Communicating probability with natural frequencies and the equivalent binomial count

Scott Ferson, University of Liverpool, United Kingdom

22 February 2019, Conference on Uncertainty in Risk Analysis

Communicating probability

- Probability density function graphs
- Box and whisker plots
- Cumulative probability function graphs
- Icon arrays
- Spinners, games
- Numerical statistics
- Percentages, odds, natural frequencies

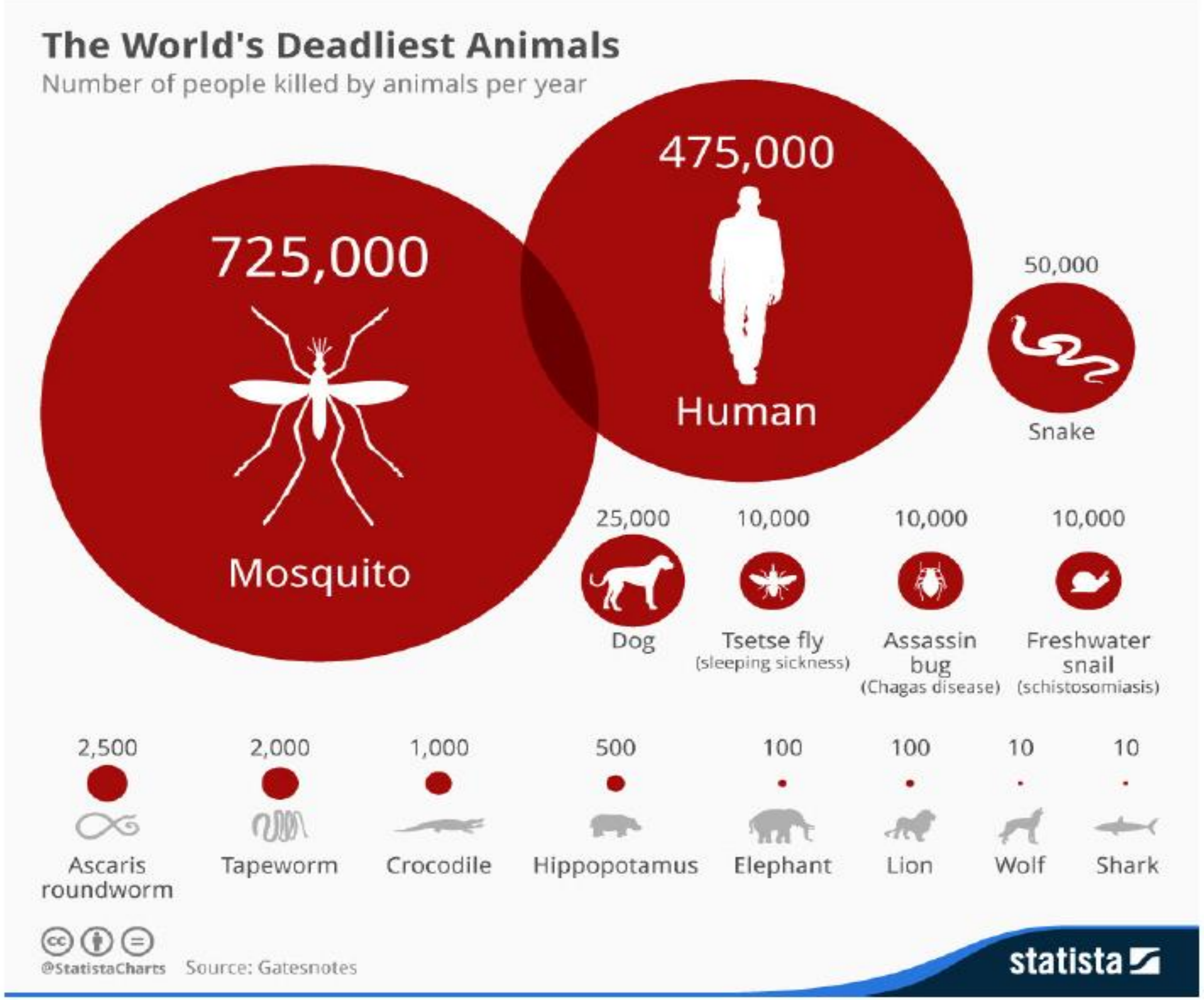
Natural frequencies and icon arrays

37 out of 100

What if it's more complex than scalar $p \in [0,1]$

- What if it is a distribution?
- What if it is a collection of distributions?
  - Second-order distribution
  - Robust Bayesian analysis
  - Probability box
  - Envelop of distributions from disagreeing experts

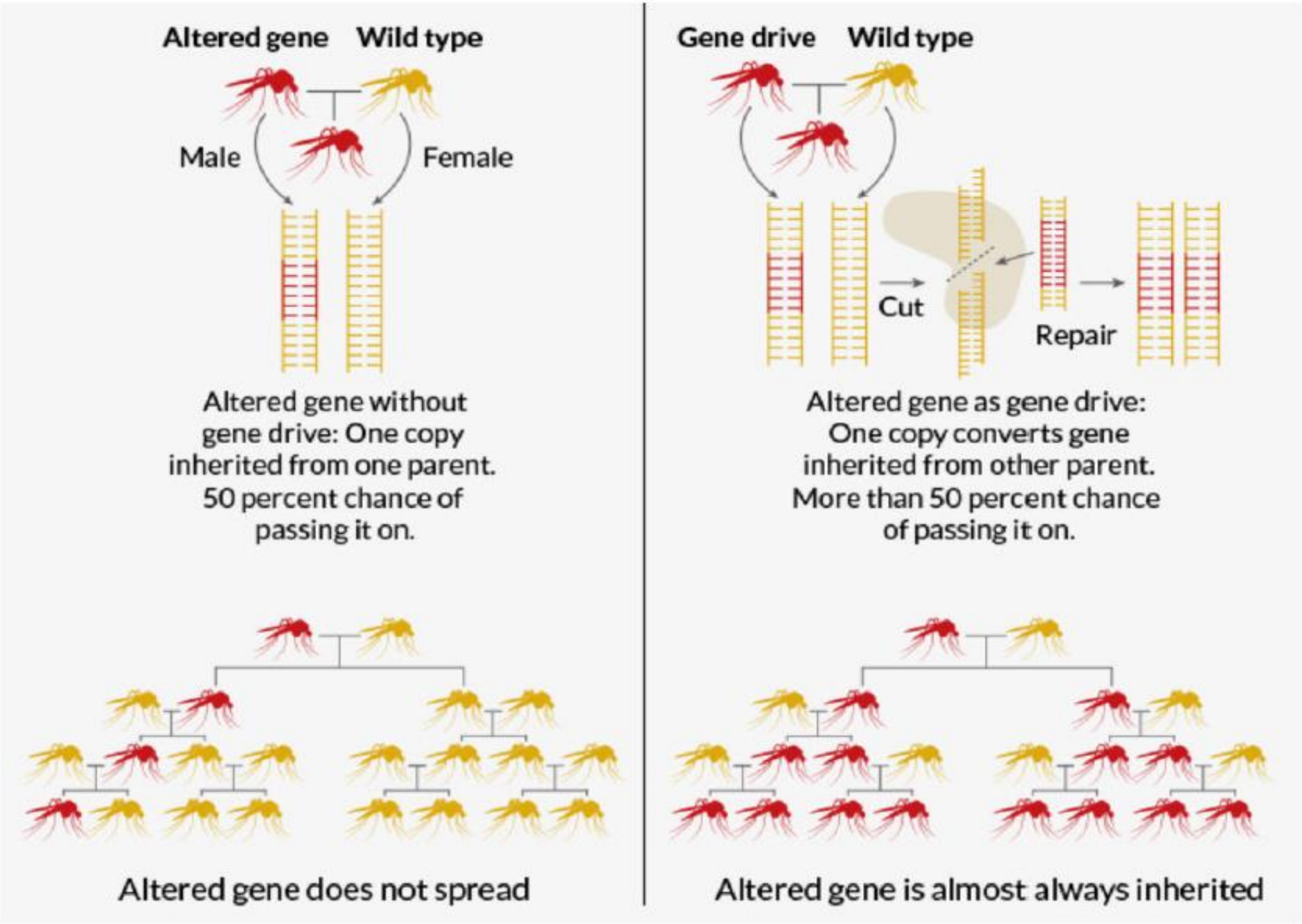
What is the world’s most dangerous animal?



