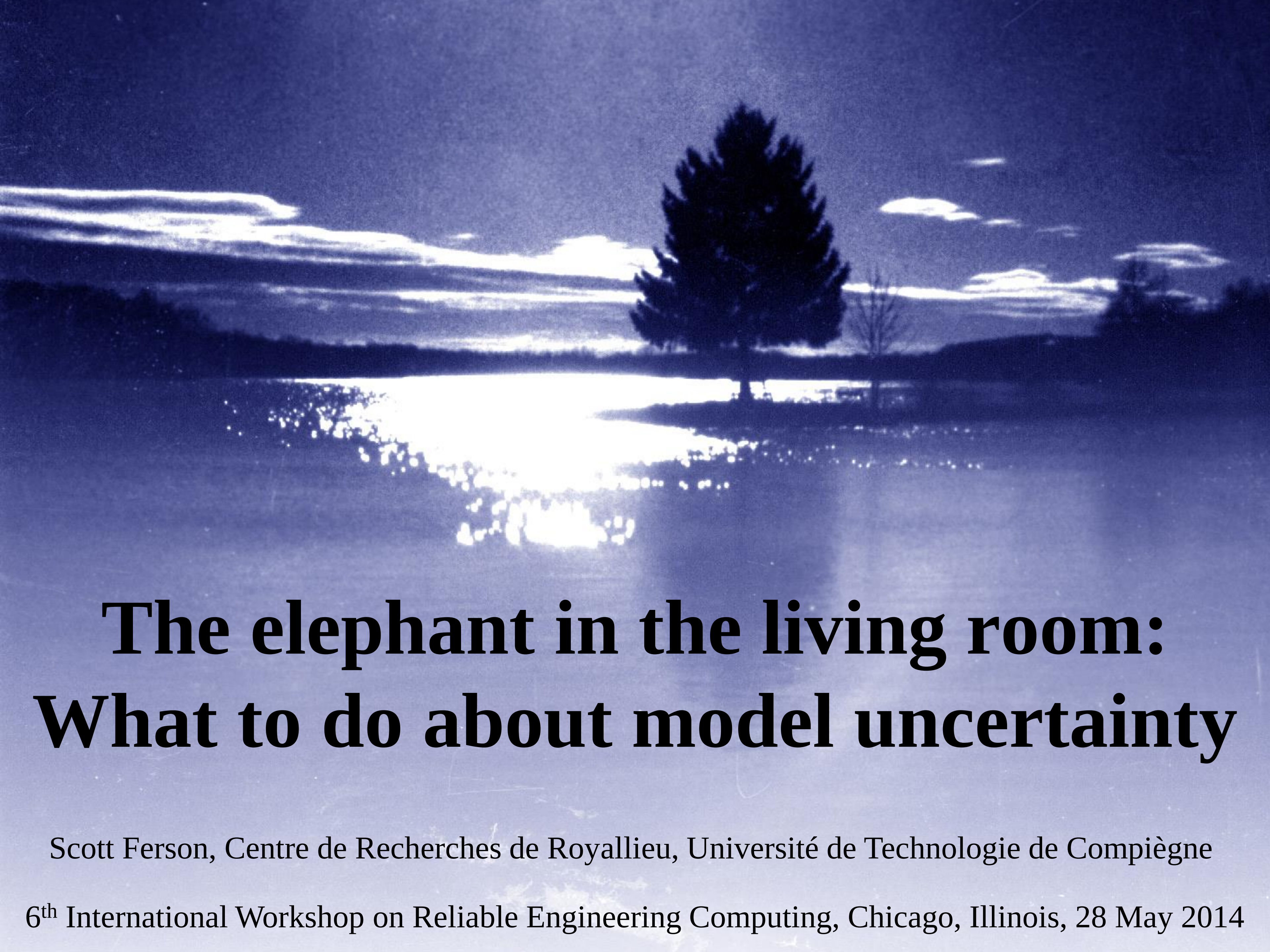


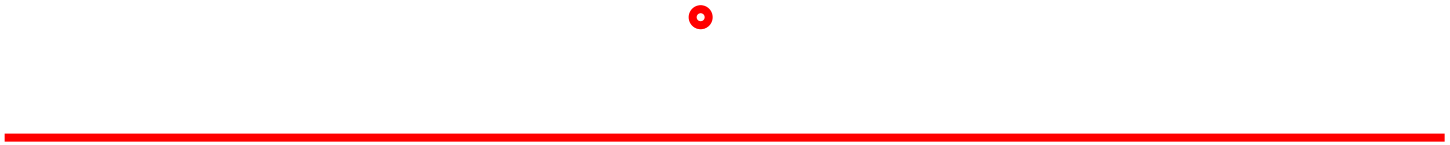
The elephant in the living room: What to do about model uncertainty

Scott Ferson, Centre de Recherches de Royallieu, Université de Technologie de Compiègne

6th International Workshop on Reliable Engineering Computing, Chicago, Illinois, 28 May 2014

Euclid

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



For over 20 centuries, the answer was ‘one’

Relax one axiom

- Non-Euclidean geometries say the answer is either ‘zero’ or ‘many’

- Controversial, but eventually accepted

- Richer mathematics and broader applications

Current tumult in uncertainty theory

- Relaxing one axiom of decision theory yields a notion of "non-Laplacian" uncertainty

That we can also compare any two gambles

- This uncertainty cannot be characterized by a single probability distribution
- Will eventually be embraced as essential

Epistemic uncertainty

- Arises from incomplete knowledge
- Incertitude arises from
  - limited sample size
  - mensurational limits (‘measurement error’)
  - use of surrogate data
- Reducible with empirical effort

Aleatory uncertainty

- Arises from natural stochasticity

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

- Not reducible by empirical effort

Model uncertainty

- Doubt about the structural form of the model
- Usually epistemic, not aleatory, uncertainty
- Often considerable in magnitude
- The elephant in the middle of the room

http://www.elephant-pictures.info/nice-elephant-drawing.jpg

Uncertainty in probabilistic analyses

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

model
uncertainty

Examples

- Structure
- Simplifications (aggregation, exclusion)
- Level of detail (e.g., mesh resolution)
- Boundary conditions
- Choice of scenarios
- Extrapolations
- Conceptual model versus reality

General strategies

1. Sensitivity (what-if) analysis

2. Monte Carlo model averaging

3. Bayesian model averaging

4. Enveloping analyses

1. Sensitivity (what-if) studies

IPCC uses

- Simply re-computes the analysis with alternative assumptions

- Keeps track of all results and presents this array to the decision maker

  – Intergovernmental Panel on Climate Change

2. Monte Carlo model averaging

NRC uses

- Identify all possible models
- Translate model uncertainty into choices about distributions
- Average probability distributions
  - Easy in Monte Carlo by selecting model randomly
- Use weights to account for different credibility (or assume equiprobability)

3. Bayesian model averaging

- Similar to the Monte Carlo model averaging

- Updates prior probabilities to get weights

- Takes account of available data

Bayesian model averaging

- Assume it's actually first model
- Compute probability distribution for $f(A,B)$
- Read off probability density of observed data
  - That's the likelihood for that model
- Repeat above steps for each model
- Compute posterior $\propto$ prior $\times$ likelihood
  - This gives the Bayes' factors
- Use the posteriors as weights for the mixture

4. Enveloping probabilities

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

Numerical example

The function *f* is one of two possibilities. Either

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

or

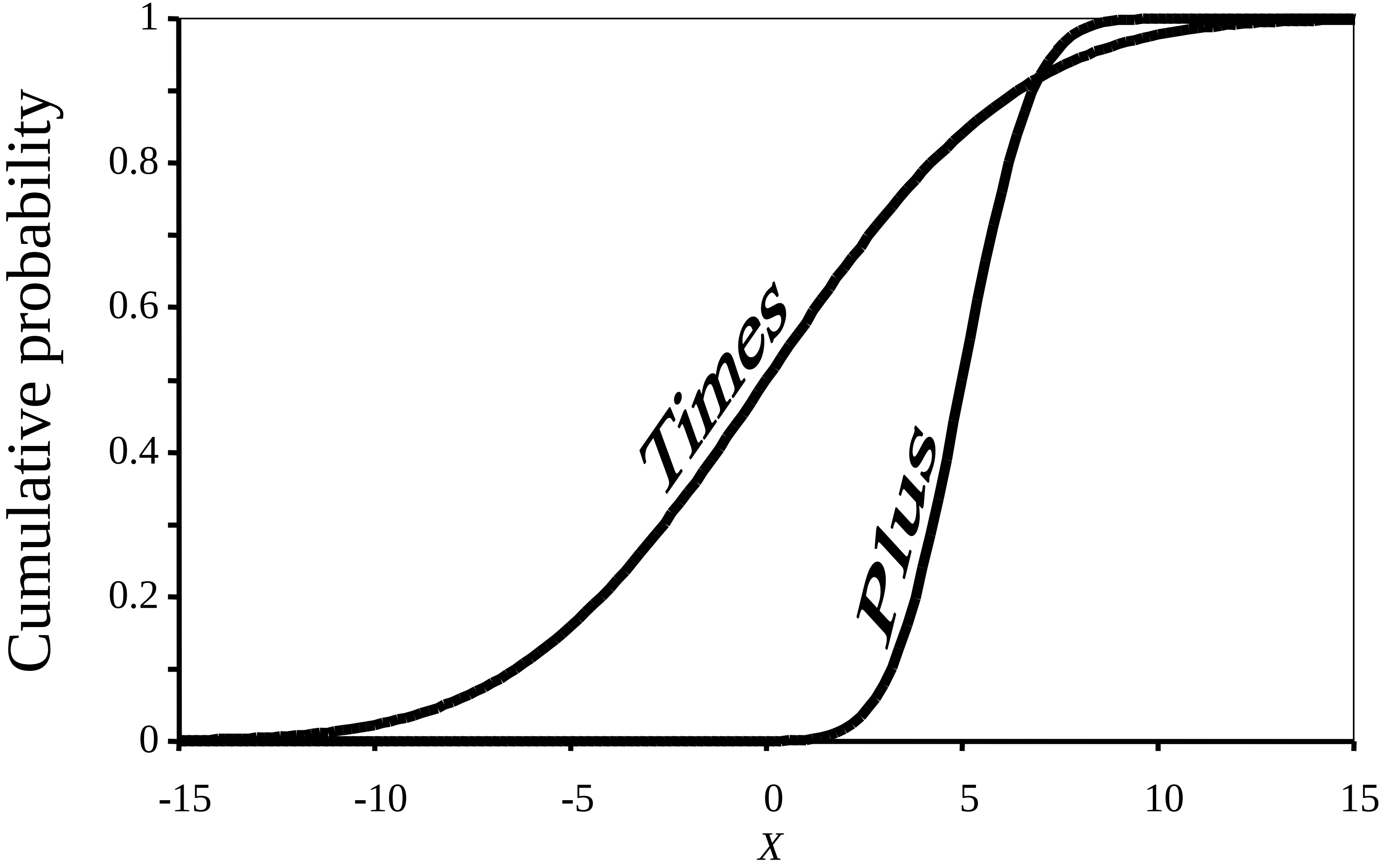
$$f(A,B) = f_{\text{Times}}(A, B) = A \times B$$

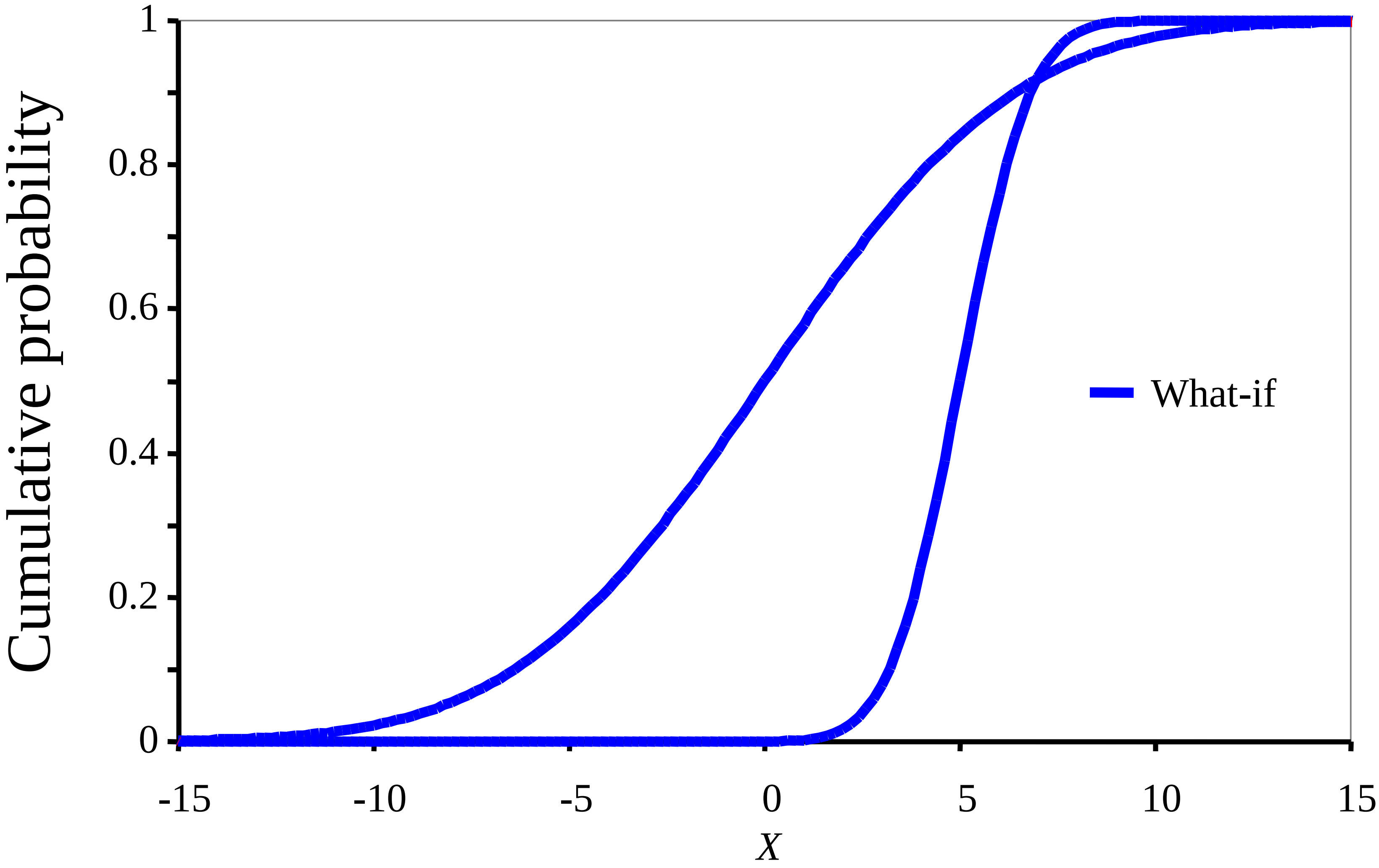
is correct, but we don't know which. Suppose

$$A \sim \text{normal}(0, 1)$$

$$B \sim \text{normal}(5, 1)$$

What can we say about $f(A, B)$?


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
What-if

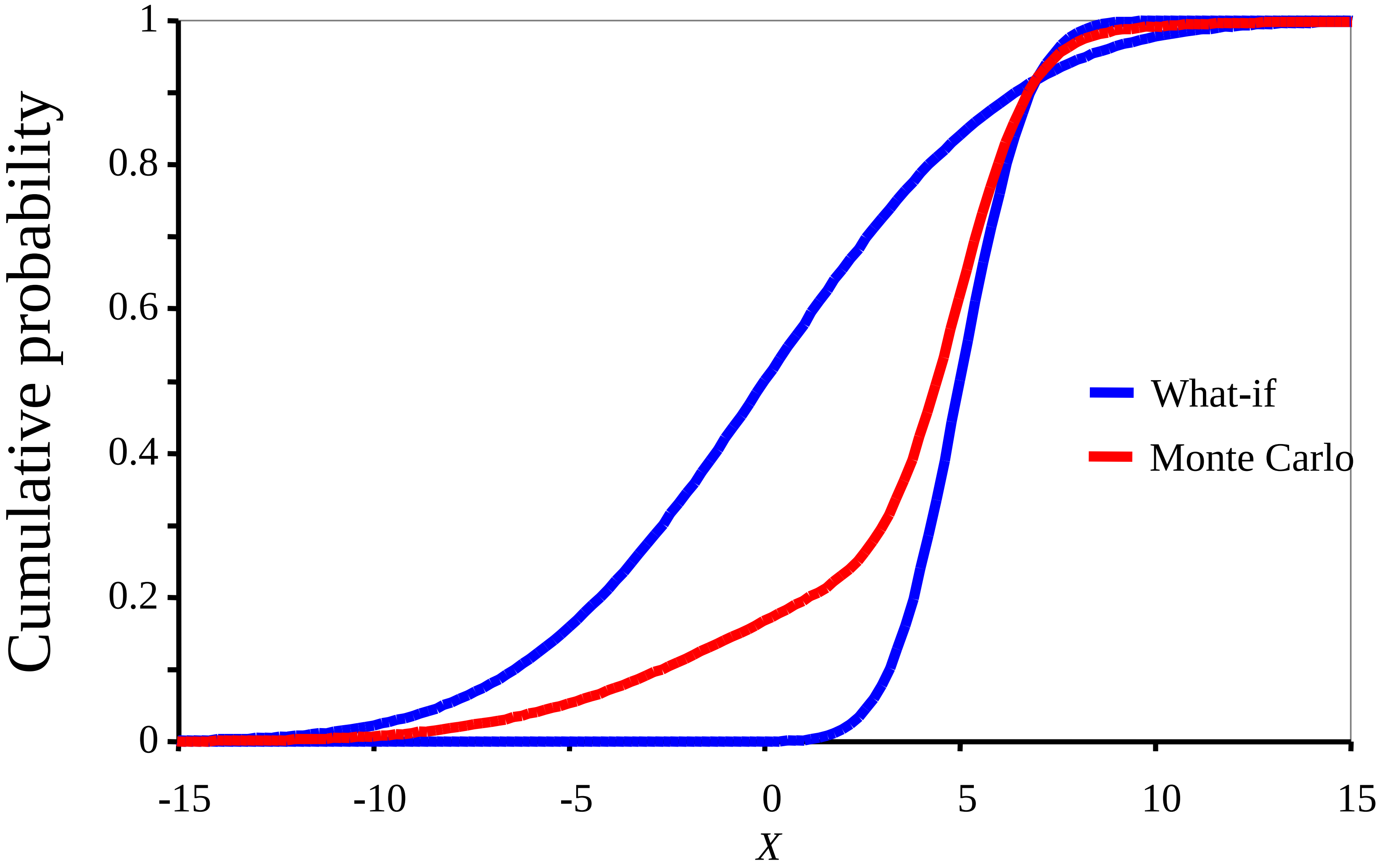
Monte Carlo model averaging

Same A and B

f is either Plus or Times

but Plus is twice as likely as Times

prob(Plus) = 2/3, prob(Times) = 1/3


Cumulative probability
1
0.8
0.6
0.4
0.2
0
-15
-10
-5
0
5
10
15
X
What-if
Monte Carlo

