w) if ((v==0) && (w==0)) env(rep(0,many),rep(1,many)) else
      if (v==0) rep(0,many) else if (w==0) rep(1,many) else sort(rbeta(many, v, w))
kn <- function(k,n) return(pairsides(env(beta(k, n-k+1), beta(k+1, n-k))))

orI <- function(x,y) return(1-(1-x)*(1-y))
andI <- function(x,y) return(x*y)

t = kn(0, 2000)
k2 = kn(3, 500)
s = kn(1, 150)
s1 = kn(0, 460)
k1 = kn(7, 10000)
r = kn(0, 380)

e1 = orI(t, orI(k2, andI(s, orI(s1, orI(k1, r)))))
```

Northeast Blackout of 2003

- 55 million people affected
- Second only to the Southern Brazil Blackout of 1999 as the most widespread in history
- Traced to a software bug in a control room alarm system in Ohio
- A national Electric Reliability Organization was created in the aftermath

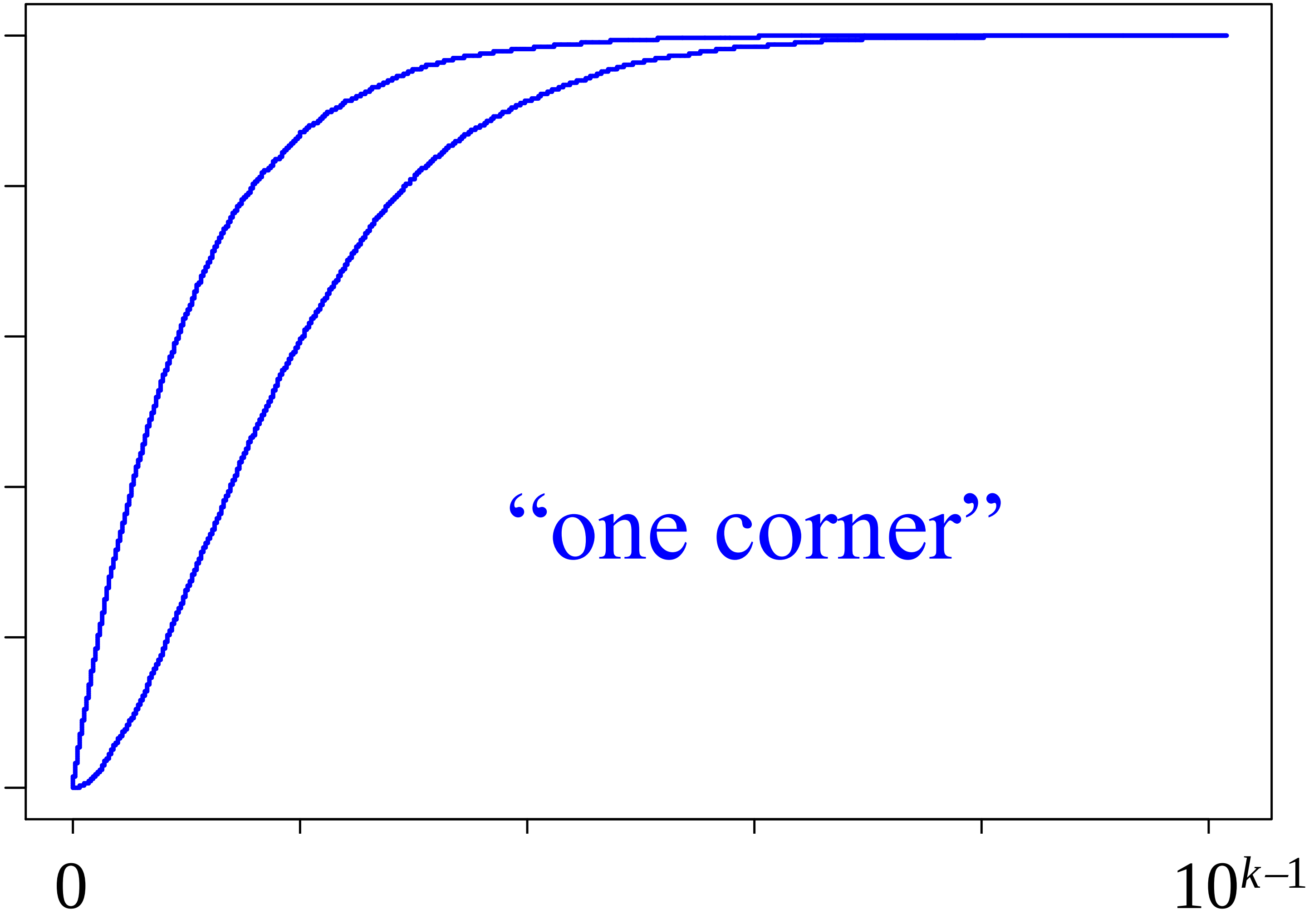


Almost 200,000 km of lines, operated by 500 separate companies

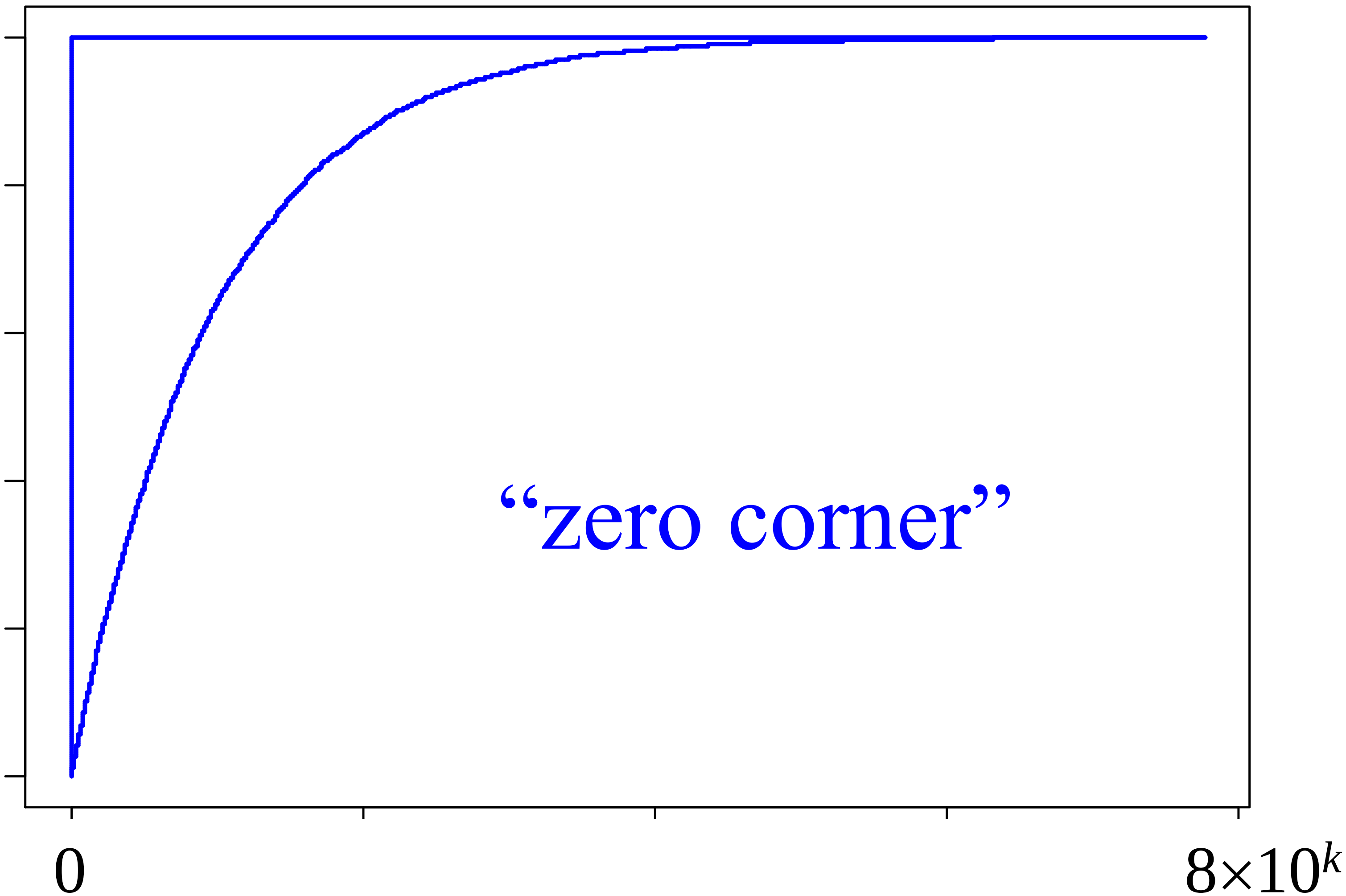
Rare event probabilities

- Hardly ever any actual data
- Sometimes we have experts
- But how should we model bald assertions?