Source:https://www.statista.com/chart/2203/the-worlds-deadliest-animals/

Slide credit: Keith Hayes, et al. 2017

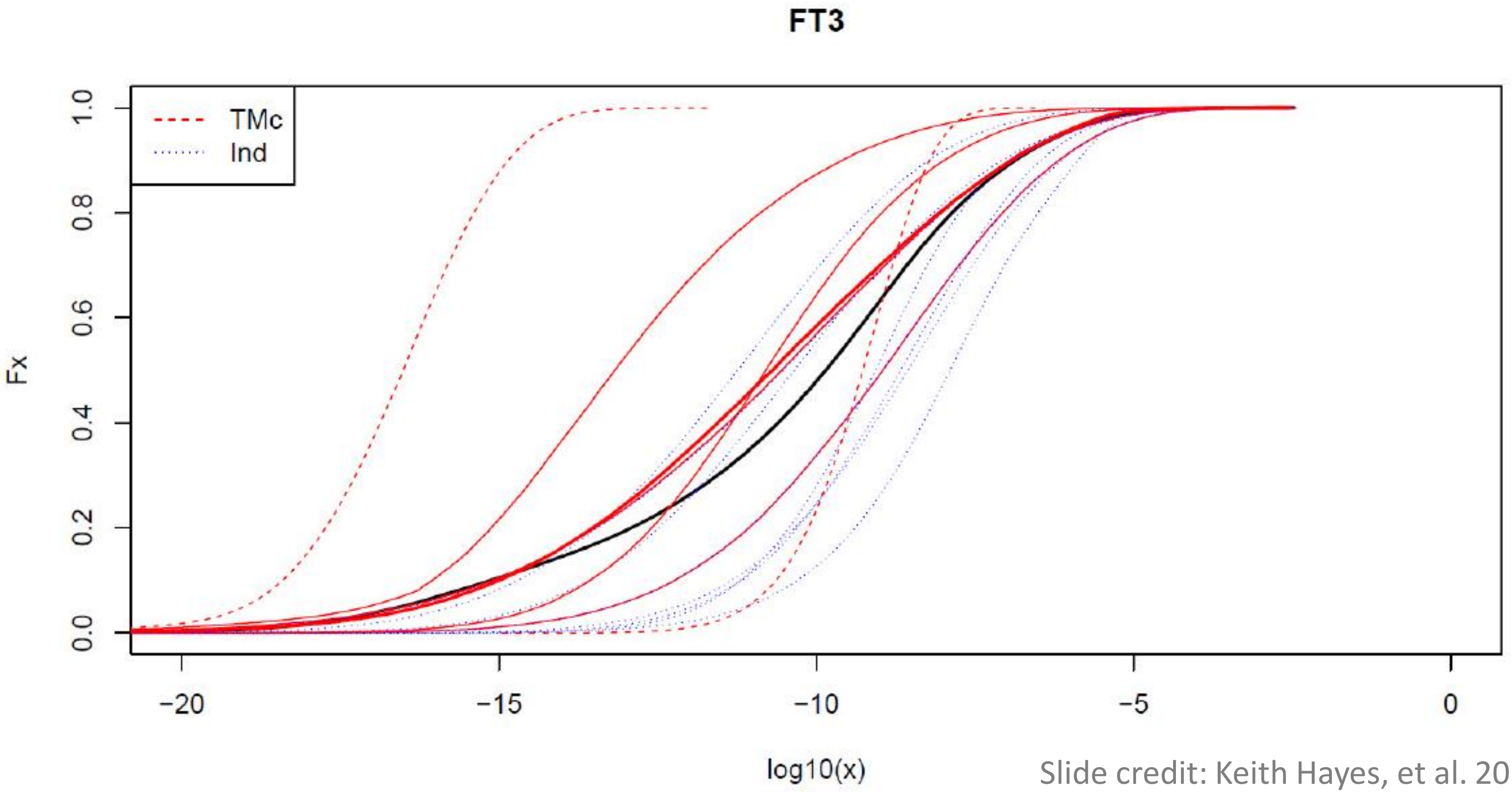
Medelian inheritance versus gene drive



Source: Saey (2015)

Slide credit: Keith Hayes, et al. 2017

Risk result: HGT to non-target Eukaryotes



Slide credit: Keith Hayes, et al. 2017

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
- Allows us to *compute with confidence*

Example: binomial rate *p* for *k* of *n* trials

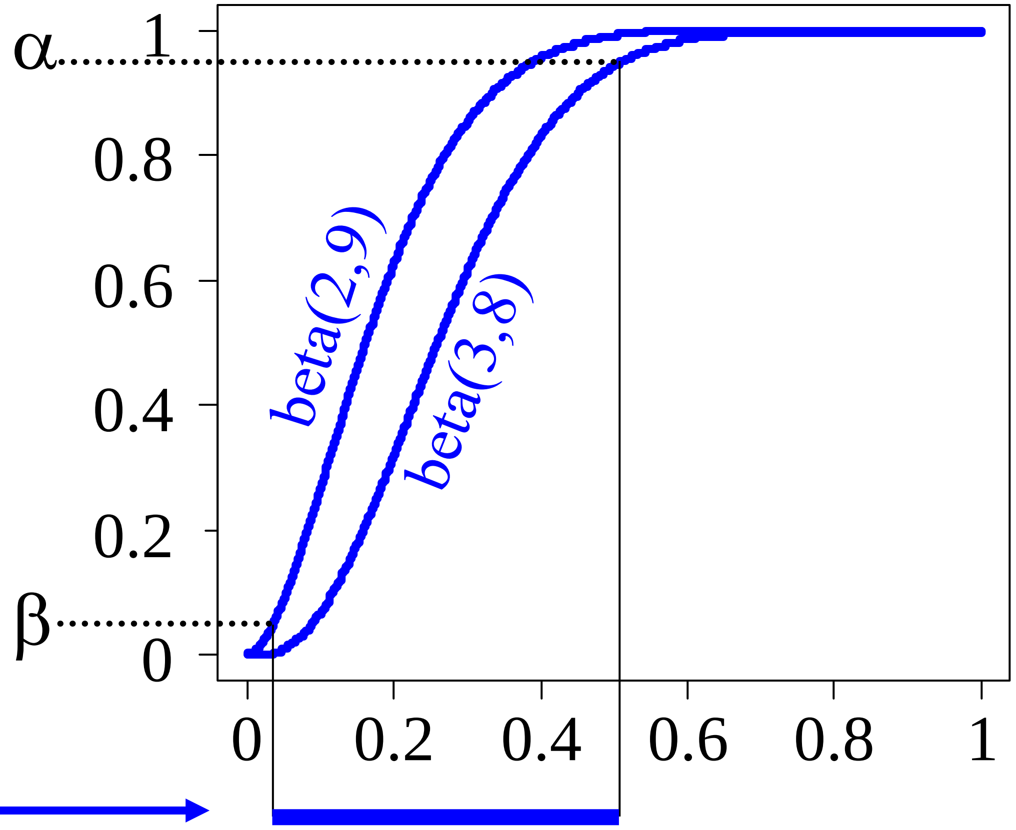
$p \sim \text{env}(\text{beta}(k, n-k+1), \text{beta}(k+1, n-k))$