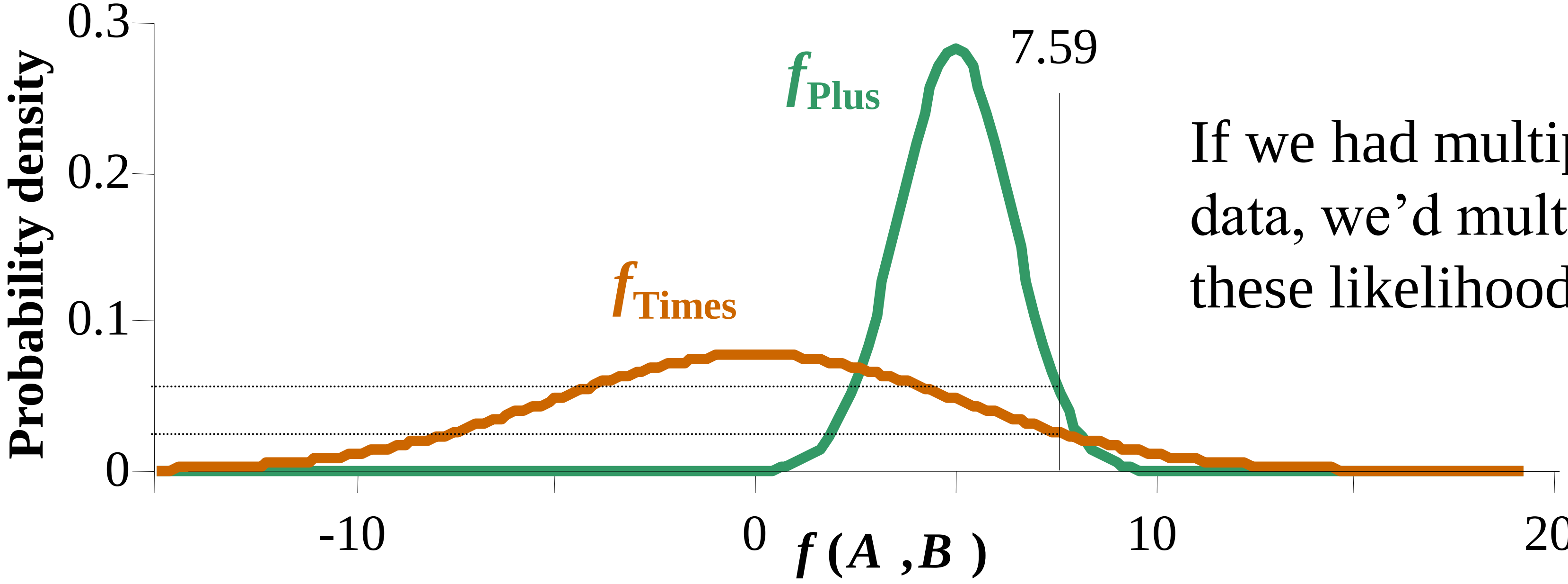
Bayesian model averaging

Same *A* and *B*

f either Plus or Times; Plus twice as likely

one observation $f(A,B) = 7.59$

Likelihoods



If we had multiple data, we'd multiply these likelihoods.

$f_{\text{Plus}}(A,B) \sim A+B \sim \text{normal}(5, \sqrt{2})$ $\quad$ $L_{\text{Plus}}(7.59) = 0.05273$

$f_{\text{Times}}(A,B) \sim A \times B \sim \text{normal}(0, \sqrt{26})$ $\quad$ $L_{\text{Times}}(7.59) = 0.02584$

R: dnorm(7.59,5,sqrt(2))

Excel: =NORMDIST(7.59, 5, SQRT(2), FALSE)

Weights

Posterior probabilities for the two models

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

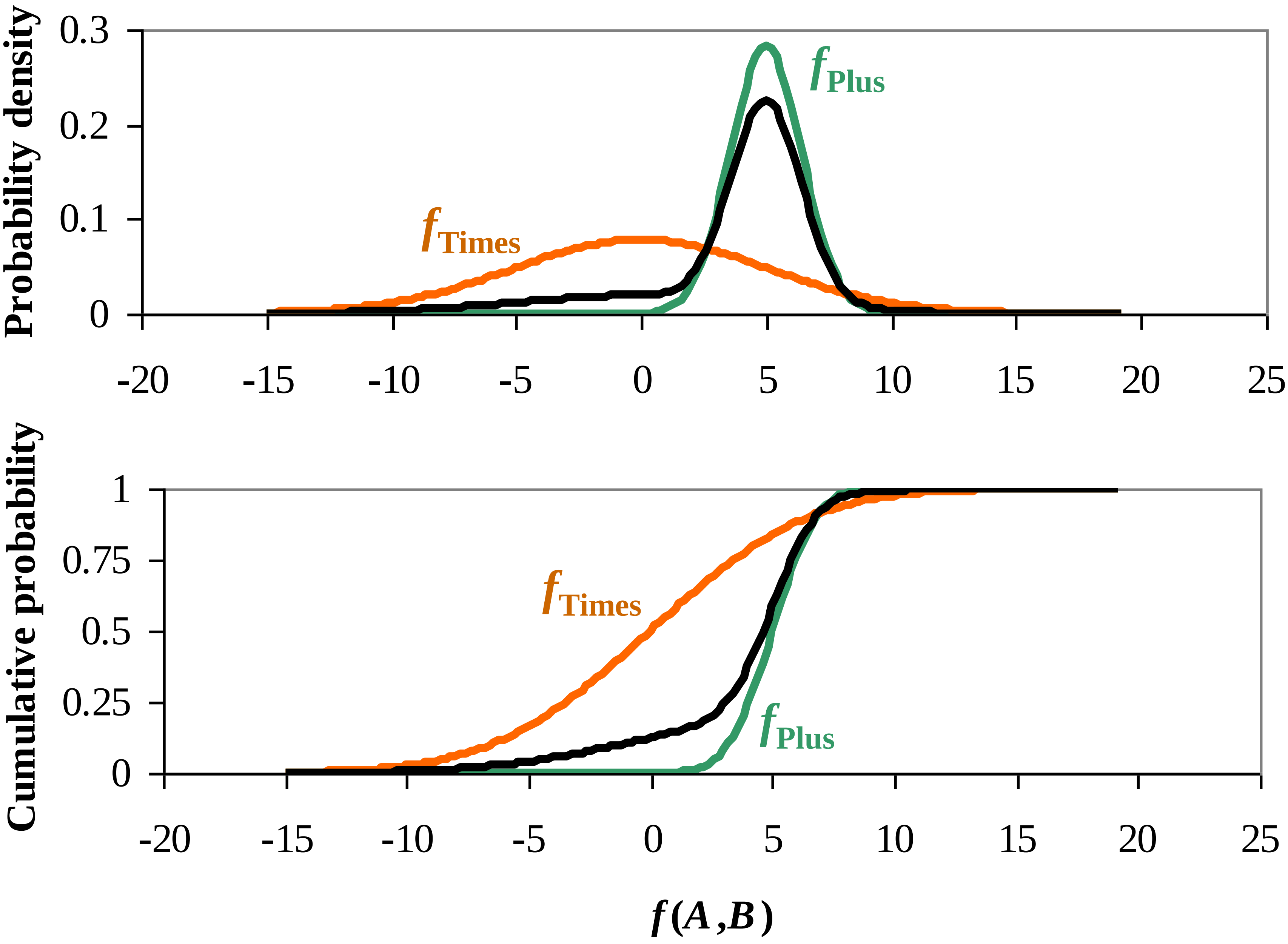
Plus

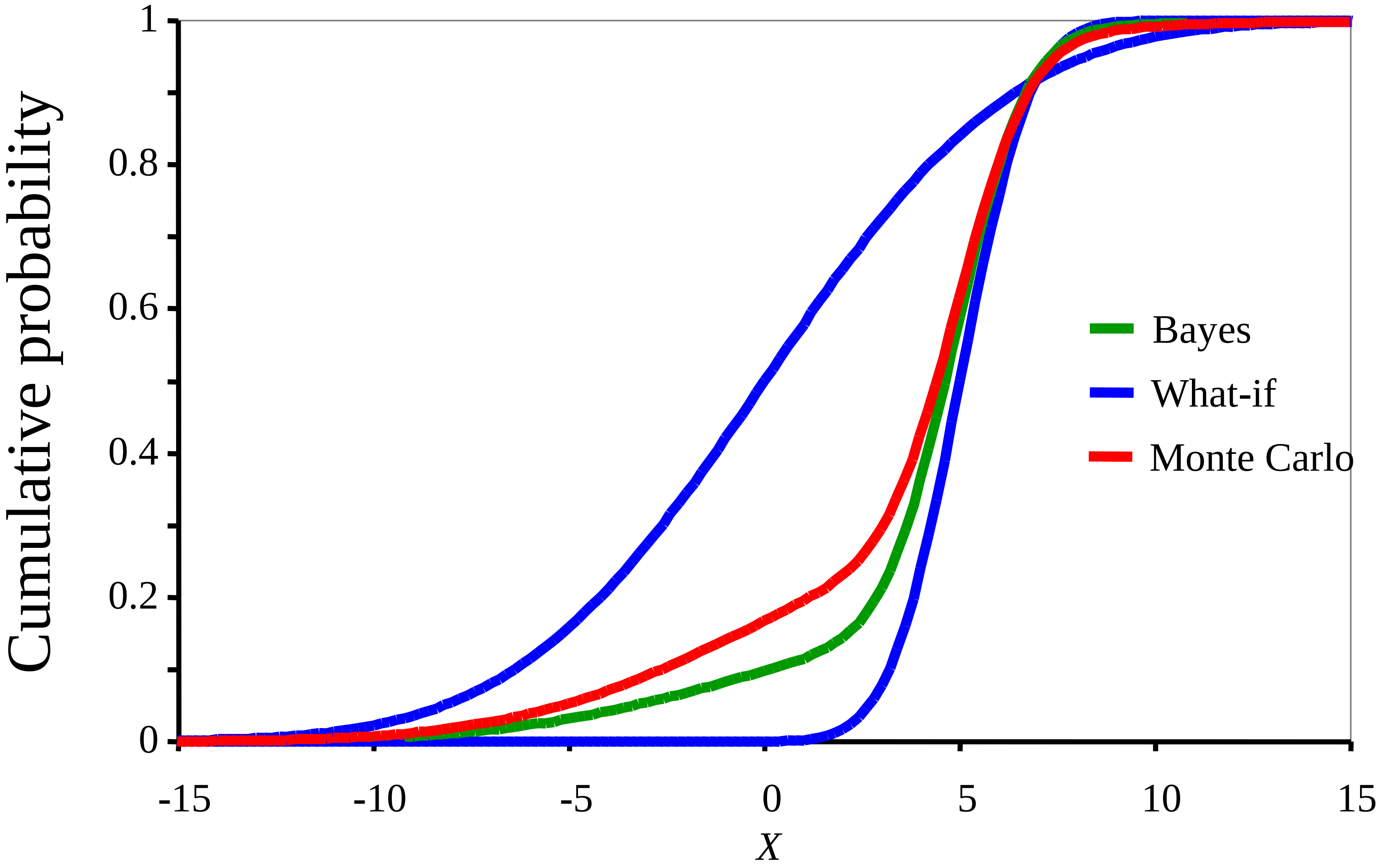
$$0.6\times0.05273/\underbrace{(0.6\times0.05273+0.4\times0.02584)}_{\text{normalization factor}}=0.7538$$

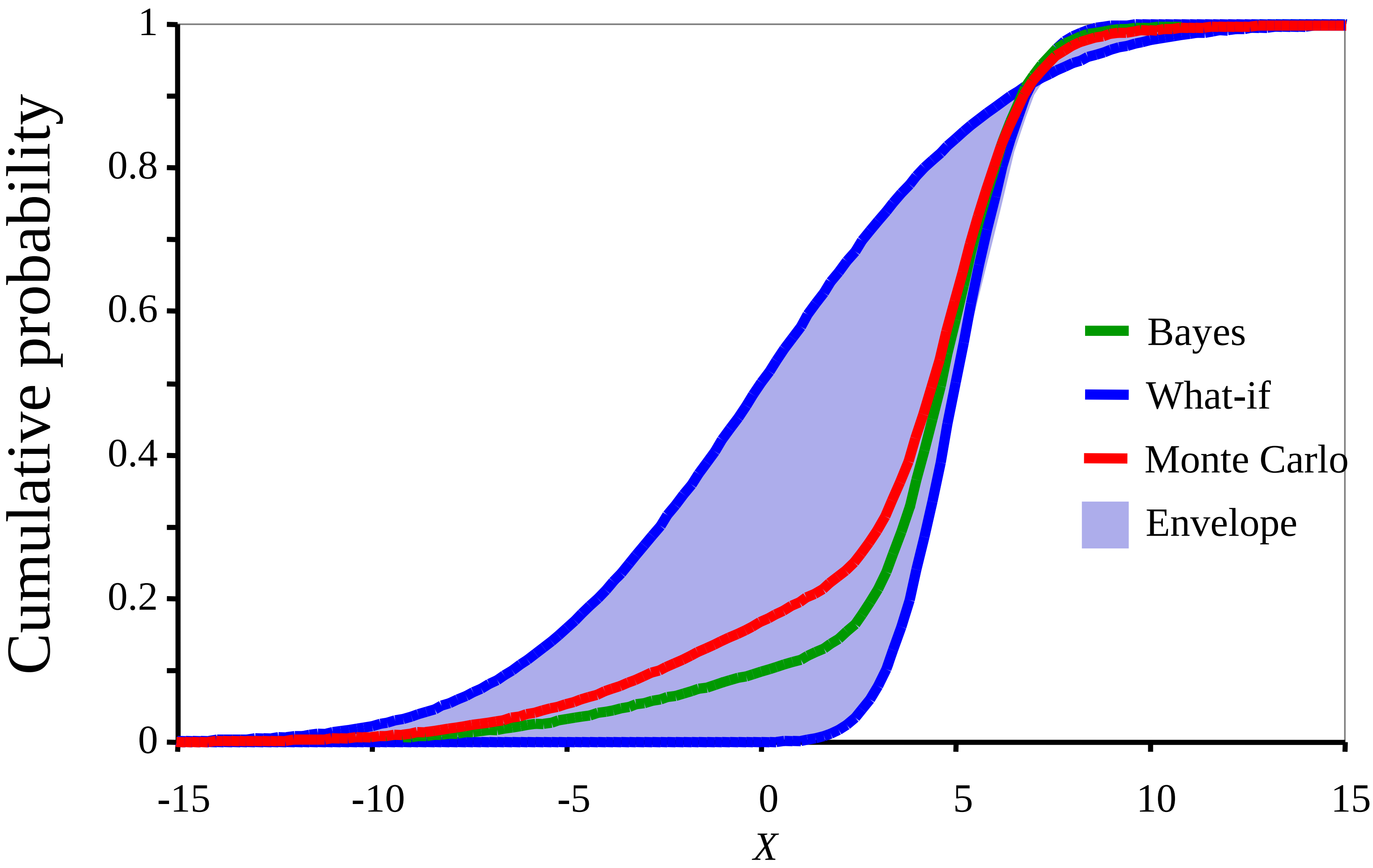
Times

$$0.4\times0.02584/\overbrace{(0.6\times0.05273+0.4\times0.02584)}=0.2462$$

These are the weights for the mixture distribution


Probability density
0.3
0.2
0.1
0
f_{Plus}
f_{Times}
-20
-15
-10
-5
0
5
10
15
20
25
Cumulative probability
1
0.75
0.5
0.25
0
f_{Times}
f_{Plus}
-20
-15
-10
-5
0
5
10
15
20
25
$f(A,B)$


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


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

Sensitivity analysis

- Simple theory
- Straightforward to implement
- Doesn’t confuse aleatory and epistemic

- Must enumerate all possible models
- Combinatorial complexity
- Hard to summarize

Drawbacks of what-if

- Consider a long-term model of the economy under global climate change 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

Monte Carlo modal averaging

- Produces single distribution as answer
- Can account for differential credibility
- (Stochastic mixture)

Monte Carlo model averaging

- State of the art in probabilistic risk analysis
  - Nuclear power plant assessments

- Need to know what all the possibilities are

- If don’t know the weights, assume equality

Drawbacks of Monte Carlo averaging

- If you cannot enumerate the possible models, you can’t use this approach

- Averages together incompatible theories and yields an answer that no theory supports

- Can underestimate tail risks

Bayesian model averaging

- Produces single distribution as answer
- Can account for differential prior credibility
- 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 probability

- Straightforward theoretically
- Yields single mathematical object as answer
- Doesn’t confuse aleatory and epistemic
- Doesn’t underestimate tail risks

Drawbacks of enveloping

- Cannot account for different model credibilities
- Can’t make use of data
- Doesn’t account for ‘holes’
  - Optimality may be computationally expensive

Bayesian model averaging

(Draper 1995)

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

Bayesian model averaging

- Assume it's actually the first model
- Compute probability distribution under that model
- Read off probability density of observed data
  - Product if multiple data; it' s the likelihood for that model
- Repeat above steps for each model
- Compute posterior ∝ prior × likelihood
- Use the posteriors as weights for the mixture

Strategy for enumerable 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, although it may be wide

But can we use envelopes?