  - “1 in 1000”
  - “1 in ten million”
  - “Never been seen in 100 years”

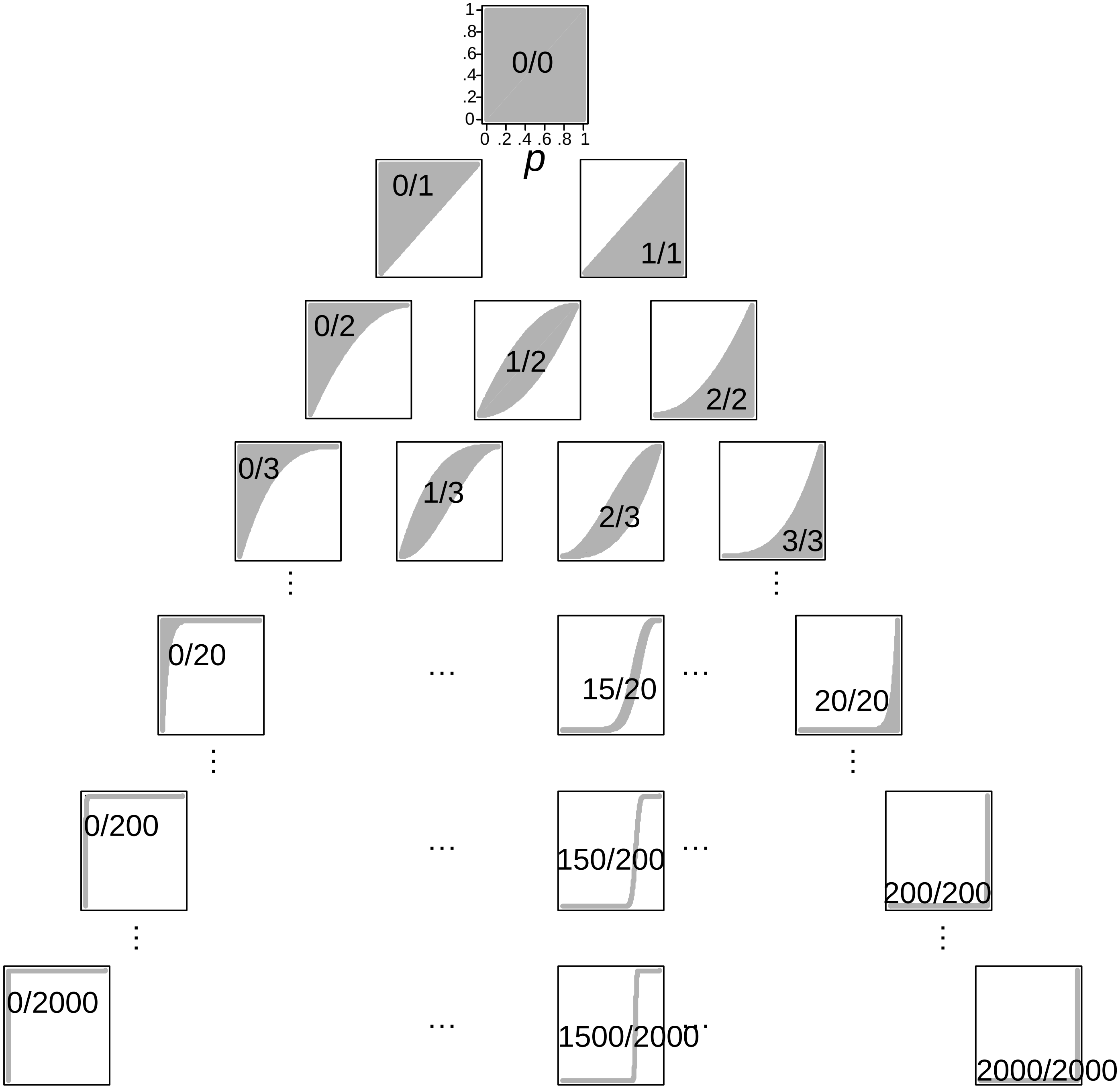
One out of 10^k trials



Zero out of 10^k trials

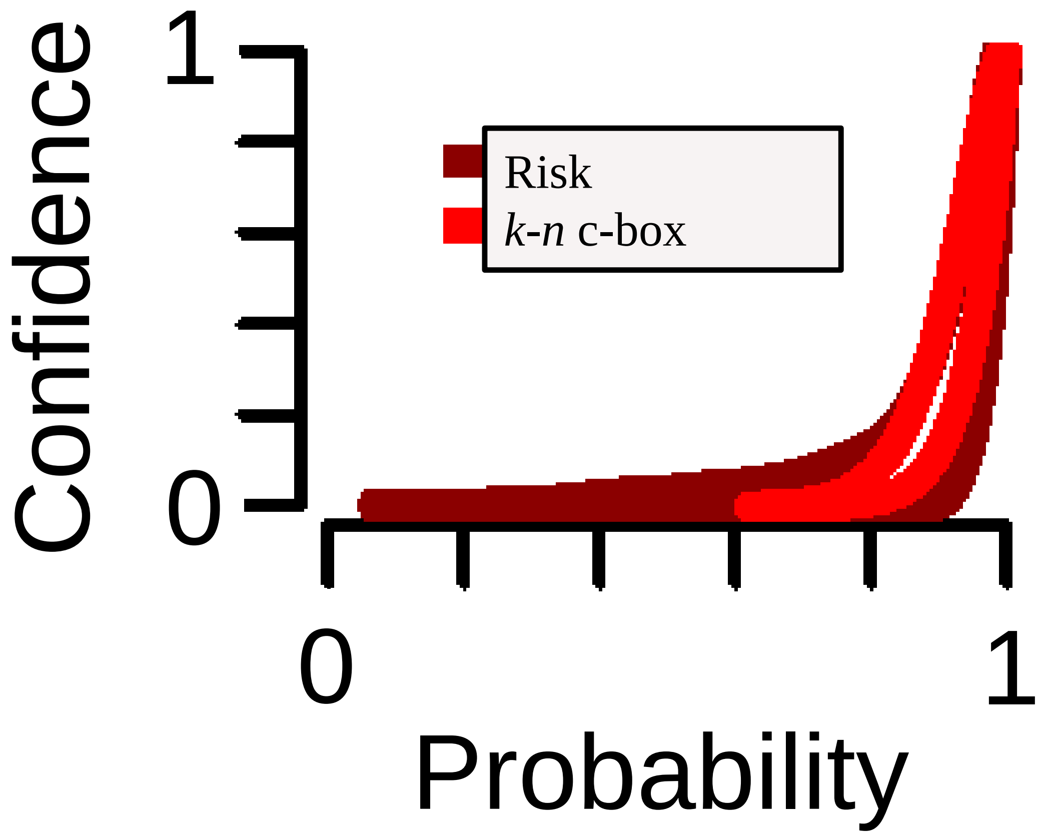


Risk communication with the “equivalent binomial count”



1
.8
.6
.4
.2
0
0 .2 .4 .6 .8 1
p
0/0
0/1
1/1
0/2
1/2
2/2
0/3
1/3
2/3
3/3
0/20
15/20
20/20
0/200
150/200
200/200
0/2000
1500/2000
2000/2000

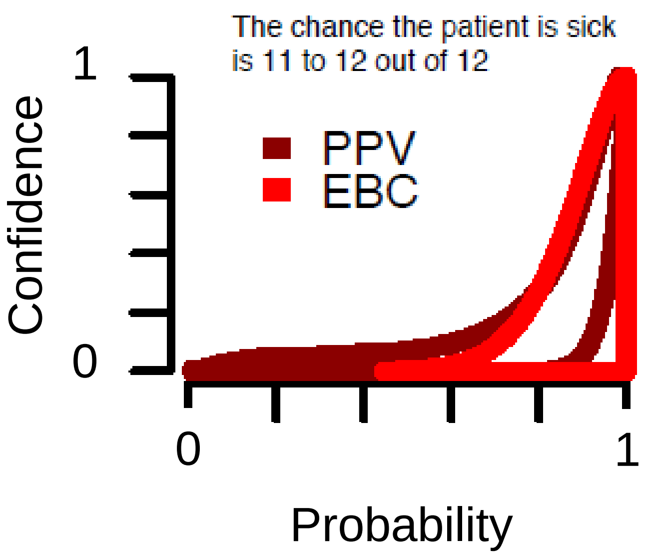
Match a calculated risk to a c-box



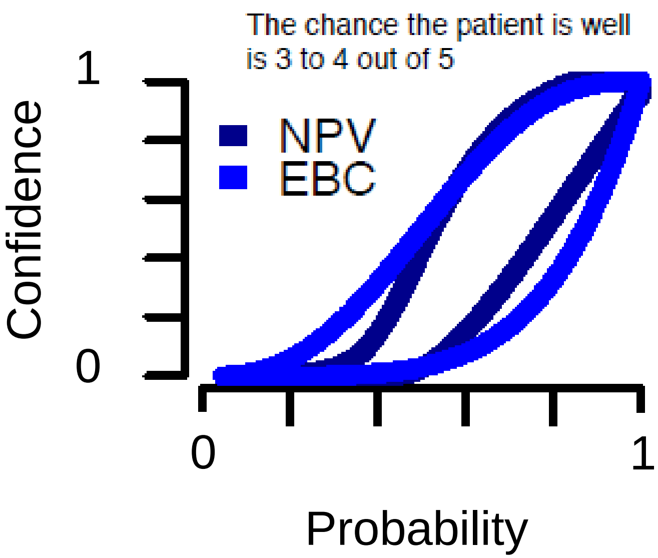
Equivalent binomial count

- An imprecisely computed risk can be expressed as a p-box over [0,1]
- This p-box can be transformed into a natural language statement of the form " *k* out of *n* "
- These are natural frequencies
  - Large uncertainties imply small denominators
  - Or even interval numerators if very large
- People can understand them

PPV


The chance the patient is sick
is 11 to 12 out of 12
PPV
EBC
Confidence
1
0
0
1
Probability

NPV


The chance the patient is well
is 3 to 4 out of 5
NPV
EBC
Confidence
1
0
0
1
Probability

Amazon Mechanical Turk

- Check preference for identical sunglasses rated by other buyers using various schemes
  - More frequent ‘excellent’ ratings should be better
  - Larger pool of buyers rating should be more reliable
- Testing whether
  - Turkers can make rational choices
  - Natural frequencies are as good as percentages
  - Larger denominators convey more reliability
  - Interval numerators can be understood

Which product is better?

- Based on the reviews left by previous customers, which product would you buy?
- Use only the customer ratings and the number of stars left by customers to guide your decision.

Pair A was rated excellent by 2 out of 4 customers.

50% of customers rated Pair B as excellent.

Which product would you buy? Pair A, or Pair B?

