Conjugacy and priors

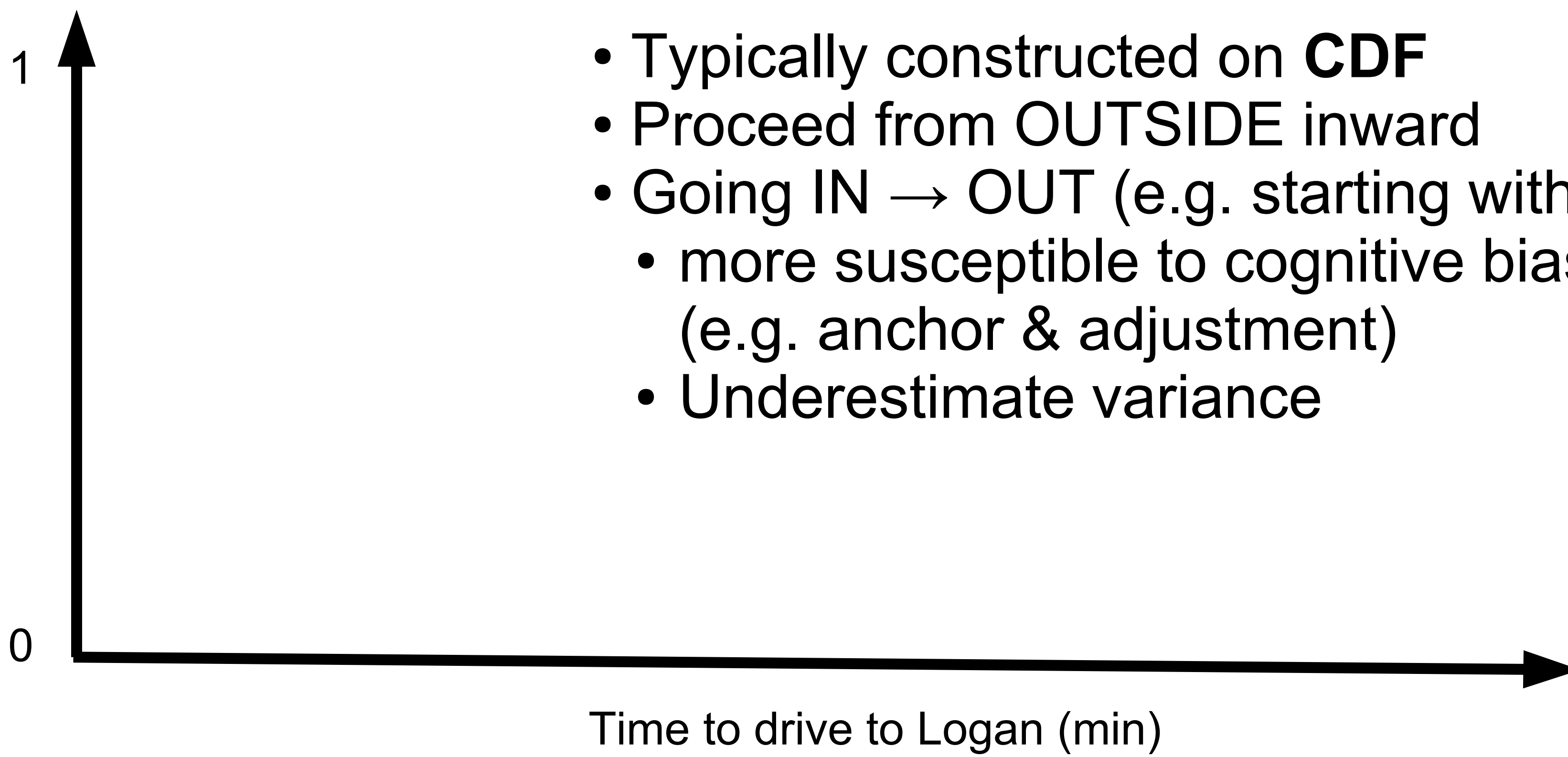
Overview

- Priors
- Conjugacy
- Non-conjugate distributions?
- Analytically-tractable vs Numerical Solutions