Notation extends the use of tilda

Data

$k = 2$ successes

$n = 10$ trials



(α–β)100% confidence interval for *p*

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

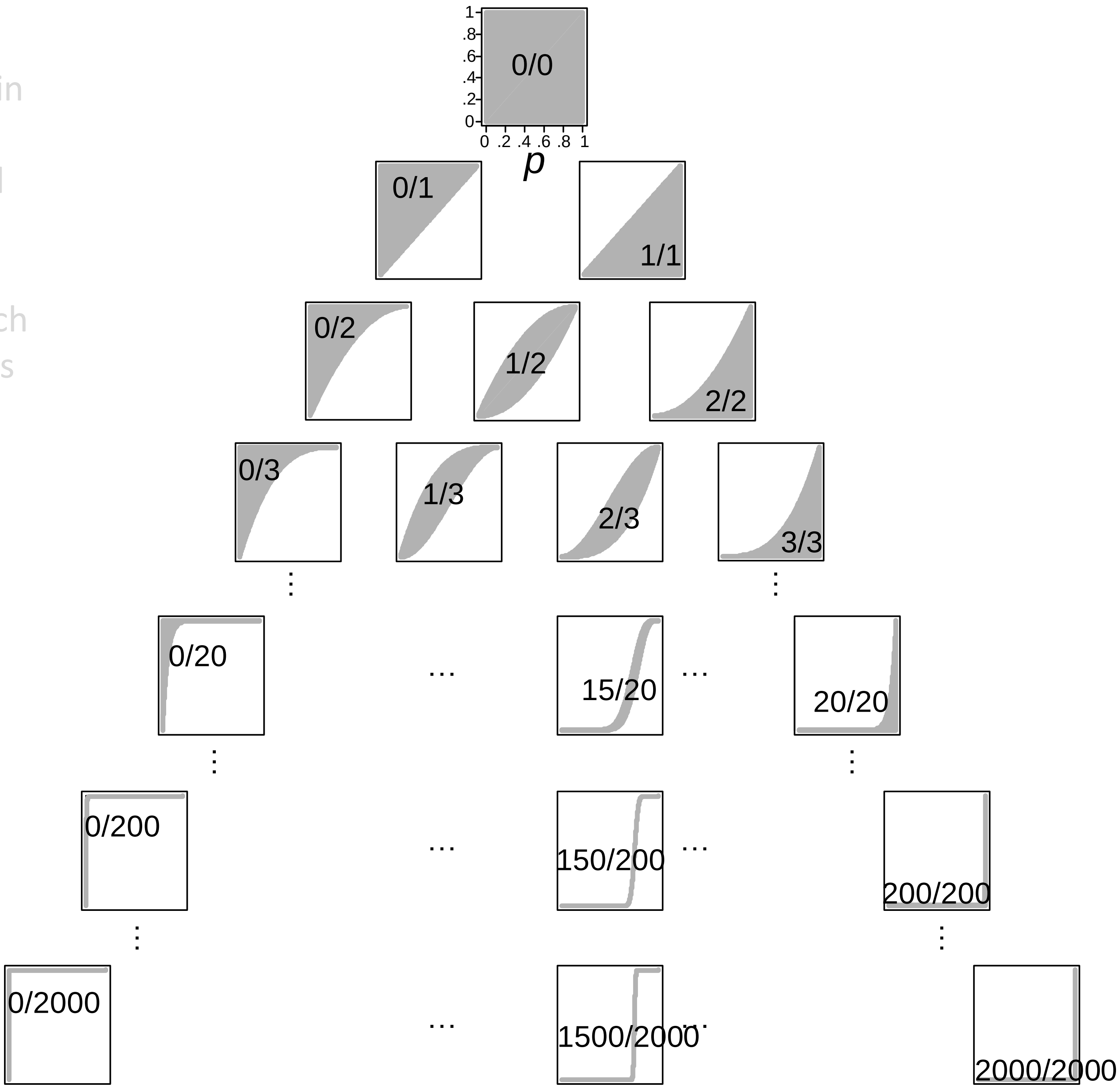
C-boxes

- Bayesian (specifies a class of priors in the uninformative case)
- But also have frequentist coverage properties
- Don’t optimize anything; they *perform*
- Characterize inputs from limited or even no information

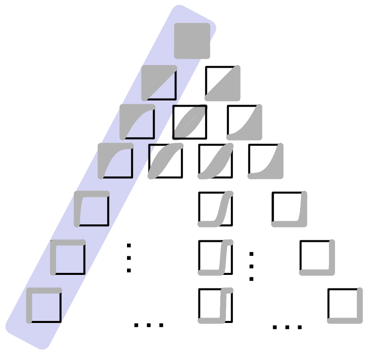
Notice that the c-boxes in every row partition the vacuous 'dunno' interval

