Bayes: Part II

Bayes' Theorem

Posterior — Likelihood — Prior

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)}$$

$$= \frac{P(y|\theta)\,P(\theta)}{\int_{-\infty}^{\infty} P(y|\theta)\,P(\theta)\,d\theta}$$

Review

- Bayes Rule
- False Positives example
- XKCD example
- Bayes' Billiard table
- Posterior
  - Is a PDF
  - Is updatable
- Priors = external information

Bayes vs Max Likelihood

	ML	Bayes
data	random	fixed
parameters	fixed	random

Random as in “has a PDF”
Not as in a random draw from a PDF

Priors

- Makes it possible to calculate a posterior density of the model parameter rather than the likelihood of the data
- Provides a way of incorporating information that is external to the data set(s) at hand
- Inherently sequential

Previous Posterior = New Prior

Where do Priors come from?

- Uninformative / vague
  - Ch… nt, allows the…
- Prev…
  - Mu…
  - Va…
- “The…
  - Me…
- Expert knowledge

Prior specification must be “blind” to the data in the analysis!!

No “double dipping” -- leads to falsely overconfident results

How do I choose a prior PDF?

- Analogous to how we choose the data model
  - Range restrictions, shape, etc.
- Conjugacy
  - A prior is conjugate to the likelihood if the posterior PDF is in the same family as the prior
  - Allow for closed-form analytical solutions to either full posterior or (in multiparameter models) for the conditional distribution of that parameter.
  - Modern computational methods no longer require conjugacy

Example: Tree mortality rate

- Data: observed n=4 trees, y=1 died this year

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

- Prior: last year observed n_0=2 trees, y_0=1 died

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

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

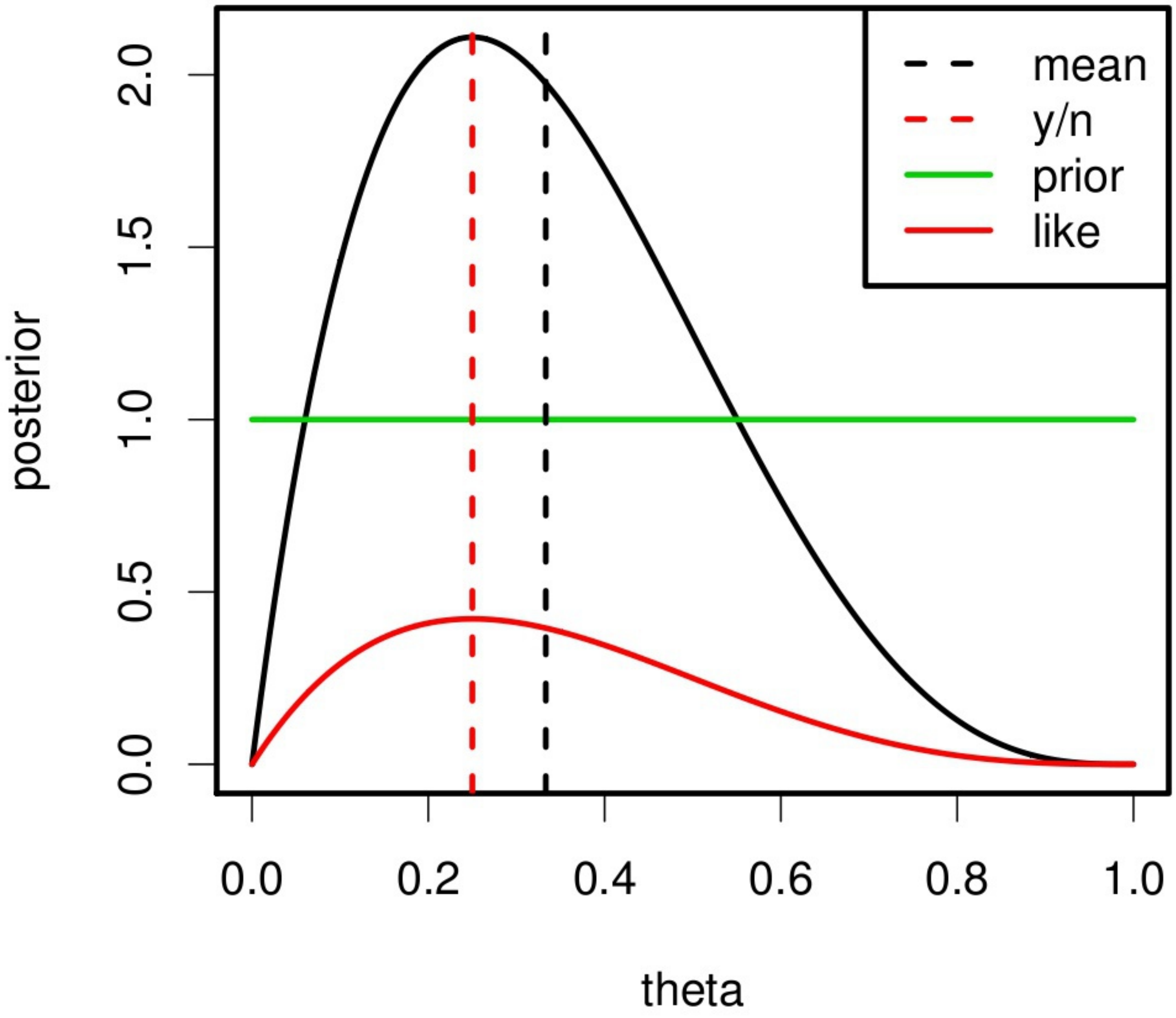
$$P(\theta|y) \propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y} \times \frac{\theta^{y_0-1}(1-\theta)^{n_0-y_0-1}}{\mathrm{B}(y_{0,} n_0 - y_0)}$$

