Beyond Probability

A pragmatic approach to uncertainty quantification in engineering

Scott Ferson, Applied Biomathematics, scott@ramas.com

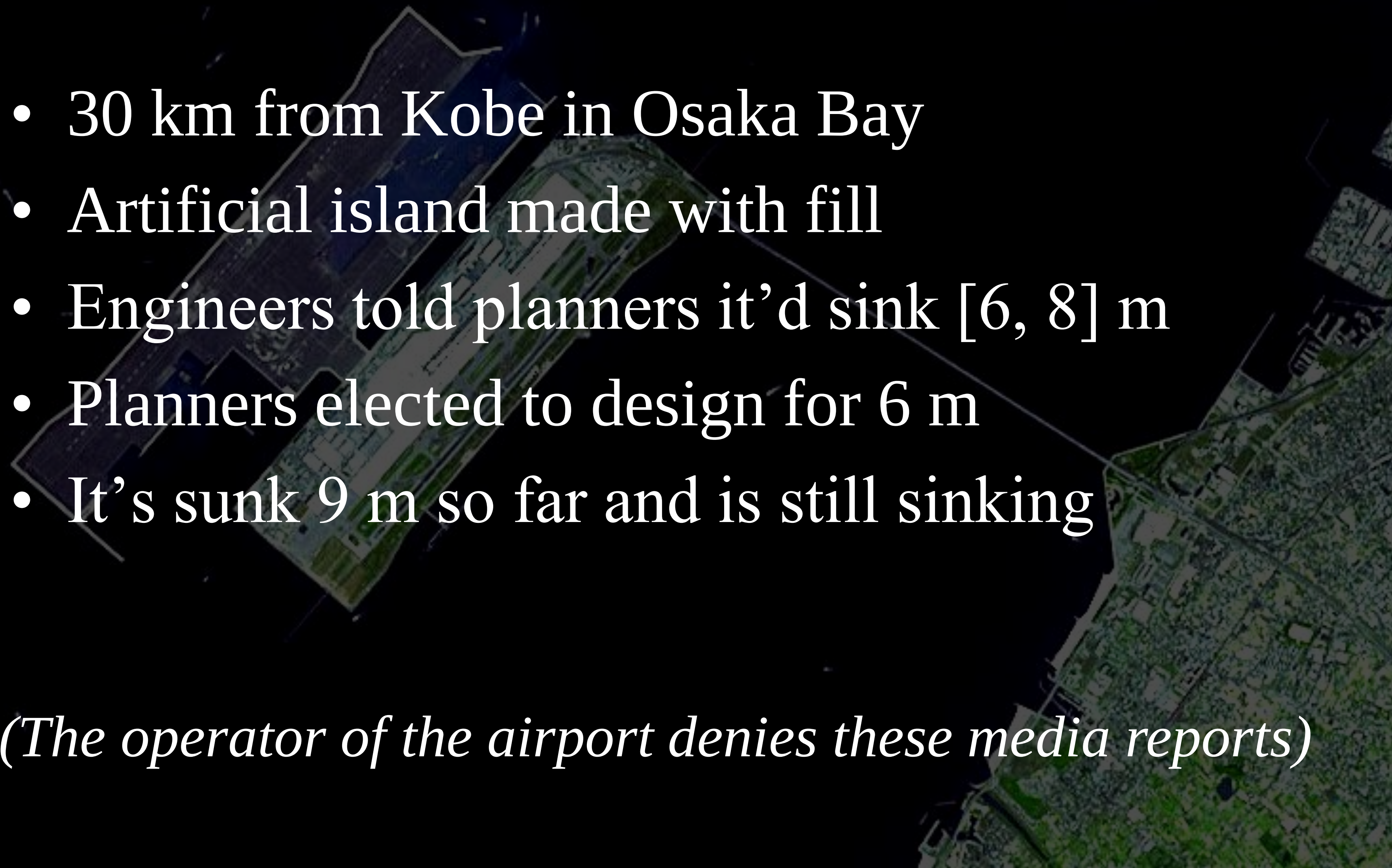
NASA Statistical Engineering Symposium, Williamsburg, Virginia, 4 May 2011

Wishful thinking

- Using inputs or models because they are convenient, or because you hope they’re true

Kansai International Airport

- 30 km from Kobe in Osaka Bay
- Artificial island made with fill
- Engineers told planners it’d sink [6, 8] m
- Planners elected to design for 6 m
- It’s sunk 9 m so far and is still sinking



(The operator of the airport denies these media reports)

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 error’)
  - use of surrogate data

- Reducible with empirical effort

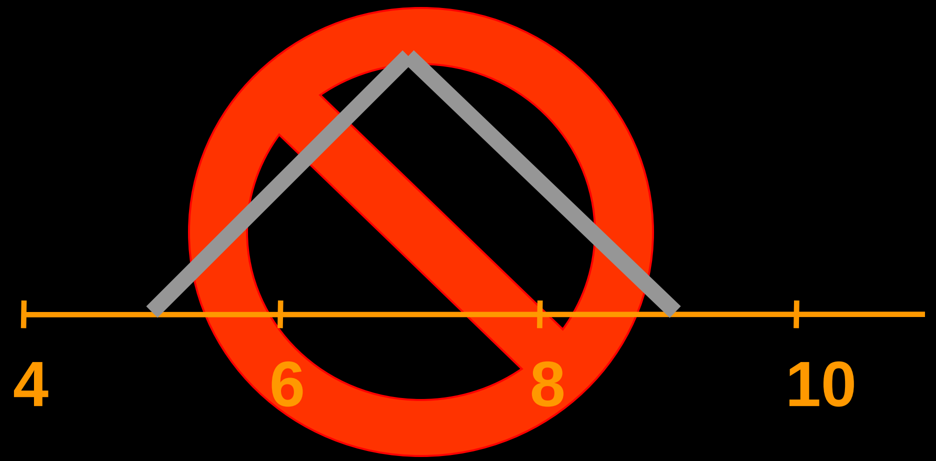
Propagating incertitude

Suppose

A is in [2, 4]

B is in [3, 5]

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



The right answer for engineering is [5,9]

They must be treated *differently*

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

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

- Imprecise probabilities can do both at once

Incertitude is common in engineering

- 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

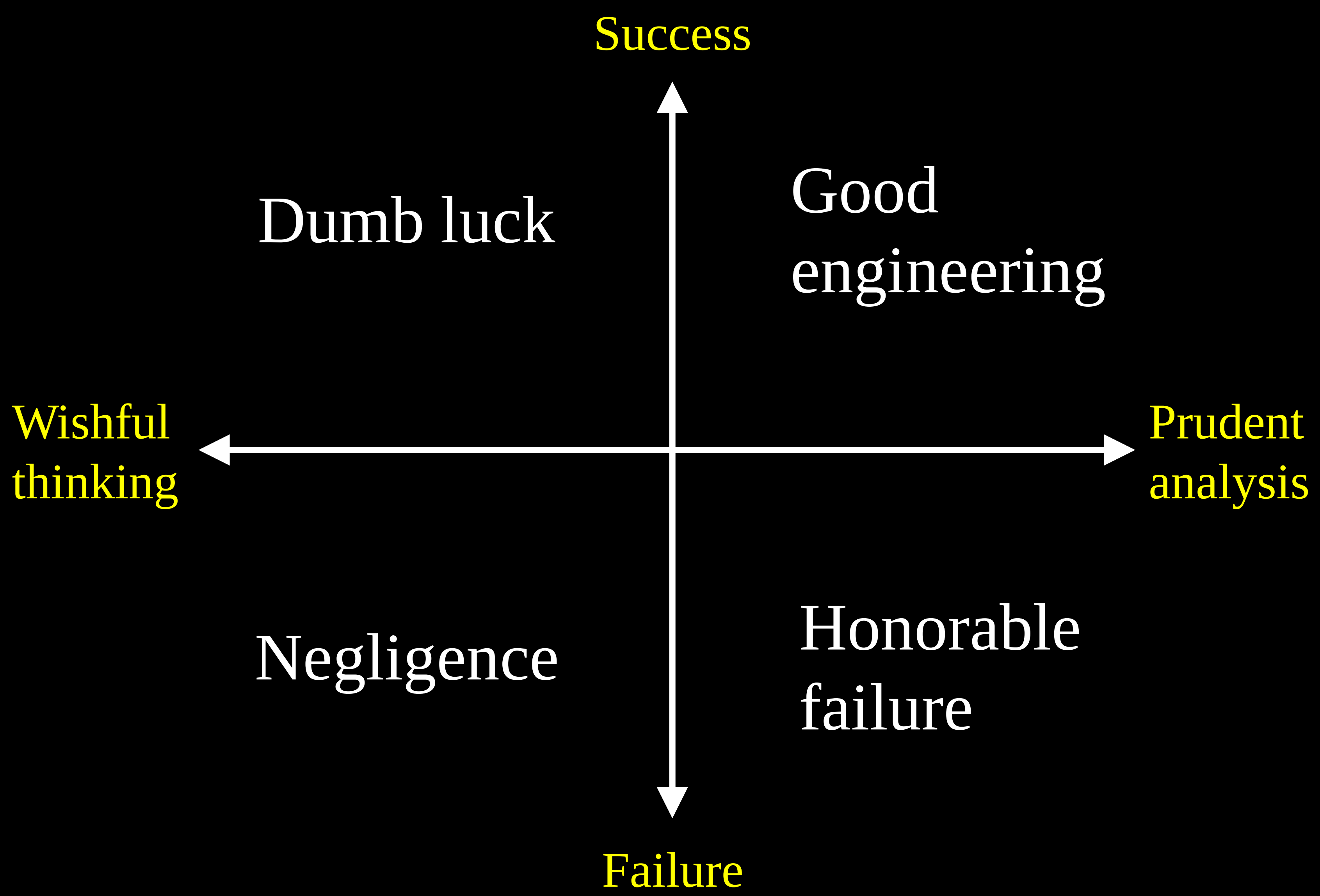
  Presumed or hypothetical limits in what-if calculations

Wishful thinking

- Pretending you know the
  - Value
  - Distribution function
  - Dependence
  - Model

  when you don't is wishful thinking

- Uncertainty analysis makes a prudent analysis


Success
Dumb luck
Good engineering
Wishful thinking
Prudent analysis
Negligence
Honorable failure
Failure

Traditional uncertainty analyses

- Interval analysis
- Taylor series approximations (delta method)
- Normal theory propagation (ISO/NIST)
- Monte Carlo simulation
- Stochastic PDEs
- Two-dimensional Monte Carlo

Untenable assumptions

- Uncertainties are small
- Distribution shapes are known
- Sources of variation are independent
- Uncertainties cancel each other out
- Linearized models good enough
- Underlying physics is known and modeled

Need ways to relax assumptions

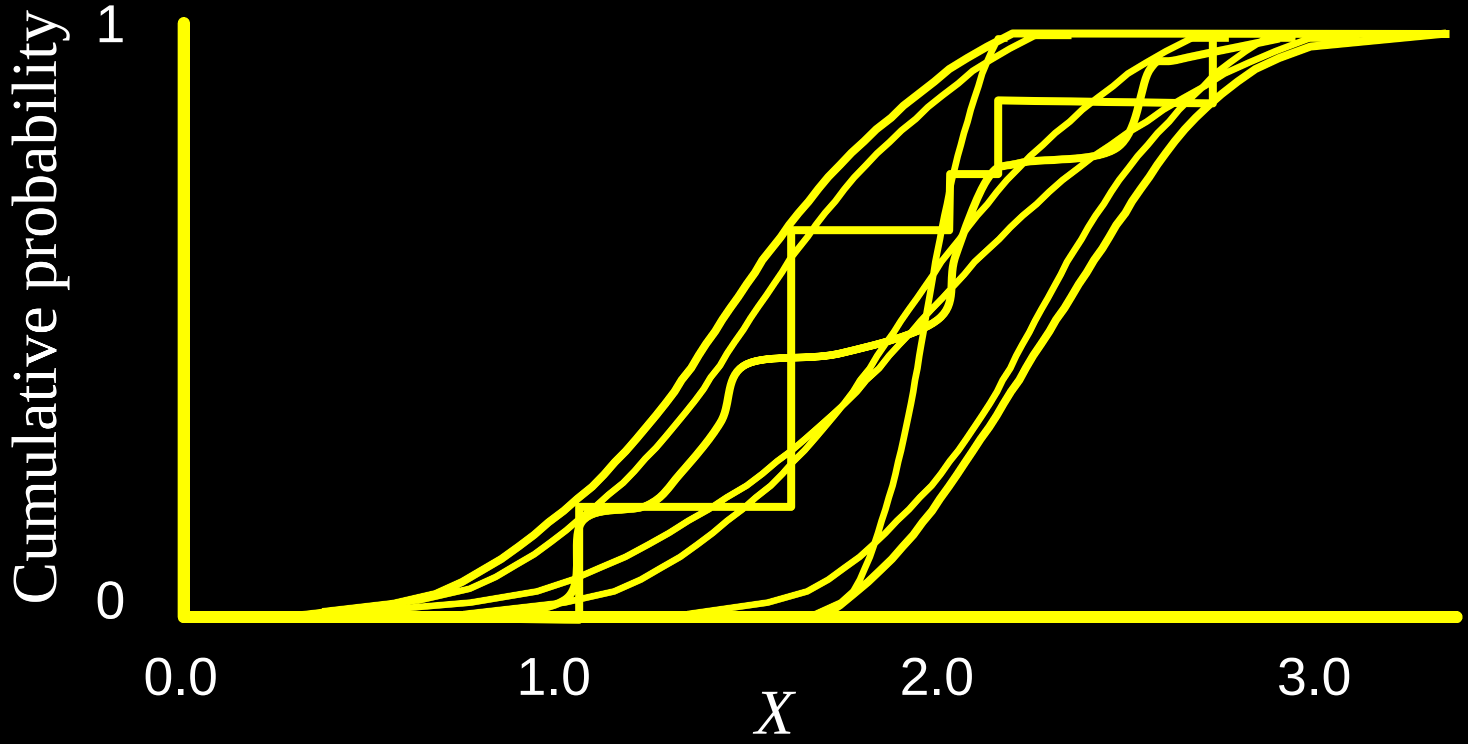
- Hard to say what the distribution is precisely
- Non-independent, or *unknown* dependencies
- Uncertainties that may not cancel
- Possibly large uncertainties
- Model uncertainty

Probability bounds analysis (PBA)

- Sidesteps the major criticisms
  - Doesn't force you to make any assumptions
  - Can use only whatever information is available
- Bridges worst case and probabilistic analysis
- Distinguishes variability and incertitude
- Acceptable to both Bayesians and frequentists

Probability box (p-box)

Interval bounds on a cumulative distribution function

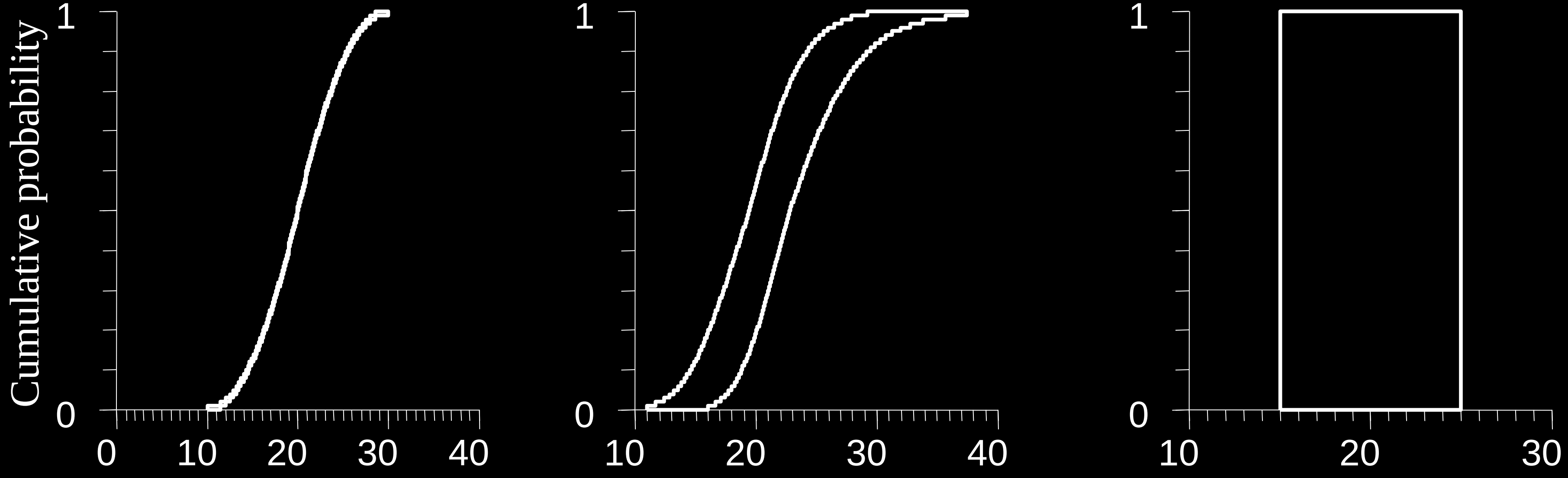


Uncertain numbers

Probability distribution

Probability box

Interval



Not a uniform distribution

Uncertainty arithmetic

- We can do math on p-boxes
- When inputs are distributions, the answers conform with probability theory
- When inputs are intervals, the results agree with interval (worst case) analysis

Calculations

- All standard mathematical operations
  - Arithmetic (+, –, , , ^, min, max)
  - Transformations (exp, ln, sin, tan, abs, sqrt, etc.)
  - Magnitude comparisons $(<, \leq, >, \geq, \subseteq)$
  - Other operations (nonlinear ODEs, finite-element methods)
- Faster than Monte Carlo
- Guaranteed to bound the answer
- Optimal solutions often easy to compute

Probability bounds analysis

- Special case of imprecise probabilities
- Addresses many problems in risk analysis
  - Input distributions unknown
  - Imperfectly known correlation and dependency
  - Large measurement error, censoring
  - Small sample sizes
  - Model uncertainty

Better than sensitivity analysis

- Unknown distribution is hard for sensitivity analysis since infinite-dimensional problem
- Analysts usually fall back on a maximum entropy approach, which erases uncertainty rather than propagates it
- Bounding seems very reasonable, so long as it reflects all available information