Priors

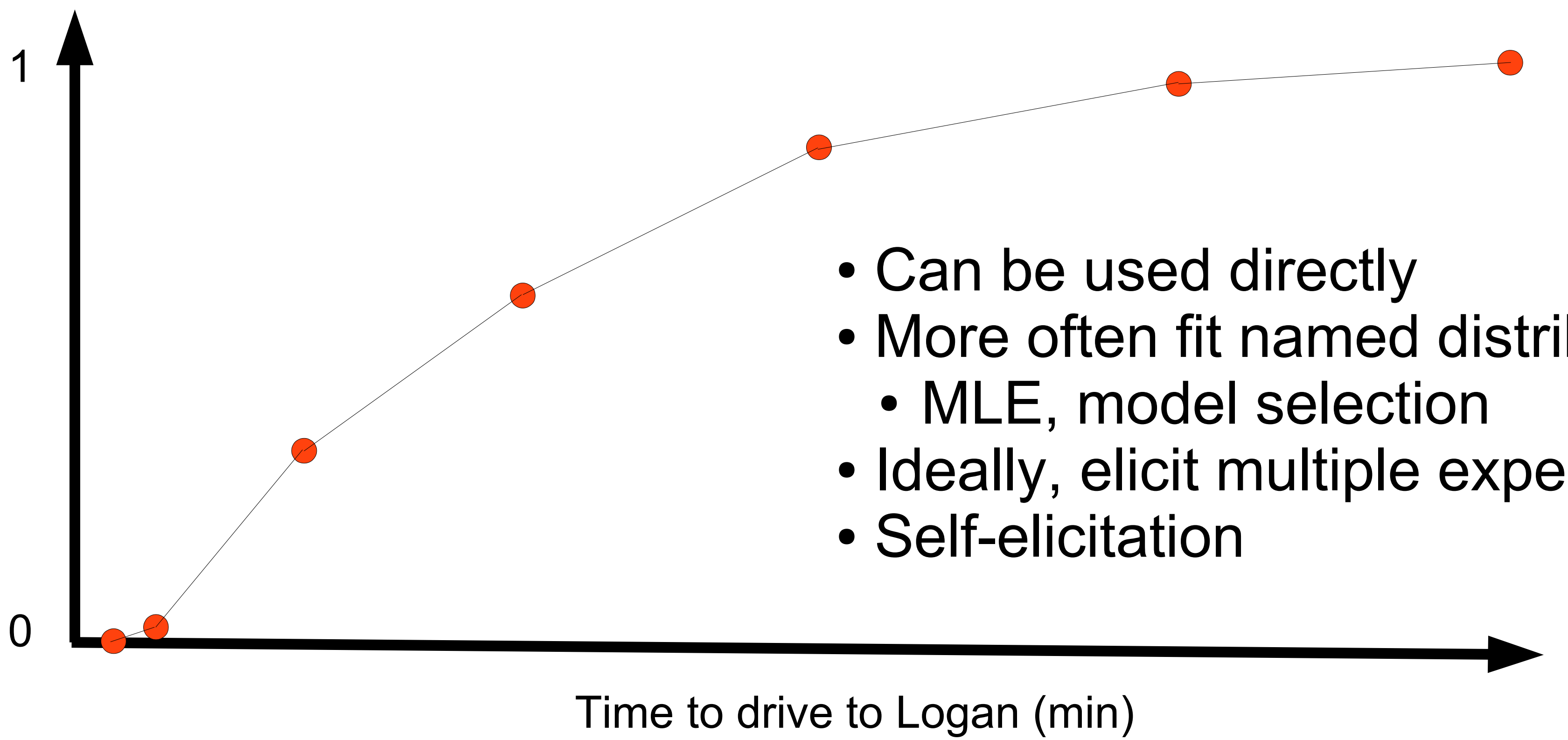
- Informative
  - Expert Elicitation

- Typically constructed on **CDF**
- Proceed from OUTSIDE inward
- Going IN → OUT (e.g. starting with mean)
  - more susceptible to cognitive biases (e.g. anchor & adjustment)
  - Underestimate variance



Priors

- Informative
  - Expert Elicitation



- Can be used directly
- More often fit named distributions
  - MLE, model selection
- Ideally, elicit multiple experts
- Self-elicitation

Priors

- Informative
- Uninformative
  - Proper
    - Integrate to 1
  - Improper
    - Does not have a finite integral
    - Posterior may or may not integrate to 1
    - Improper Posteriors
      - **Invalid**
      - **Hard to catch with numerical methods**
      - Source of most jokes among Bayesians