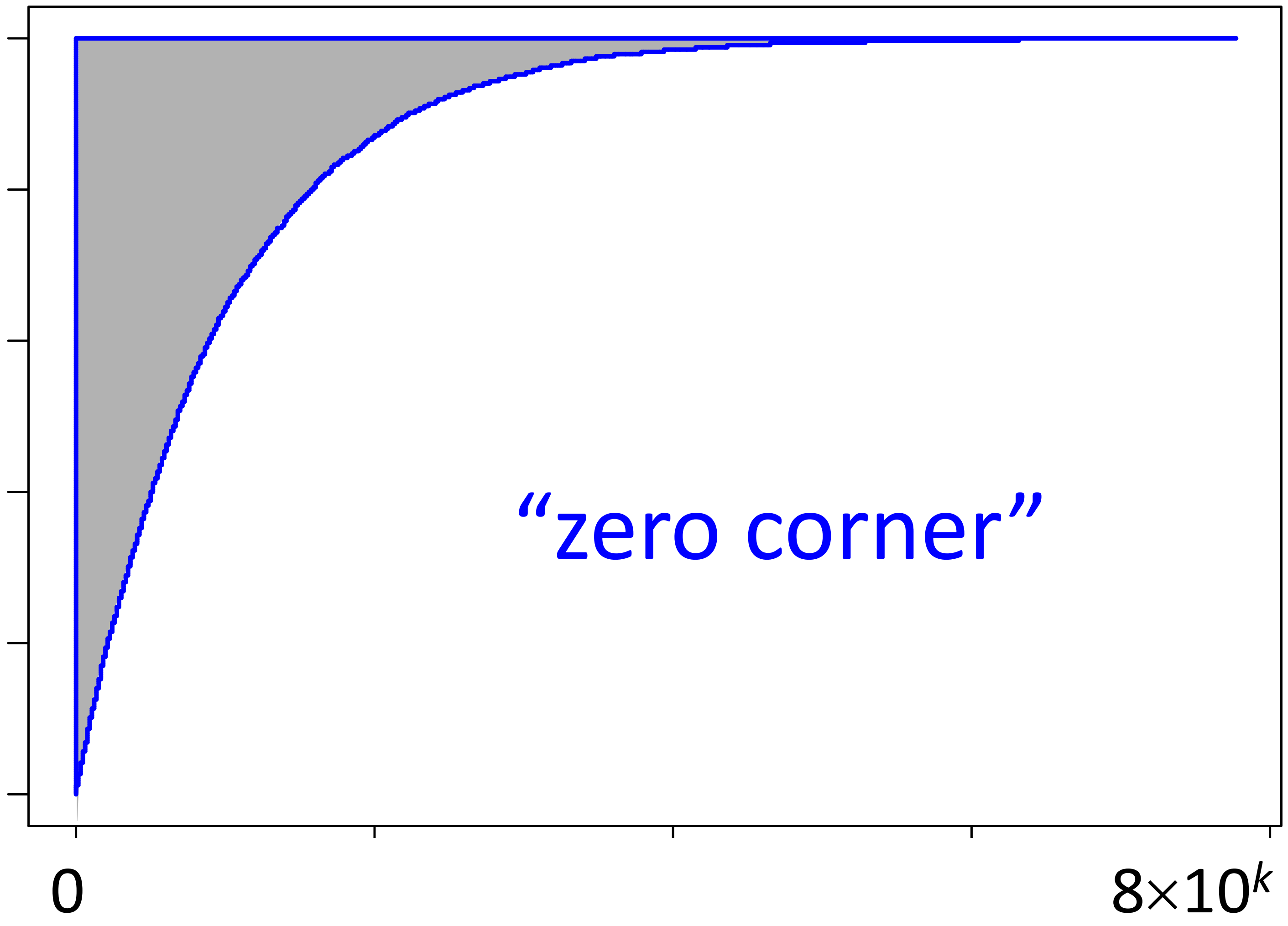
We call the first (or first and second) c-box in each row the "corner" c-boxes