- Yes, we can compute with them *directly*

- Several related approaches
  - Dempster-Shafer evidence theory
  - Probability bounds analysis
  - Robust Bayes methods
  - Imprecise probabilities
  - others

Special case: distribution shape

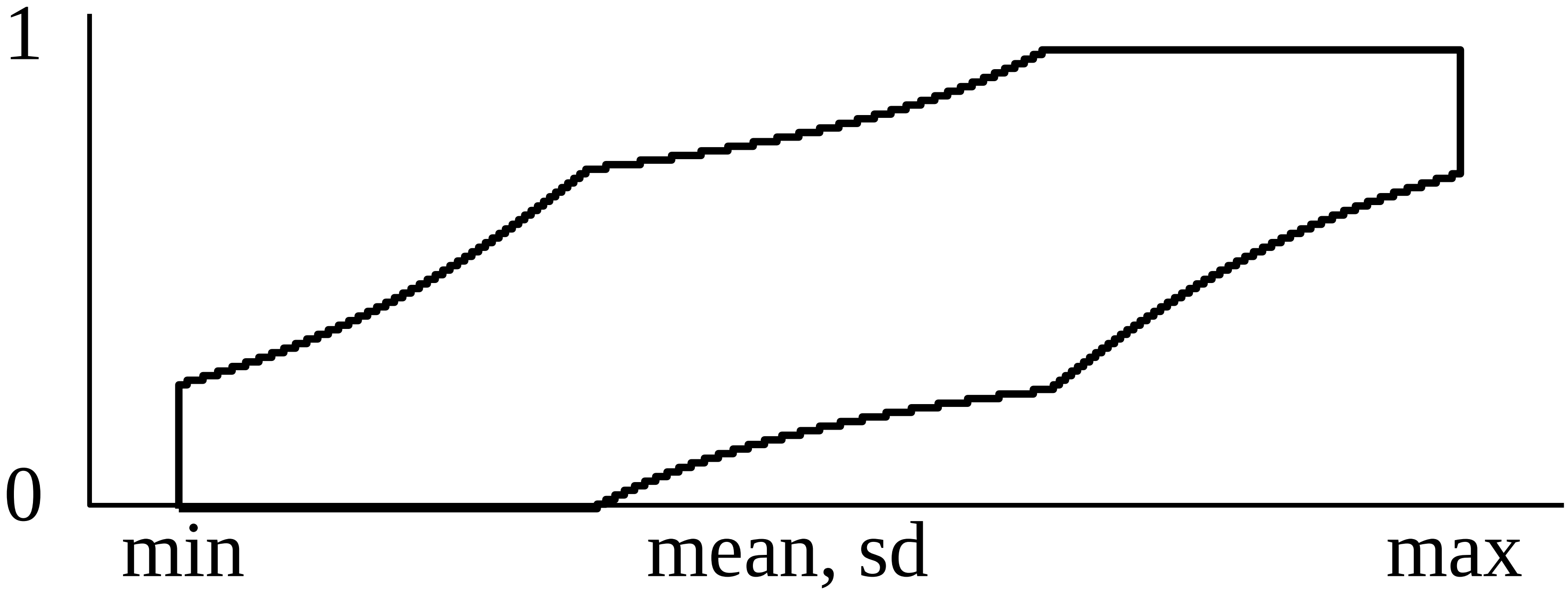
Uncertainty about distribution shape

- Can we use normal distributions for everything?
- Is this distribution gamma, Weibull or lognormal?
- Could it be a Gumbel distribution?
- Could it be some *unnamed* distribution?

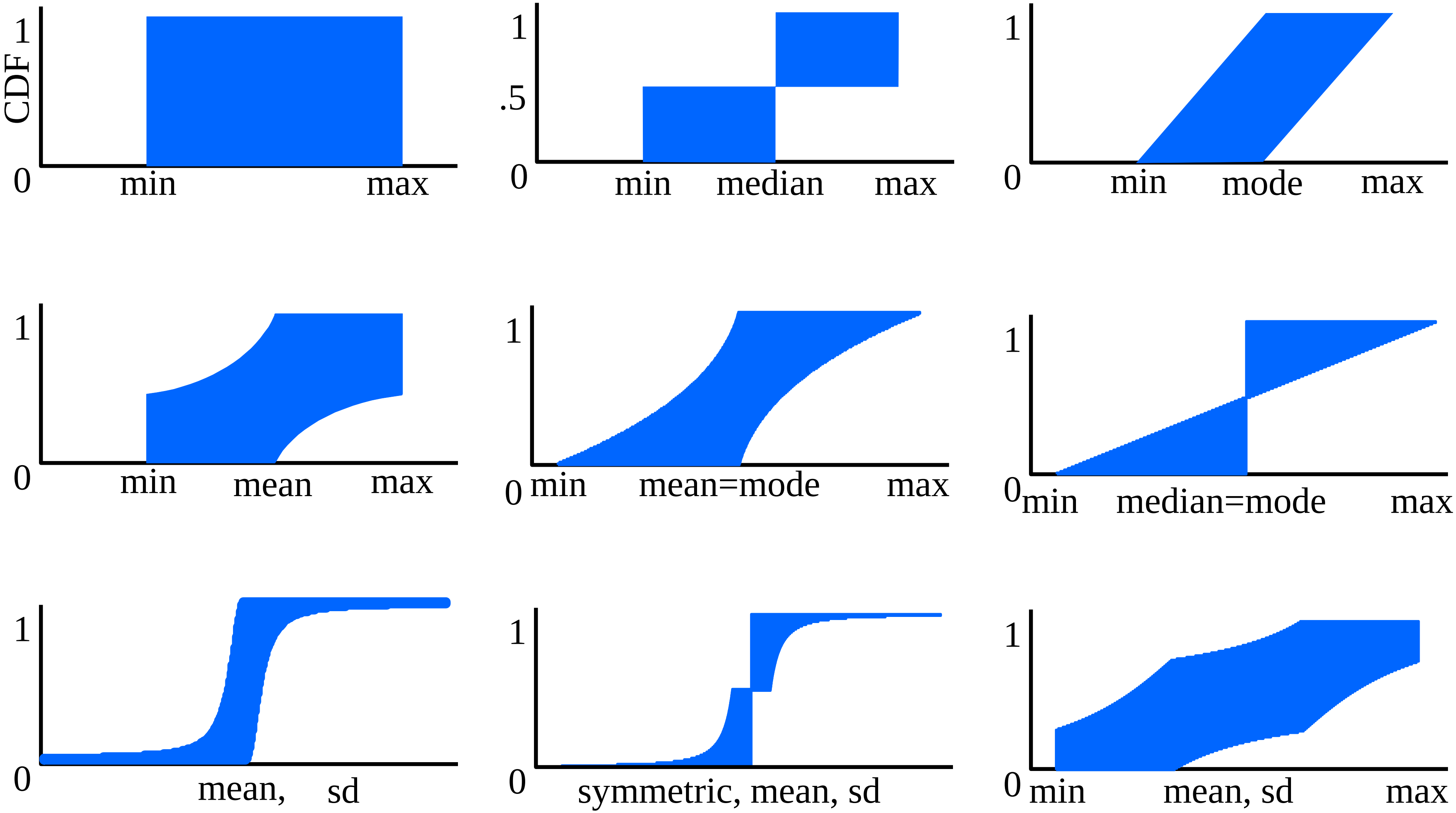
- Some analysts just try several distribution shapes but this is unsatisfactory because there are uncountably many possible shapes

P-boxes

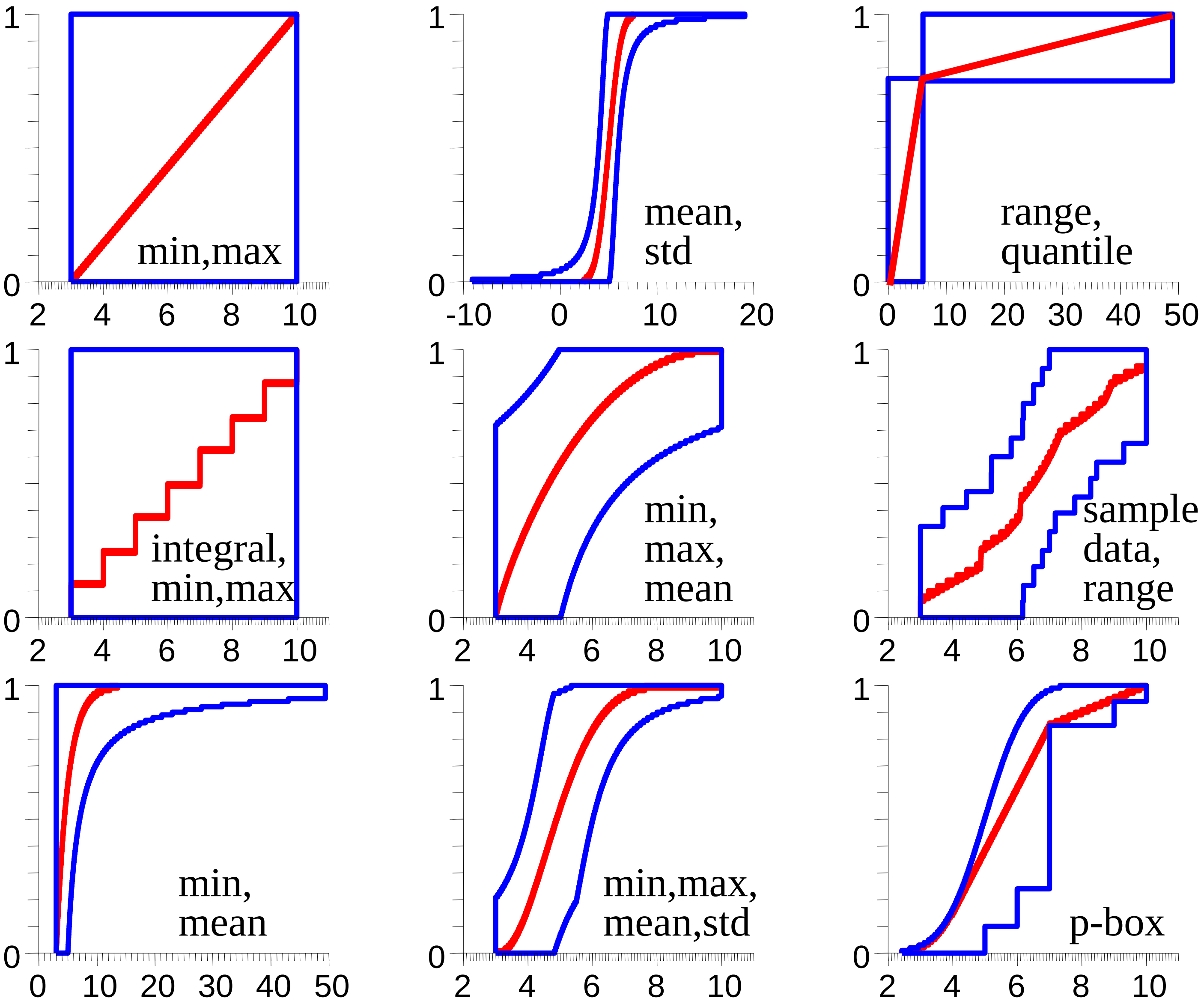
- P-boxes were invented to address this issue
- Can define p-boxes by specifying constraints



Ready solutions for many problems



Comparing p-boxes with maximum entropy distributions

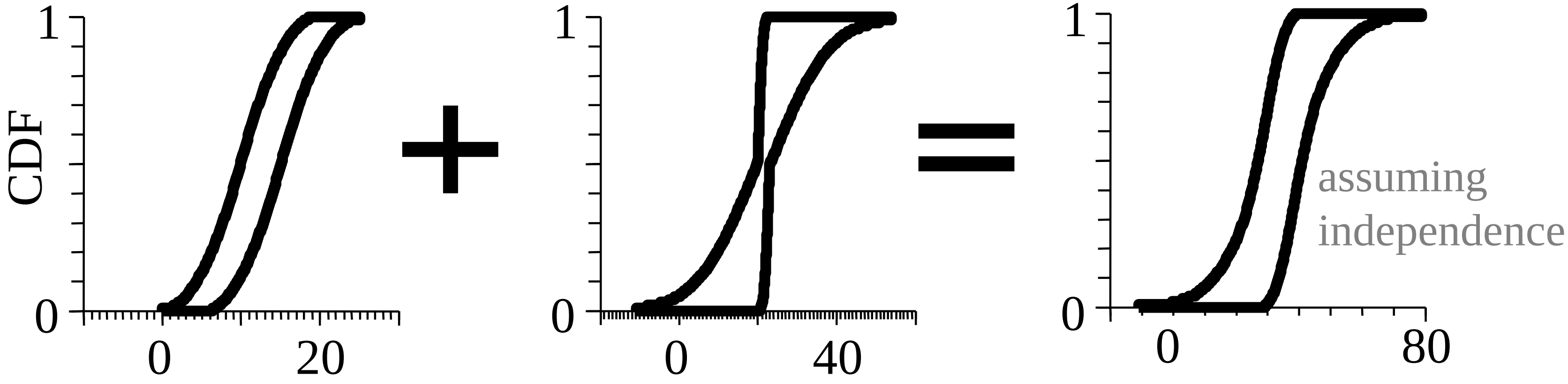


Maximum entropy’s problem

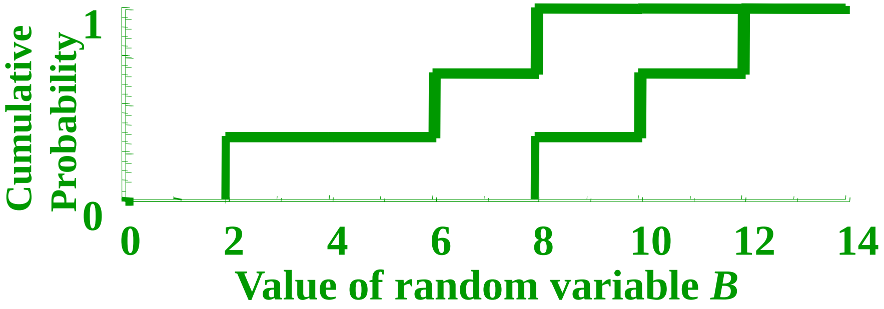
- Depends on the choice of scale
- A solution in terms of degradation rate is incompatible with one based on half life even though the information is equivalent
- P-boxes are the same whichever scale is used

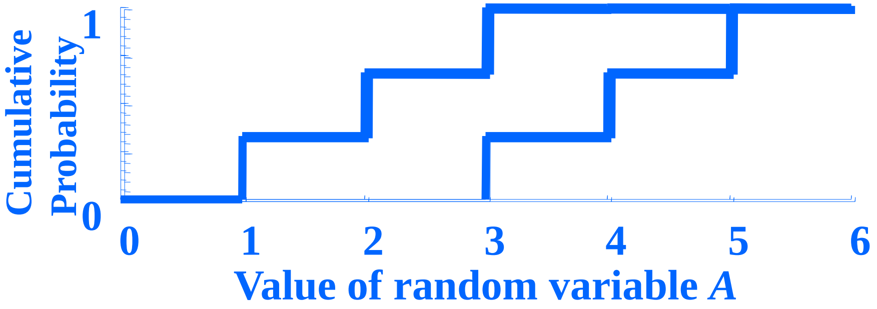
Warner North interprets Ed Jaynes as saying that “two states of information that are judged to be equivalent should lead to the same probability assignments”. Maxent doesn’t do this! But PBA does.