Examples

- Informative
  - N(0,10) Unif(-3,3) Beta(5,5)
- Uninformative but proper
  - $N(0,10^{32})$ $Unif(-10^{32},10^{32})$ Beta(1,1)
- Improper
  - $N(0,\infty)$ 1 Beta(0,0)

Beta-Binomial

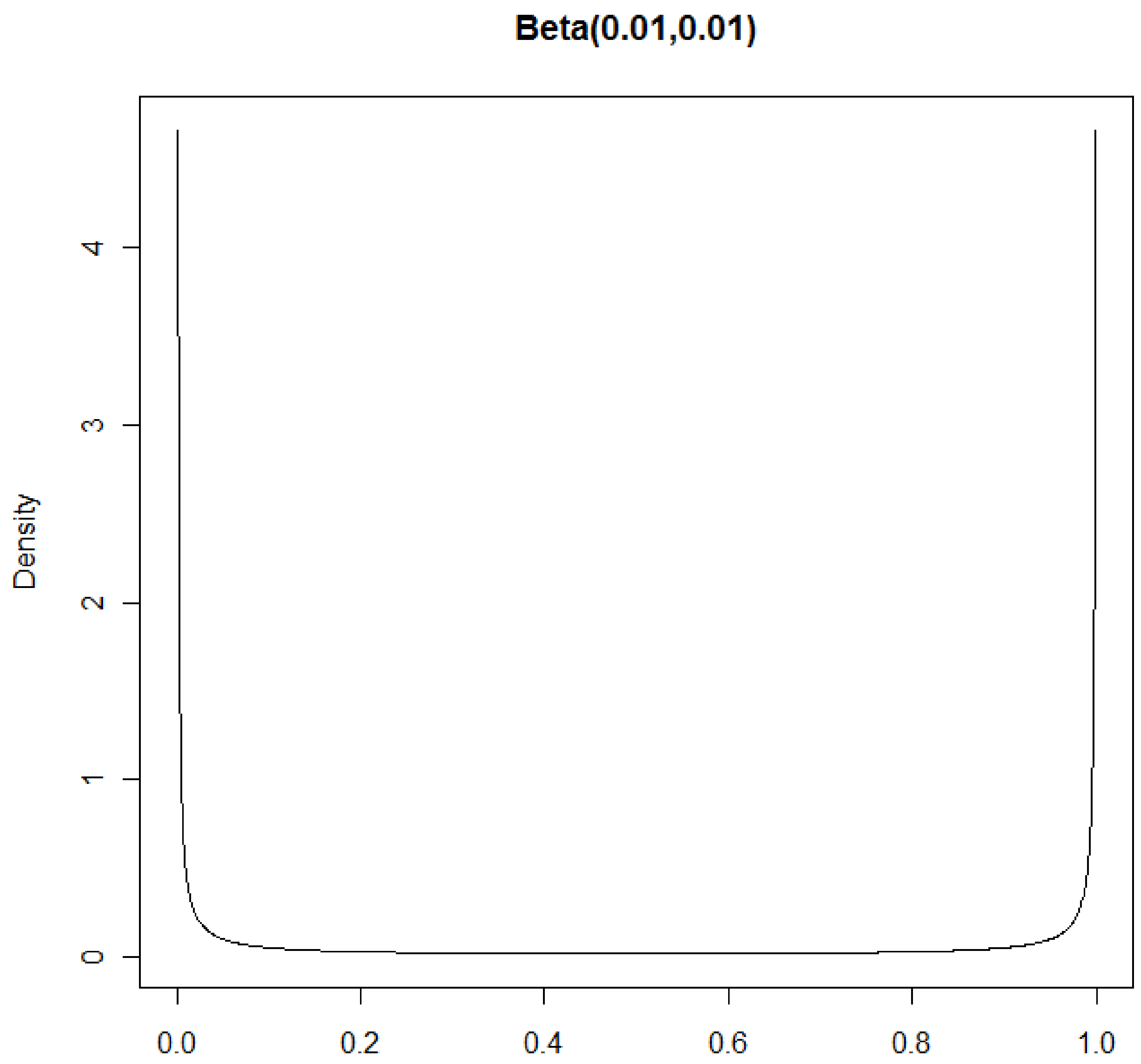
$$L = P(y|\theta) = \mathrm{Binom}(y|\theta, n)$$