They correspond to the rare events of concern

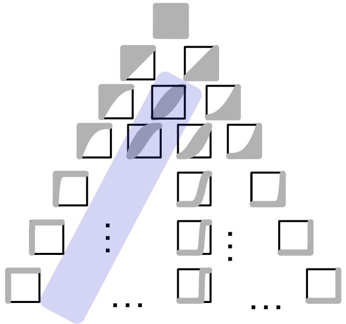


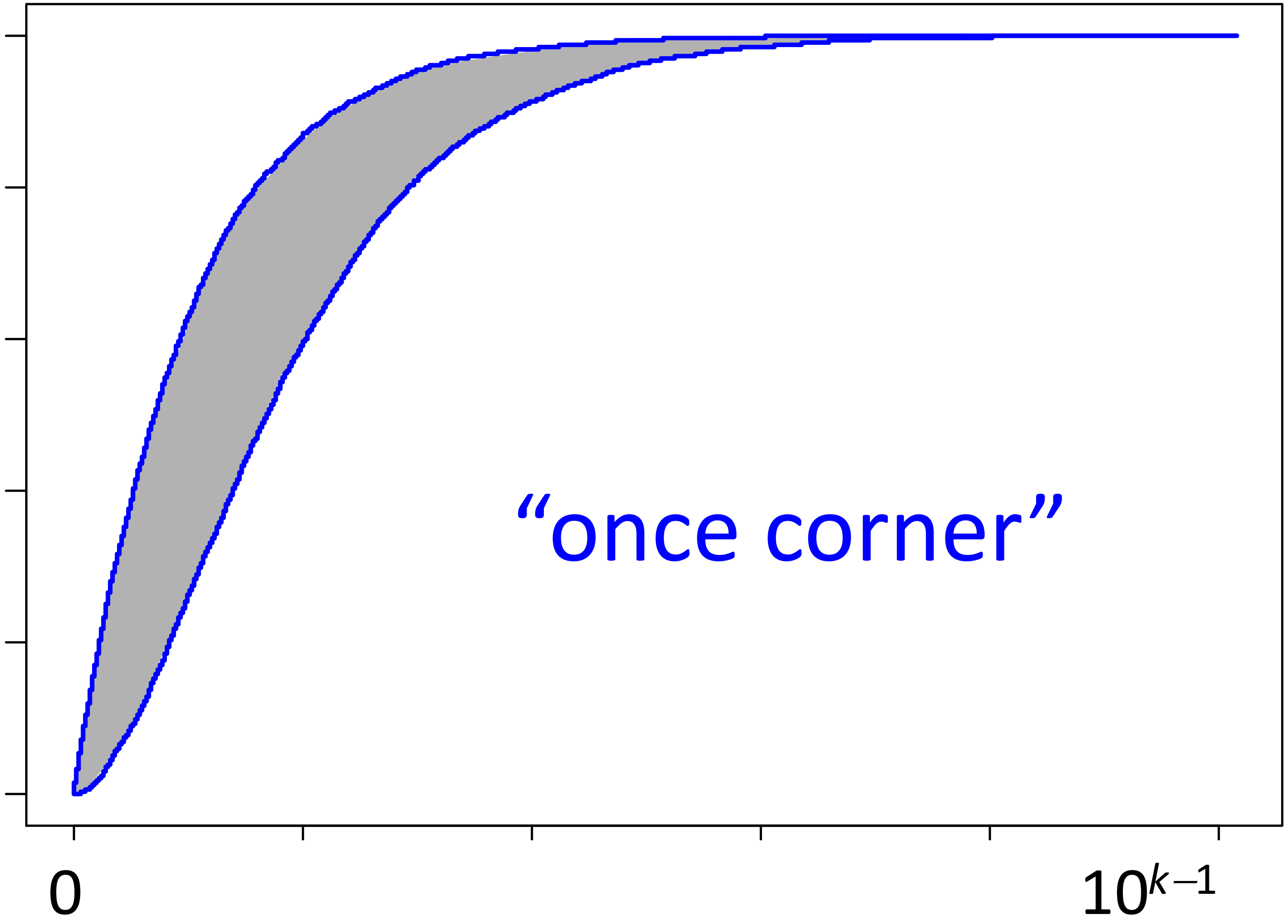
Zero out of 10^k trials



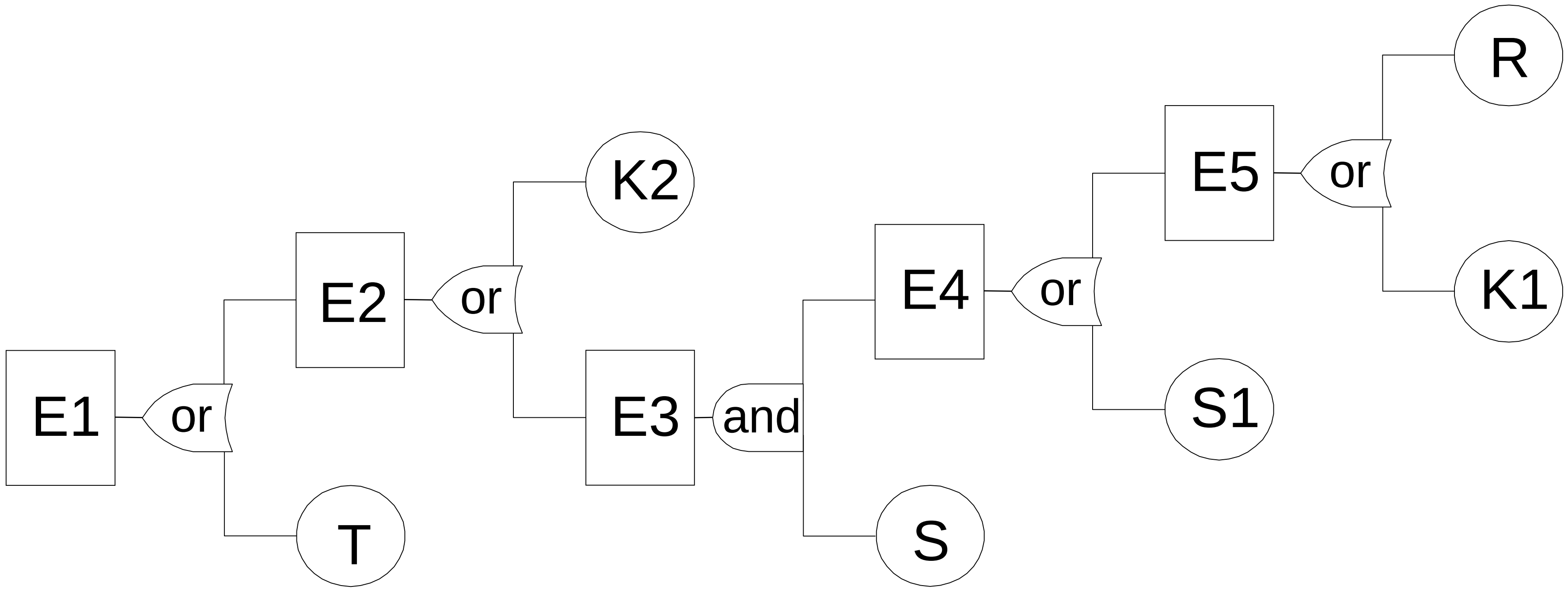


One out of 10^k trials



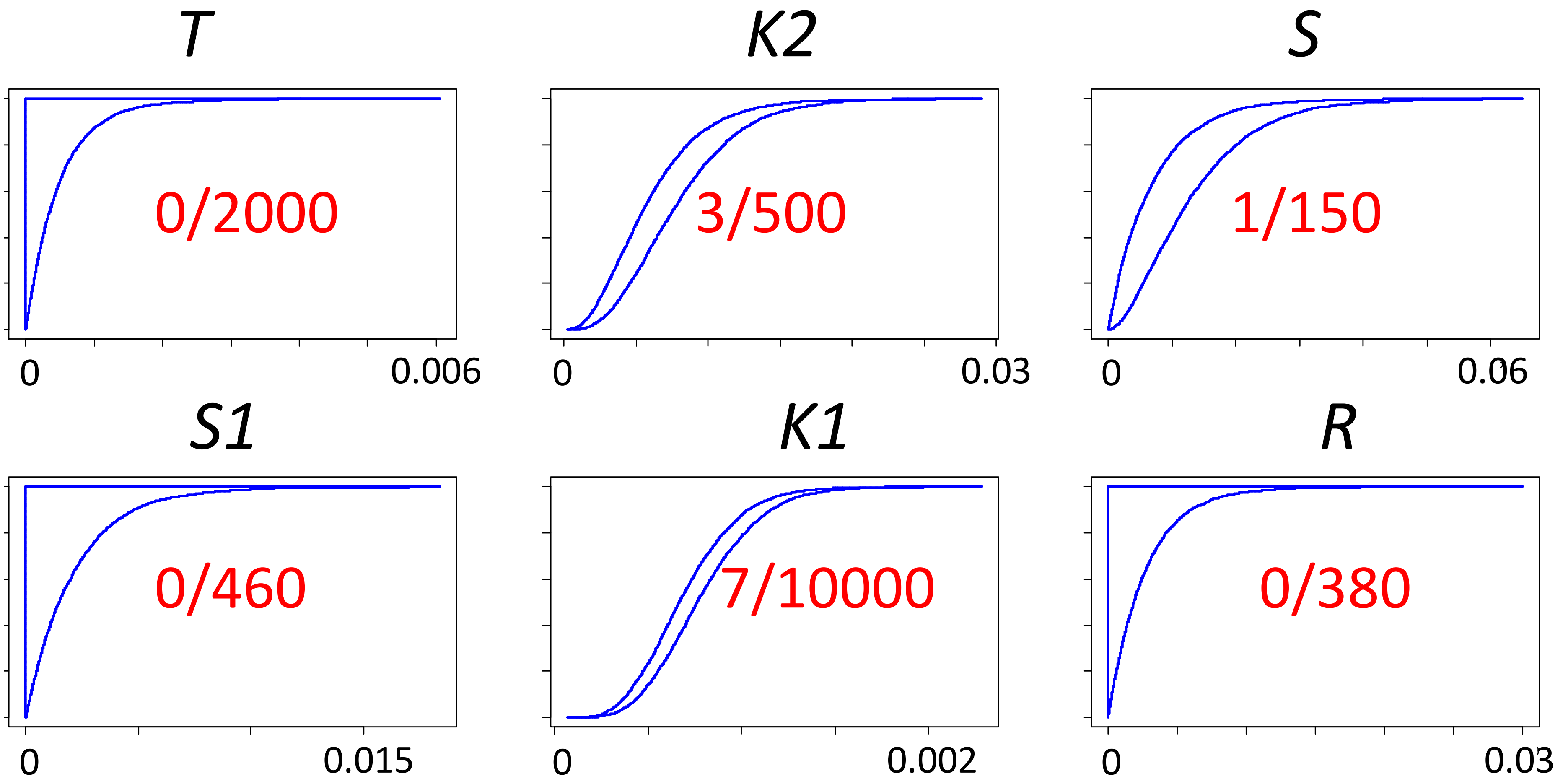