Probability bounds analysis



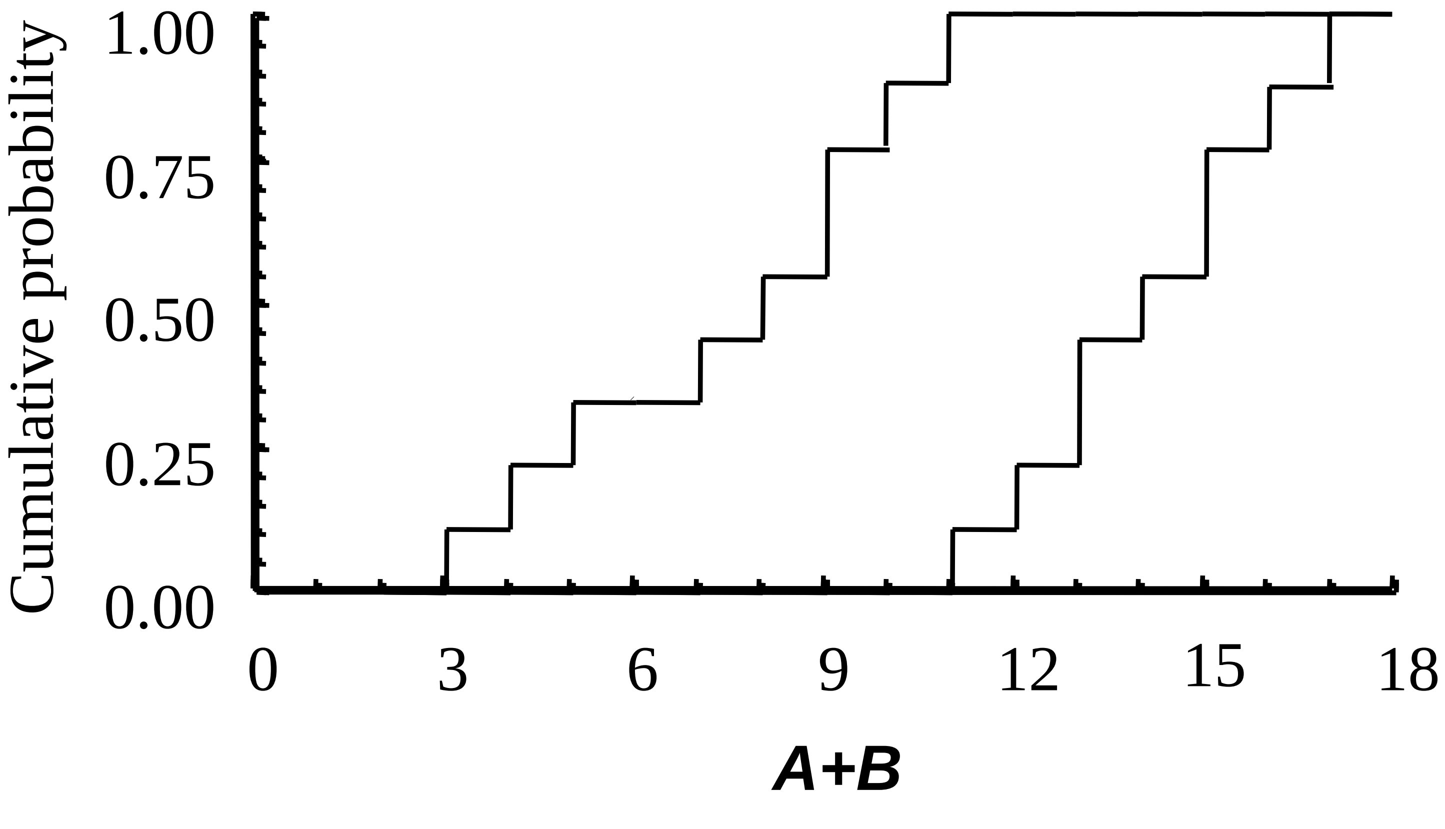
assuming independence





$A+B$ 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



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

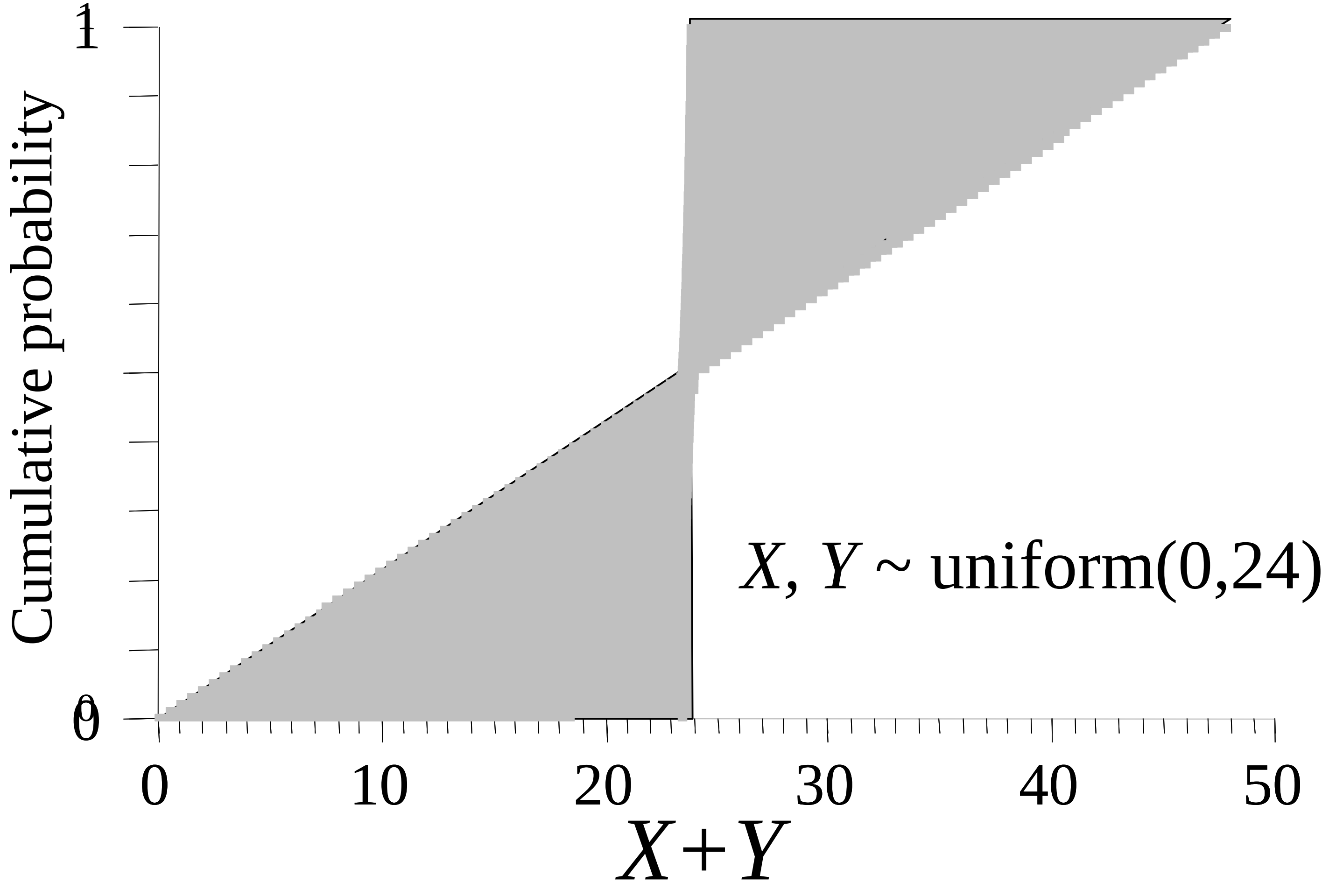
Special case: distribution shape

Uncertainty about dependence

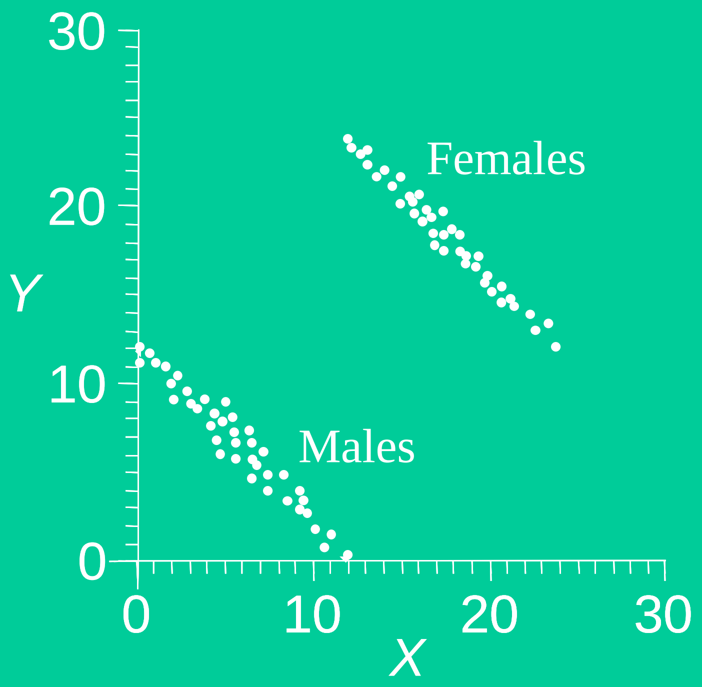
- Sensitivity analyses usually used
  - – Vary correlation coefficient between –1 and +1

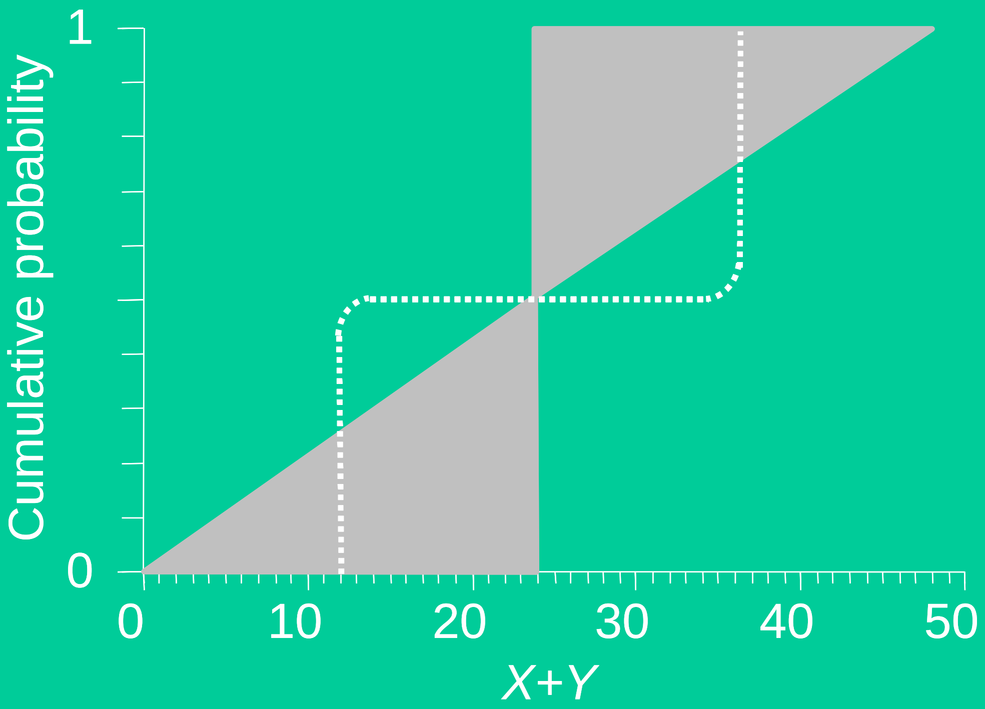
- But this *underestimates* the true uncertainty
  - – Example: suppose X, Y ~ uniform(0,24) but we don't know the dependence between X and Y

Varying the correlation coefficient

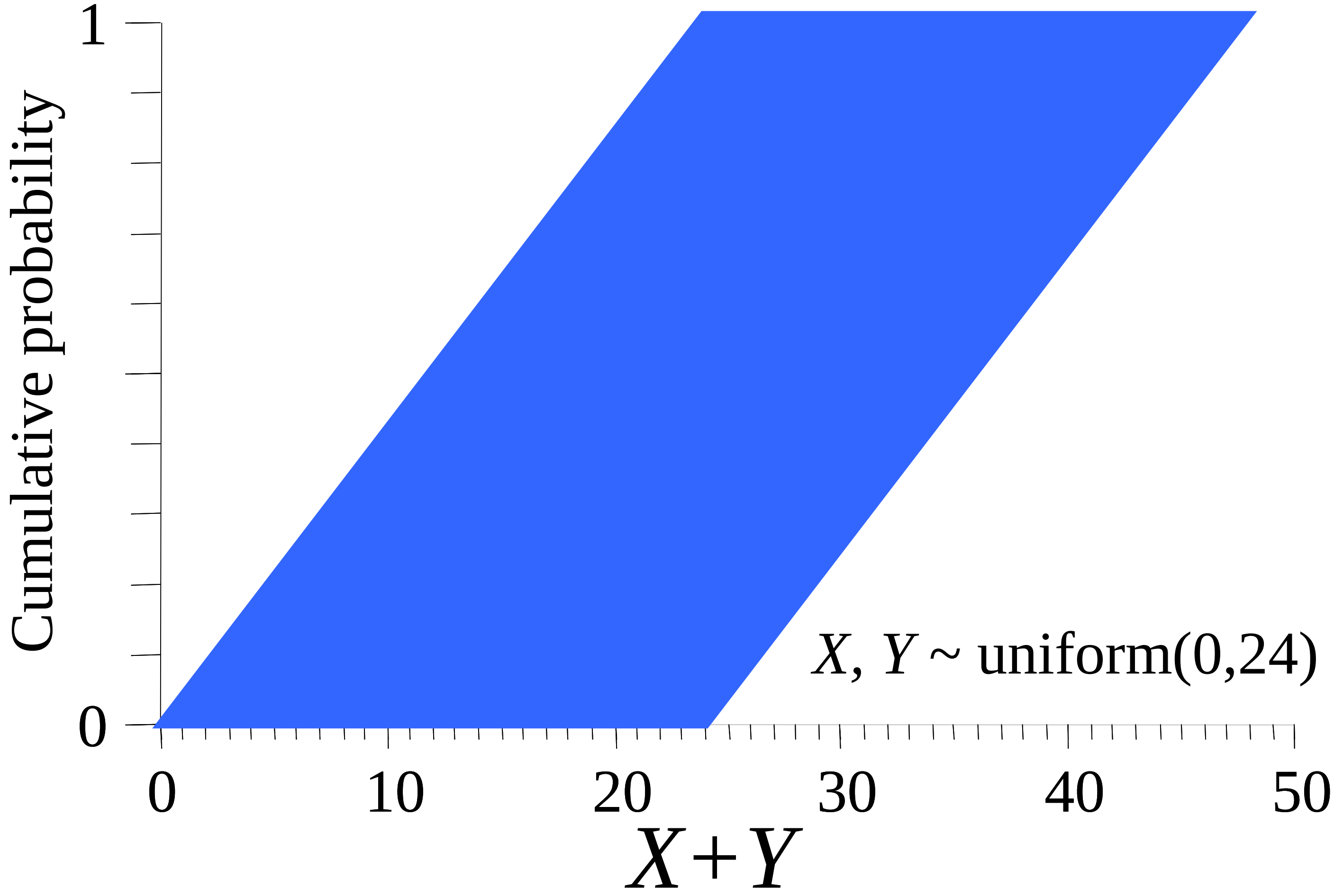


Counterexample: outside the cone!