$$Prior = P(\theta) = \mathrm{Beta}(\theta|y_{0,} n_0 - y_0)$$

$$P(\theta|y) = \mathrm{Beta}(\theta|y + y_{0,} n + n_0 - y - y_0)$$

$\mathrm{Beta}(y_0 = 0, n_0 - y_0 = 0)$ is improper

$\mathrm{Beta}(y_0 = 1, n_0 - y_0 = 1)$ is proper and flat
but equivalent to two observations


Beta(0.01,0.01)
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0

Normal-Inverse Gamma

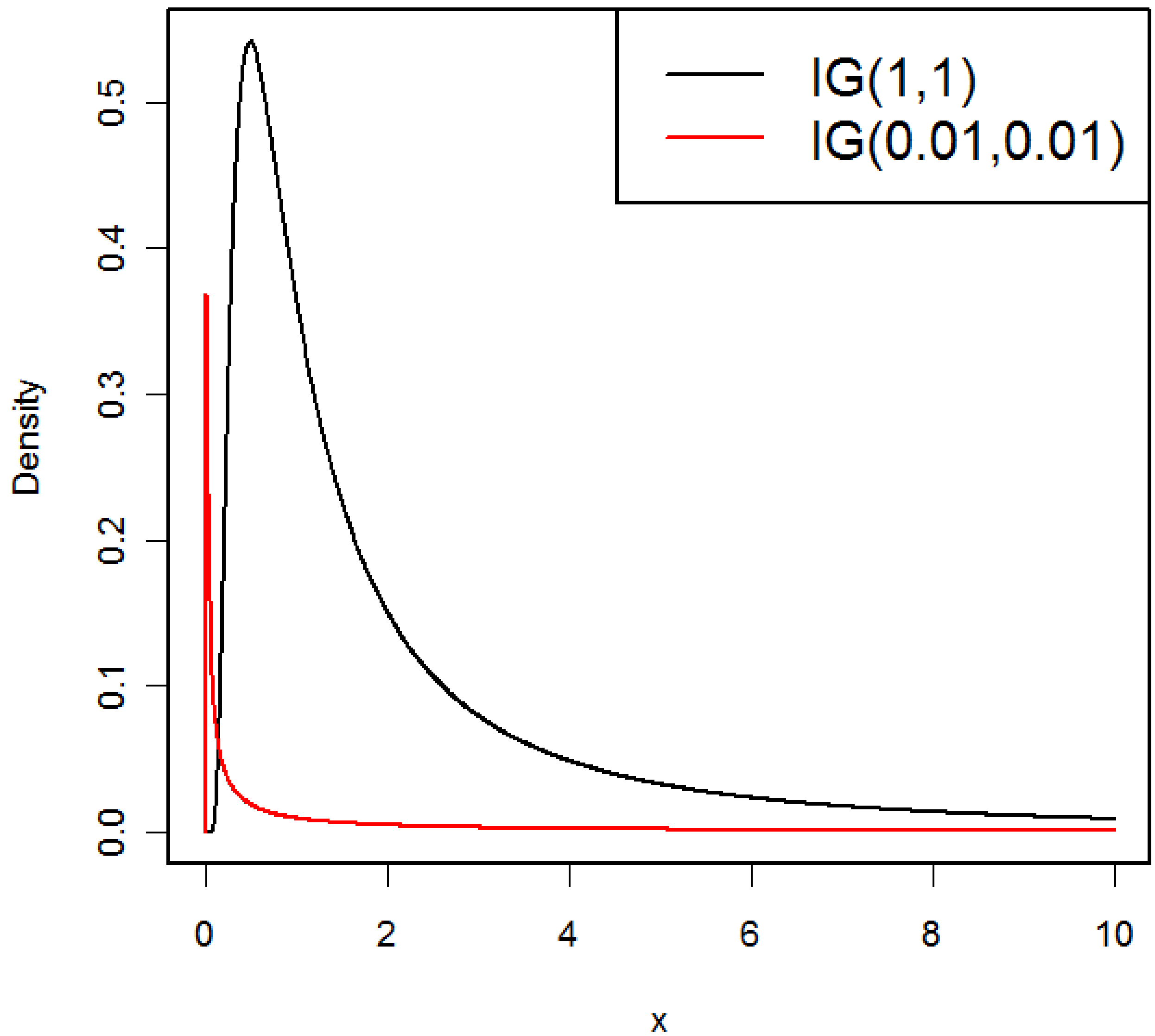
$$L = p(\vec{y}|\sigma^2) = N(\vec{y}|\mu, \sigma^2)$$