Fault tree

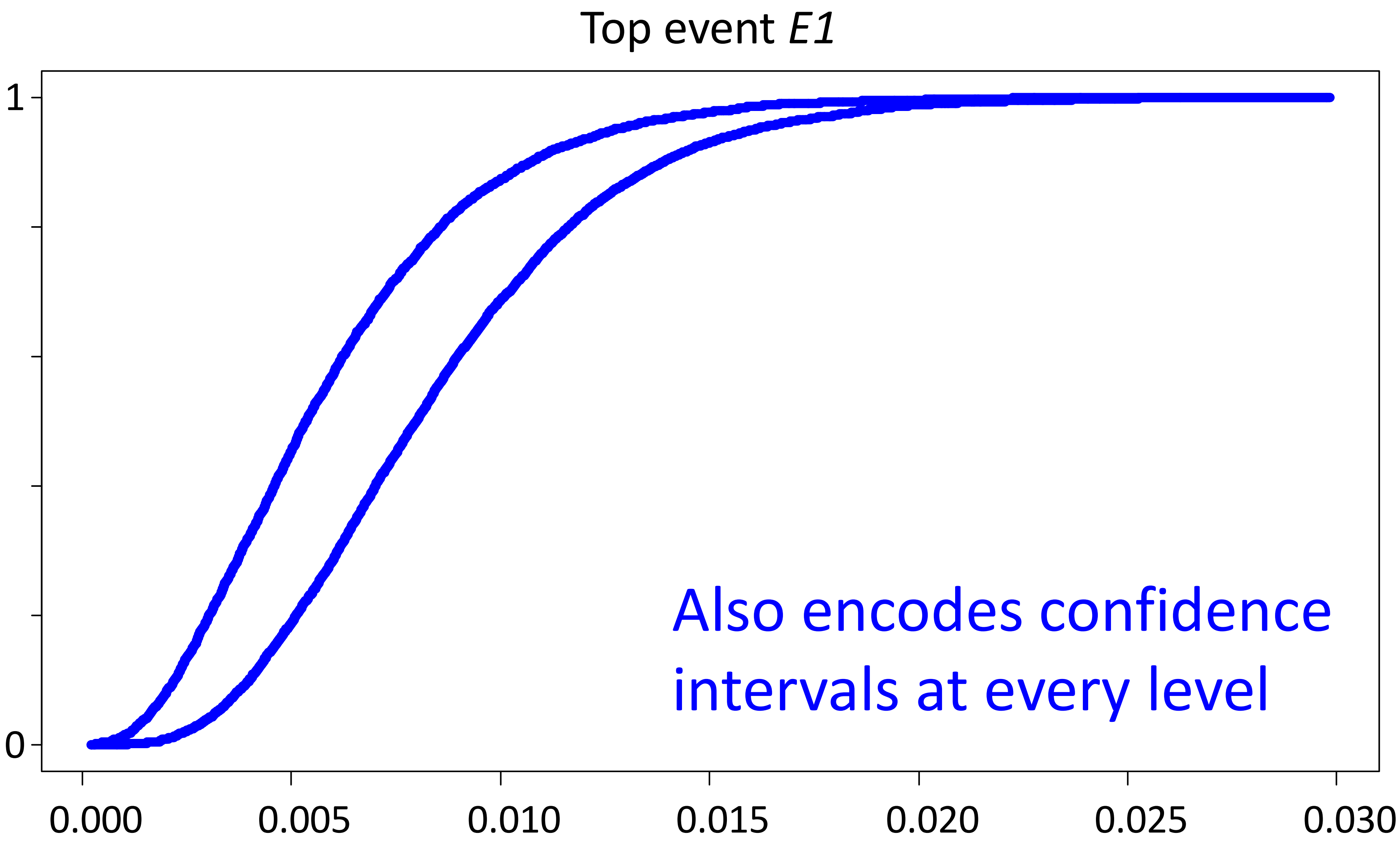


$$E1 = T \vee (K2 \vee (S \,\&\, (S1 \vee (K1 \vee R))))$$

Fault tree inputs

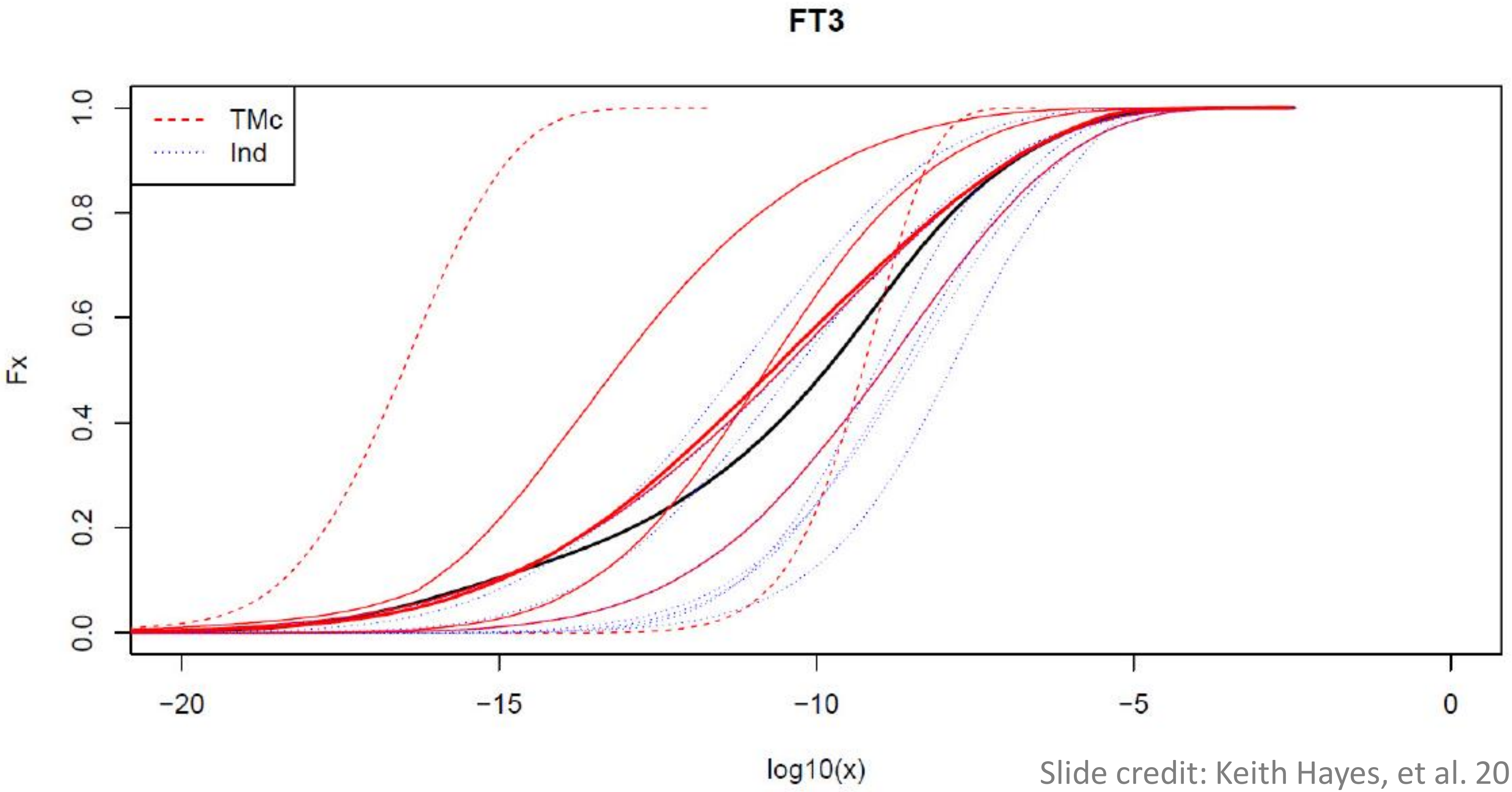


The blue c-boxes are posteriors for the inputs from a robust Bayes analysis based on the red data