Unknown dependence



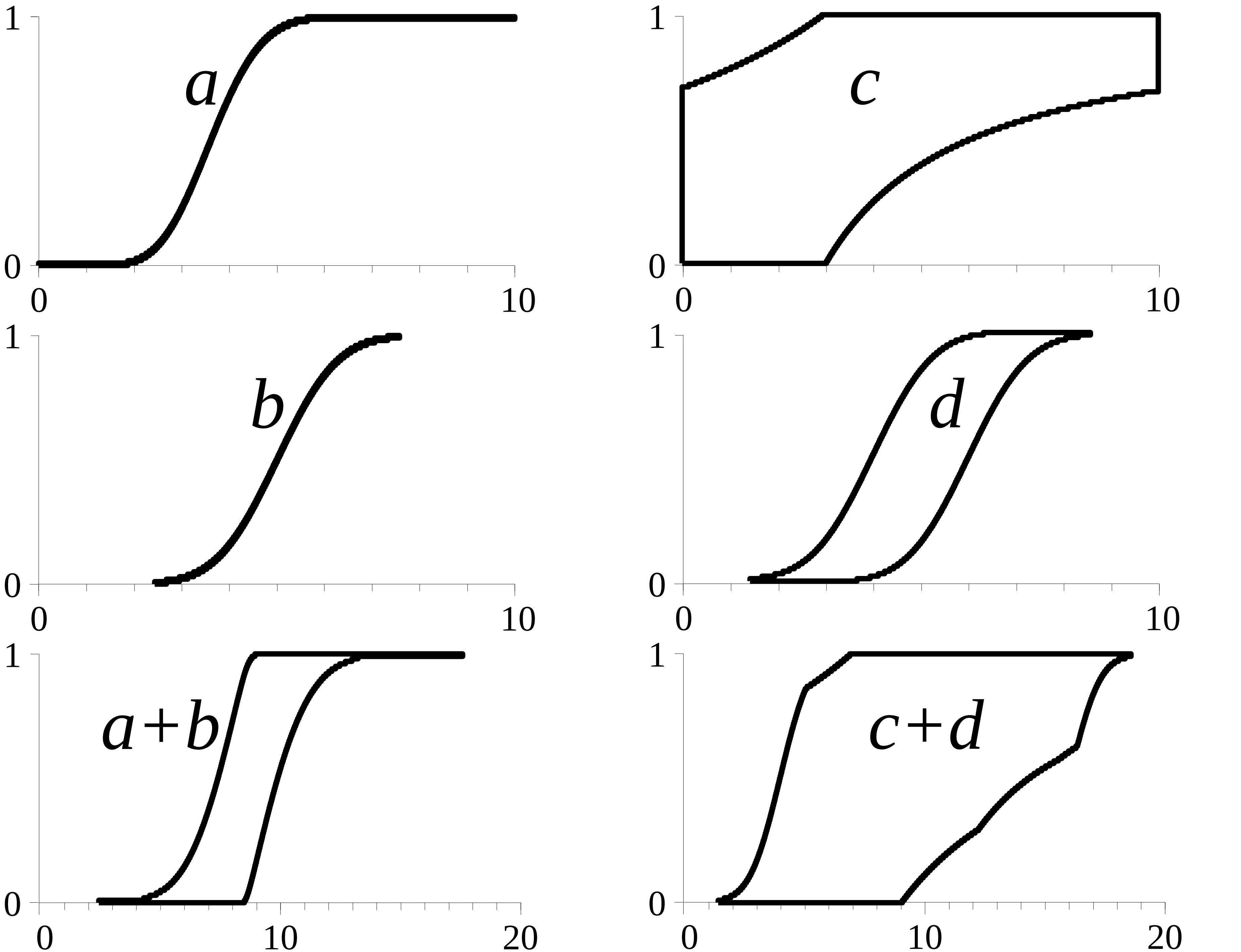
Fréchet bounds

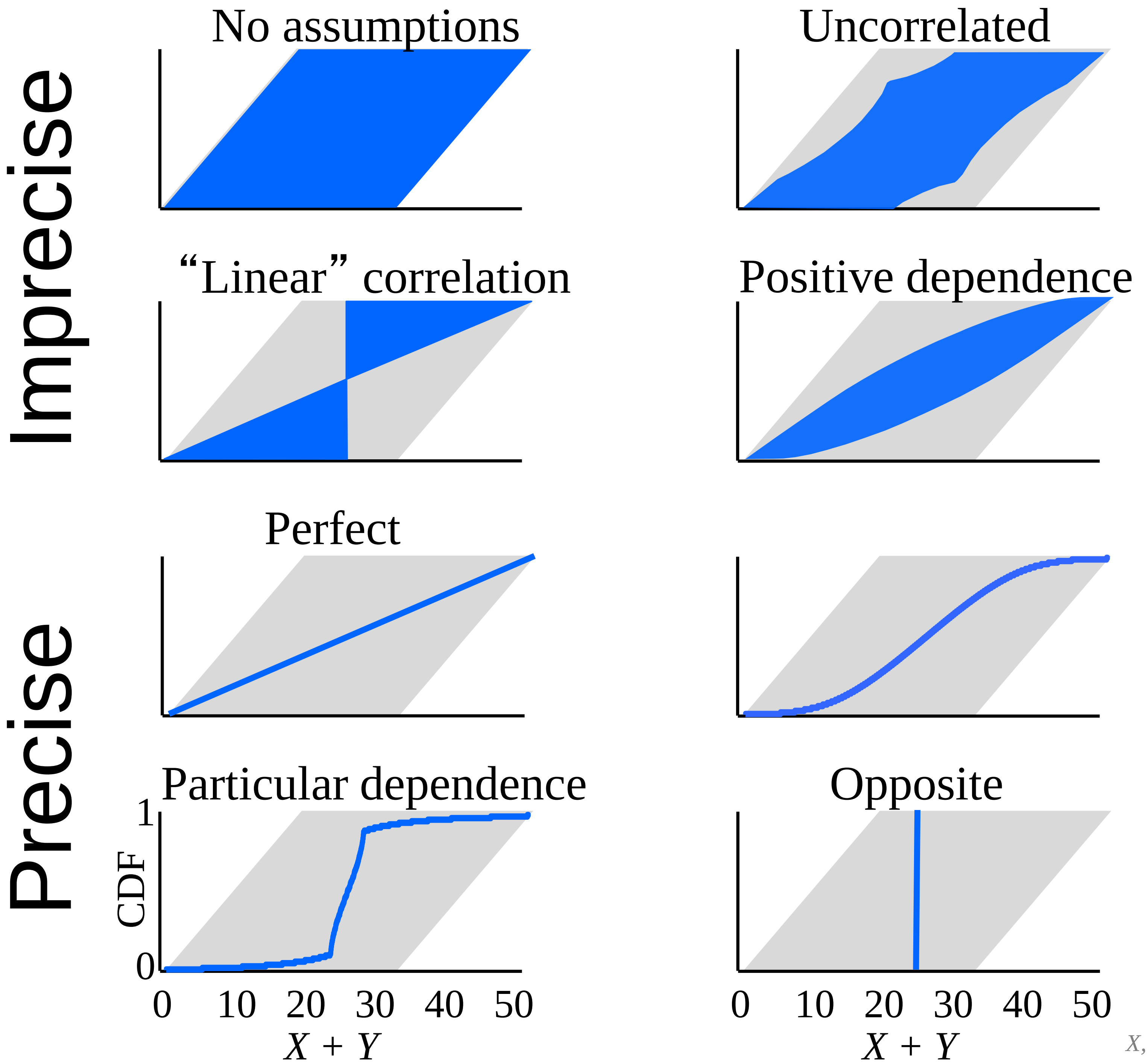
- If $X \sim F$ and $Y \sim G$, the distribution of $X+Y$, is bounded by

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

and these bounds are pointwise best-possible.

Plus, they're simpler to compute than the distribution under independence, which involve integrals


1
0
0
10
a
1
0
0
10
c
1
0
0
10
b
1
0
0
10
d
1
0
0
10
20
a+b
1
0
0
10
20
c+d



No assumptions

Uncorrelated

“Linear” correlation

Positive dependence

Perfect

Particular dependence

Opposite

X,*Y* ~ uniform(1,24)

Uncertainty about dependence

Solved

- Neither sensitivity studies nor Monte Carlo simulation can comprehensively assess it
- Bayesian model averaging can’t even begin
- Only bounding strategies work
- Fréchet bounding lets you be sure

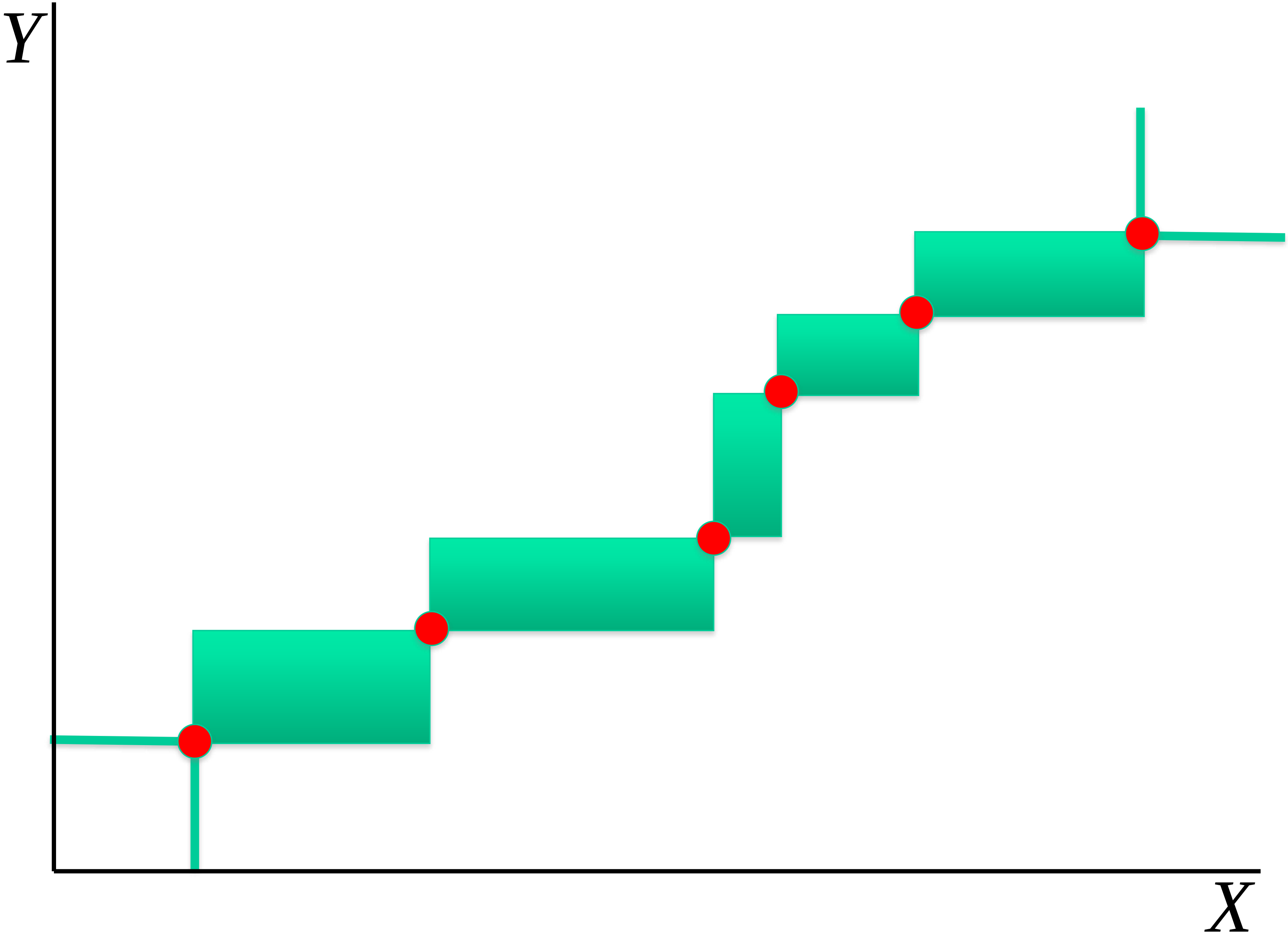
Interpolations

(deterministic functions)

Constrained family of functions

- Sometimes we have some information about a function

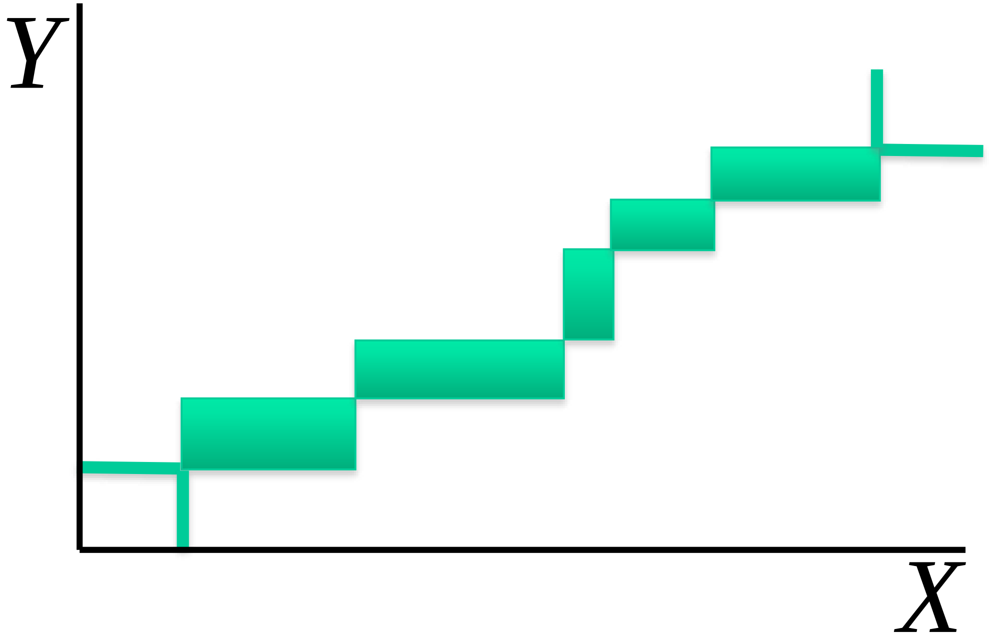
- For example,
  - Deterministic
  - Some function points
  - Monotonicity


Y
X

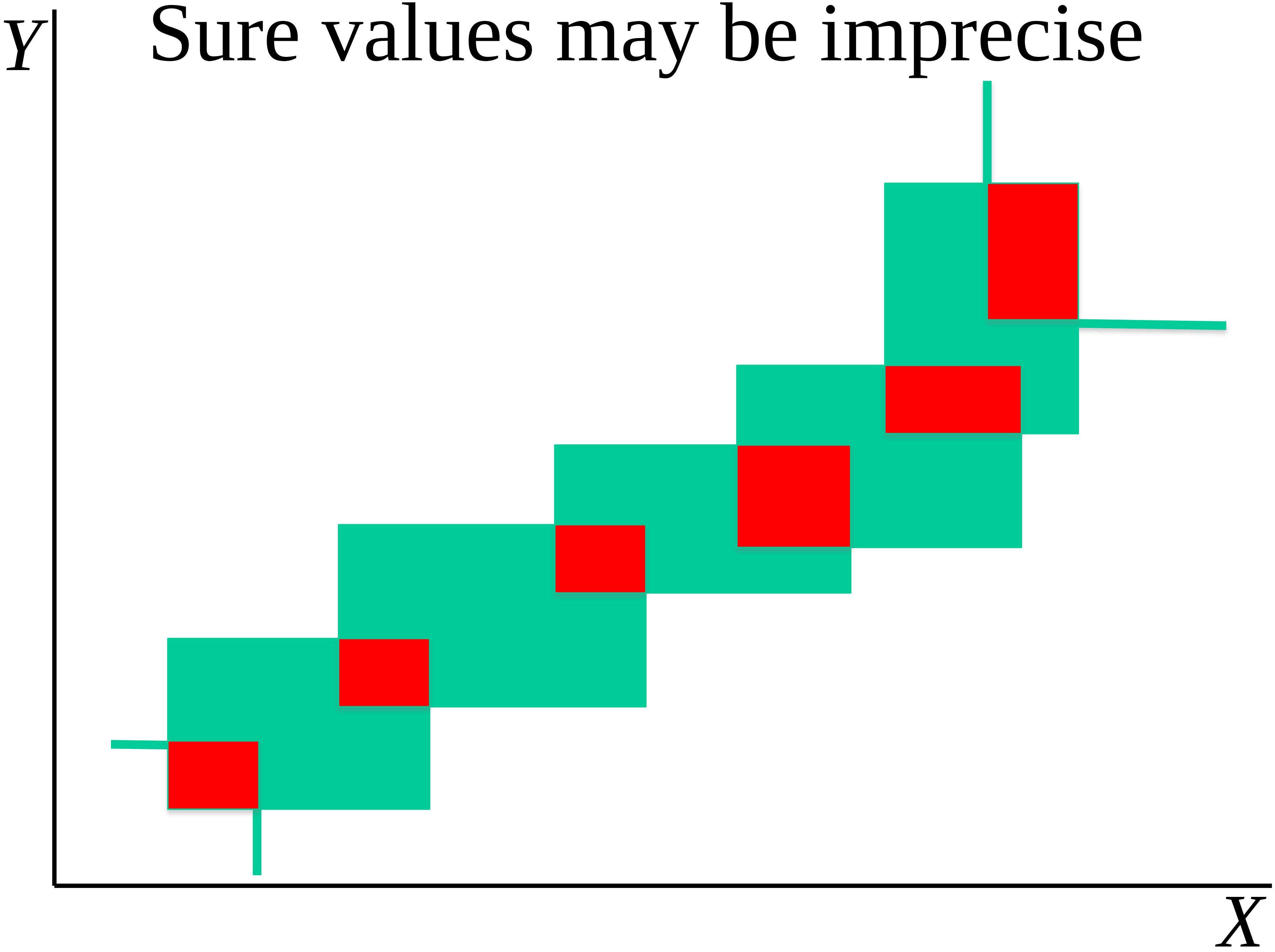
Projects function uncertainty

- Uncertainty about the function is propagated into the uncertainties about the *Y*-values

points ➔ intervals, and distributions ➔ p-boxes



Sure values may be imprecise



Kriging (GP regression)

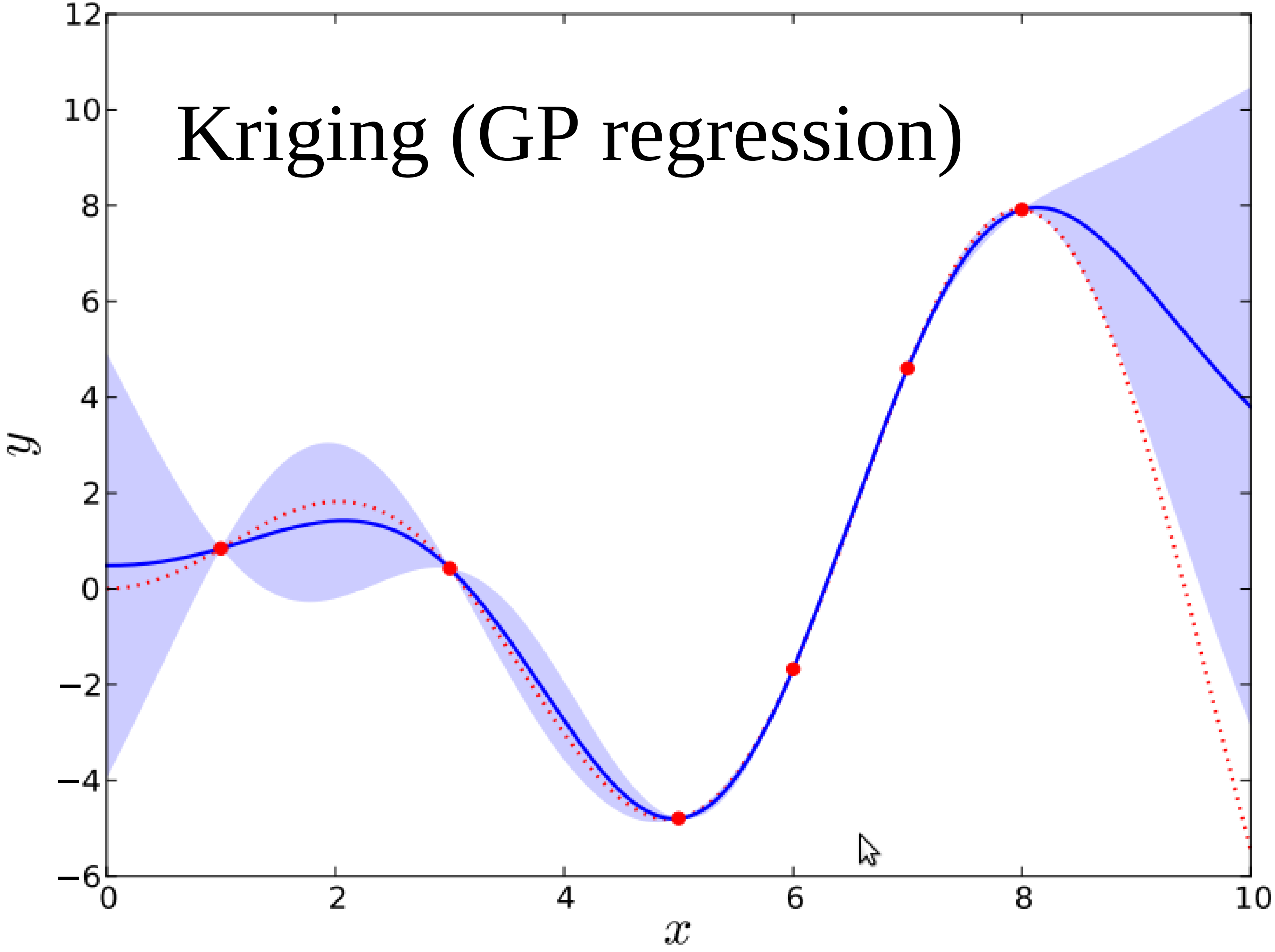


Figure from Vincent Dubourg's dissertation *Adaptive surrogate models for reliability analysis and reliability-based design optimization*

Kriging (Gaussian process regression)

- Sideways Gaussians; σ varies at each point
- Not good for extrapolations, or if *n* is small
- Assumes there’s a fixed function
- Measured values may actually be recorded with error (imprecise Gaussian process)

“Nonparametric”

- These constrained function families are nonparametric methods

- Still have assumptions (and model uncertainty)

- Uncertainty attached to choice of kernel shape or smoothing function

Regressions

(stochastic functions)