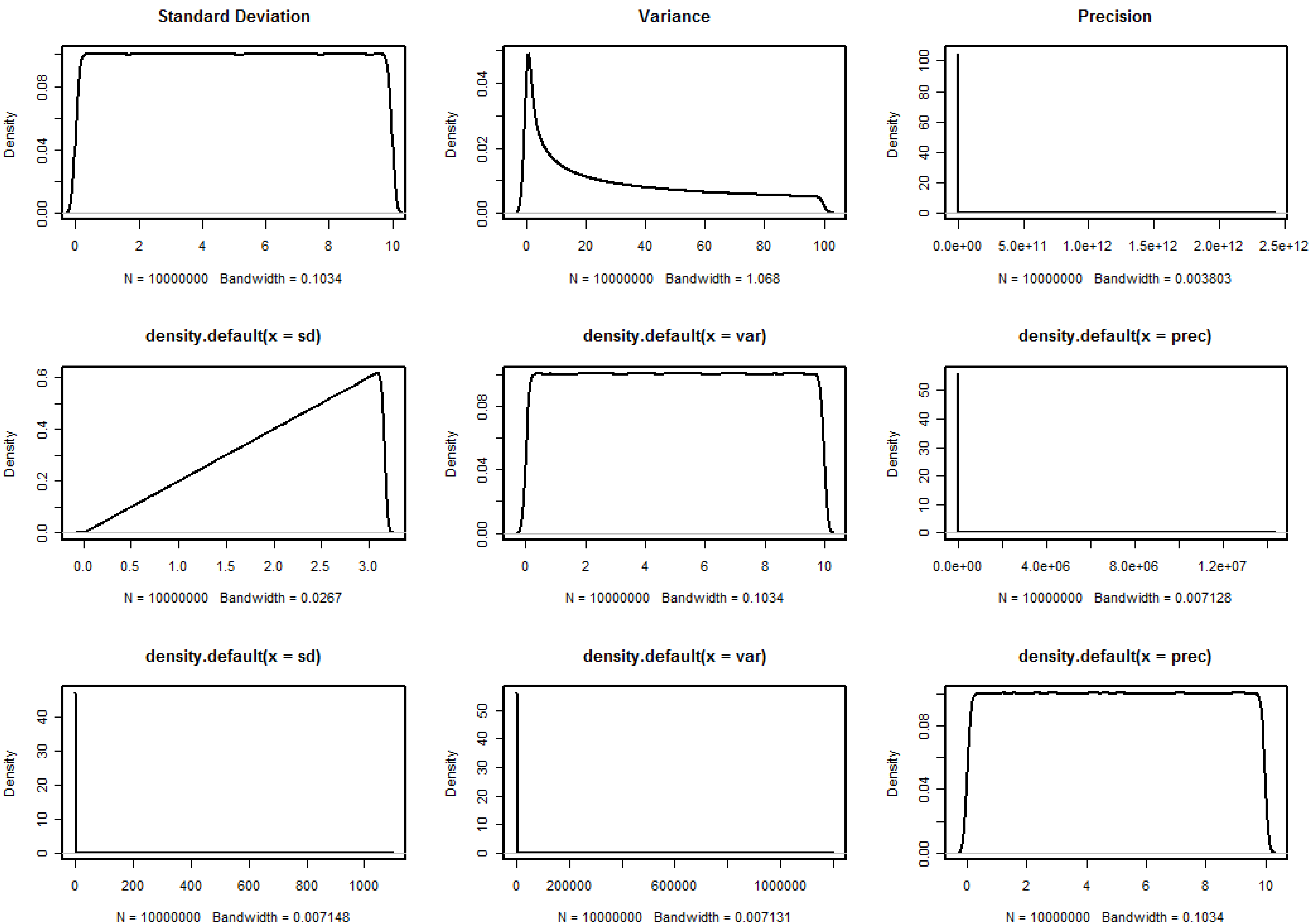
$$\text{prior} = p(\sigma^2) = IG(\sigma^2|\alpha, \beta)$$

$$p(\sigma^2|y) = IG\left(\sigma^2 \middle| \alpha + \frac{n}{2},\ \beta + \frac{1}{2}\sum (y_i - \mu)^2\right)$$