Top event E1
1
0
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Also encodes confidence
intervals at every level

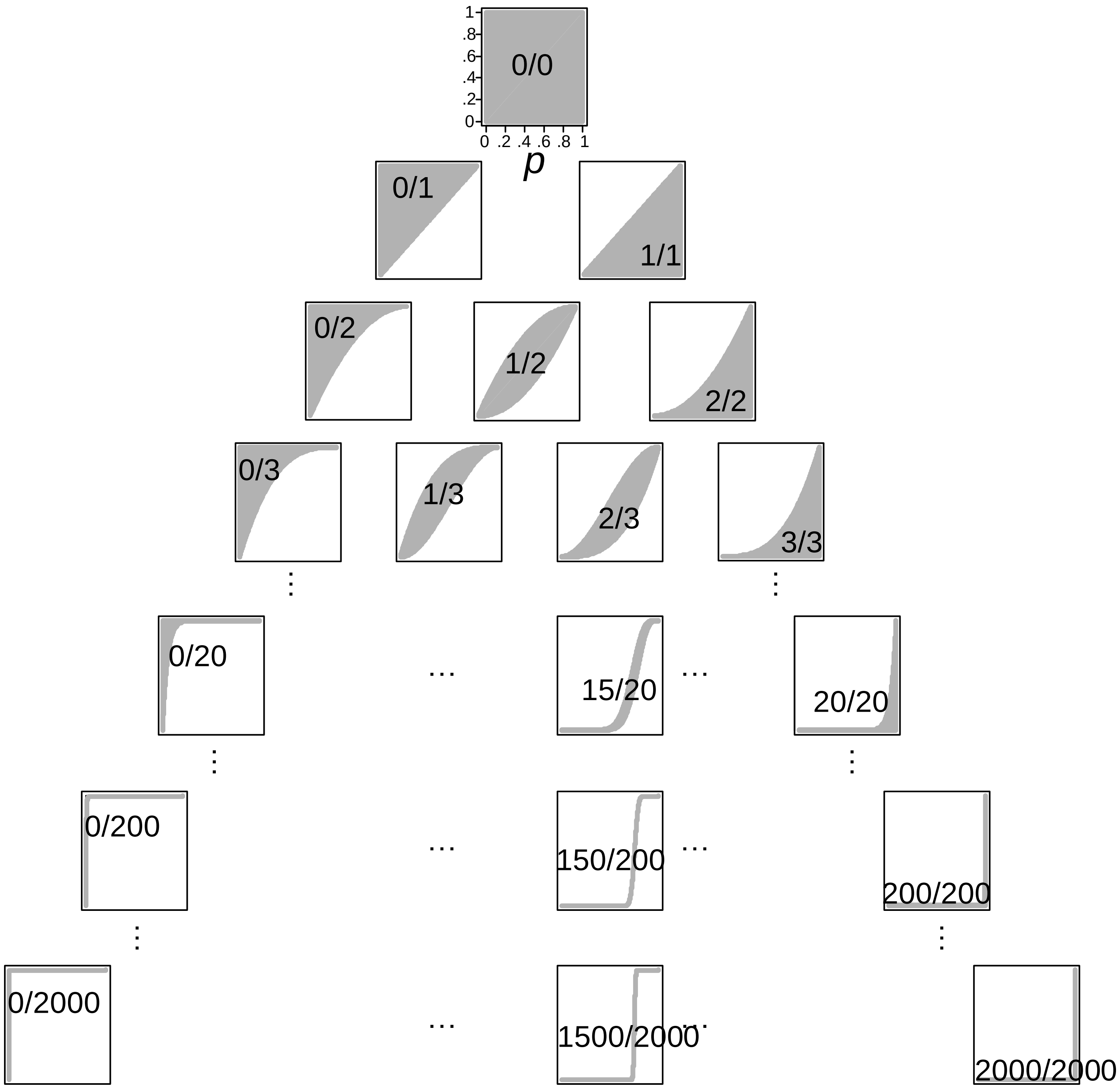
Risk result: HGT to non-target Eukaryotes



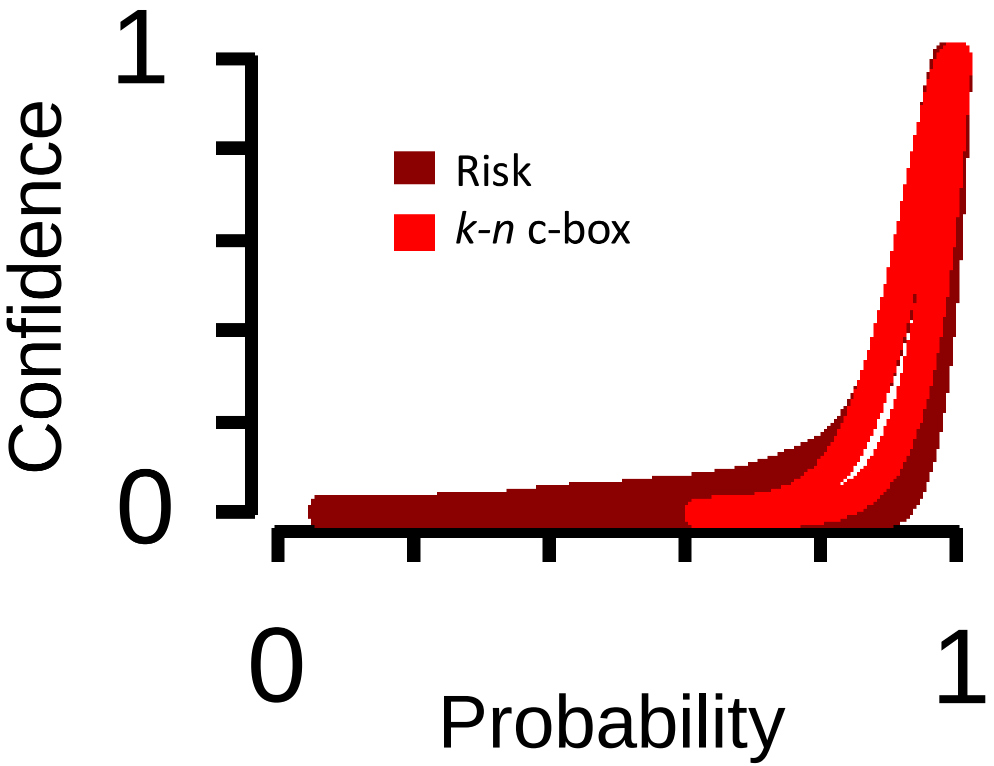
Slide credit: Keith Hayes, et al. 2017

Equivalent binomial count

- An imprecisely computed risk can be expressed as a p-box on [0,1]
- Transform it into a natural language expression “ *k* out of *n* ”
- These are *natural frequencies*
  - Ratio *k/n* implies magnitude of risk
  - Large uncertainties imply small denominators
- People can understand them