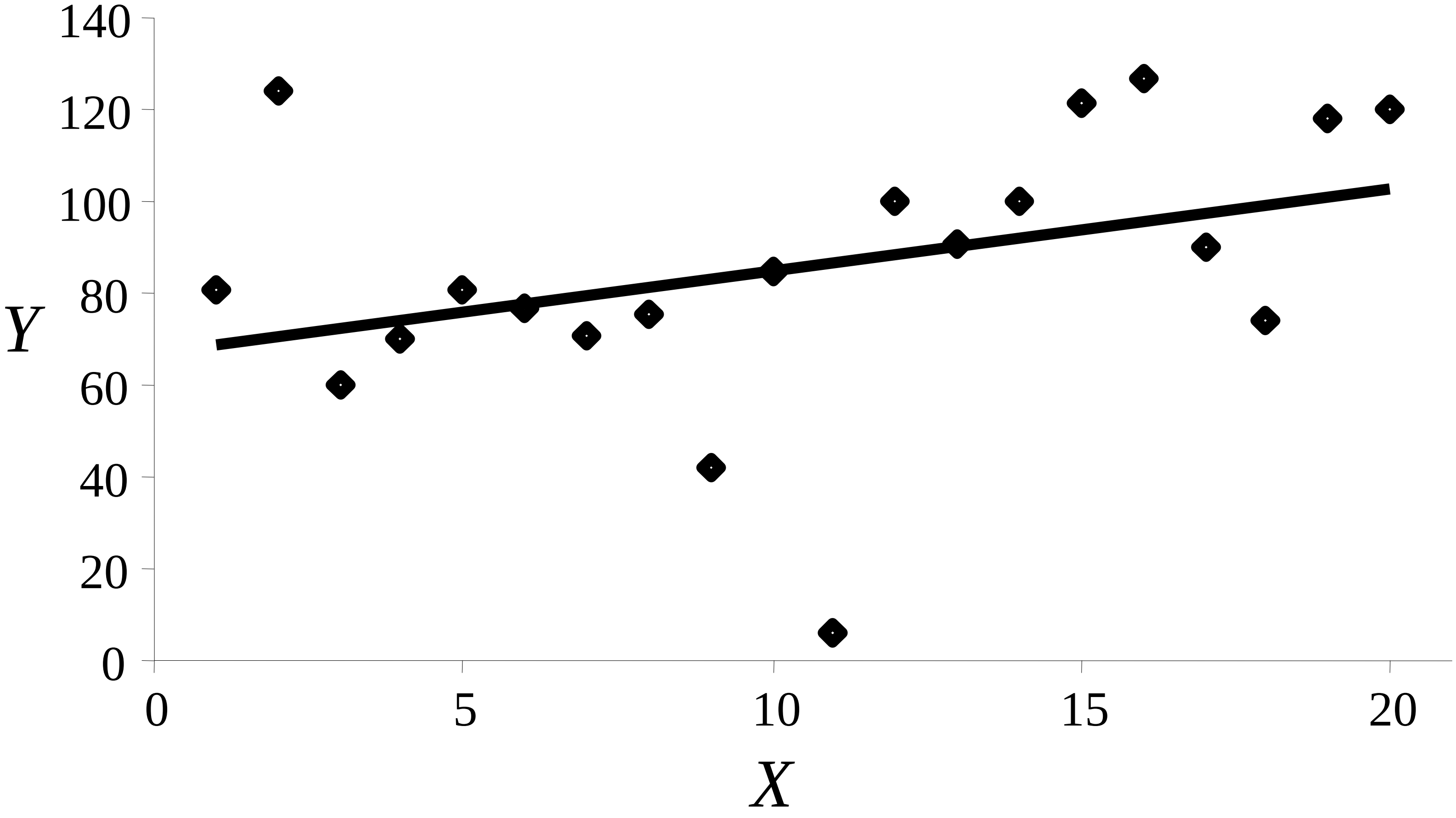
Regressions in risk analysis

- You need variable *Y* from variable *X*, but *X* is a random variable.
- You have a paper from the literature that gives a regression of *Y* on *X*.
- What do you do?

Here’s one way

- Realize *X*, apply regression to get *Y*=*a*+*bX*
- The distribution of *Y* is then a linear scaling of the distribution of *X*
- But this ignores any uncertainty about the regression itself

The line neglects the scatter



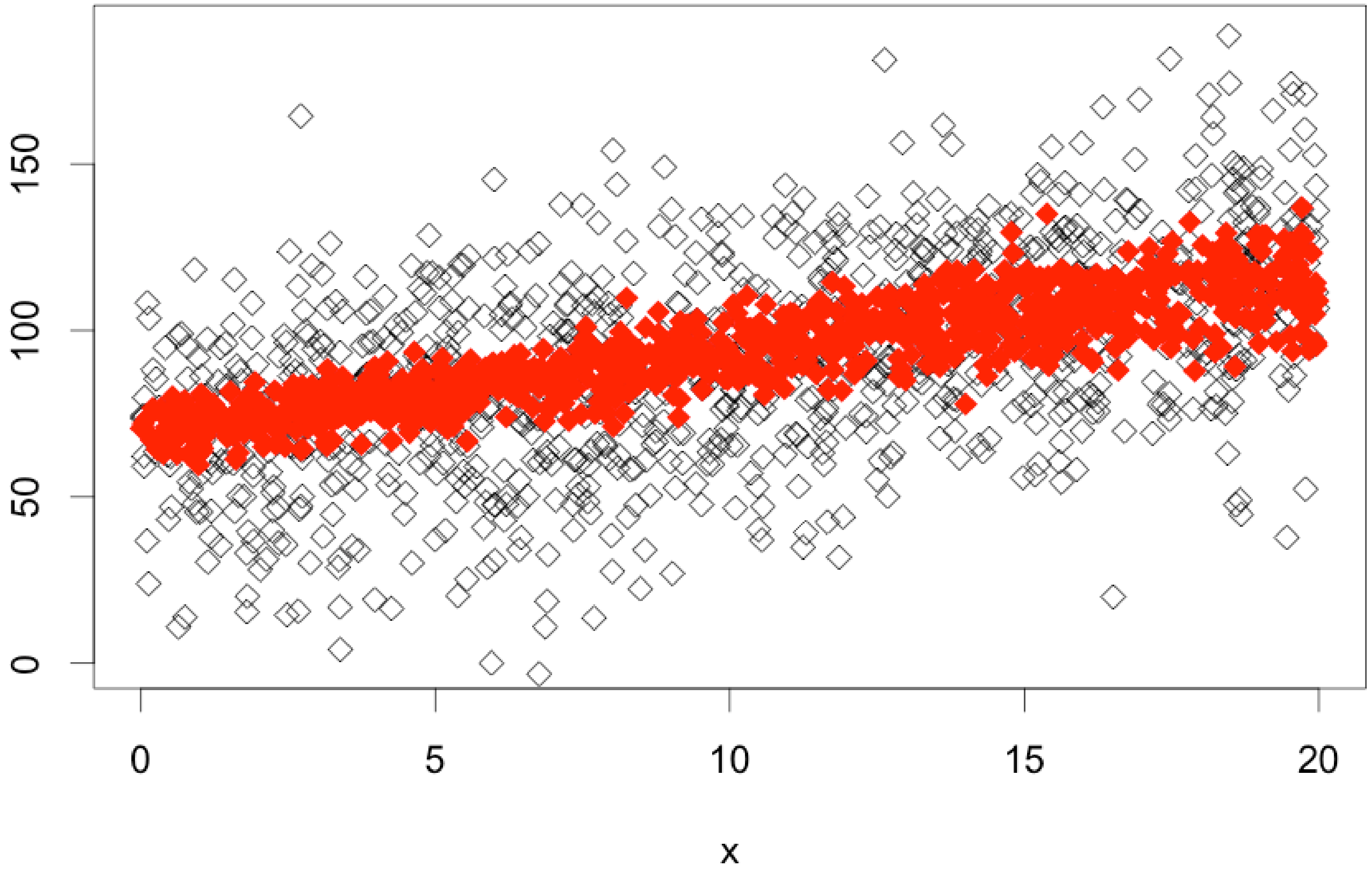
Here’s another way

- Realize *a*, *b* and *X* from their distributions (*a*,*b* ~ normal), then get *Y*=*a*+*bX*

- The distribution of *Y* is then a convolution of the distributions for *a*, *b*, and *X*

- But this still understates the uncertainty about the regression

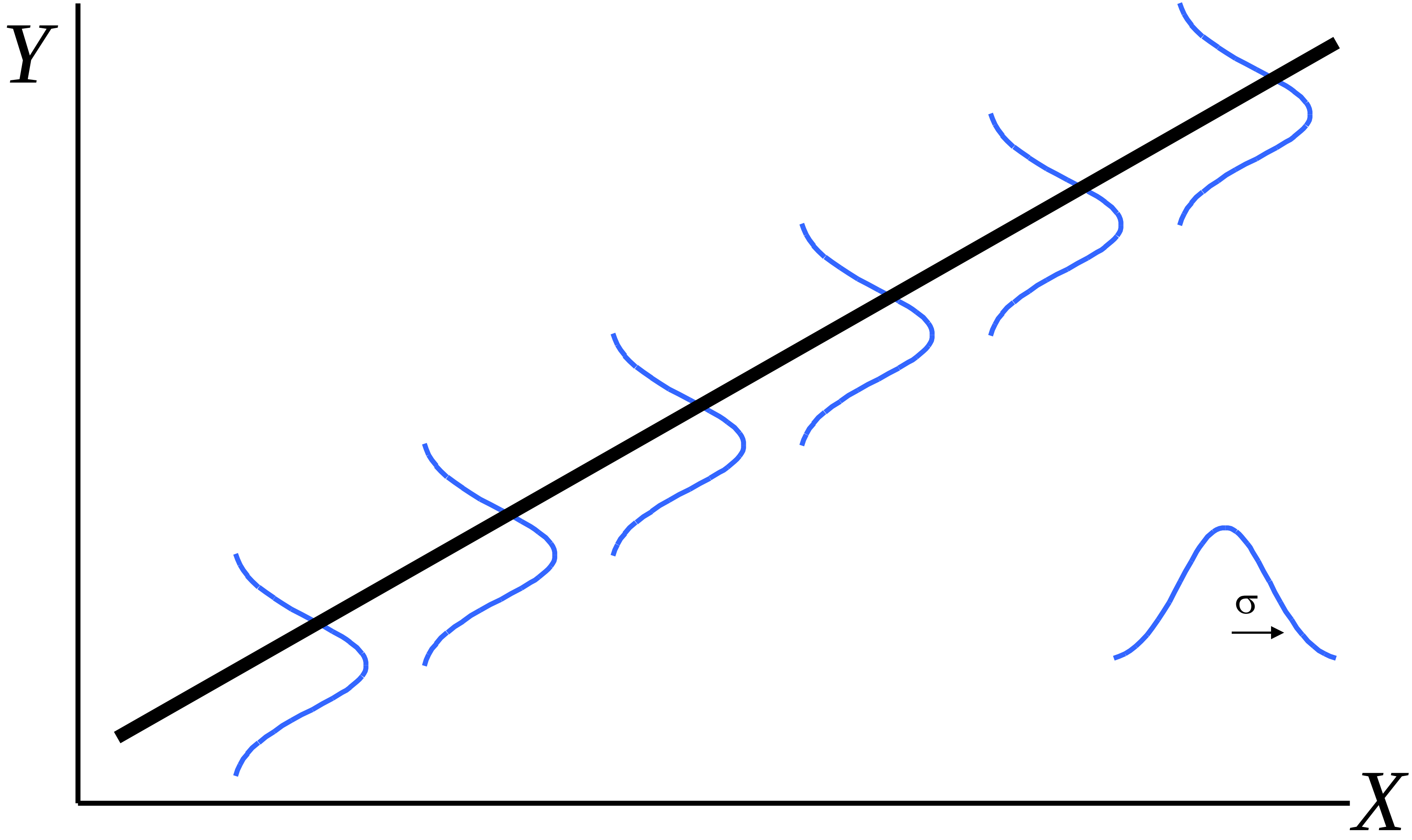
Standard errors for *a* and *b*



Assumptions of regression

- No error in the X value
- Linear in the mean, $E(Y(X)) = \alpha + \beta X$
- Y_i are independent and normal for any X
- Homoscedastic

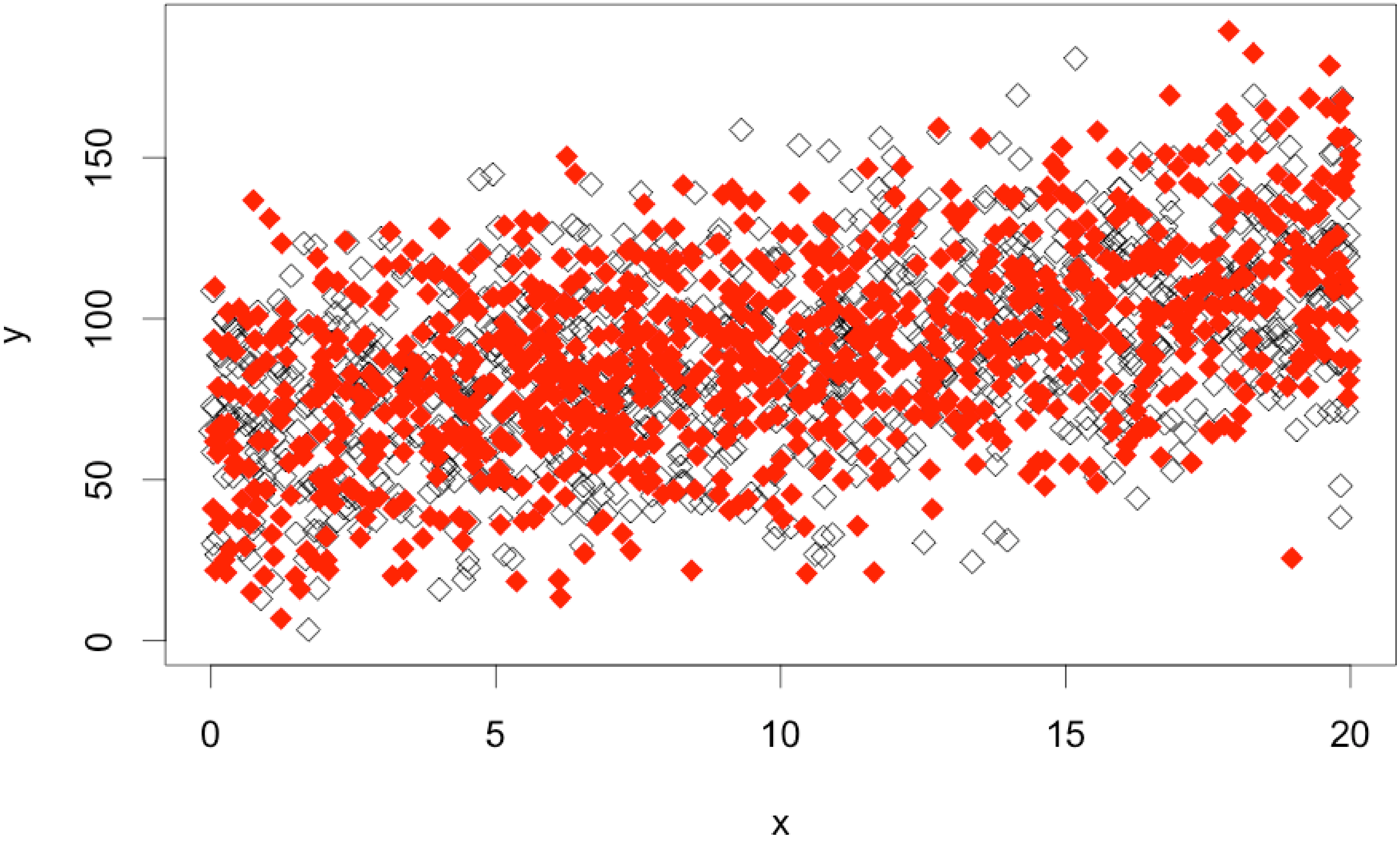
Linear regression



A third way

- Realize X
- Realize $\varepsilon \sim \text{normal}(0, \sigma)$
- Get $Y = a + bX + \varepsilon$
- This just follows the regression model

Reconstructs the scatter



The three ways

✗ 1) $Y = a + b\,X$

✗ 2) $Y = \mathrm{N}(a, \sigma_a) + \mathrm{N}(b, \sigma_b)\,X$

3) $Y = a + b\,X + \mathrm{N}(0, \sigma)$

Recovering the σ

- Raw data
- Six quantities (n, $\sum X$, $\sum X^2$, $\sum Y$, $\sum Y^2$, $\sum XY$)
- Standard error of the regression
- Mean sum of squares unexplained
- Sum of squares unexplained, sample size
- Mean sum of squares explained, F value
- Mean sum of squares explained, P value
- Regression parameters a and b and their standard errors, sample size
- R^2, a, b, sample size

σ *rules*

- The ‘re-add error’ approach works for linear multivariate and polynomial regressions too

- The dependent variable is estimated as a deterministic function of the independent variable(s) plus an independent error term

Model uncertainty about the regression

- We used a linear regression
- What if it is a quadratic, or cubic, or higher?
- Regression analyses does not reveal the true order, even in careful step-wise studies
- Could our inferences be wrong if we guess the wrong order of the function?

What is good practice?

- Analysts generally don’t know the correct order of function that relates *Y* and *X*
- They cannot determine it from data
- Can we account for uncertainty about the form of the regression equation?
- How can we ensure our results are conservative against this uncertainty?