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

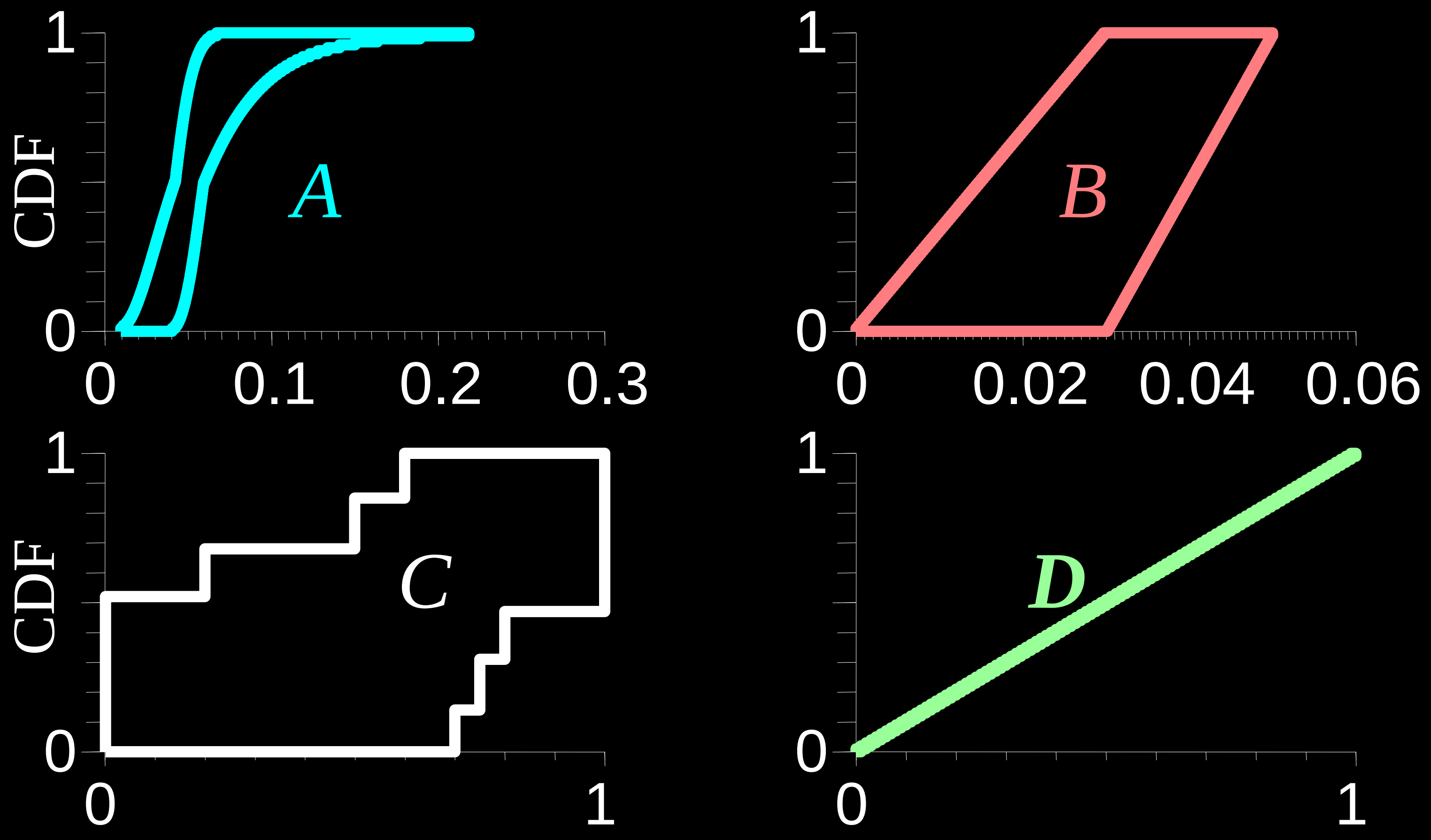
- Calculate A & B & C & D, with partial information:
  - A's distribution is known, but not its parameters
  - B's parameters known, but not its shape
  - C has a small empirical data set
  - D is known to be a precise distribution
- Bounds assuming independence?
- Without any assumption about dependence?

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

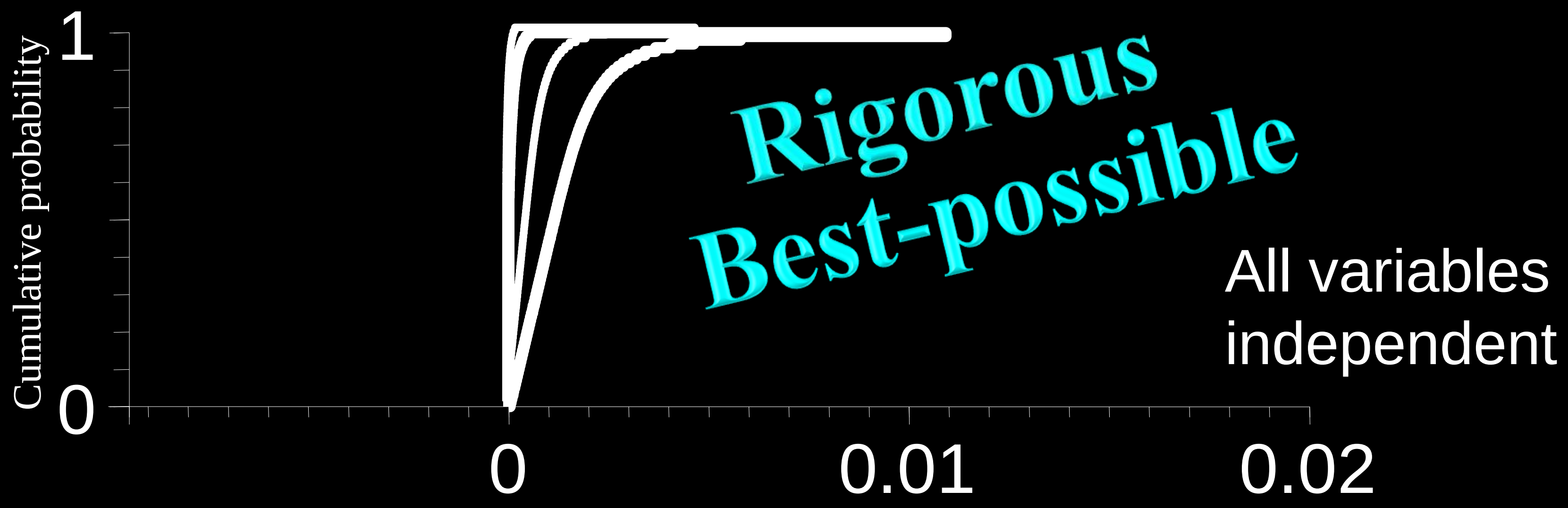
B = {min = 0, max = 0.05, mode = 0.03}

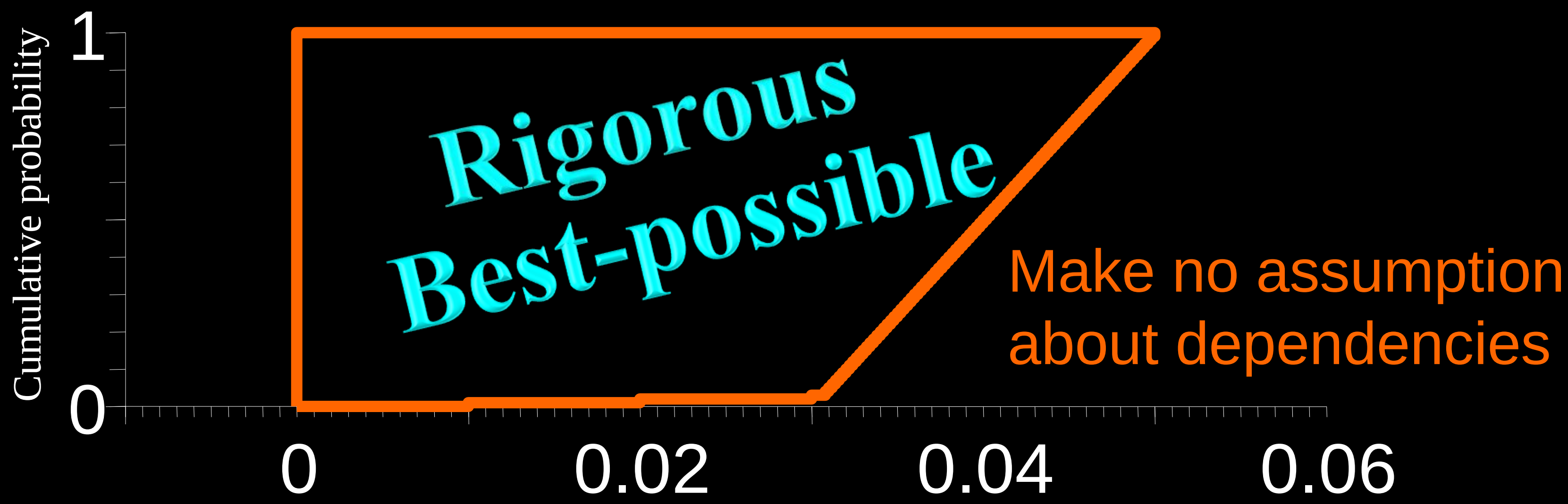
C = {sample data = 0.2, 0.5, 0.6, 0.7, 0.75, 0.8}

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 swamps the decision

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

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

Justifying further empirical effort

- If incertitude is too wide for decisions, and bounds are best possible, more data is needed

- Strong argument for collecting more data

Advantages

- Computationally efficient
  - No simulation or parallel calculations needed
- Fewer assumptions
  - Not just different assumptions, *fewer* of them
  - Distribution-free probabilistic risk analysis
- Rigorous results
  - Built-in quality assurance
  - Automatically verified calculation

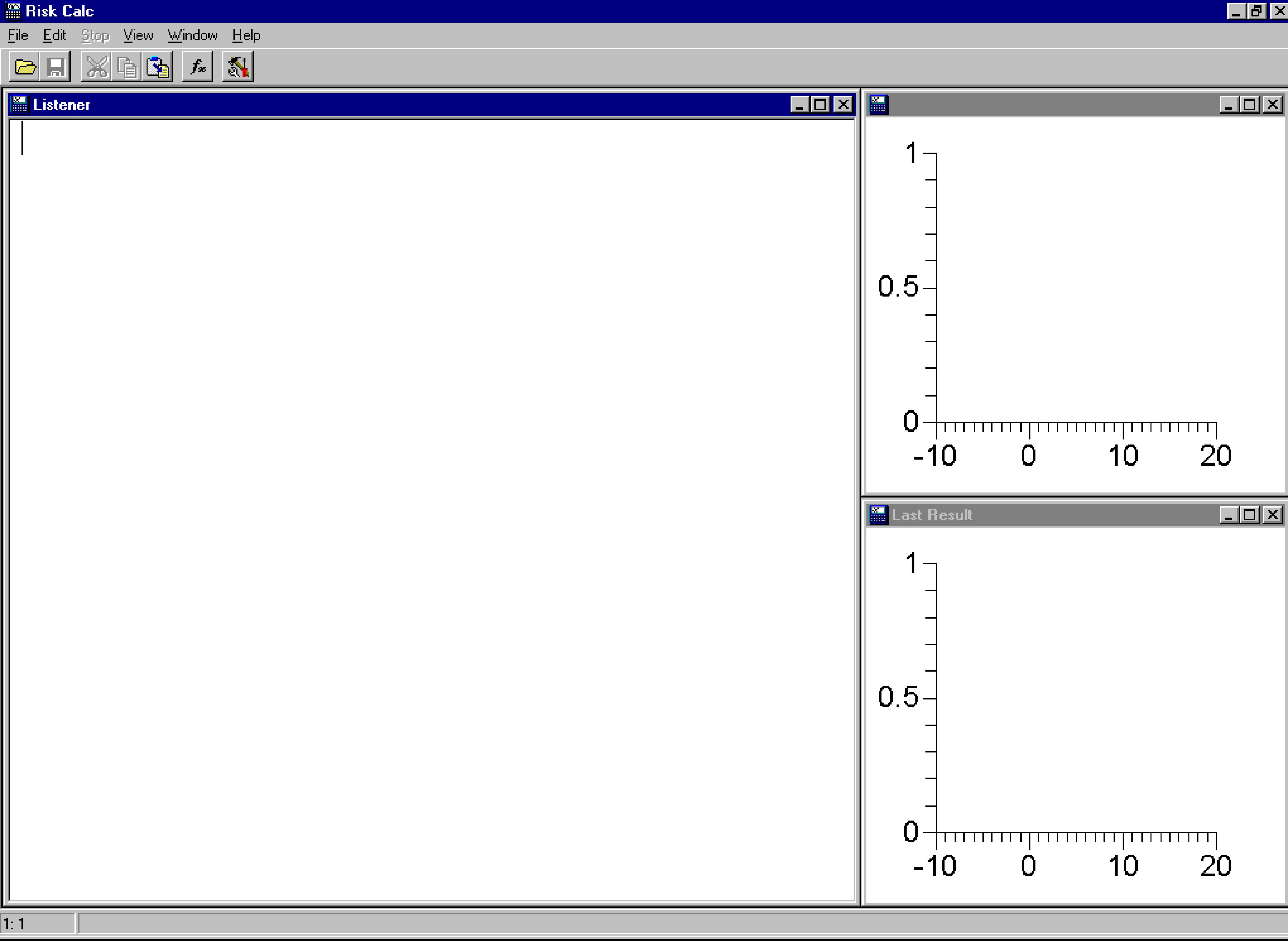
Disadvantages

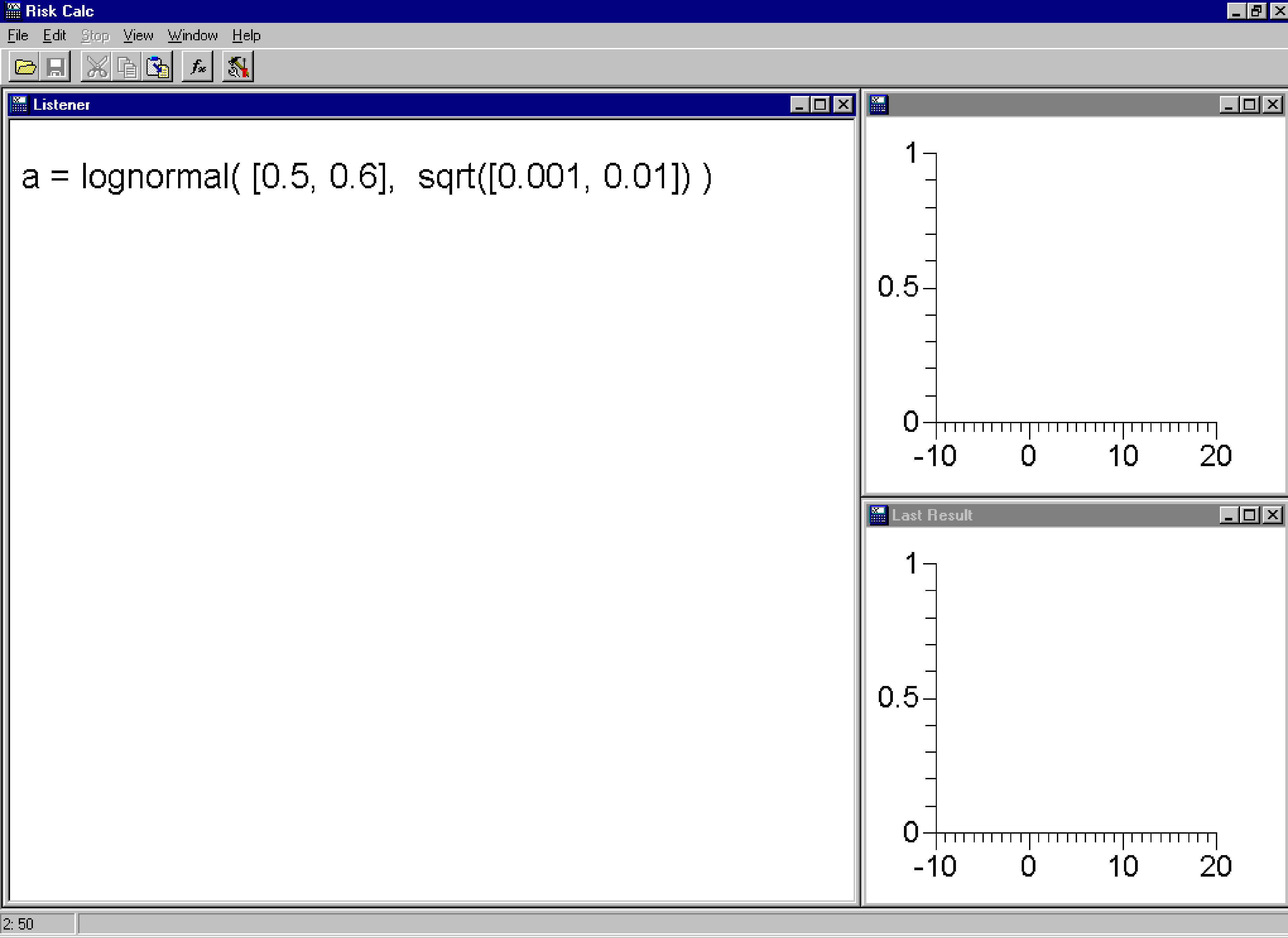
- P-boxes don’t say what outcome is most likely
- Hard to get optimal bounds on non-tail risks
- Some technical limits (e.g., sensitive to repeated variables, tricky with black boxes)
- A p-box may not express the tightest possible bounds given all available information (although it often will)