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

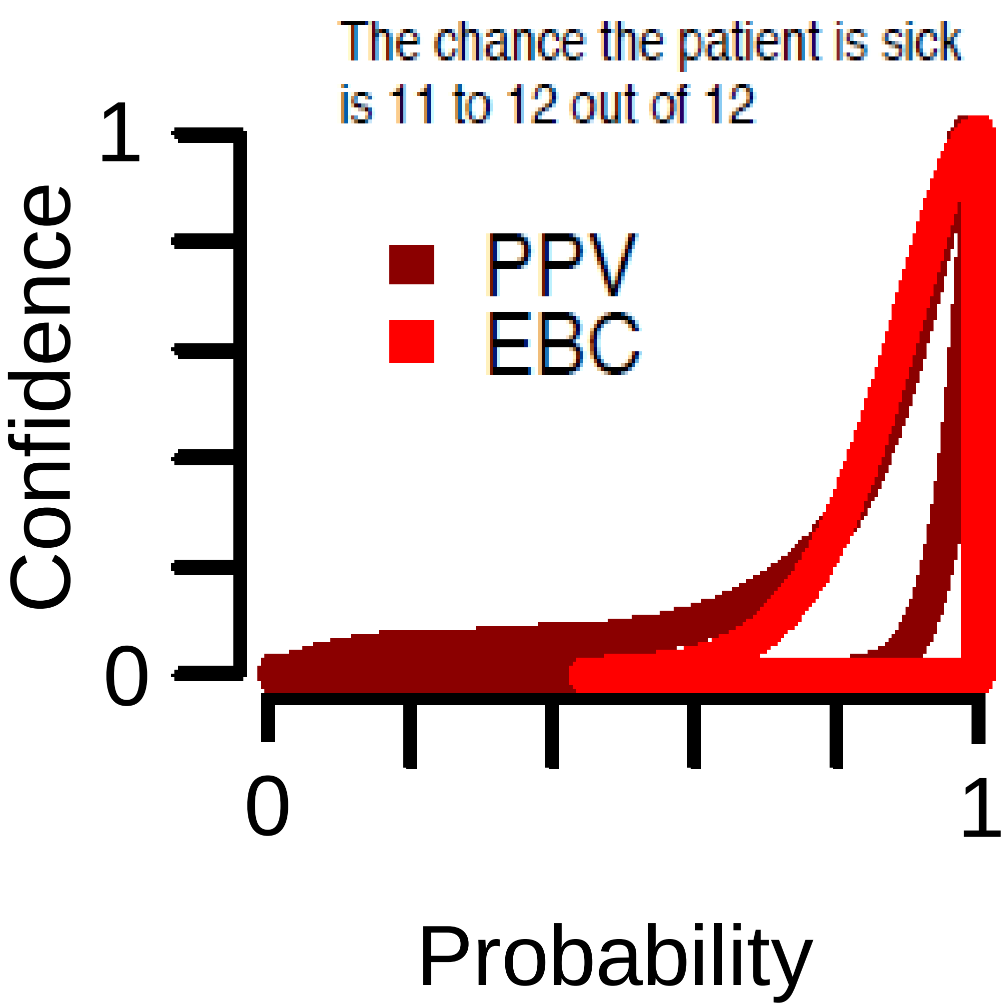


Express risk as "*k* out of *n*"

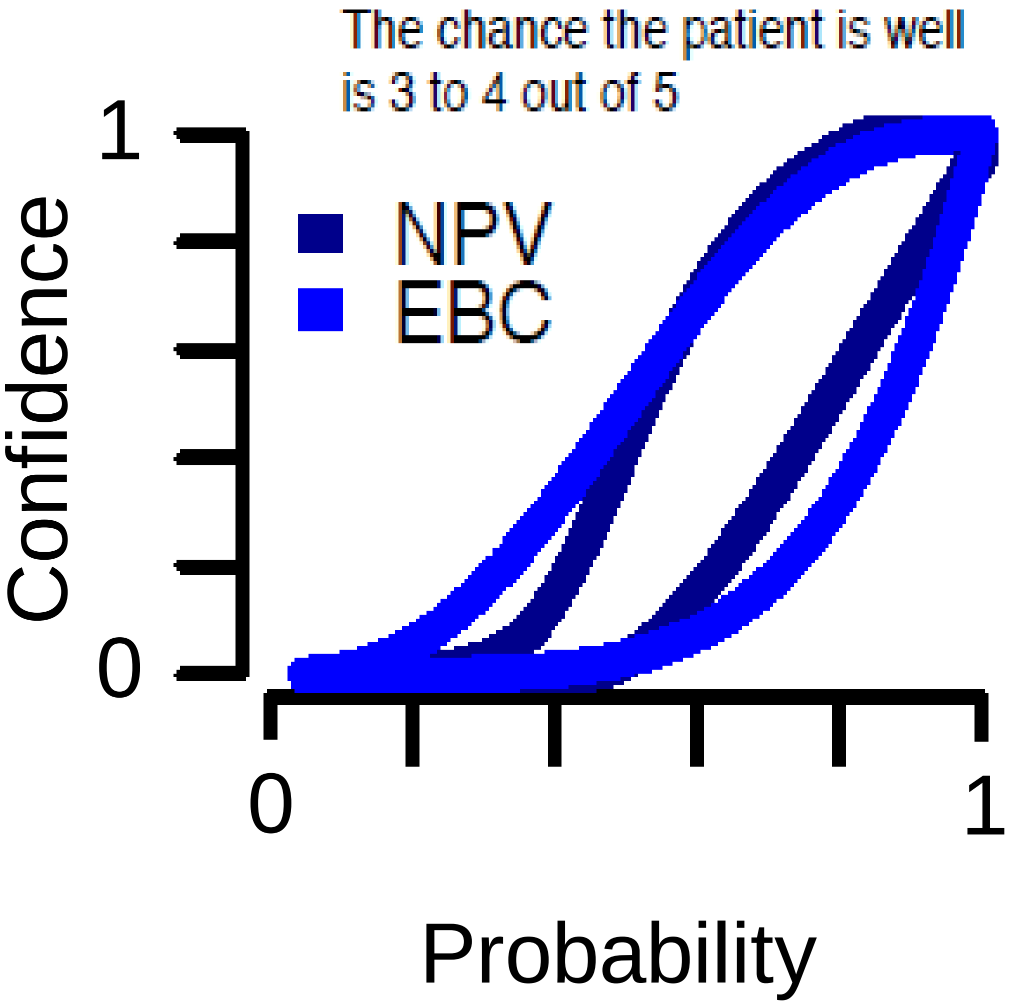
When there is a lot of epistemic uncertainty

- It might be possible to use intervals as numerators *k*

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

Do people understand?

- We used Amazon Mechanical Turk to check this
- We showed >300 “turkers” several mock sunglasses comparisons
- We checked the turkers’ preferences for identical sunglasses rated by other buyers using various schemes
- More frequent ‘excellent’ ratings should be preferred
- Larger pool of buyers rating should be more reliable

We tested whether

- Turkers can make rational choices
- Natural frequencies are as good as or better than 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

These are exemplar comparisons from the study with “master turkers”

Conclusions

- People make rational choices when given natural frequencies “*k* out of *n*”
- Analysts can translate results into *k*-out-of-*n* equivalent binomial counts
- Natural frequencies express probabilities so humans can understand them
- Also embody uncertainty about the risks which humans also care about

Acknowledgments

- Keith Hayes
- Jason O’Rawe
- Michael Balch
- National Institutes of Health

End