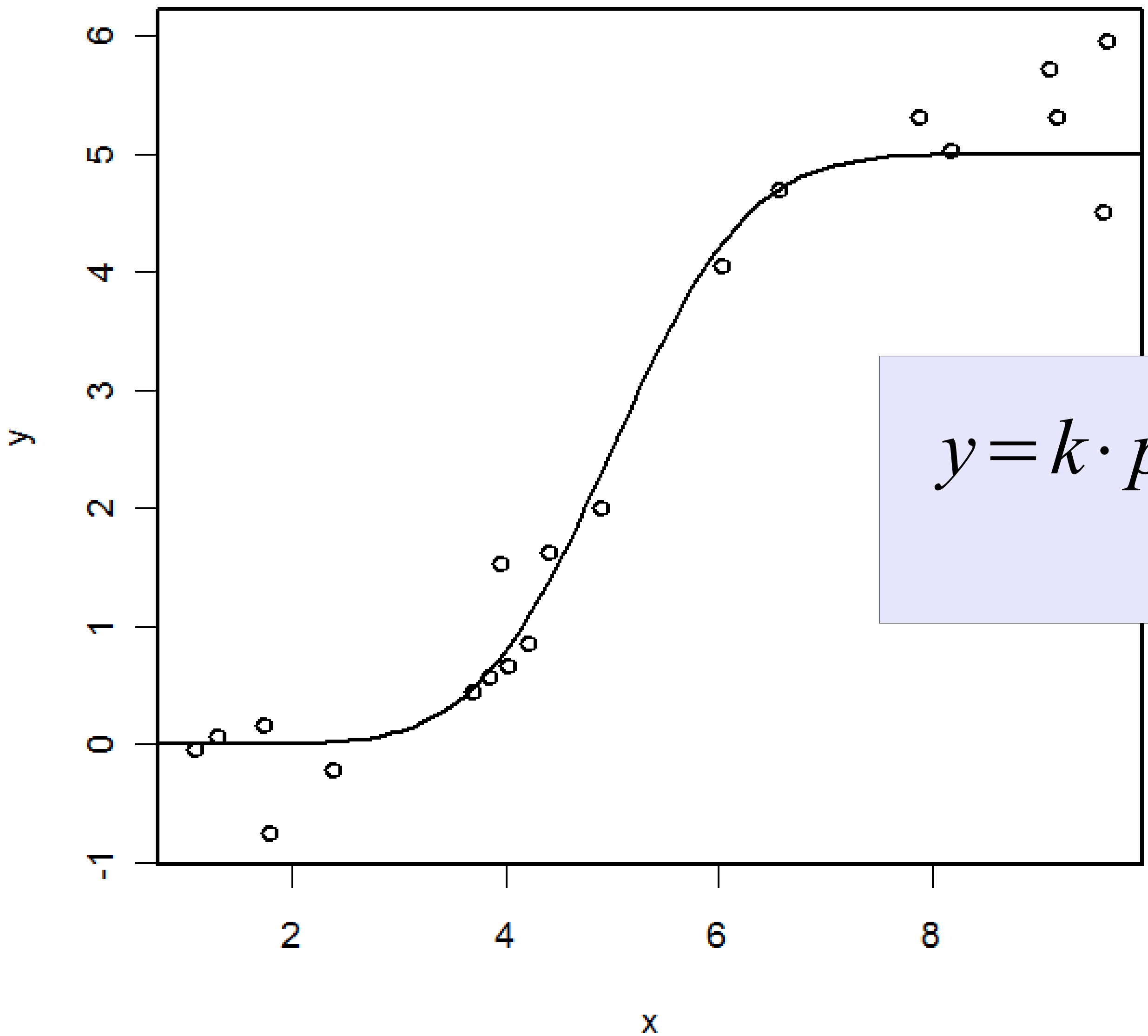
- IG(0,0) is an improper prior
- IG(1,1) is proper but equiv. to two observations

- Weak IG prior
  - Mode ≈ 0
  - Mean undef.
  - Dangerous if data is small n or small SS
  - Can occur when used as hyperpriors
  - UNITS!