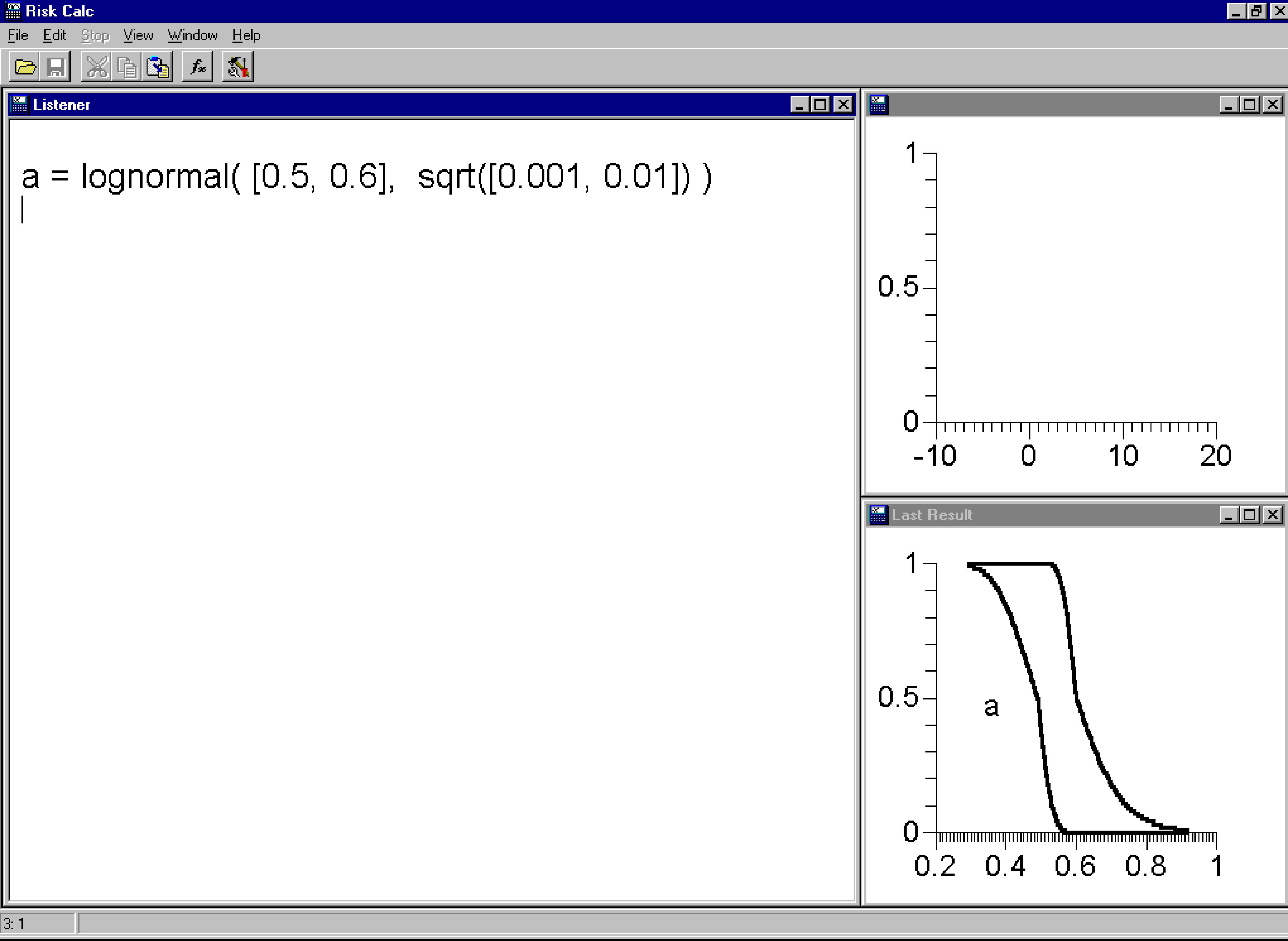
Software

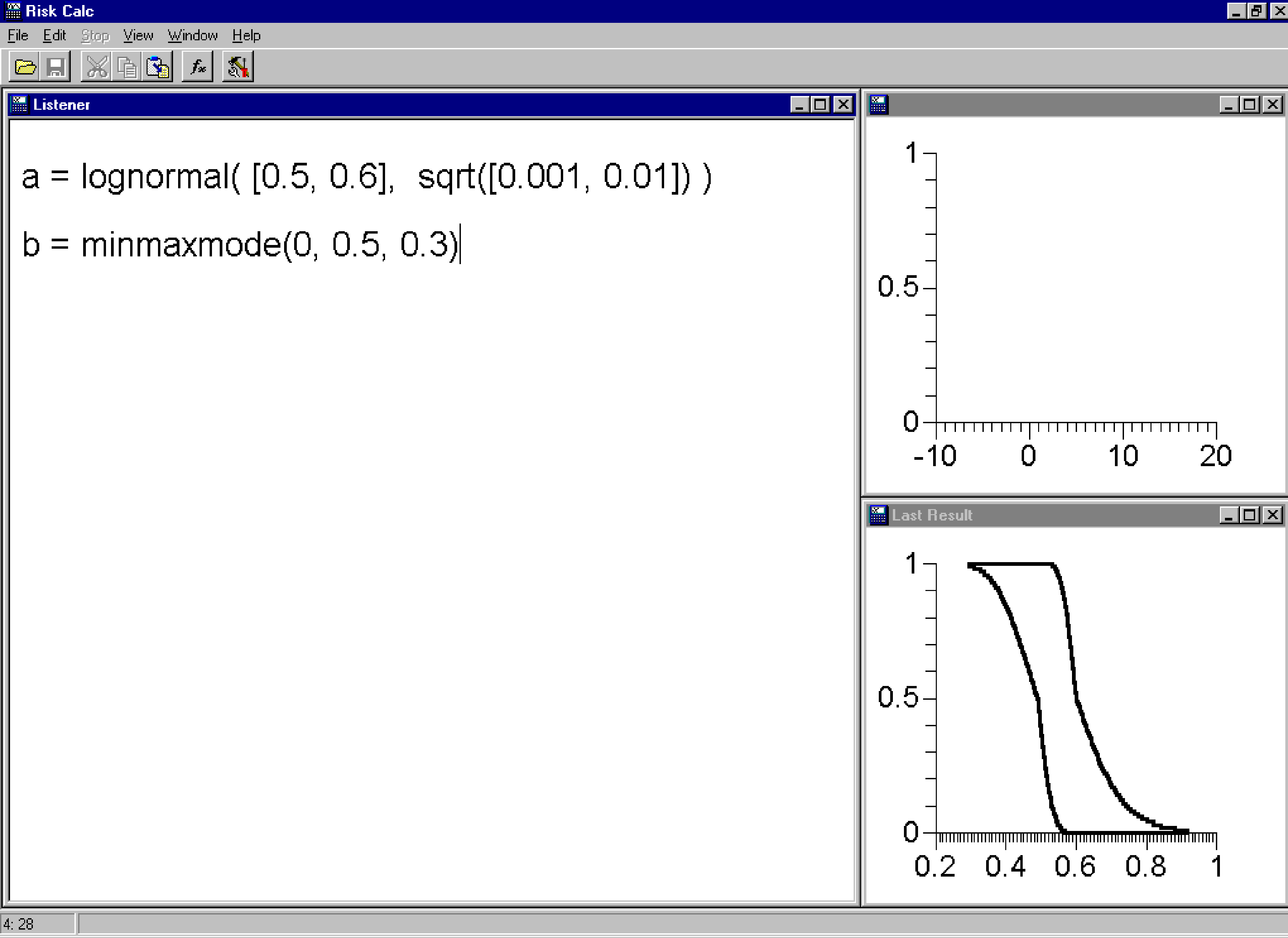
Need β testers

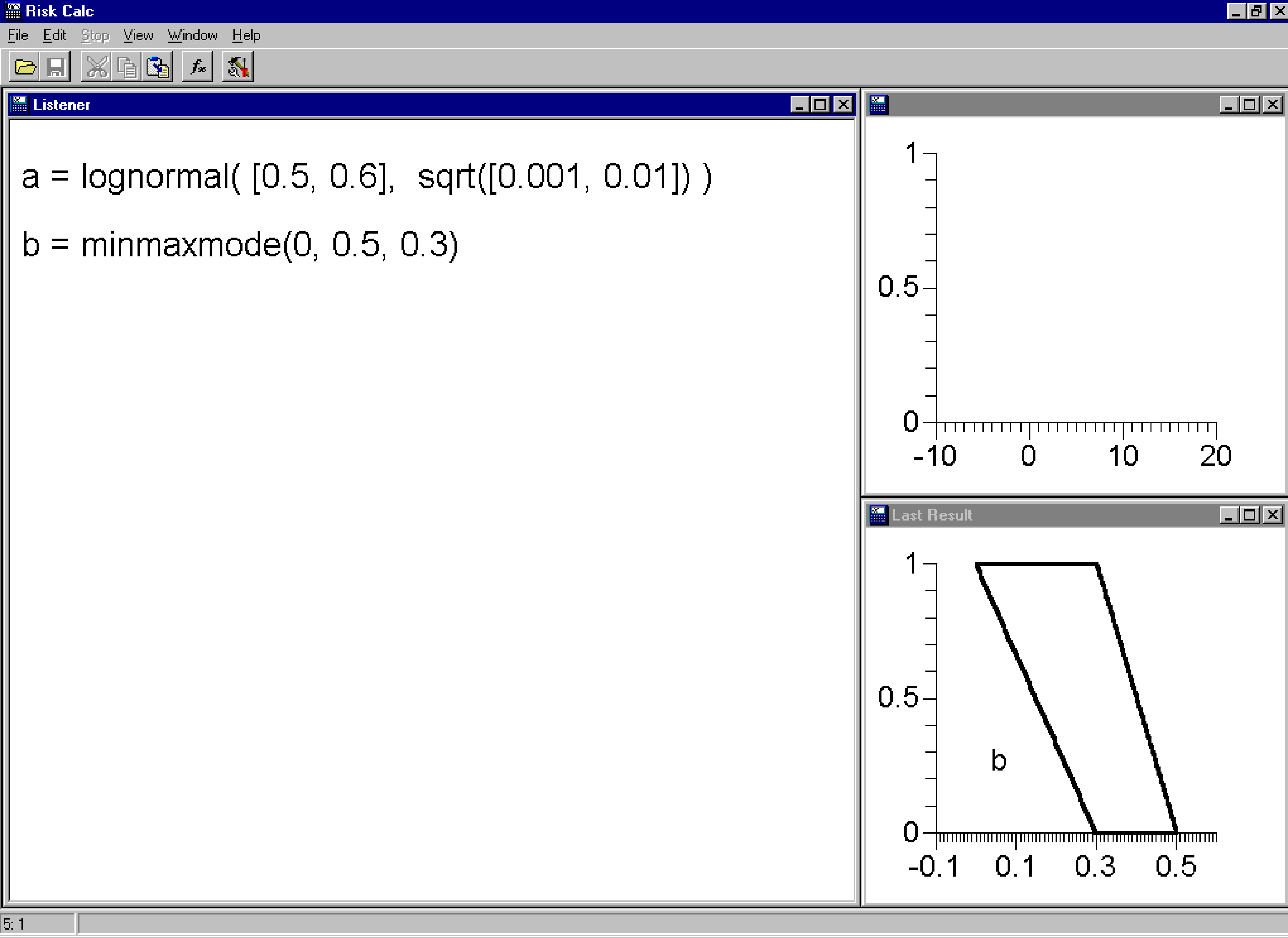
- UC add-in for Excel (NASA, beta 2011)
- RAMAS Risk Calc 4.0 (NIH, commercial)
- Statool (Dan Berleant, freeware)
- Constructor (Sandia and NIH, freeware)
- Pbox.r library for R
- PBDemo (freeware)
- Williamson and Downs (1990)

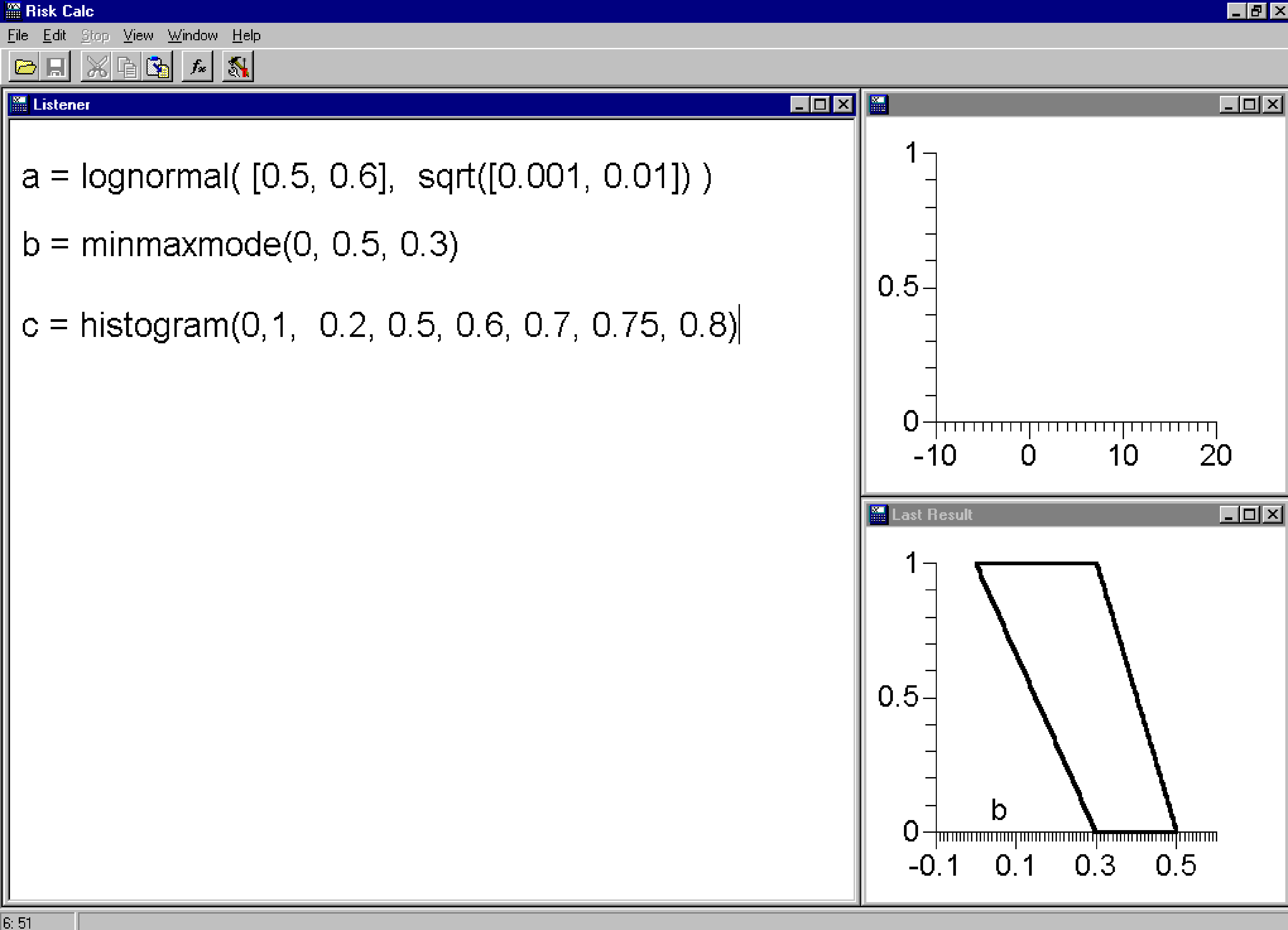

Risk Calc
File Edit Stop View Window Help
Listener
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
-10 0 10 20
1: 1

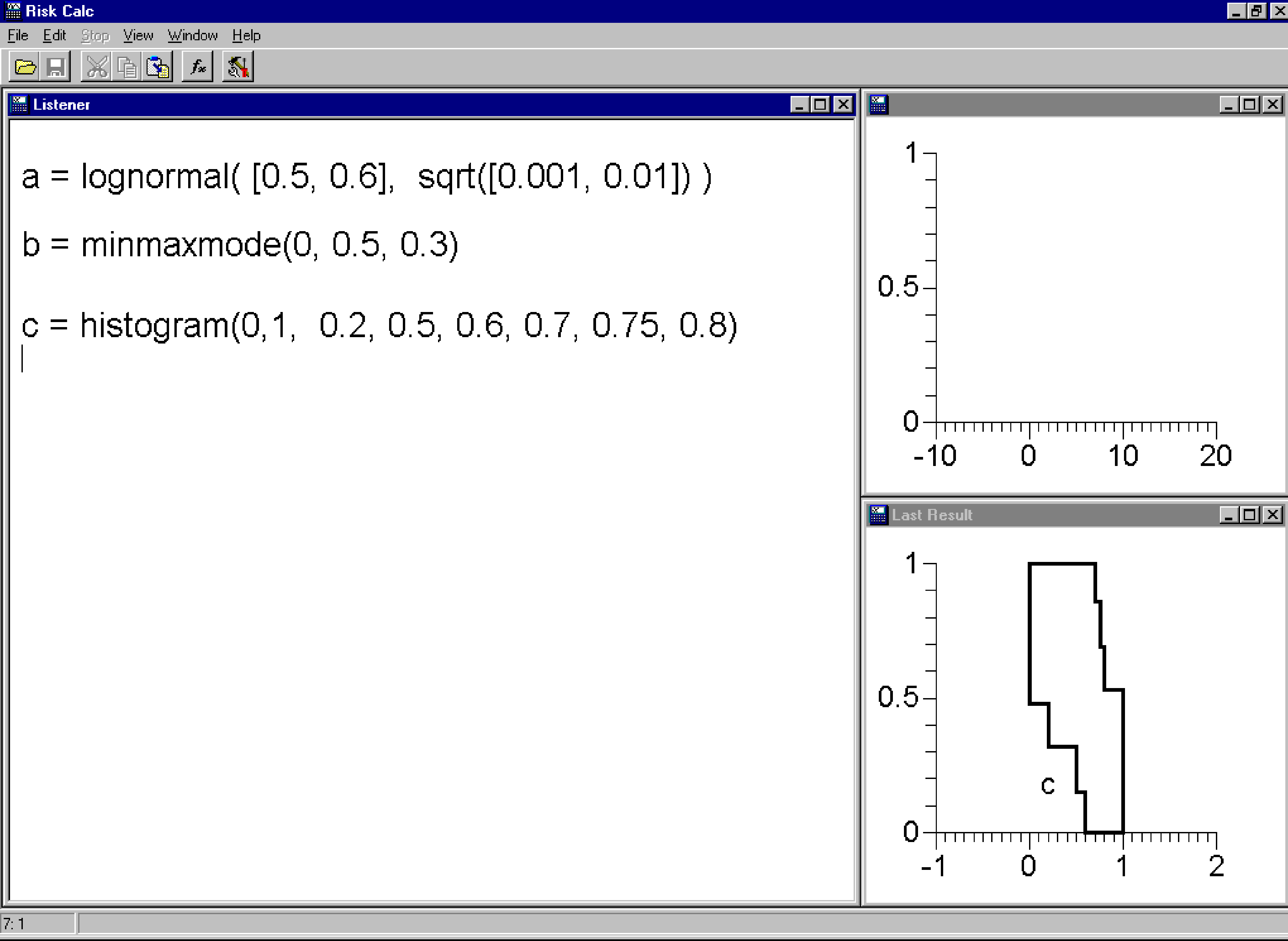

Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
-10 0 10 20
2: 50


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
a
0.2
0.4
0.6
0.8
1
3: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
1
0.5
0
-10 0 10 20
Last Result
1
0.5
0
0.2 0.4 0.6 0.8 1
4: 28