Findings

80%	50%	rational
10/100	80/100	same sample size
66/198	2/6	same magnitude
[66,88]/198	33/100	even with ambiguity
50%	50/100	prefer natural frequency
2/4	50%	unless very uncertain

Sample sizes ~300 “master turkers”

Confidence boxes

- Structures that let you infer confidence intervals for a parameter, at any confidence level
- Can be propagated just like p-boxes
- Allow us to *compute with confidence*

Next steps

- How to incorporate other constraint information besides the distribution shape

- Big, multi-parameter problems

- C-box approach for estimating copulas

More information

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

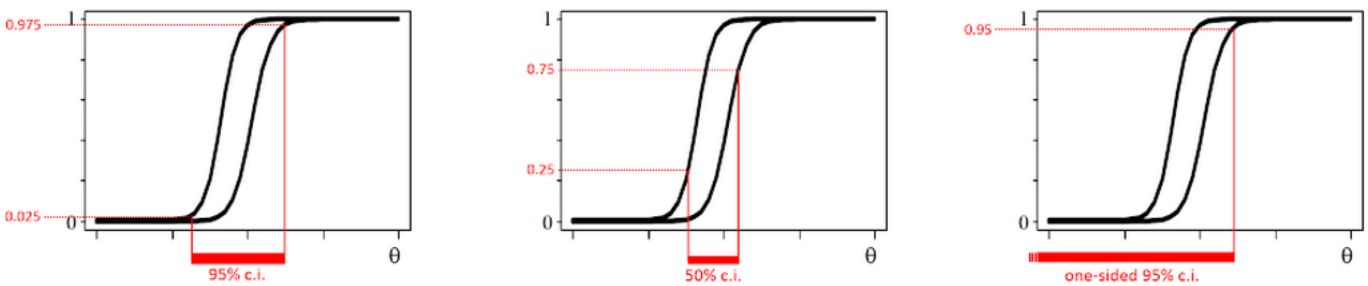
https://sites.google.com/site/reliabilityuncertainty

- Papers
- Slide presentations
- Free software

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

Computing with Confidence

Confidence boxes ("c-boxes") tell you confidence intervals for a parameter at any confidence level you like. For instance, the confidence box depicted below yields several confidence intervals for the parameter θ. Although you can't generally compute with confidence intervals, *you can compute with confidence boxes*, and you can get arbitrary confidence intervals for the results.



Confidence boxes can be computed in a variety of ways directly from random sample data. There are confidence boxes for both parametric problems where the family of the underlying distribution from which the data were randomly generated is known (including normal, lognormal, exponential, binomial, Poisson, etc.), and nonparametric problems in which the shape of the underlying distribution is unknown. Confidence boxes account for the uncertainty about a parameter that comes from the inference from observations, including the effect of small sample size, but also the effects of imprecision in the data and demographic uncertainty which arises from trying to characterize a continuous parameter from discrete data observations.

When confidence boxes have the form of probability boxes, they can be propagated through mathematical expressions using the ordinary machinery of probability bounds analysis, and this allows analysts to compute with confidence, both figuratively and literally, because the results also have this confidence interpretation.

This website is a portal to several papers and presentations about confidence boxes, including

1. Slide presentations and posters,
2. Introductory paper on c-boxes with R functions and a comparison with the Imprecise Beta Model,
3. Paper with several c-box formulas and a comparison between c-box results and analogous Bayesian and maximum likelihood results,
4. Application of c-boxes in common inference problems arising in risk analysis,
5. Application of the c-box for nonparametric difference in a calibration/validation study,
6. Original paper on confidence structures, and
7. Review paper on confidence distributions; another one.

Confidence boxes are imprecise generalizations of traditional confidence distributions. Like Student's *t* distribution, they encode frequentist confidence intervals for parameters of interest at every confidence level. They are analogous to Bayesian posterior distributions in that they characterize the inferential uncertainty about distribution parameters estimated from sparse or imprecise sample data, but they have a purely frequentist interpretation that makes them useful in engineering because they offer a guarantee of statistical performance through repeated use. Unlike traditional confidence intervals which cannot usually be propagated through mathematical calculations, c-boxes can be used in calculations to yield results that also admit the same confidence interpretation. For instance, they can be used to compute probability boxes for both prediction and tolerance distributions. They are easy to construct and use in calculations; see the software page for R functions to construct several c-boxes.

Note that c-boxes are completely different from confidence bands such as the Kolmogorov-Smirnov distributional bands which are nonparametric confidence limits *at some particular confidence level* for the distribution from which sample data were randomly drawn. C-boxes encode confidence intervals at all possible confidence levels at the same time.

Confidence boxes

About
This site collects papers describing the use of statistical confidence structures in risk analysis.