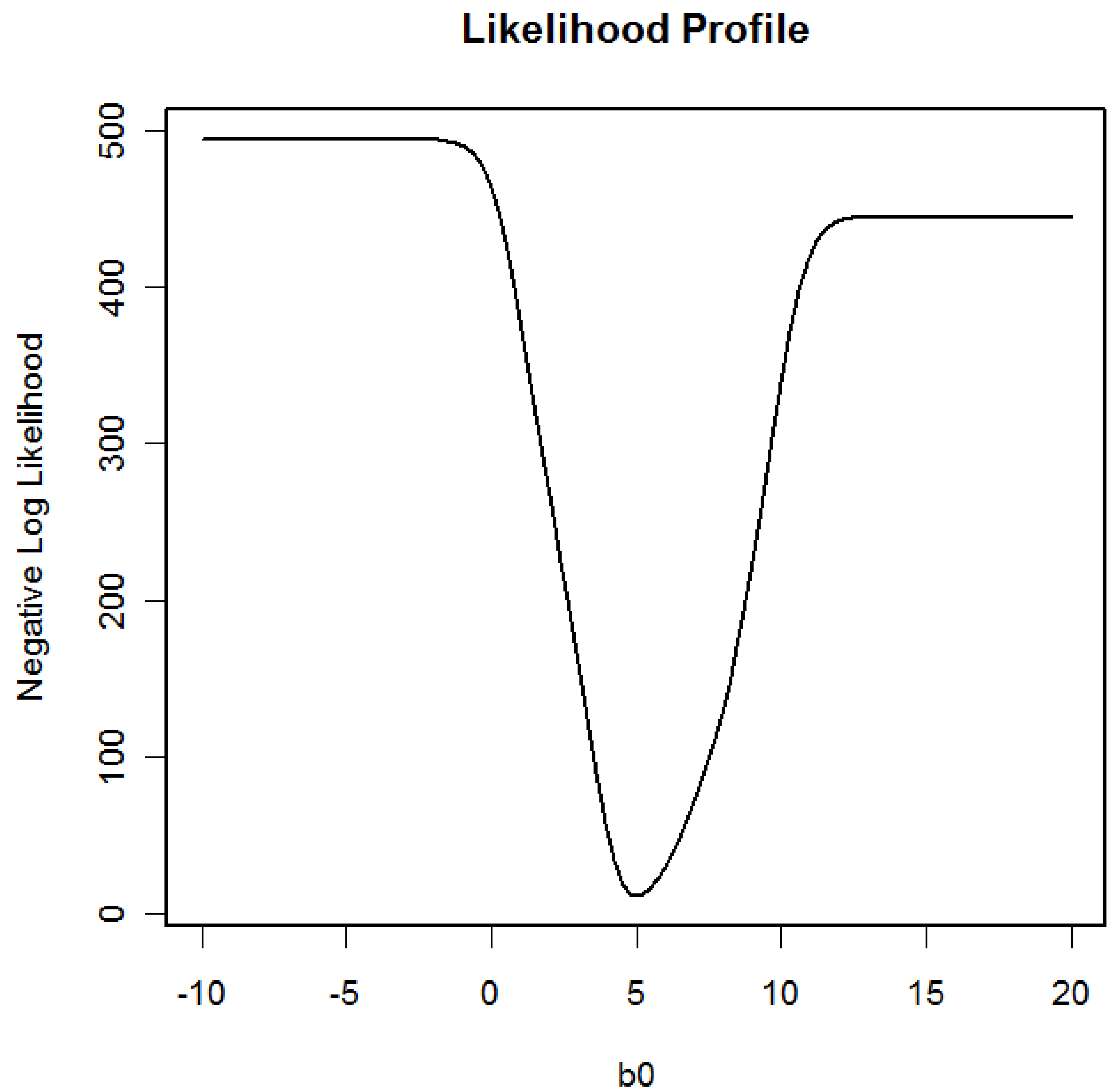

Standard Deviation
Density
0.00 0.04 0.08
0 2 4 6 8 10
N = 10000000 Bandwidth = 0.1034
Variance
Density
0.00 0.02 0.04
0 20 40 60 80 100
N = 10000000 Bandwidth = 1.068
Precision
Density
0 20 40 60 80 100
0.0e+00 5.0e+11 1.0e+12 1.5e+12 2.0e+12 2.5e+12
N = 10000000 Bandwidth = 0.003803
density.default(x = sd)
Density
0.0 0.2 0.4 0.6
0.0 0.5 1.0 1.5 2.0 2.5 3.0
N = 10000000 Bandwidth = 0.0267
density.default(x = var)
Density
0.00 0.04 0.08
0 2 4 6 8 10
N = 10000000 Bandwidth = 0.1034
density.default(x = prec)
Density
0 10 20 30 40 50
0.0e+00 4.0e+06 8.0e+06 1.2e+07
N = 10000000 Bandwidth = 0.007128
density.default(x = sd)
Density
0 10 20 30 40
0 200 400 600 800 1000
N = 10000000 Bandwidth = 0.007148
density.default(x = var)
Density
0 10 20 30 40 50
0 200000 600000 1000000
N = 10000000 Bandwidth = 0.007131
density.default(x = prec)
Density
0.00 0.04 0.08
0 2 4 6 8 10
N = 10000000 Bandwidth = 0.1034

Improper Posteriors



$$y = k \cdot probit(x, b0, b1) + \epsilon$$
$$\epsilon \sim N(0, \sigma^2)$$

- Likelihood integrates to infinity
- Proper prior required to ensure proper posterior



Conjugacy

- Posterior same distribution as prior
  - Beta - Binomial
  - Normal – Normal
  - Normal – Inverse Gamma
- Allowed us to “cheat” on math by matching terms
- Analytically-tractable (though only sometime conditionally)

Other Conjugate Distributions

- Discrete
  - Poisson – Gamma
  - Negative Binomial – Beta
  - Multinomial - Dirichlet
- Continuous
  - Exponential – Gamma
  - Gamma – Gamma
  - Normal – Gamma