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
b
-0.1
0.1
0.3
0.5
5: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
b
-0.1
0.1
0.3
0.5
6: 51


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
c
-1
0
1
2
7: 1

Risk Calc

File Edit Stop View Window Help

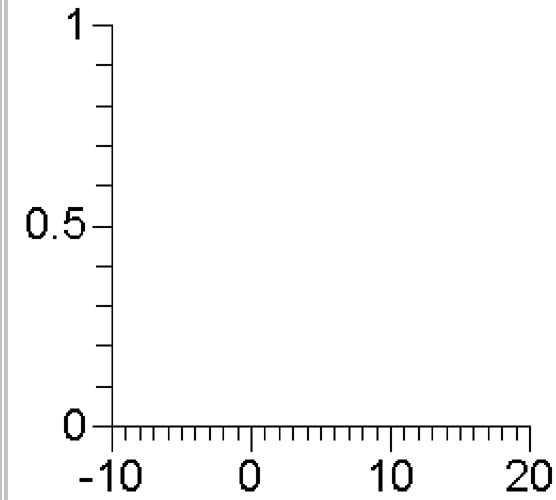
Listener

a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))

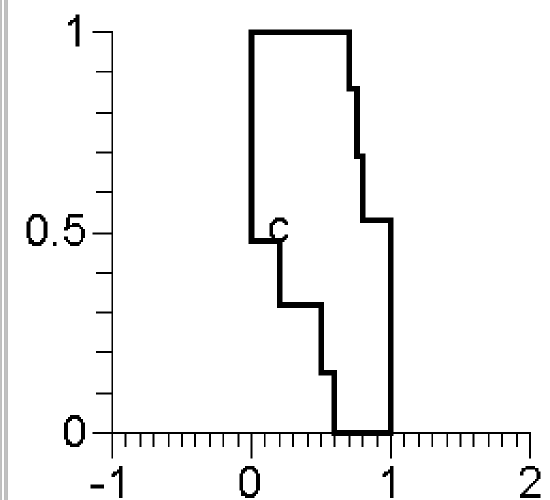
b = minmaxmode(0, 0.5, 0.3)

c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)

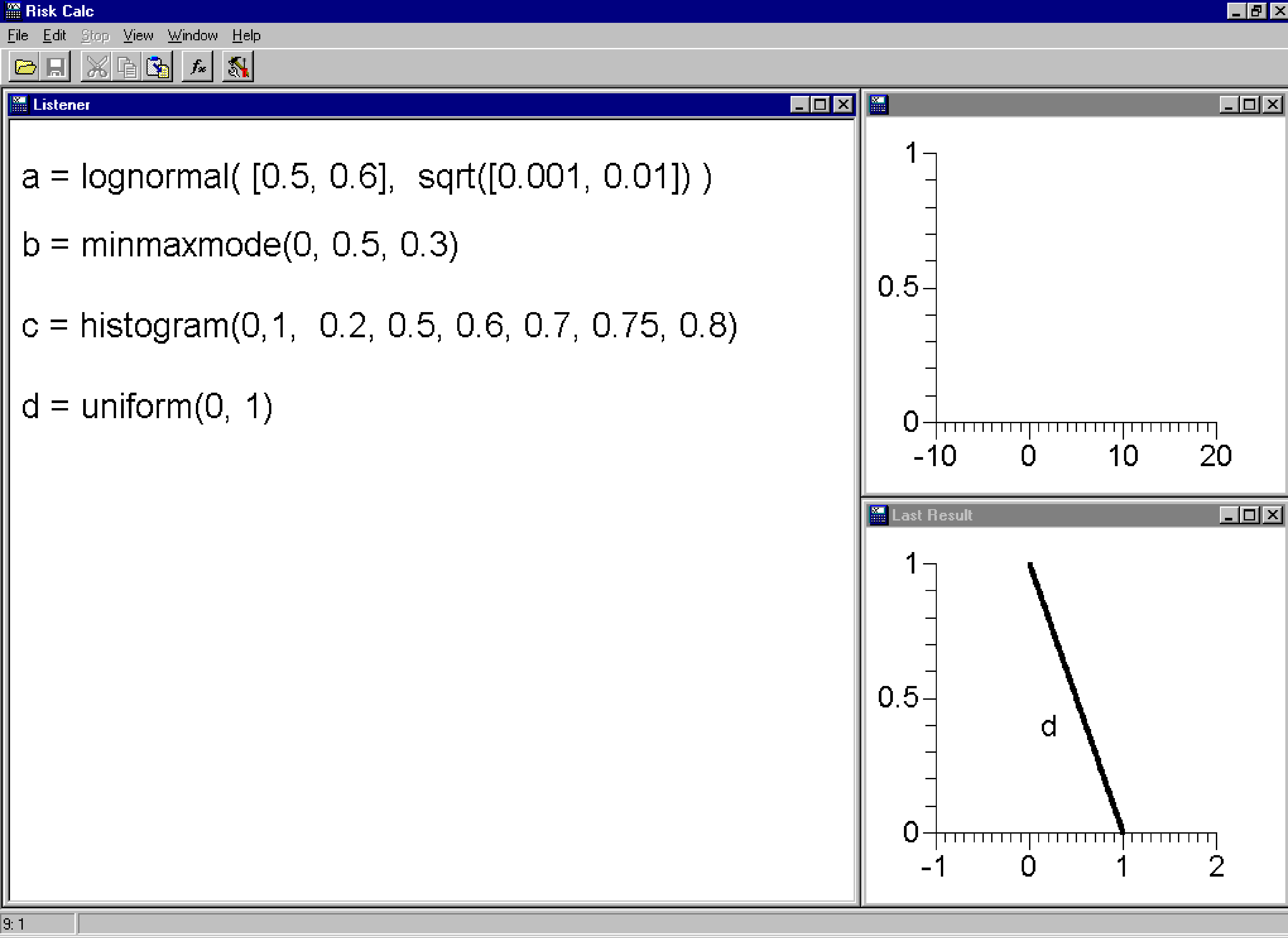
d = uniform(0, 1)

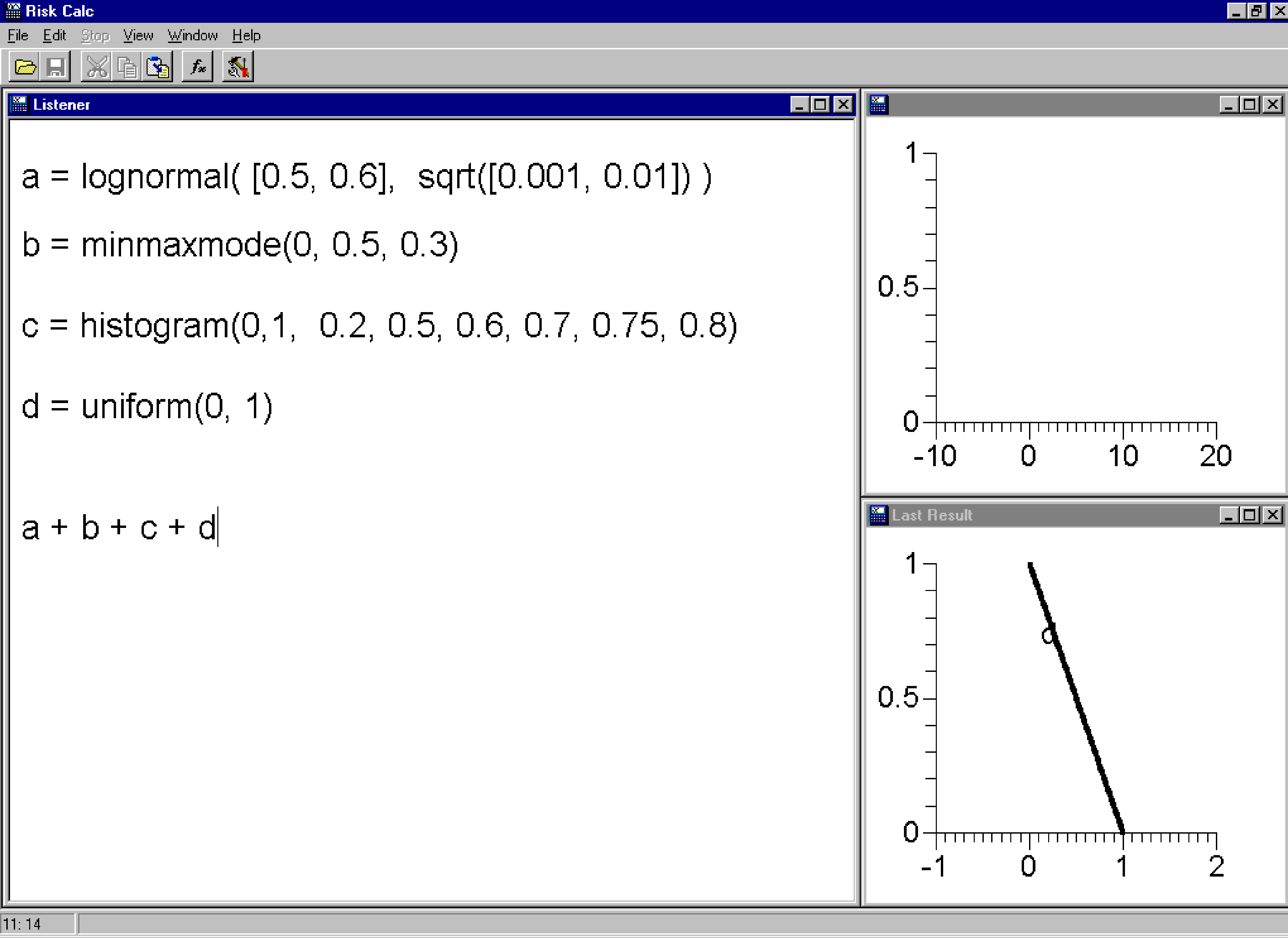


Last Result



8: 18


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
d
-1
0
1
2
9: 1

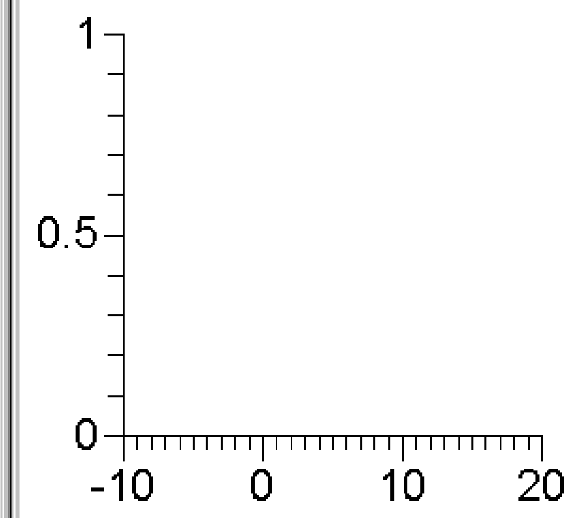

Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
-1
0
1
2
11: 14

Risk Calc

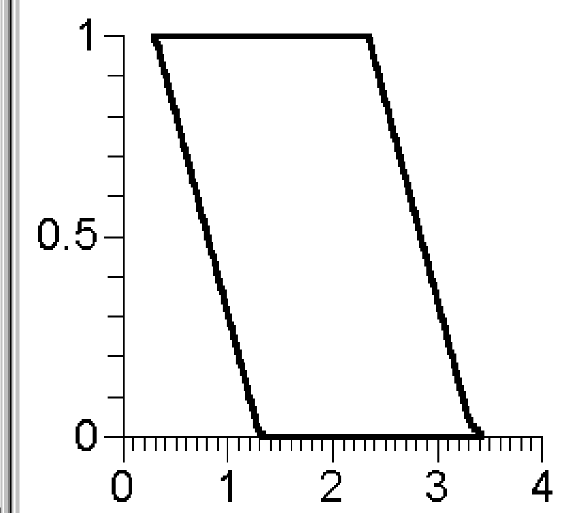
File Edit Stop View Window Help

Listener

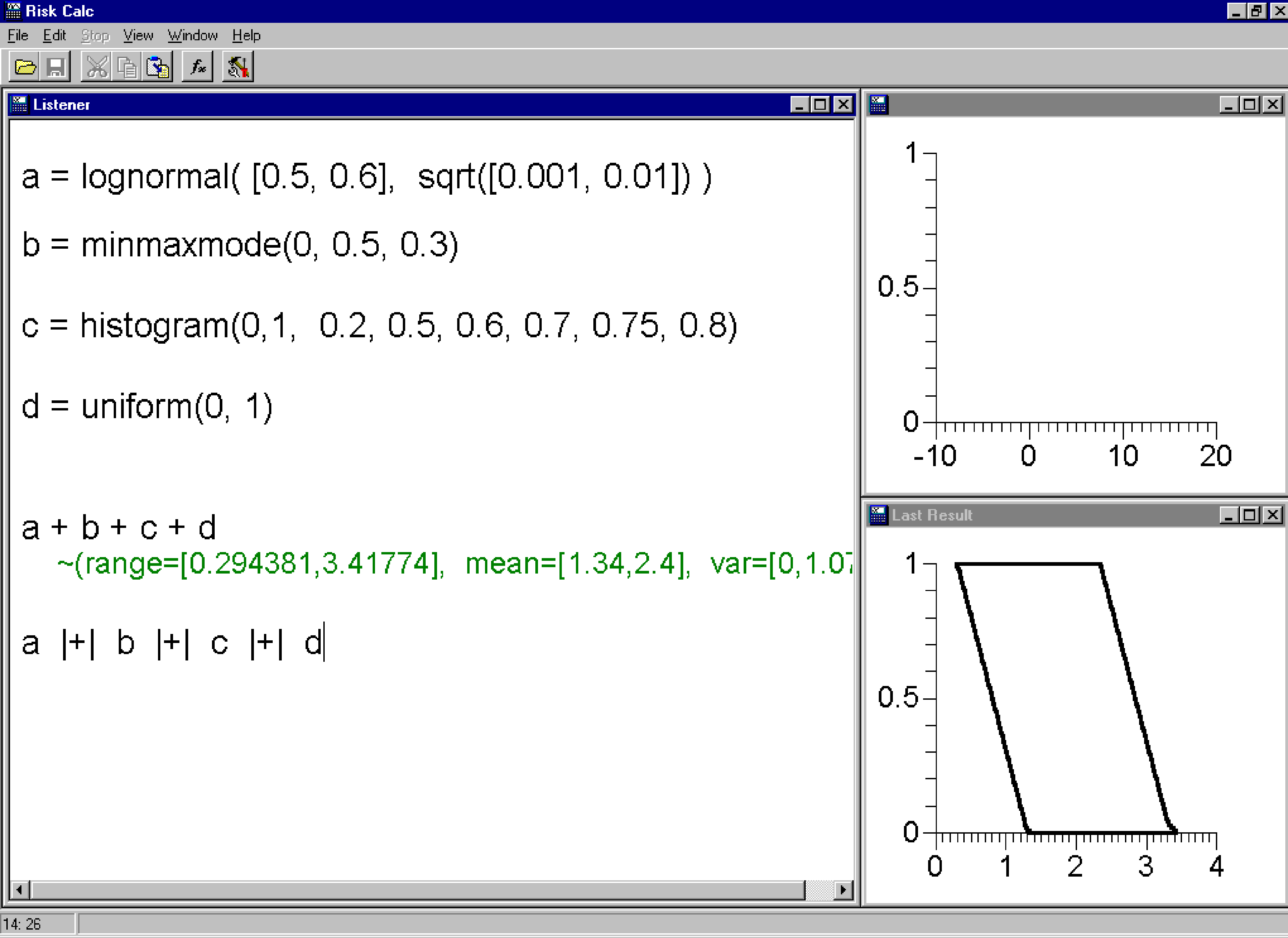
```
a = lognormal( [0.5, 0.6],  sqrt([0.001, 0.01]) )

b = minmaxmode(0, 0.5, 0.3)

c = histogram(0,1,  0.2, 0.5, 0.6, 0.7, 0.75, 0.8)

d = uniform(0, 1)


a + b + c + d
   ~(range=[0.294381,3.41774],  mean=[1.34,2.4],  var=[0,1.07
```