Numerical experiments

- We assume *Y* really is a polynomial function $Y = a + bX + cX^2 + dX^3 + \ldots +\mathrm{N}(0, \sigma)$

- We know the real values *a*, *b*, *c*, *d*,…, σ

- *X* has some distribution, or p-box

- So we can compute the actual distribution of *Y*

Sample data

- We draw *n* random samples from this hypothetical relationship (with error)

- We fit *n*+1 regression analyses
  - Zeroth order is the average of *Y*
  - First order is a linear regression
  - Second, quadratic…
  - *N*th order analysis goes through every point

Brown carpet

- Models of *all* orders yield conservative values for the variance of *Y*

- Models of *all* orders give (reasonably) conservative characterizations of the tail risks of *Y*

- The *envelope* of results from all orders yields a conservative characterization of *Y* (i.e., the p-box encloses the true *Y*)

Conclusions

Model uncertainty

- Commonly significant, sometimes huge
- Rarely explored or even discussed
- Very rarely systematically addressed

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

Special problems

1. Enumerable models
2. Parameterized family of models
3. Distribution shape
4. Unknown dependence
5. Constrained family of models
6. Regression analysis
7. Surrogacy (knowing *X* but needing *Y*)
8. Non-random sampling

Meta-conclusions

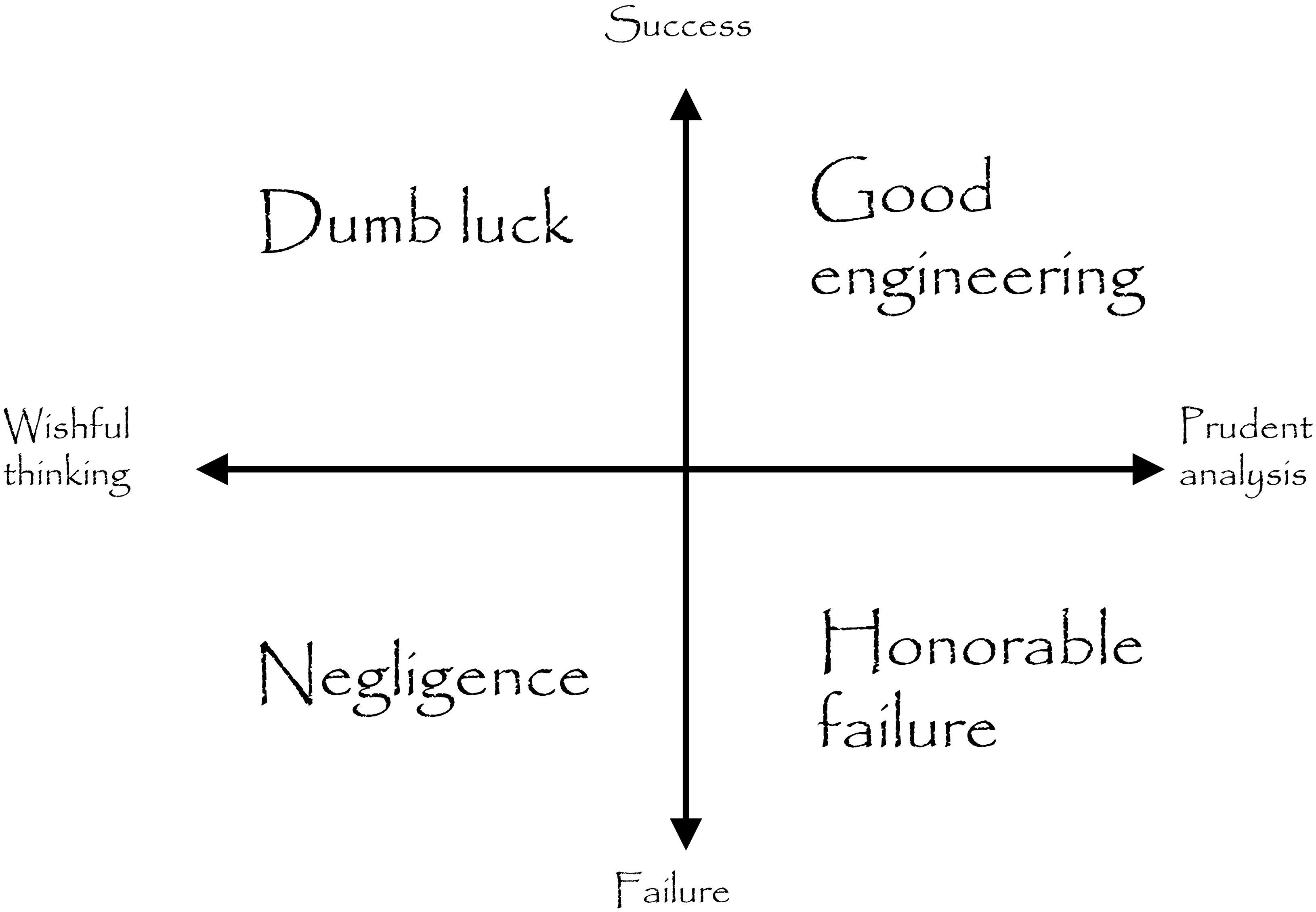
- Because model uncertainty is usually epistemic, enveloping seems to be the best approach, especially when data are sparse

- When the inference is clear despite the enveloping, it gives strong assurance for the conclusion

Acknowledgments

Thanks to Bill Oberkampf, Tony Cox and Chris Frey, Lev Ginzburg

End



Success

Dumb luck

Good engineering

Wishful thinking

Prudent analysis

Negligence

Honorable failure

Failure

One more

- Even if model uncertainty is so big that it swamps everything, you may still assess robustness of different designs

Challenge problems

- In each of the following 5 problems, we want to compute what can be inferred about the distribution $f(A,B)$, where A and B are random numbers but f is imperfectly specified.
- Display your answer graphical if possible
- Specify any assumptions you must make

1. Enumerable models

The function f is one of two possibilities. Either

$$f(A,B) = f_1(A,B) = \mathrm{A} + \mathrm{B}$$

or

$$f(A,B) = f_2(A,B) = \mathrm{A} \times \mathrm{B}$$

is the correct model, but the analyst does not know which. One and only one is correct. Suppose there is one sample value for $f(A,B) = 7.59$, and that f_1 is twice as likely as f_2. Suppose that the random variables A ~ triangular(–2.6, 0, 2.6) and B ~ triangular(2.4, 5, 7.6).

2. Parameterized family of models

The true model is one of a family of models parameterized by the real quantity $\alpha \in [0,1]$,

$$f(A,B) = \alpha(A + B) + (1 - \alpha)(A \times B).$$

The analyst feels confident that α has a fixed value between zero and one, but is not sure what it is. Suppose A ~ triangular(–2.6, 0, 2.6) and B ~ triangular(2.4, 5, 7.6).

3. Distribution shape

The correct model is known to be $f(A,B) = A+B$, but the distributions for A and B are not precisely known. A ~ triangular([–2.6,–5.2], 0, [2.6,5.2]), and B ~ minmaxmeanvar(0, 12, 5, 1).

4. Unknown dependence

The correct model is known to be $f(A,B) = A + B$, where A ~ triangular(–2.6, 0, 2.6) and B ~ triangular(2.4, 5, 7.6), but the dependence between A and B is not known.

5. Constrained family of models

Suppose again that A ~ triangular(–2.6, 0, 2.6) and B ~ triangular(2.4, 5, 7.6). The function $f(A,B)$ is known to be non-decreasing, and $f(A,B)$ cannot be smaller than –10 or larger than +10. Suppose the probability that $f(A,B) < 0$ is between 0.25 and 0.5, and the function f is quadratic.

Not even sure about the structure

Profound uncertainty

- Not sure what level of abstraction to address
- Not sure about what inputs should be present
- Not even sure about outputs
  - Health
  - Ecosystem ‘health’

Many partial strategies

- Stone soup: you already have ideas
- Going for correct, be happy with productive

- Avoid regression analyses as exploratory tools
- Study variables should be selected at random Cattell

- Dimensional reasoning
- Thinking about causality
- Non-metric cluster analysis
- Data mining

Non-metric cluster analysis

Matthews' Riffle program

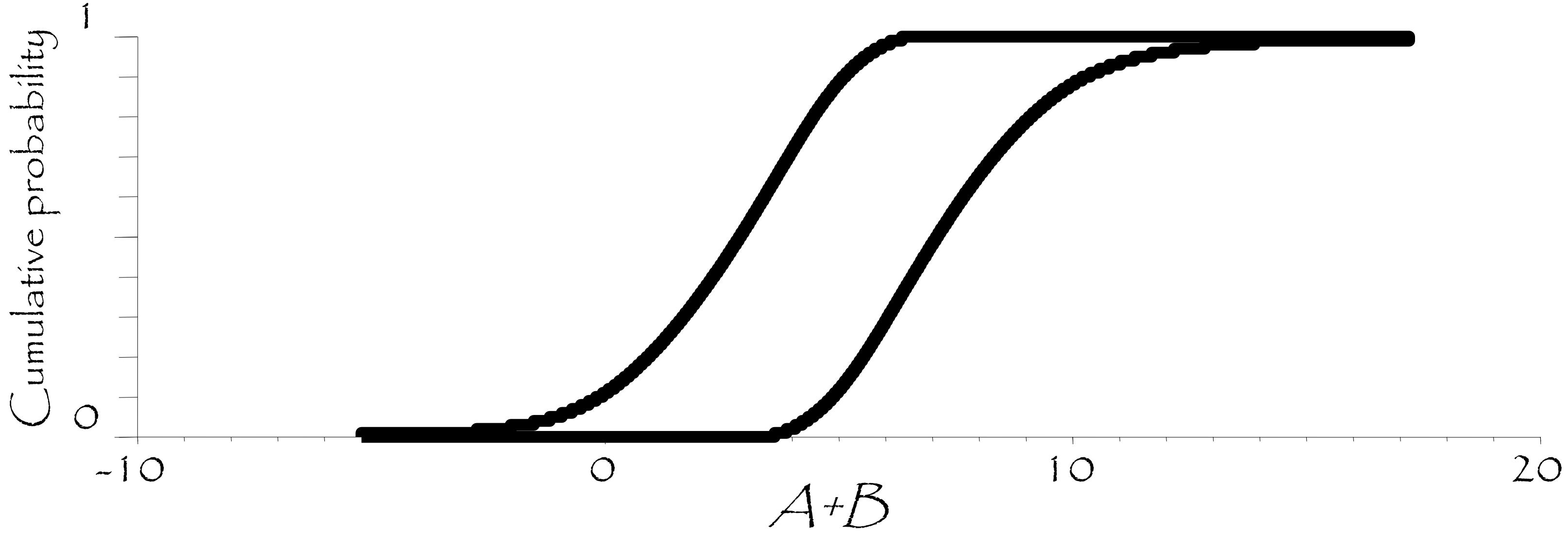
- Doesn't use distance metrics to define similarity
- Nonparametric (uses ranks and medians)
- Each variable is examined independently
- Data are searched iteratively to find good clusters that consist of samples with many similar features
  - Ignores noisy variables to identify extant patterns
  - Often better than KMeans or hierarchical clustering

Data mining

- Nontrivial extraction of implicit, previously unknown, and potentially useful information from data
    - Anecdote
    - Observations
    - Data
    - Information
    - Knowledge
    - Understanding

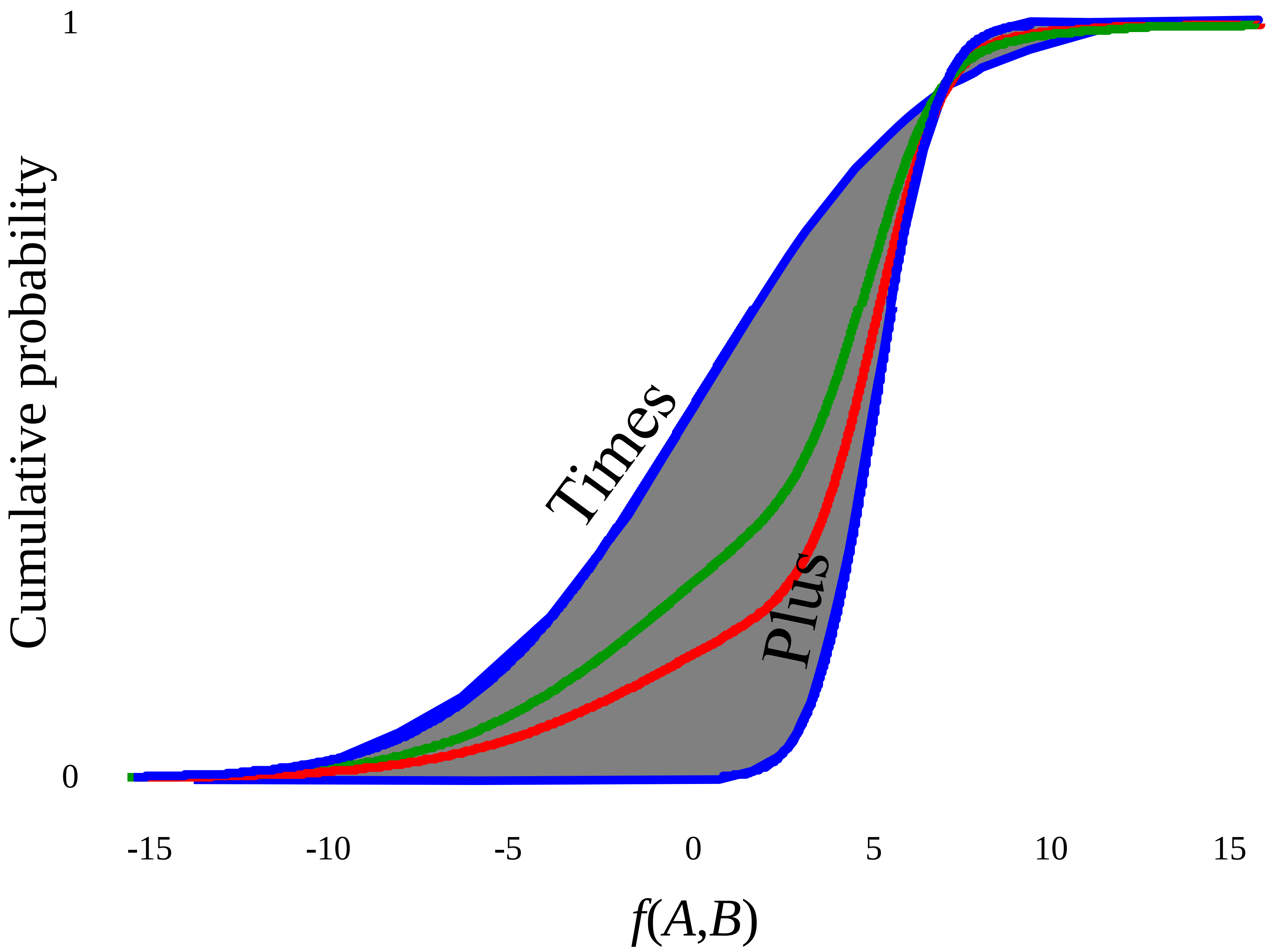
    –Repetition
    –Structure
    –Organization
    –Context
    –Implication
- Exploratory data analysis
- Unsupervised learning and feature extraction
- Overfitting (‘data dredging’)
    - For example, if we examine correlations among enough variables, some are bound to seem interesting