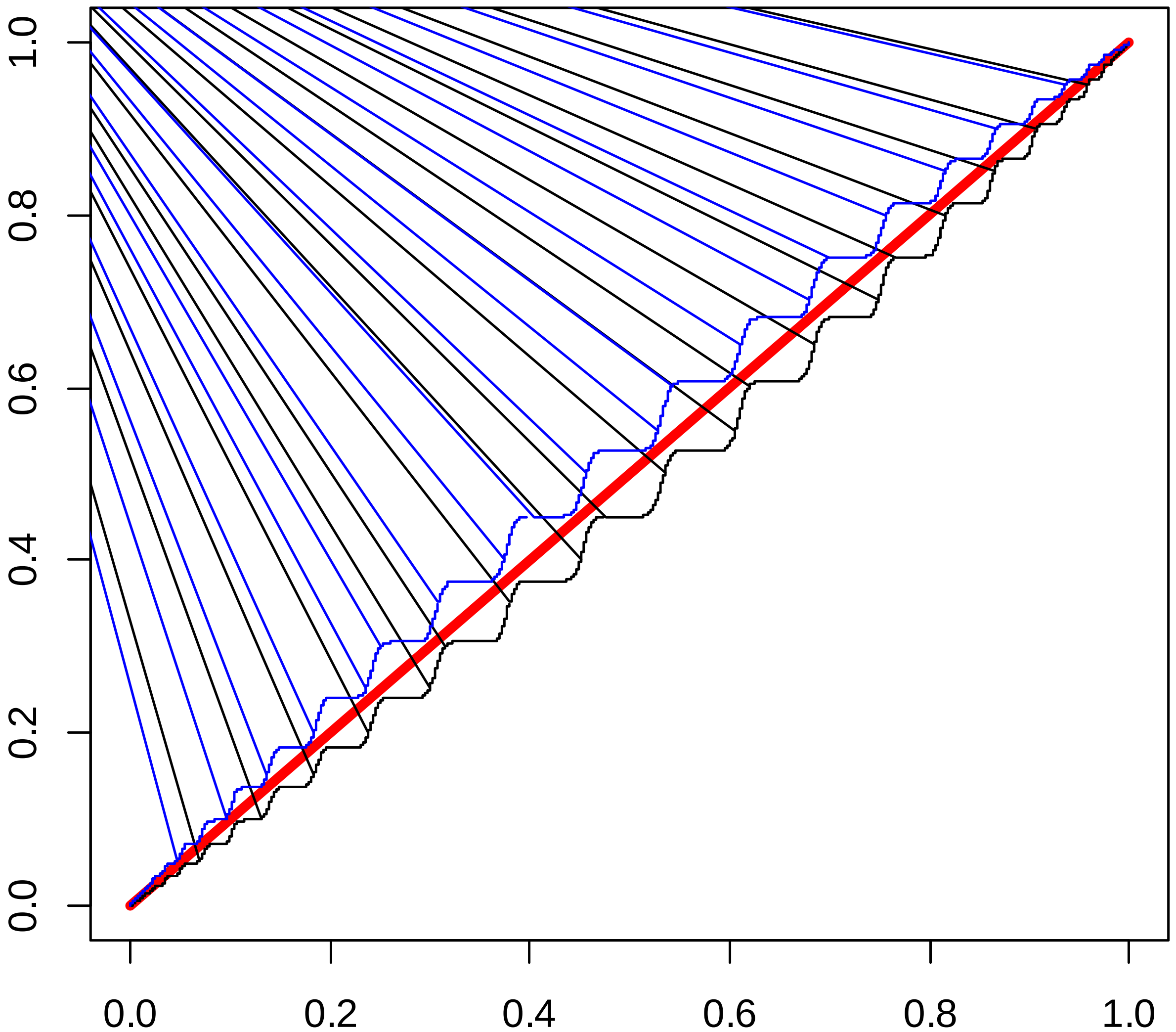
Acknowledgments
Support for this project was provided by the National Library of Medicine, a component of the National Institutes of Health (NIH), through a Small Business Innovation Research grant (award number RC3LM010794) to Applied Biomathematics funded under the American Recovery and Reinvestment Act.

Links to related sites
Other AB NIH ARRA links
Michael Balch's IJAR paper
Nonparametric difference
Sandia report on p-boxes
Sandia report on interval statistics
Sandia report on dependence
Sitemap

Acknowledgments

- Michael Balch
- Jason O’Rawe, Cold Spring Harbor
- Kari Sentz, Los Alamos National Lab
- Matthias Troffaes, Durham
- Mohamed Sallak, UTC
- Teddy Seidenfeld, Carnegie Melon
- National Institutes of Health

Questions?



0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Stopping rule

- C-boxes can depend on the stopping rule
  - But not knowing the stopping rule may just mean the c-box is wider
  - Knowing the stopping rule tightens the c-box