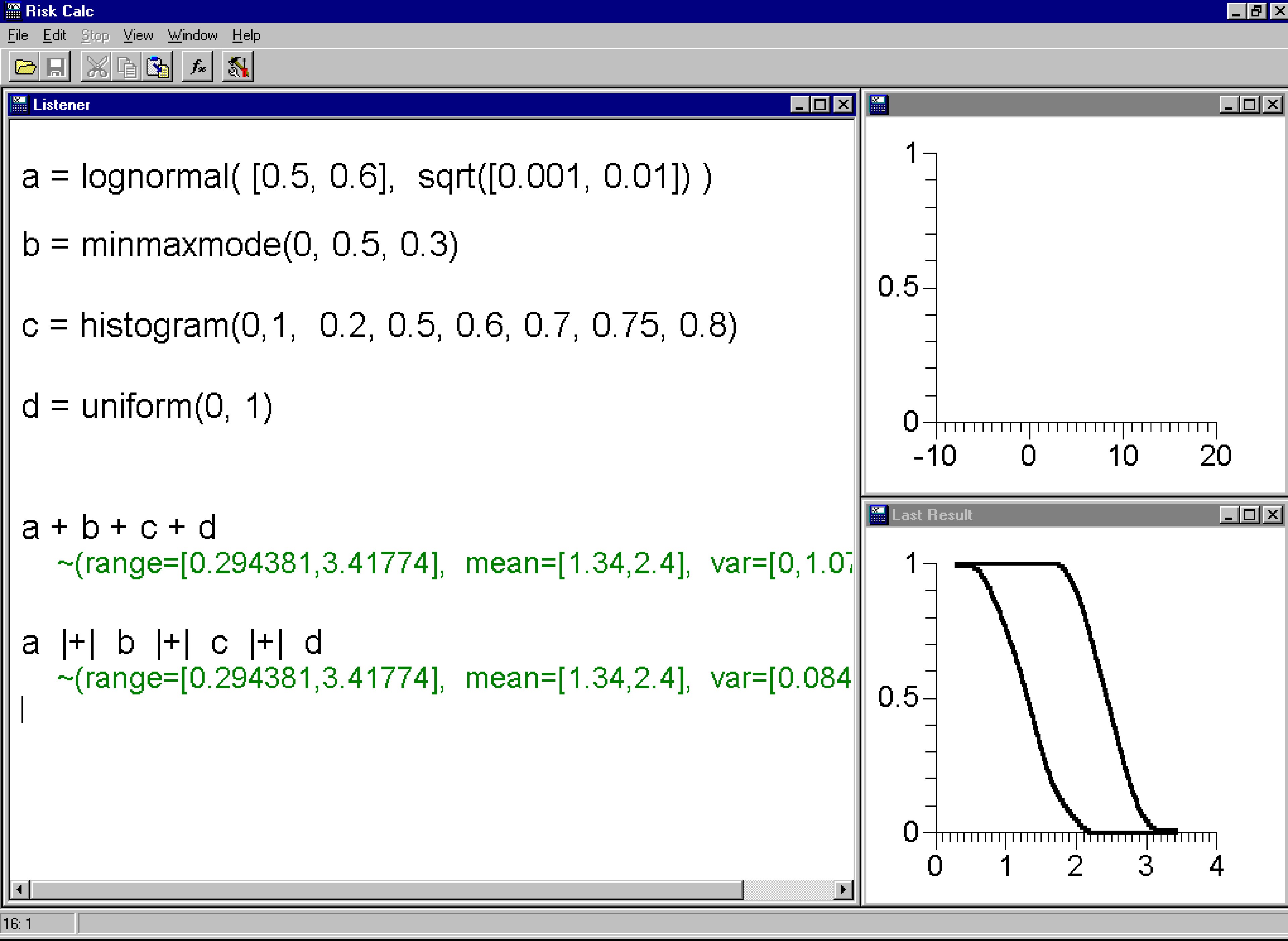


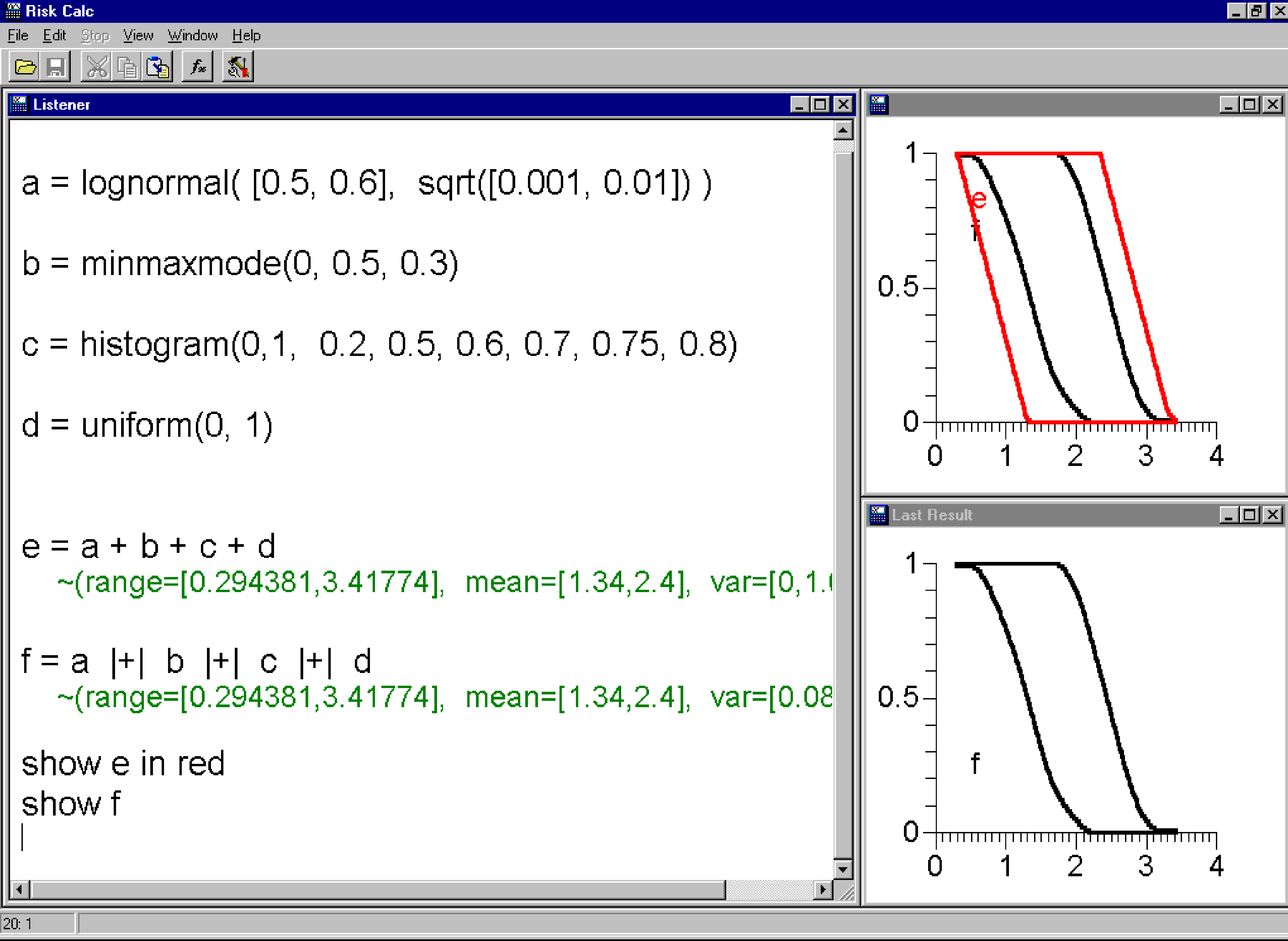
Last Result



13: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.07
a |+| b |+| c |+| d
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
0
1
2
3
4
14: 26


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.07
a |+| b |+| c |+| d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0.084
1
0.5
0
-10
0
10
20
Last Result
1
0.5
0
0
1
2
3
4
16: 1


Risk Calc
File Edit Stop View Window Help
Listener
a = lognormal([0.5, 0.6], sqrt([0.001, 0.01]))
b = minmaxmode(0, 0.5, 0.3)
c = histogram(0,1, 0.2, 0.5, 0.6, 0.7, 0.75, 0.8)
d = uniform(0, 1)
e = a + b + c + d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0,1.
f = a |+| b |+| c |+| d
~(range=[0.294381,3.41774], mean=[1.34,2.4], var=[0.08
show e in red
show f
e
f
1
0.5
0
0 1 2 3 4
Last Result
f
1
0.5
0
0 1 2 3 4
20: 1

Diverse applications

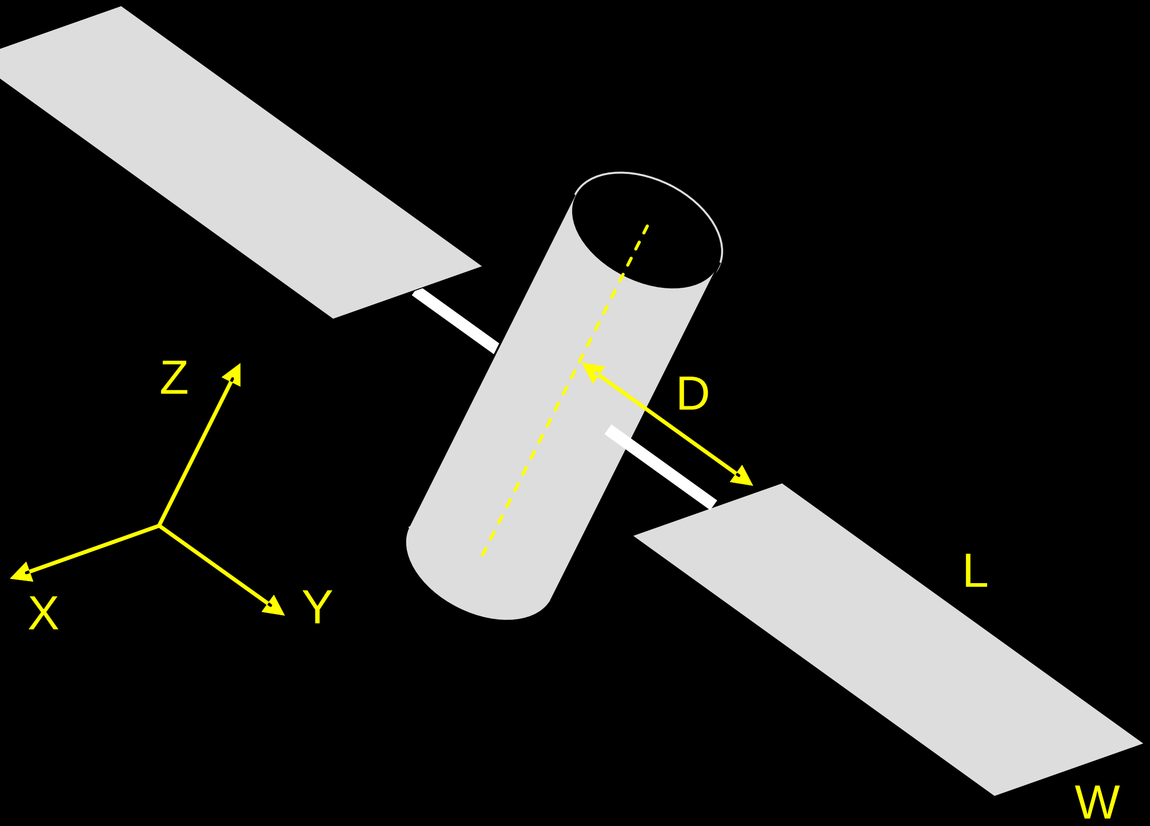
- Superfund risk analyses
- Conservation biology extinction/reintroduction
- Occupational exposure assessment
- Food safety
- Chemostat dynamics
- Global climate change forecasts
- Safety of engineered systems
- Engineering design

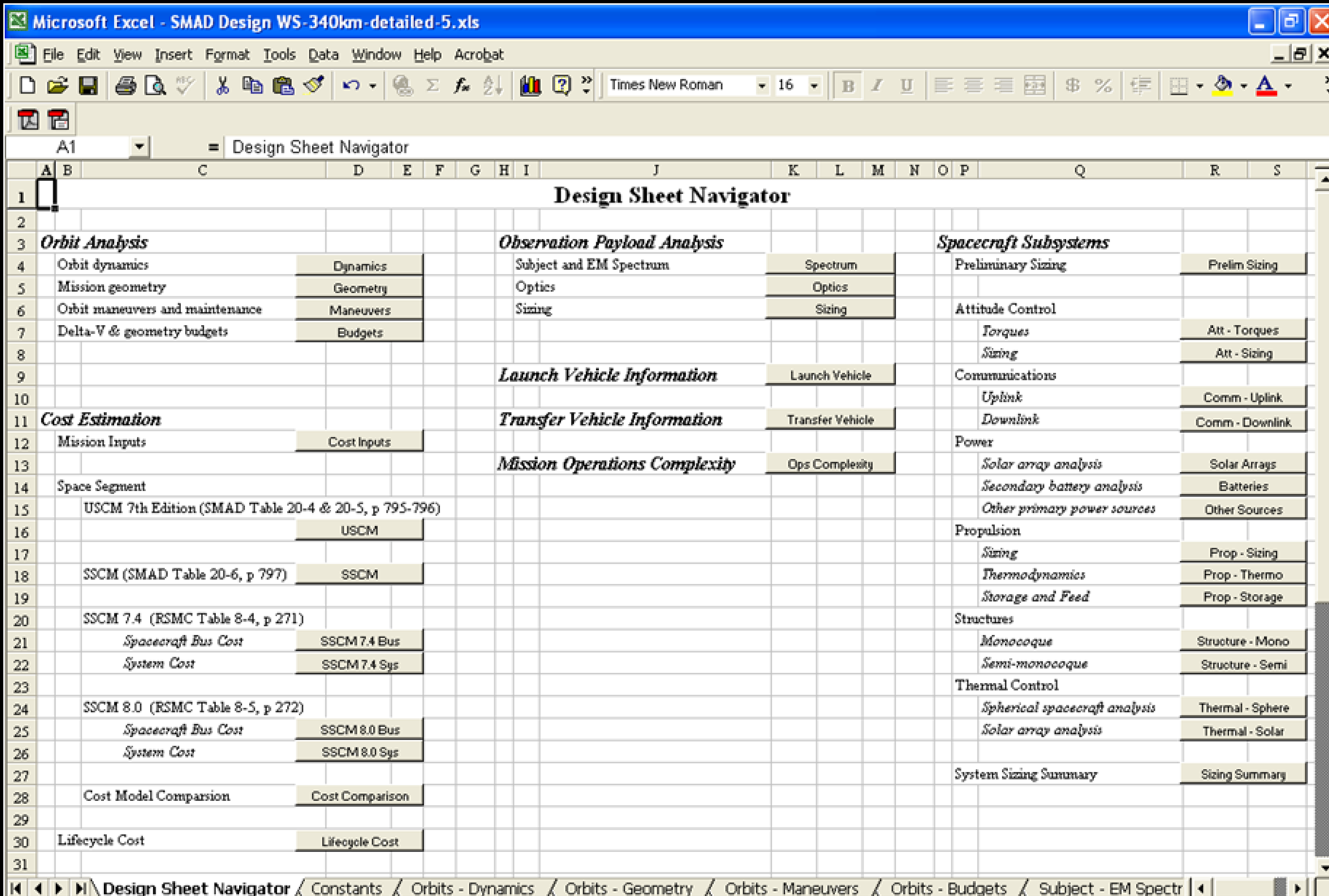
Case study:

Spacecraft design under mission uncertainty

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

M17 =

Return to Navigator

Attitude Control - Torque Estimates

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

Orbit characteristics				***Environmental torques***			
Altitude		340.000	km	Gravity gradient	1.794E-03	N-m	
Satellite velocity		7.703	km/s	Solar radiation	2.565E-05	N-m	
				Magnetic	5.250E-05	N-m	
Atmospheric density		1.983E-11	kg/m^3	Aerodynamic	6.473E-03	N-m	
Sun incidence angle		0.00	deg				
Maximum deviation from local vertical		10.00	deg	Total (RSS)	6.717E-03	N-m	
Spacecraft characteristics				***Slewing torque***	3.360E-02	N-m	
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				

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

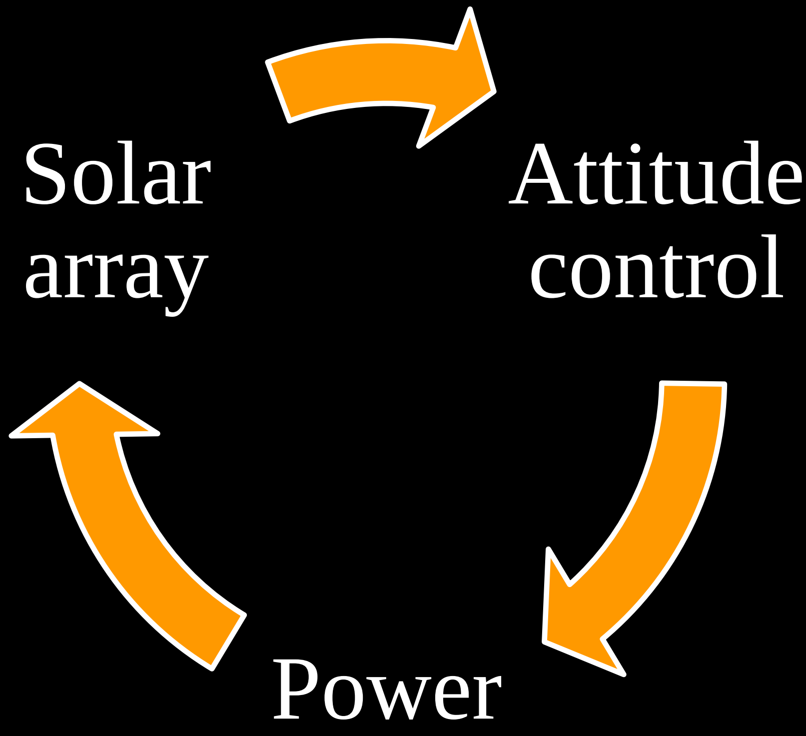
Typical subsystems

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

Demonstrations

- 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
  - $\Delta t_{\mathrm{orbit}}$ = 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*