  - Normal – Inverse Chi-Square
  - Multivariate Normal - Wishart

Poisson - Gamma

$$L = p(\vec{y}|\lambda) = Pois(\vec{y}|\lambda) \propto \lambda^{\sum y} \exp(-n\lambda)$$

$$\text{prior} = p(\lambda) = Gamma(\lambda|\alpha, \beta) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

$$p(\lambda|y) = Gamma\left(\lambda\middle|\alpha + \sum y\ ,\ \beta + n\right)$$

What about non-conjugate priors?

$$L = p(\vec{y}|\lambda) = logN(\vec{y}|\mu, \sigma^2) \propto \frac{1}{y} \exp\left(-\frac{(\ln(y) - \mu)^2}{2\sigma^2} \right)$$

$$\text{prior} = p(\mu) = Gamma(\mu|\alpha, \beta) \propto \mu^{\alpha - 1} \exp(-\beta\mu)$$

$$p(\mu|y) = \frac{logN(\vec{y}|\mu, \sigma^2)\, Gamma(\mu|\alpha, \beta)}{\int_{-\infty}^{\infty} logN(\vec{y}|\mu, \sigma^2)\, Gamma(\mu|\alpha, \beta)\, d\mu}$$

$$p(\mu|y) = \frac{logN(\vec{y}|\mu,\sigma^2)\,Gamma(\mu|\alpha,\beta)}{\int_{-\infty}^{\infty} logN(\vec{y}|\mu,\sigma^2)\,Gamma(\mu|\alpha,\beta)\,d\mu}$$

$$p(\mu|y) = \frac{\exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right)\cdot\exp(-\beta\mu)\cdot\mu^{\alpha-1}}{\int_{-\infty}^{\infty}\exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}-\beta\mu\right)\mu^{\alpha-1}\,d\mu}$$

$$p(\mu|y)=\frac{\exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right)\cdot\exp(-\beta\mu)\cdot\mu^{\alpha-1}}{\int_{-\infty}^{\infty}\exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}-\beta\mu\right)\mu^{\alpha-1}d\mu}$$

- Does not match any known/standard distribution
- Integral of denominator *very* daunting!
- How to proceed numerically???