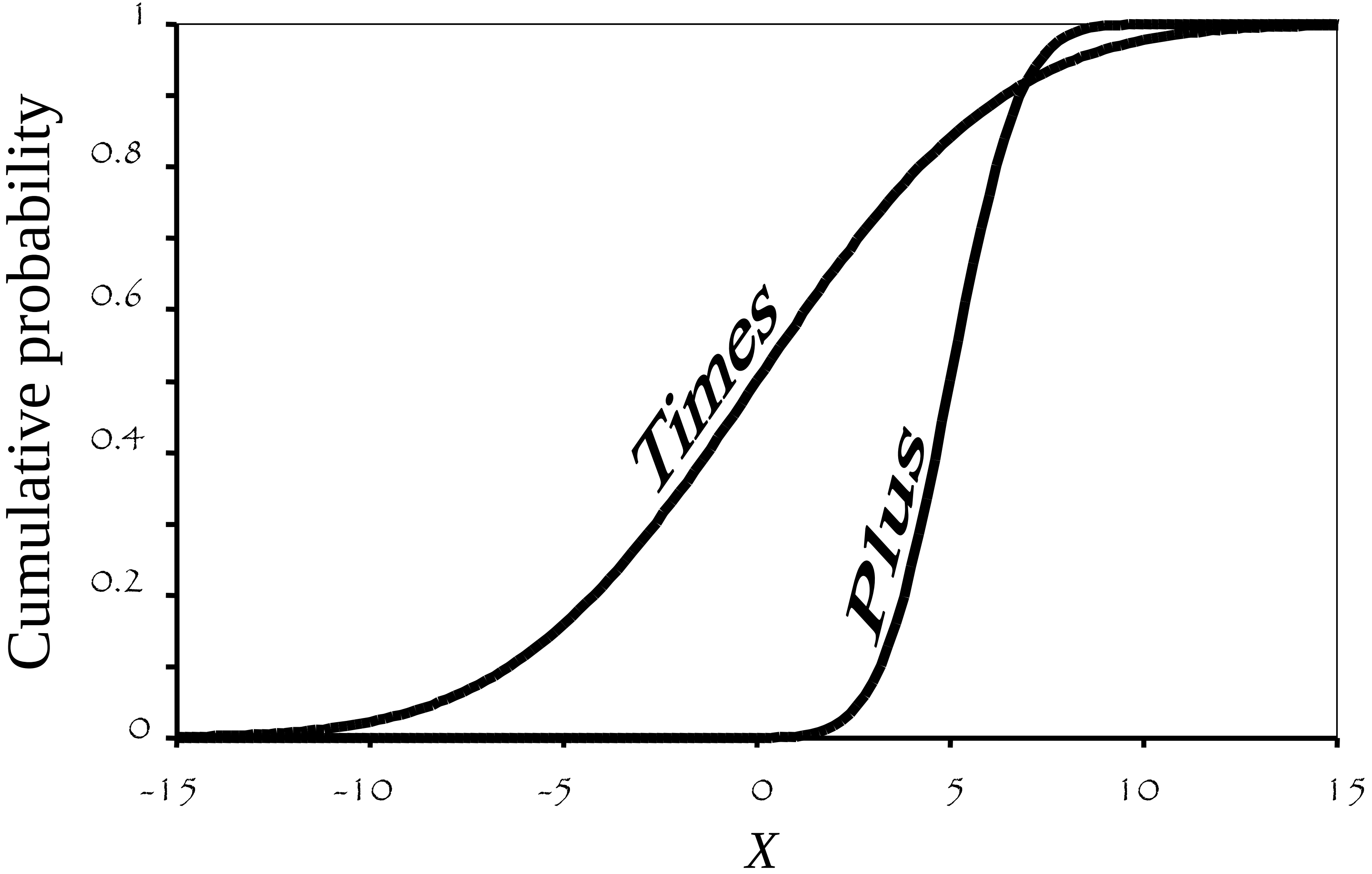
Sum under independence

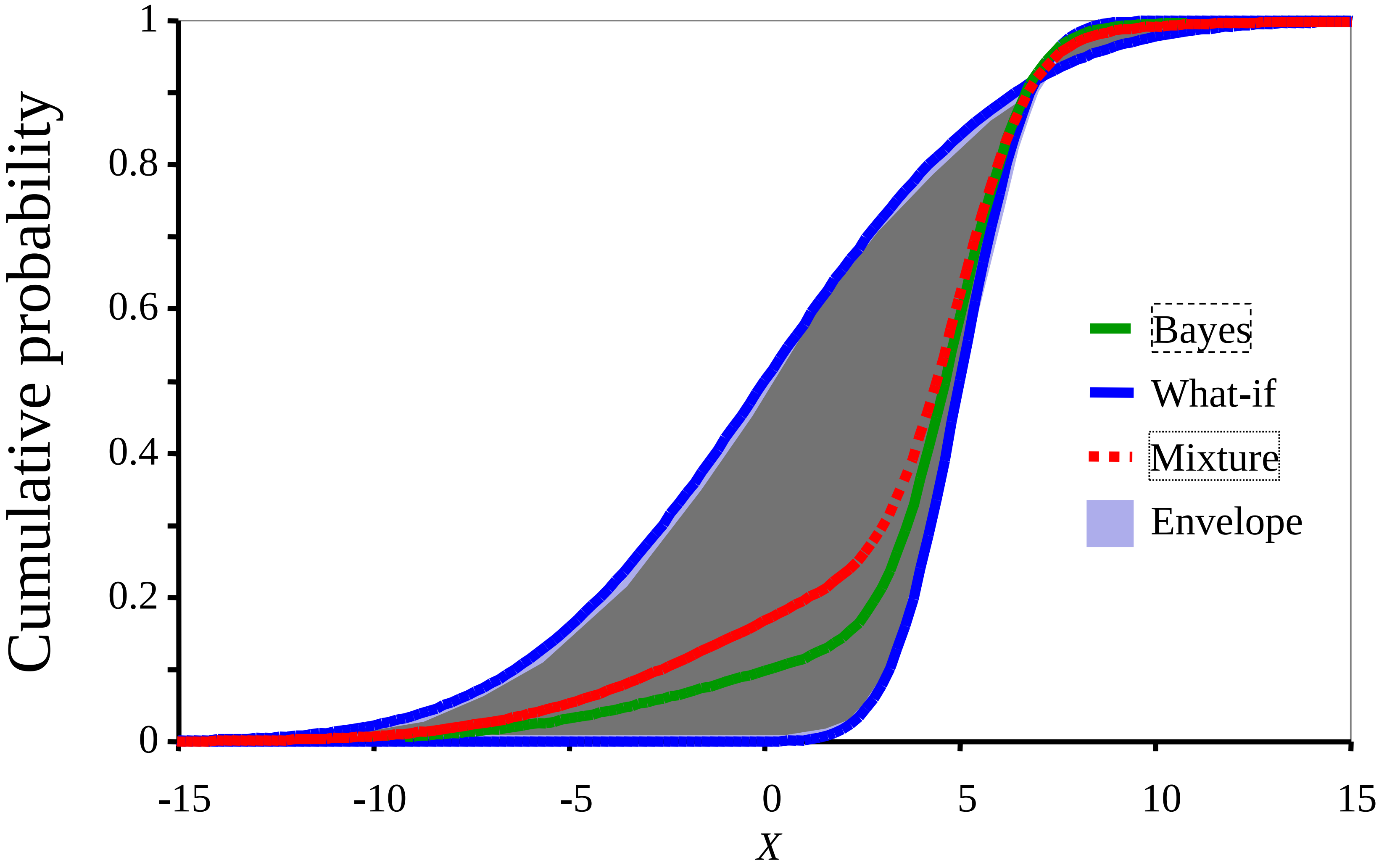


These bounds are rigorous (guaranteed) and often best-possible (as narrow as can be justified given what is known).

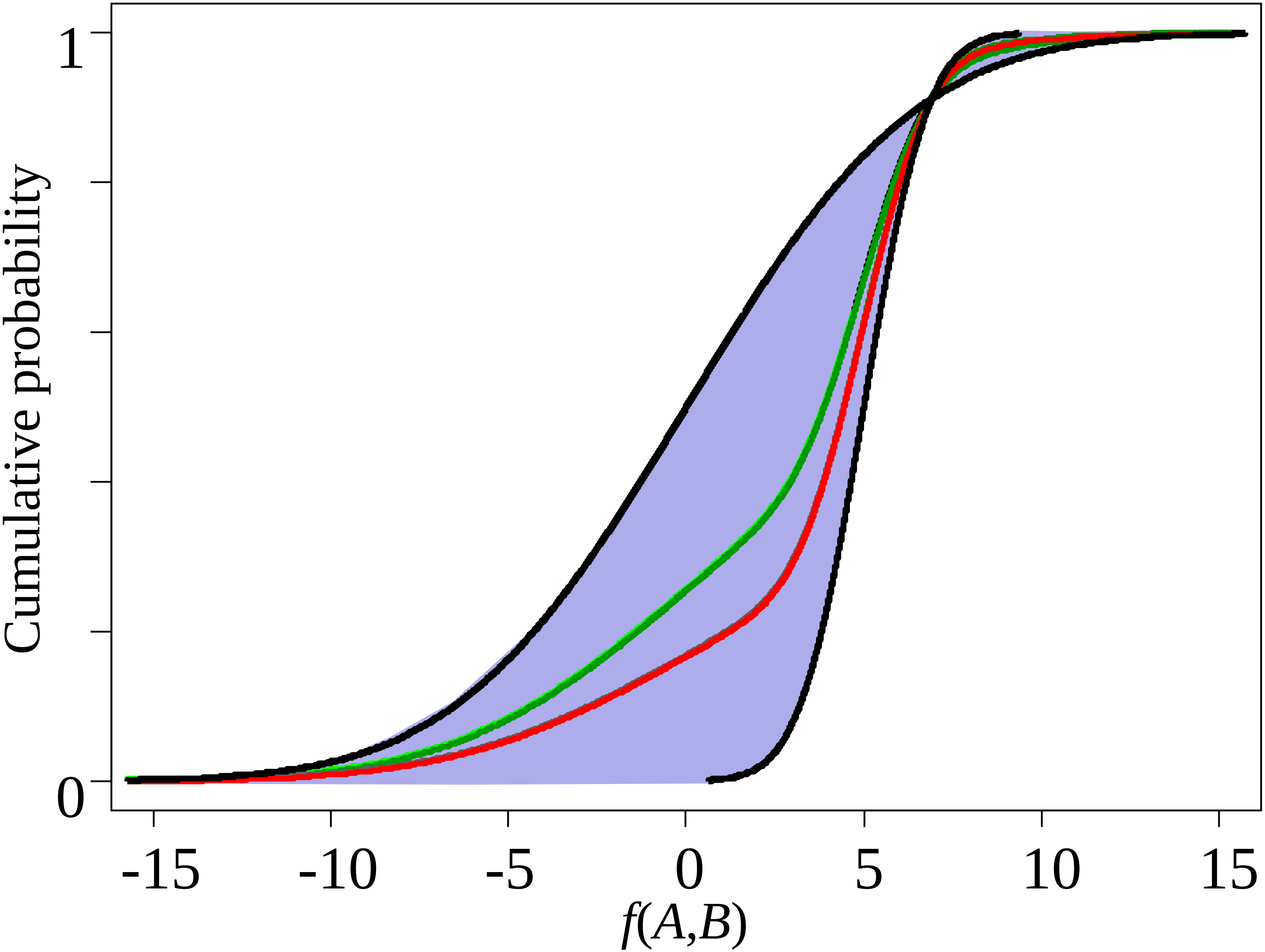
```
# Reconstructing the scatter as PREDICTIONS

n = 100    # harder to see if n is small
N = 1000
reps = N/n

# true relationship
x = runif(N,0,20)
y = 65 + x * (50/20) + rnorm(length(x),0,26)
par(cex=1.45)
plot(x,y,pch=5,xlim=c(-10,30),ylim=c(-50,250),col='white')
#lines(predict(lm(y~x), newdata=data.frame(x=1:20), interval='none'))

# a smallish sample
y = y[1:n]
x = x[1:n]

# regression
r <- lm(y~x)
s = summary(r)
#abline(r)

X = 1:400/10 - 10
p = predict(r, newdata=data.frame(x=X), se.fit = TRUE, interval='prediction')



s$sigma
s$coefficients[1,]
s$coefficients[2,]

n = length(x)
c = s$coefficients

par(cex=2)

for (i in 1:reps) {
  # reconstruct scatter
```

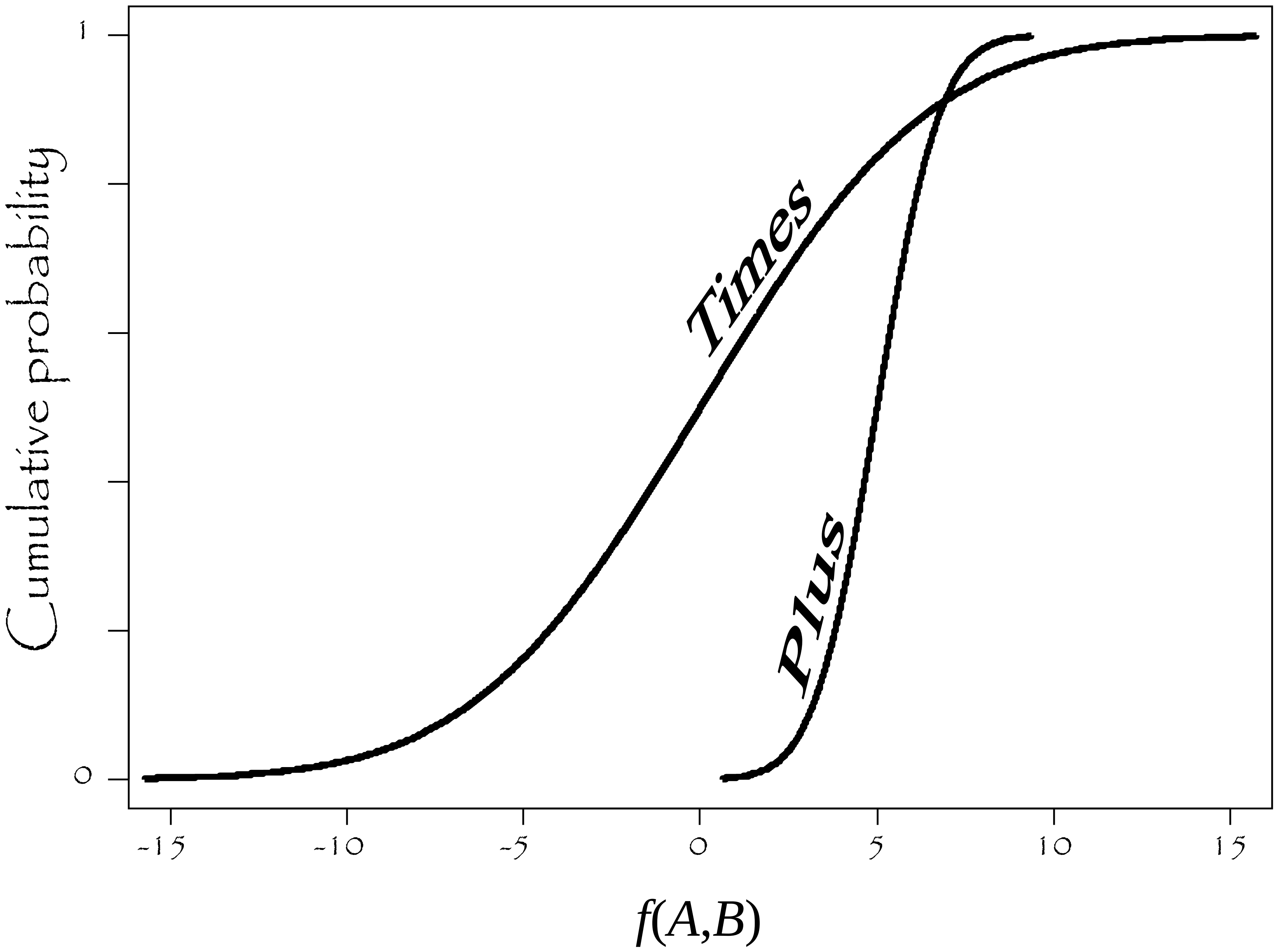

1
0
Cumulative probability
Times
Plus
-15
-10
-5
0
5
10
15
$f(A,B)$

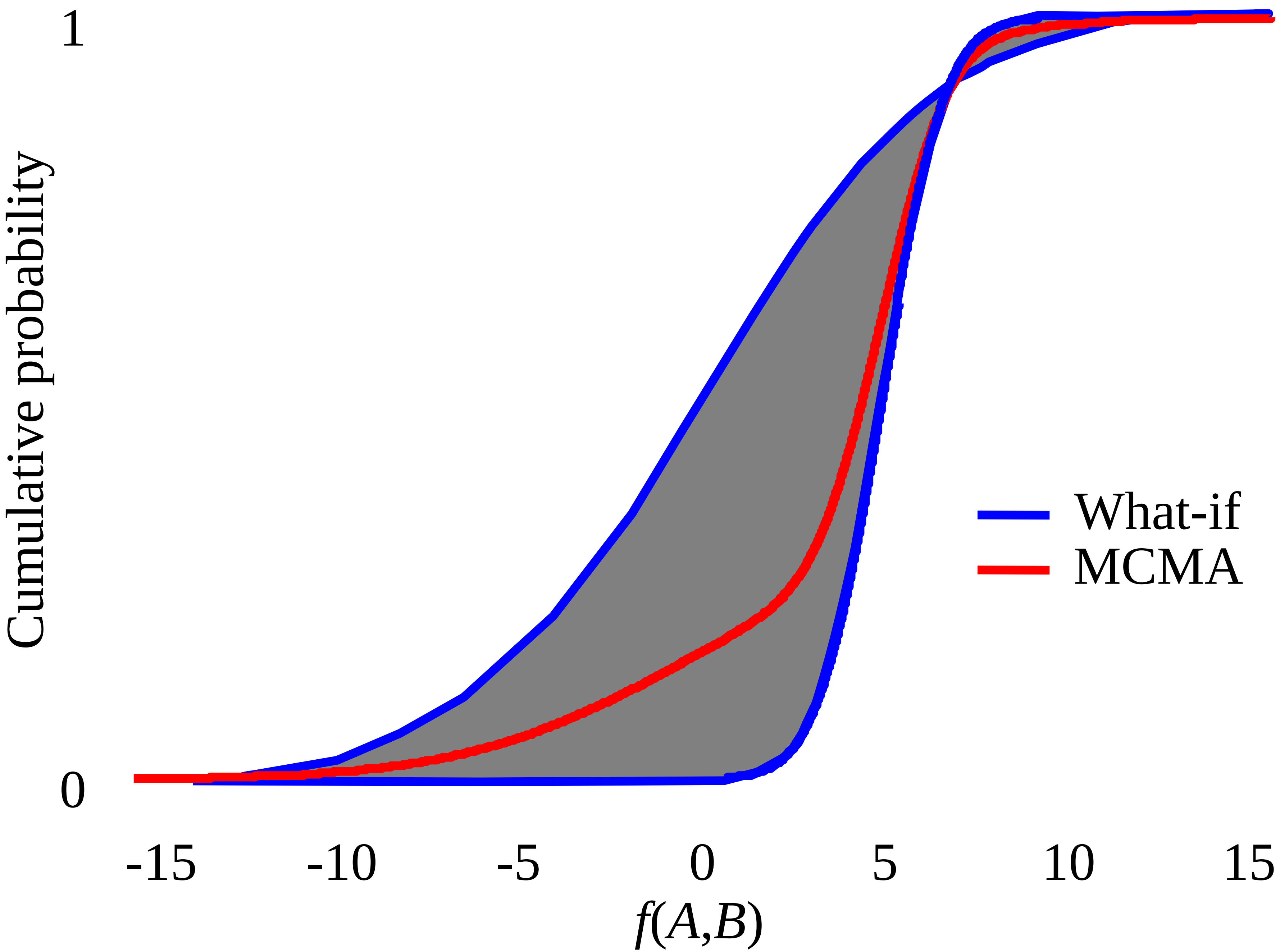

Cumulative probability
1
0.8
0.6
0.4
0.2
0
-15
-10
-5
0
5
10
15
X
Times
Plus


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

```
A = normal(0,1)
B = normal(5,1)

fplus = A + B
ftimes = A * B

fPlus = normal(5, sqrt(2))
fTimes = normal(0, sqrt(26))

par(mfrow=c(2,1))
plot(fPlus)
lines(fplus)
plot(fTimes)
lines(ftimes)

# fPlus is twice as likely as fTimes
w = 2/3

m = mix(fPlus, fTimes, w=c(w,1-w))

datum = 2.1

Lp = dnorm(datum, 5, sqrt(2))
Lt = dnorm(datum, 0, sqrt(26))
Lp
Lt

wp = w*Lp/(w*Lp+(1-w)*Lt)
wt = (1-w)*Lt/(w*Lp+(1-w)*Lt)

b = mix(fPlus, fTimes, w=c(wp,wt))

Pbox$steps = 800
par(mfrow=c(1,1))
plot(fPlus,col='black',xlim=c(-15,15))
lines(fTimes,col='black')
lines(m,col='red')
```


Cumulative probability
1
0
-15
-10
-5
0
5
10
15
f(A,B)


Cumulative probability
1
0
-15
-10
-5
0
5
10
15
f(A,B)
Times
Plus


1
0
Cumulative probability
-15
-10
-5
0
5
10
15
$f(A,B)$
What-if
MCMA