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

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

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

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

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

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

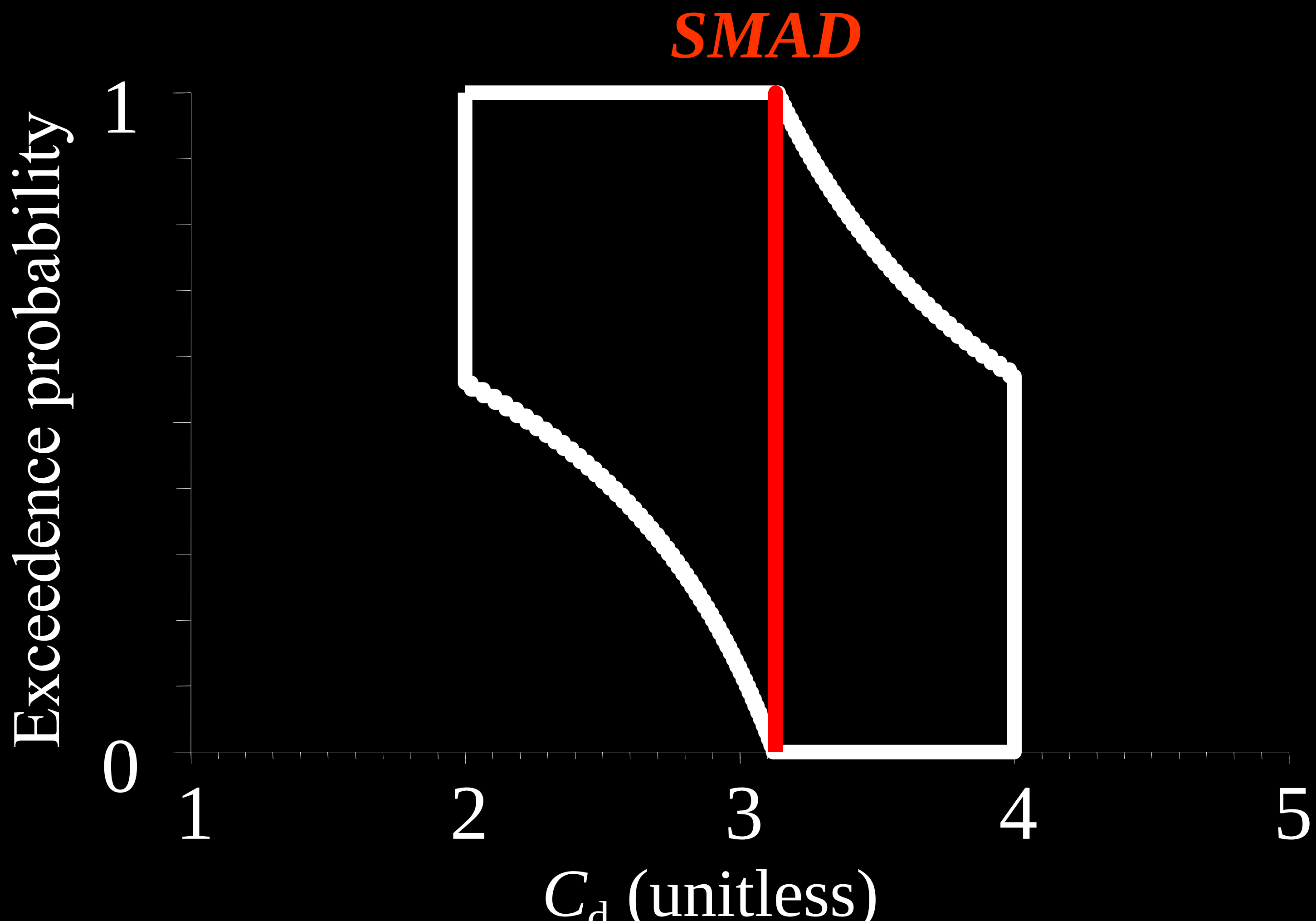
$$\tau_{\mathrm{m}} = \frac{2MD}{(R_{\mathrm{E}} + H)^3}$$

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

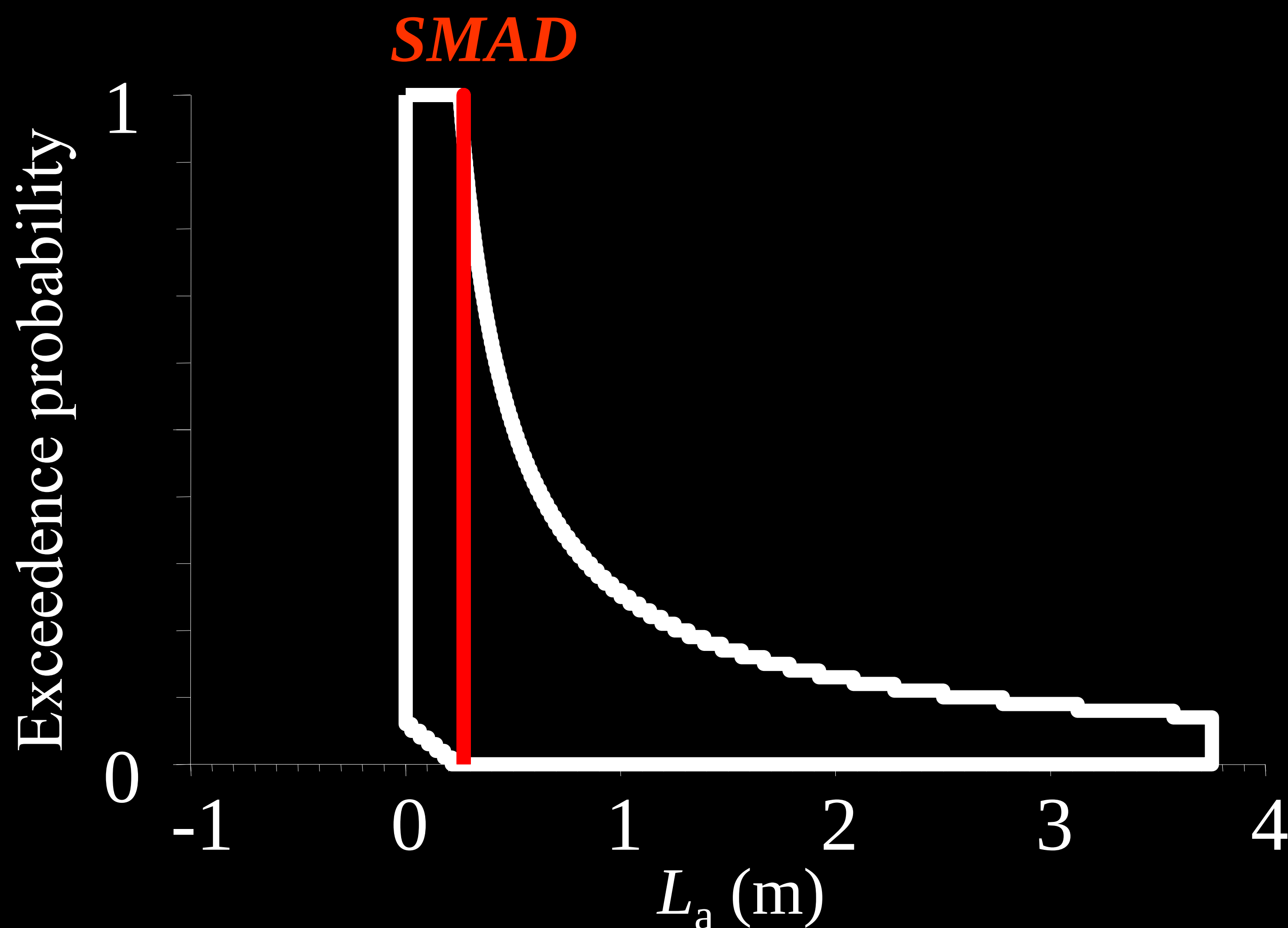
Attitude control input variables

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_r	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

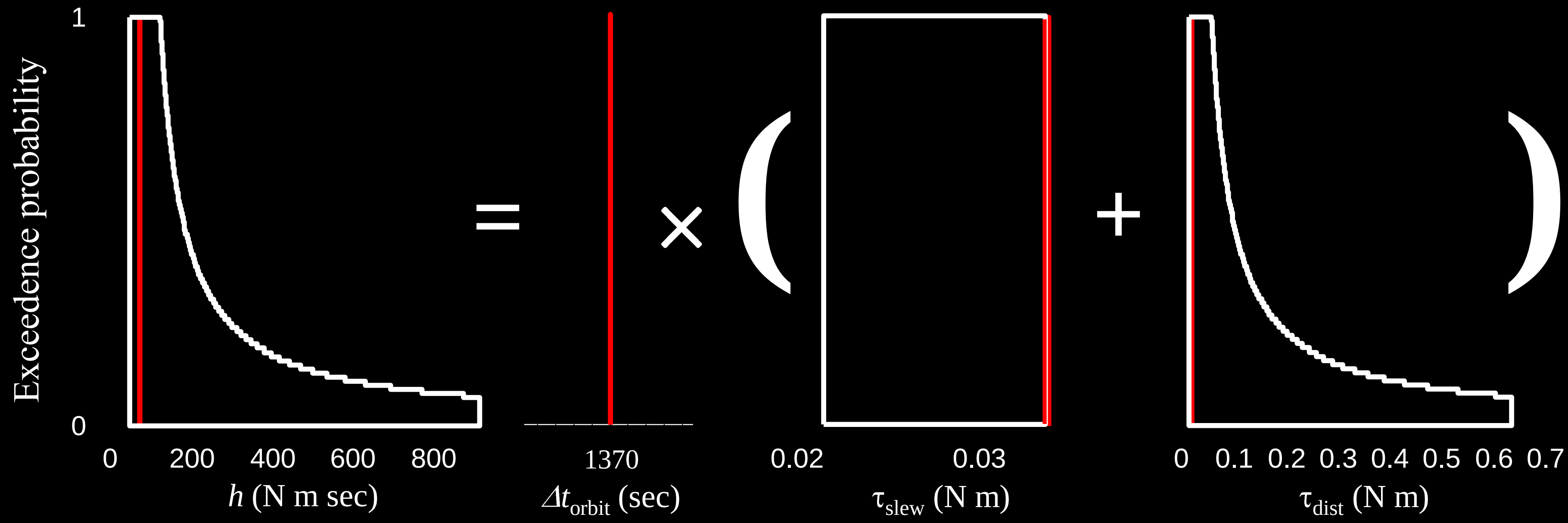
Coefficient of drag, C_d



Aerodynamic drag torque moment, L_a

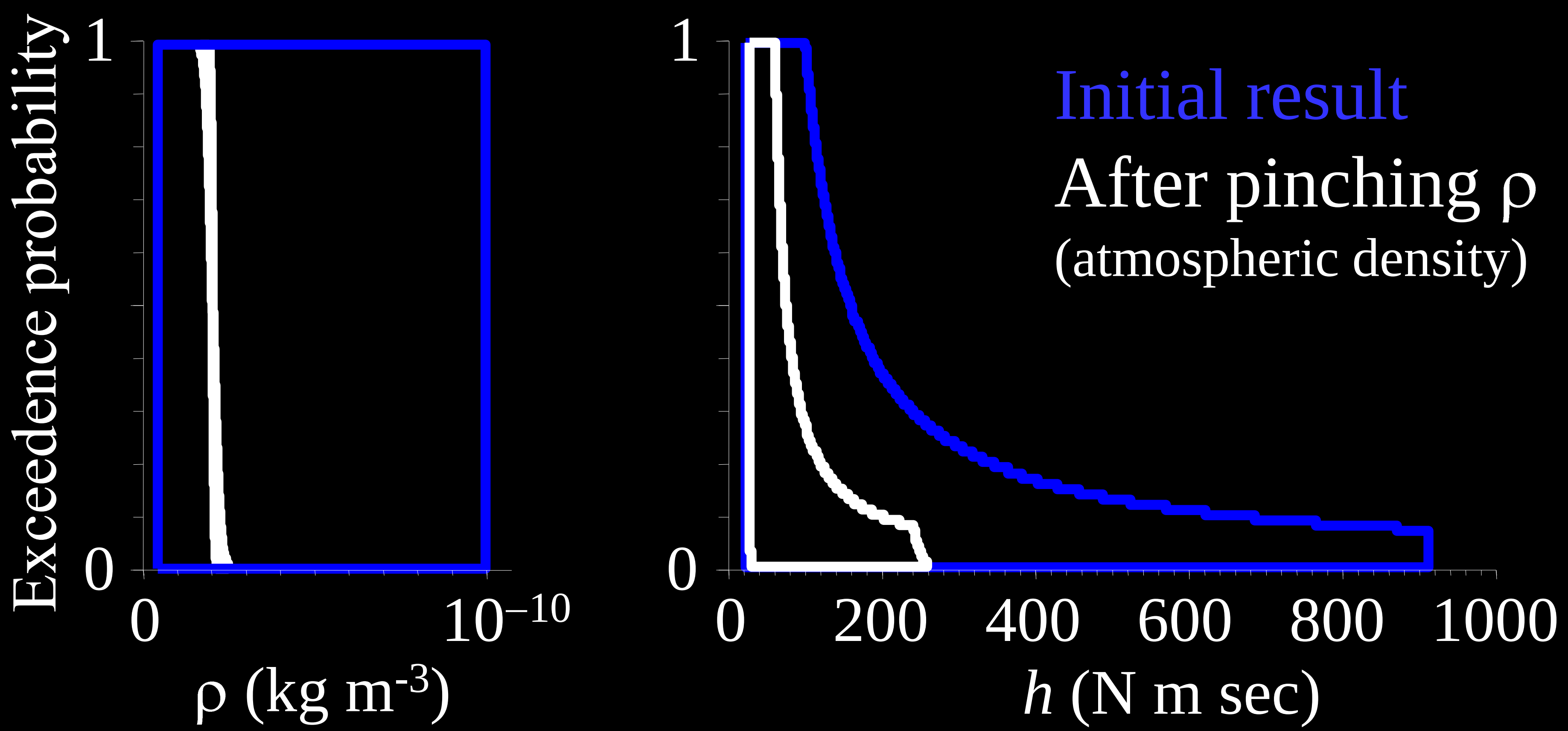


Required angular momentum *h*

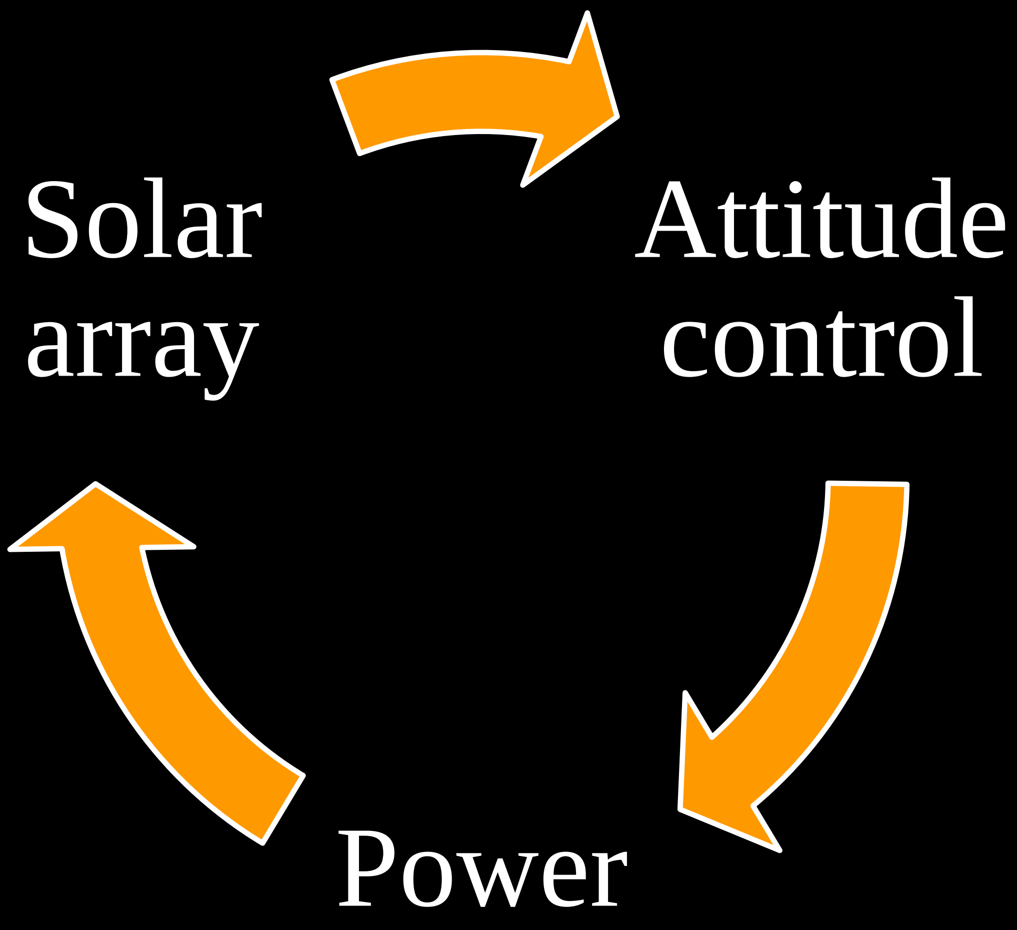


$$h = \Delta t_{orbit} \times \left(\tau_{slew} + \tau_{dist}\right)$$

Value of information: pinching ρ



Three linked subsystems



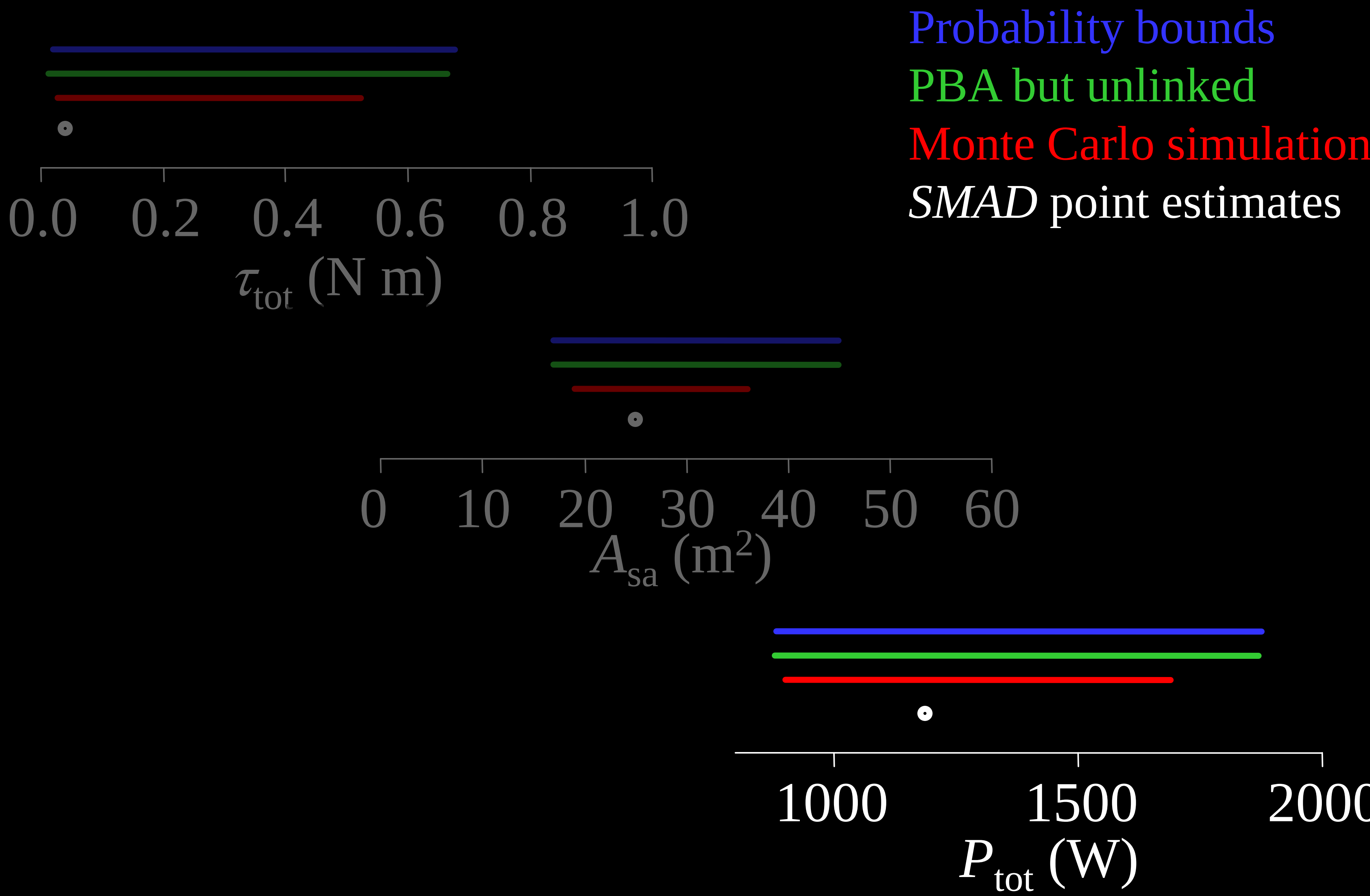
Variables passed iteratively

- 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}

Analysis of calculations

- Need to check that original *SMAD* values and all Monte Carlo simulations are enclosed by p-boxes
- Need to ensure iteration through links doesn't cause runaway uncertainty growth (or reduction)
- Four parallel analyses
  - *SMAD*'s point estimates
  - Monte Carlo simulation
  - P-boxes but without linkage among subsystems
  - P-boxes with fully linked subsystems

Supports of results



Case study findings

- Different answers are consistent
  - Point estimates match the *SMAD* results
  - P-boxes span the points and the Monte Carlo intervals
- Calculations workable
  - No runaway inflation (or loss) of uncertainty
  - Easier than with Monte Carlo
- Practical and interesting results
  - Uncertainty can affect engineering decisions
  - Reducing uncertainty about ρ (by picking a launch date) strongly reduces design uncertainty

For more information, consult SBIR project report to NASA, July 2009

Accounting for epistemic and aleatory uncertainty in early system design

NASA SBIR Phase 2 Final Report

Applied Biomathematics
100 North Country Road
Setauket, NY 11733

Phone: 1-631-751-4350
Fax: 1-631-751-3435
Email: scott@ramas.com
www.ramas.com

Order Number: NNL07AA06C

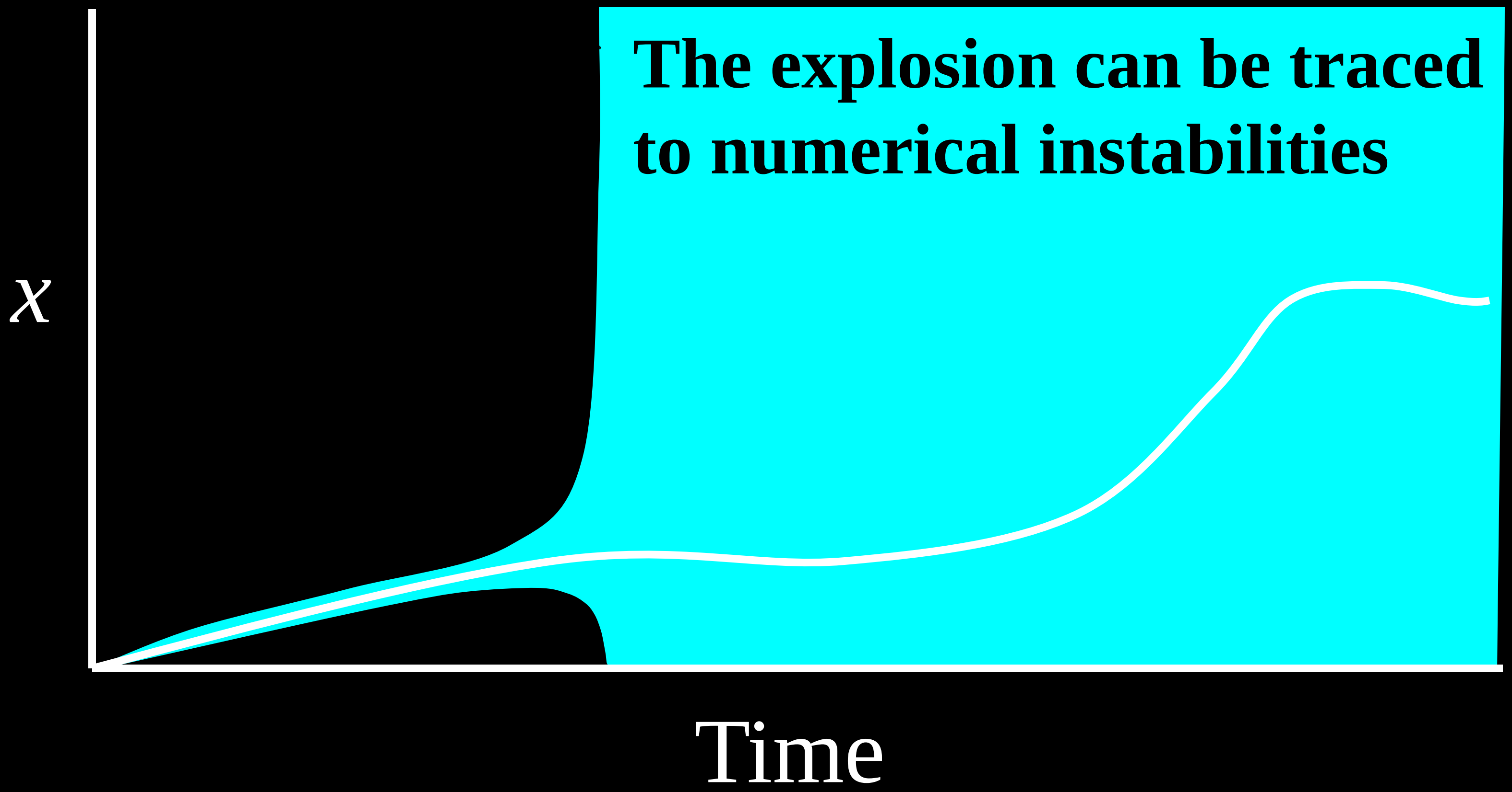
July 2009

Uses for probability bounds analysis

- Uncertainty propagation
- Risk assessment
- Sensitivity analysis (for control and study)
- Reliability theory
- Engineering design ✓
- Validation
- Decision theory
- Regulatory compliance
- Finite element modeling
- Differential equations ✓

Differential equations

Uncertainty usually explodes



Uncertainty

- Artifactual uncertainty
  - Too few polynomial terms
  - Numerical instability
  - Can be reduced by a better analysis
- Authentic uncertainty
  - Genuine unpredictability due to input uncertainty
  - Cannot be reduced by a better analysis

Only by more information, data or assumptions

Uncertainty propagation

- We *want* the prediction to ‘break down’ if that’s what should happen

- But we don’t want artifactual uncertainty
  - Numerical instabilities
  - Wrapping effect
  - Dependence problem
  - Repeated parameters

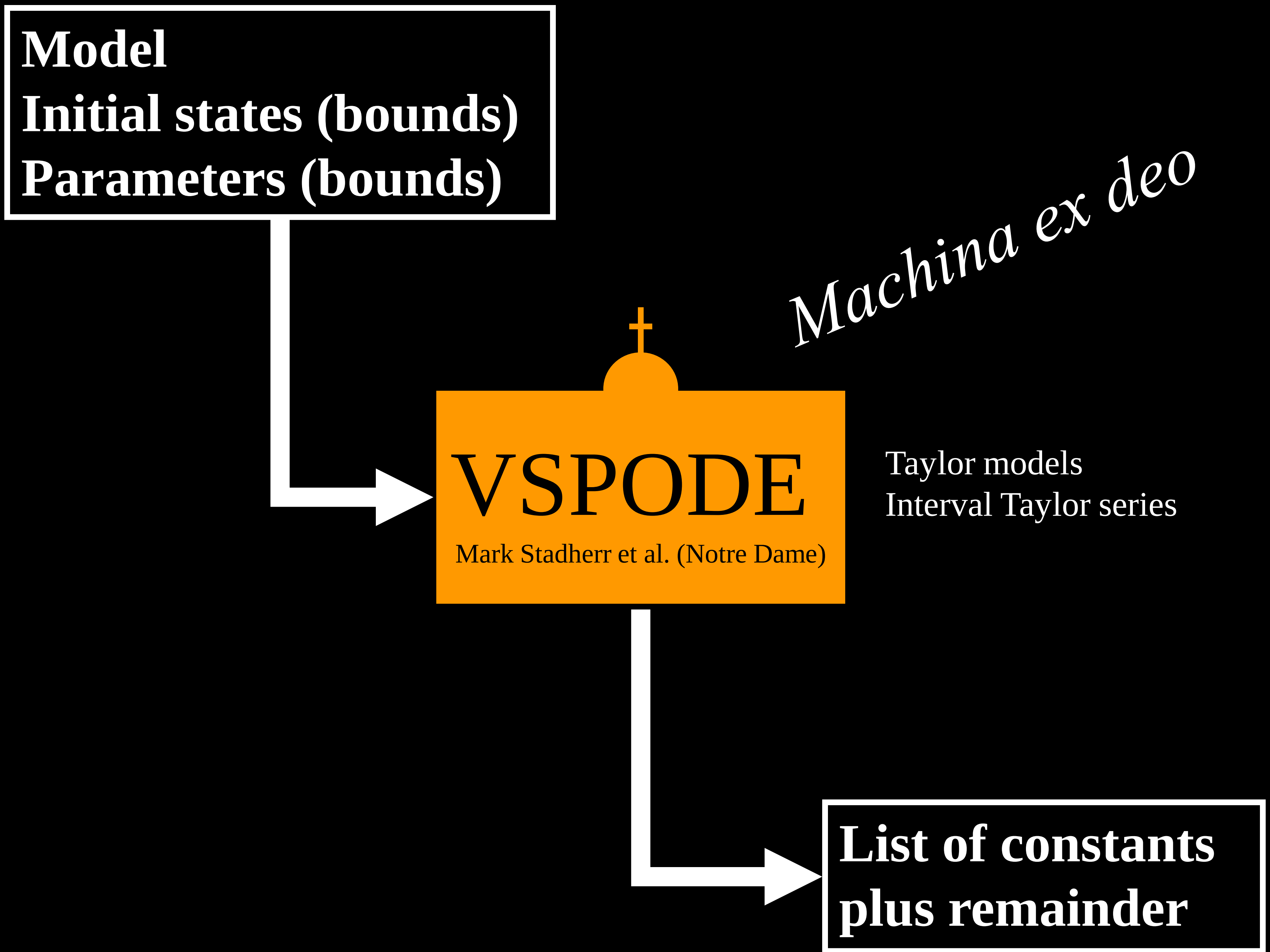
Problem

- Nonlinear ordinary differential equation (ODE)

$$\mathrm{d}\boldsymbol{x}/\mathrm{d}t = f(\boldsymbol{x}, \theta)$$

  with uncertain θ and uncertain initial state x_0

- Information about θ and x_0 comes as
  - Interval ranges
  - Probability distributions
  - Probability boxes


Model
Initial states (bounds)
Parameters (bounds)
Machina ex deo
VSPODE
Mark Stadherr et al. (Notre Dame)
Taylor models
Interval Taylor series
List of constants
plus remainder

Example ODE

$$\mathrm{d}x_1/\mathrm{d}t = \theta_1\, x_1(1 - x_2)$$
$$\mathrm{d}x_2/\mathrm{d}t = \theta_2\, x_2(x_1 - 1)$$

What are the states at $t = 10$?

$$x_0 = (1.2,\ 1.1)^{\mathrm{T}}$$
$$\theta_1 \in [2.99,\ 3.01]$$
$$\theta_2 \in [0.99,\ 1.01]$$

VSPODE

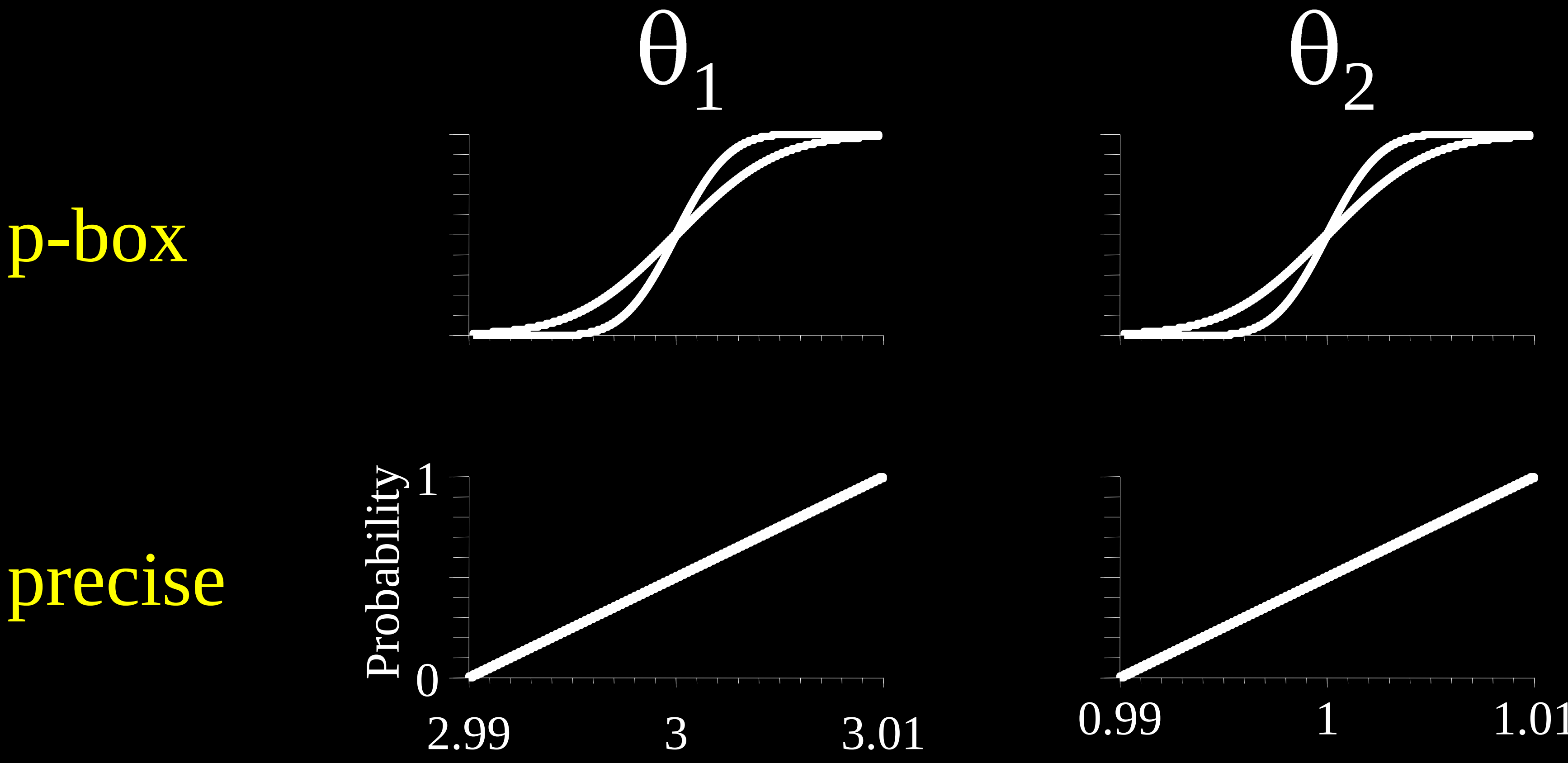
- Constant step size $h = 0.1$, Order of Taylor model $q = 5$,
- Order of interval Taylor series $k = 17$, QR factorization

VSPODE tells how to compute x_1

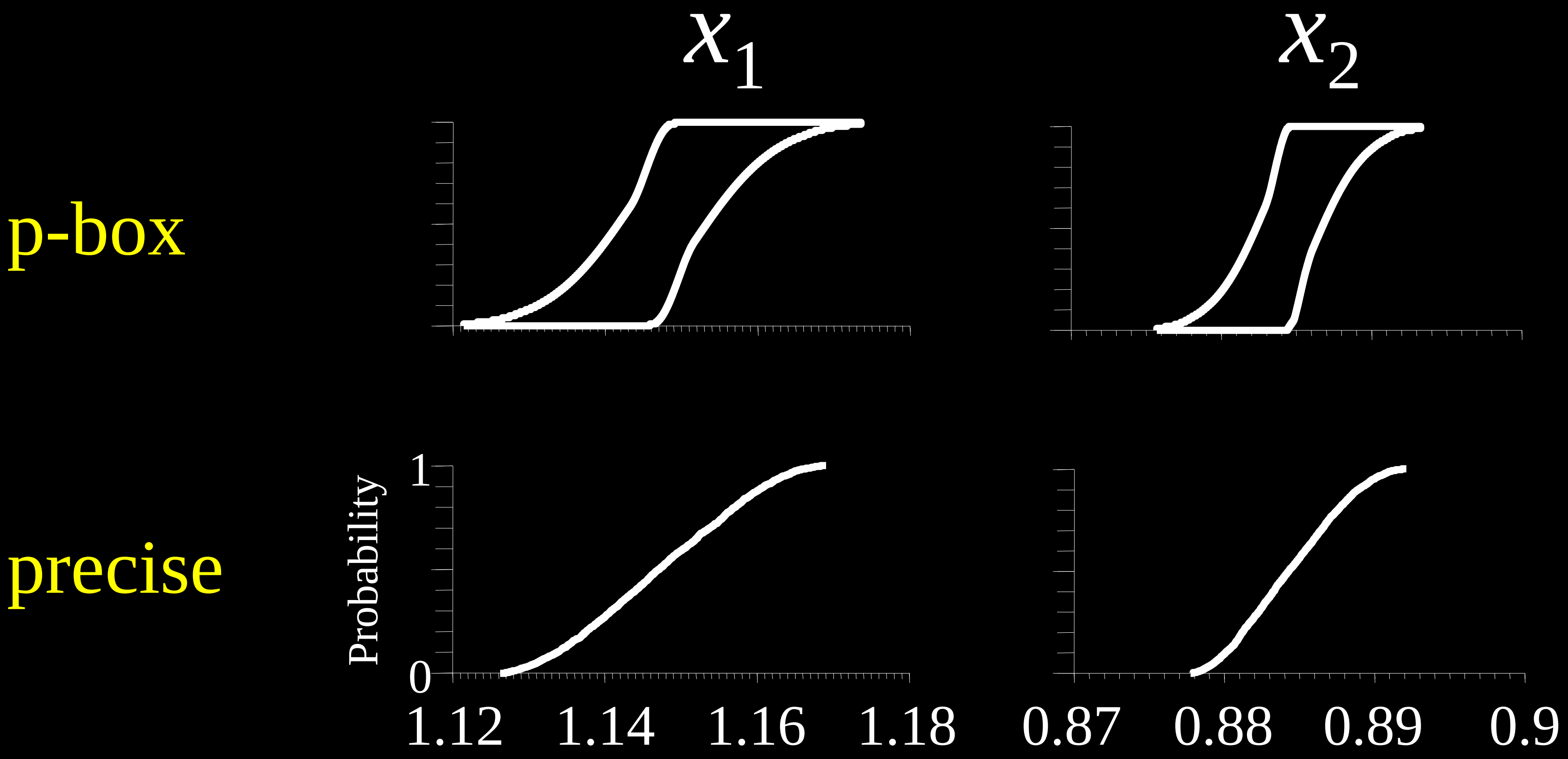
$$
\begin{array}{rr}
1.916037656181642 \times \theta_1^{\,0} \times \theta_2^{\,1} \;+ & 0.689979149231081 \times \theta_1^{\,1} \times \theta_2^{\,0} \;+ \\
-4.690741189299572 \times \theta_1^{\,0} \times \theta_2^{\,2} \;+ & -2.275734193378134 \times \theta_1^{\,1} \times \theta_2^{\,1} \;+ \\
-0.450416914564394 \times \theta_1^{\,2} \times \theta_2^{\,0} \;+ & -29.788252573360062 \times \theta_1^{\,0} \times \theta_2^{\,3} \;+ \\
-35.200757076497972 \times \theta_1^{\,1} \times \theta_2^{\,2} \;+ & -12.401600707197074 \times \theta_1^{\,2} \times \theta_2^{\,1} \;+ \\
-1.349694561113611 \times \theta_1^{\,3} \times \theta_2^{\,0} \;+ & 6.062509834147210 \times \theta_1^{\,0} \times \theta_2^{\,4} \;+ \\
-29.503128650484253 \times \theta_1^{\,1} \times \theta_2^{\,3} \;+ & -25.744336555602068 \times \theta_1^{\,2} \times \theta_2^{\,2} \;+ \\
-5.563350070358247 \times \theta_1^{\,3} \times \theta_2^{\,1} \;+ & -0.222000132892585 \times \theta_1^{\,4} \times \theta_2^{\,0} \;+ \\
218.607042326120308 \times \theta_1^{\,0} \times \theta_2^{\,5} \;+ & 390.260443722081675 \times \theta_1^{\,1} \times \theta_2^{\,4} \;+ \\
256.315067368131281 \times \theta_1^{\,2} \times \theta_2^{\,3} \;+ & 86.029720297509172 \times \theta_1^{\,3} \times \theta_2^{\,2} \;+ \\
15.322357274648443 \times \theta_1^{\,4} \times \theta_2^{\,1} \;+ & 1.094676837431721 \times \theta_1^{\,5} \times \theta_2^{\,0} \;+ \\
[\, \underline{1.147753}7620811058,\ \underline{1.147753}9164945061 \,] &
\end{array}
$$

where θ's are centered forms of the parameters; $\theta_1 = \theta_1 - 3$, $\theta_2 = \theta_2 - 1$

Input p-boxes



Results


x_1
x_2
p-box
precise
Probability
1
0
1.12
1.14
1.16
1.18
0.87
0.88
0.89
0.9

Still repeated uncertainties

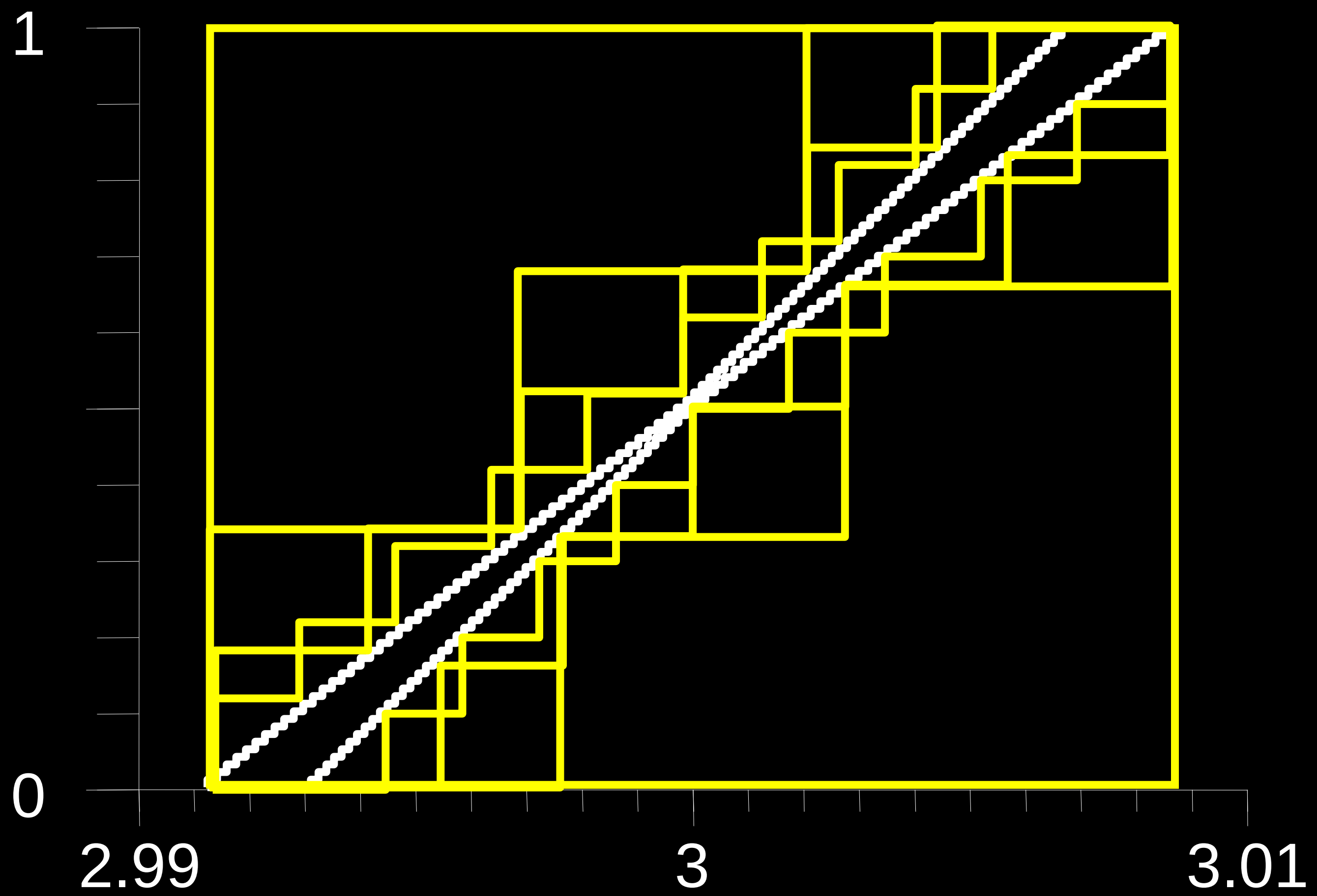
$$
\begin{array}{rr}
1.916037656181642 \times \theta_1^{0} \times \theta_2^{1} + & 0.689979149231081 \times \theta_1^{1} \times \theta_2^{0} + \\
-4.690741189299572 \times \theta_1^{0} \times \theta_2^{2} + & -2.275734193378134 \times \theta_1^{1} \times \theta_2^{1} + \\
-0.450416914564394 \times \theta_1^{2} \times \theta_2^{0} + & -29.788252573360062 \times \theta_1^{0} \times \theta_2^{3} + \\
-35.200757076497972 \times \theta_1^{1} \times \theta_2^{2} + & -12.401600707197074 \times \theta_1^{2} \times \theta_2^{1} + \\
-1.349694561113611 \times \theta_1^{3} \times \theta_2^{0} + & 6.062509834147210 \times \theta_1^{0} \times \theta_2^{4} + \\
-29.503128650484253 \times \theta_1^{1} \times \theta_2^{3} + & -25.744336555602068 \times \theta_1^{2} \times \theta_2^{2} + \\
-5.563350070358247 \times \theta_1^{3} \times \theta_2^{1} + & -0.222000132892585 \times \theta_1^{4} \times \theta_2^{0} + \\
218.607042326120308 \times \theta_1^{0} \times \theta_2^{5} + & 390.260443722081675 \times \theta_1^{1} \times \theta_2^{4} + \\
256.315067368131281 \times \theta_1^{2} \times \theta_2^{3} + & 86.029720297509172 \times \theta_1^{3} \times \theta_2^{2} + \\
15.322357274648443 \times \theta_1^{4} \times \theta_2^{1} + & 1.094676837431721 \times \theta_1^{5} \times \theta_2^{0} +
\end{array}
$$

[1.147537620811058, 1.147753916494506 1]

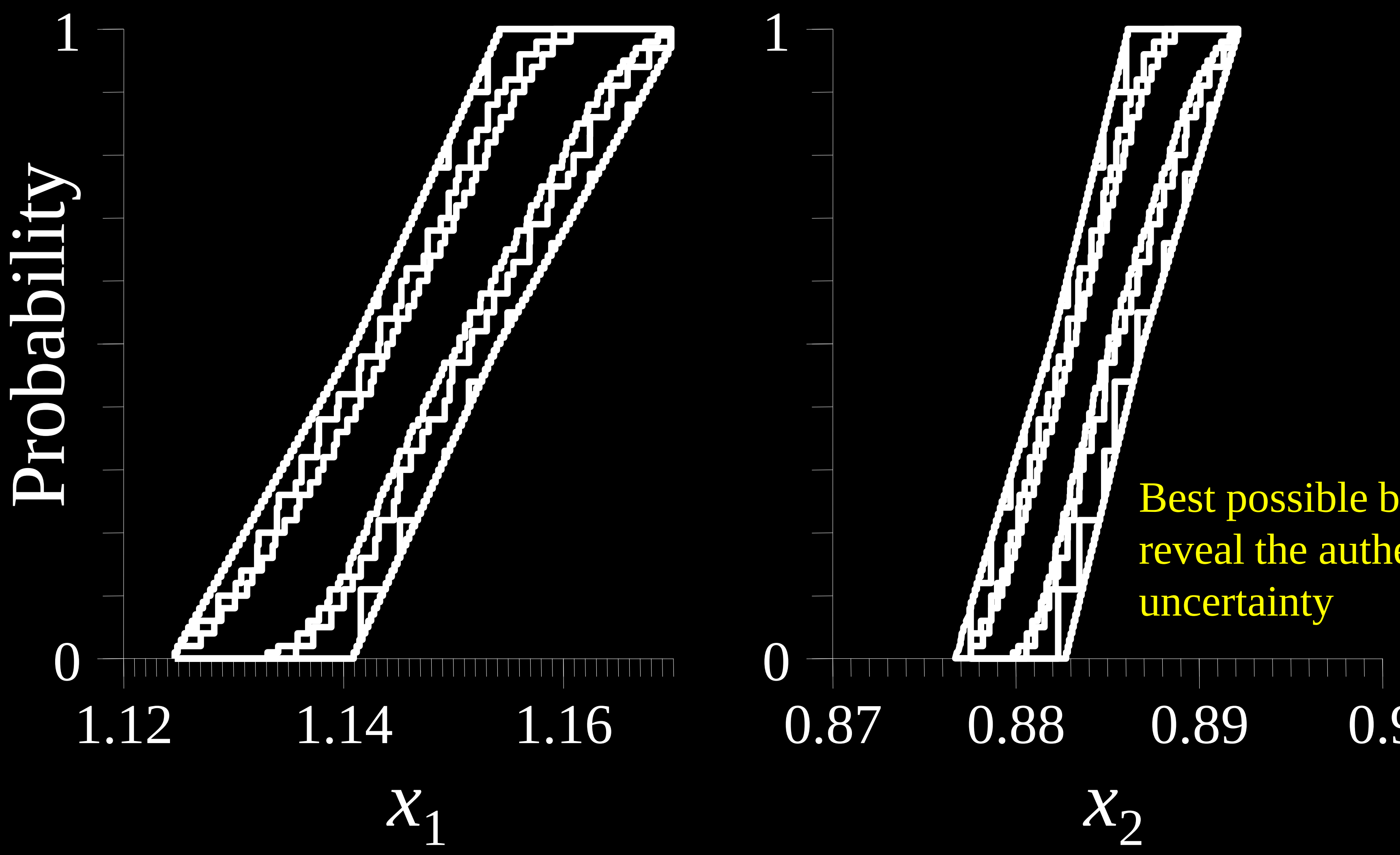
Subinterval reconstitution

- Subinterval reconstitution (SIR)
  - Partition the inputs into subintervals
  - Apply the function to each subinterval
  - Form the union of the results
- Still rigorous, but often tighter
  - The finer the partition, the tighter the union
  - Many strategies for partitioning
- Apply to *each cell* in the Cartesian product

Discretizations



Contraction from SIR



Best possible bounds reveal the authentic uncertainty

Monte Carlo is more limited

- Monte Carlo cannot propagate incertitude
- Monte Carlo *cannot produce validated results*
  - Though can be checked by repeating simulation
- Validated results from distributions can be obtained by modeling inputs with (narrow) p-boxes and applying probability bounds analysis
- Results converge to narrow p-boxes obtained from infinitely many Monte Carlo replications

Results

- Probability bounds analysis with VSPODE are useful for bounding solutions of nonlinear ODEs

- They rigorously propagate uncertainty

about	*in the form of*
Initial states	Intervals
Parameters	Distributions
	P-boxes

Paper in *AIChE Journal* [*American Institute of Chemical Engineers*], February 2011 (on line May 2010)

AIChE

Probability Bounds Analysis for Nonlinear Dynamic Process Models

Joshua A. Enszer and Youdong Lin
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

Scott Ferson
Applied Biomathematics, Setauket, NY 11733

George F. Corliss
Dept. of Electrical and Computer Engineering, Marquette University, Milwaukee, WI 53201

Mark A. Stadtherr
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

DOI 10.1002/aic.12278
Published online May 4, 2010 in Wiley Online Library (wileyonlinelibrary.com).

Dynamic process models frequently involve uncertain parameters and inputs. Propagating these uncertainties rigorously through a mathematical model to determine their effect on system states and outputs is a challenging problem. In this work, we describe a new approach, based on the use of Taylor model methods, for the rigorous propagation of uncertainties through nonlinear systems of ordinary differential equations (ODEs). We concentrate on uncertainties whose distribution is not known precisely, but can be bounded by a probability box (p-box), and show how to use p-boxes in the context of Taylor models. This allows us to obtain p-box representations of the uncertainties in the state variable outputs of a nonlinear ODE model. Examples having two to three uncertain parameters or initial states and focused on reaction process dynamics are used to demonstrate the potential of this approach. Using this method, rigorous probability bounds can be determined at a computational cost that is significantly less than that required by Monte Carlo analysis. © 2010 American Institute of Chemical Engineers *AIChE J*, 57: 404–422, 2011

Keywords: design (process simulation), mathematical modeling, numerical solutions, reactor analysis, bioprocess engineering

Introduction

Systems of ordinary differential equations (ODEs) are the basis for many mathematical models in engineering and science. For example, models of reactor dynamics are based on unsteady-state material and energy balances and thus take the form of a system of first-order ODEs, which typically is nonlinear. Generally, the problem of interest is an initial value problem (IVP), in which an initial state is given and the system then integrated numerically until some final time (time horizon) is reached, thus determining numerical approximations of the final state, as well as of the trajectory followed to reach it.

Often these dynamic models involve uncertainties in parameters and/or initial states. Analysis of the impact of such uncertainties is clearly important in models of process dynamics, as used, for example, in state and parameter estimation[1,2]

Correspondence concerning this article should be addressed to M. A. Stadtherr at markst@nd.edu.
Current Address of Youdong Lin: LINDO Systems, Inc., 1415 North Dayton St., Chicago, IL 60642.

PBA relaxes assumptions

- Everyone makes assumptions, but not all sets of assumptions are equal:

Linear	Normal	Independence
Montonic	Unimodal	Known correlation
Any function	***Any distribution***	***Any dependence***

- PBA doesn’t require unwarranted assumptions

Wishful thinking

Analysts often make convenient assumptions that are not really justified:

1. Variables are independent of one another
2. Uniform distributions model gross incertitude
3. Distributions are stationary (unchanging)
4. Distributions are perfectly precisely specified
5. Measurement uncertainty is negligible

You don’t have to think wishfully

A p-box can discharge a false assumption:

1. Don’t have to assume any dependence at all
2. An interval can be a better model of incertitude
3. P-boxes can enclose non-stationary distributions
4. Can handle imprecise specifications
5. Measurement data with plus-minus, censoring

Rigorousness

- "Automatically verified calculations"
- The computations are guaranteed to enclose the true results (so long as the inputs do)
- You can still be wrong, but the *method* won't be the reason if you are

Take-home messages

- Using bounding, you don’t have to pretend you know a lot to get quantitative results
- Probability bounds analysis bridges worst case and probabilistic analyses in a way that’s faithful to both and makes it suitable for use in early design

Acknowledgments

Larry Green, LaRC

Bill Oberkampf, Sandia

Chris Paredis, Georgia Tech

NASA

National Institutes of Health (NIH)

Electric Power Research Institute (EPRI)

Sandia National Laboratories

End