$$P(\theta|y) \propto \theta^{y+y_0-1}(1-\theta)^{n-y+n_0-y_0-1}$$

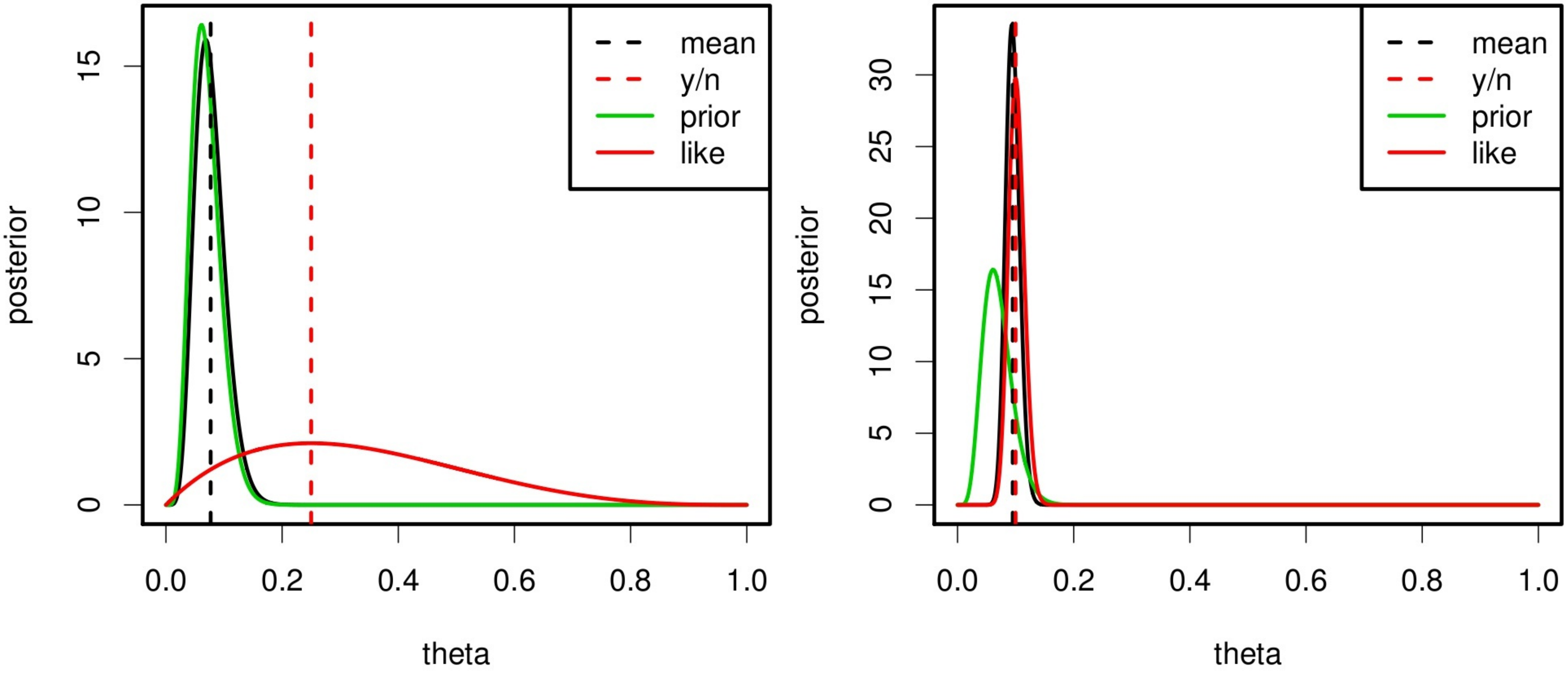
$$P(\theta|y) = \mathrm{Beta}(\theta|y+y_{0,} n+n_0-y-y_0)$$

Beta-Binomial Model

$$P(\theta|y) = \text{Beta}(\theta|y + y_{0,} n + n_0 - y - y_0)$$
$$= \text{Beta}(\theta|2,4)$$



How much impact does the prior have on the analysis?



General pattern

- Write down likelihood
  - Data model
  - Process model
- Write down priors for ALL free parameters in the likelihood
  - Parameter model
- Solve for posterior distribution of model parameters

Example: Normal mean and variance

- First case: Mean with known variance

$$L = p(y|\mu) = N(y|\mu, \sigma^2) \propto \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]$$

$$\text{prior} = p(\mu) = N(\mu|\mu_0, \tau^2) \propto \exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

Prior Mean

Prior Variance

$$p(\mu|y) \propto p(y|\mu)\, p(\mu) = N(y|\mu, \sigma^2) \cdot N(\mu|\mu_{0,} \tau^2)$$

$$\propto \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right] \cdot \exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{(y-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau^2}\right)\right]$$

$$p(\mu|y) \propto \exp\left[-\frac{1}{2}\left(\frac{(y-\mu)^2}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\tau^2}\right)\right]$$

Looks like a Normal PDF

$$p(\mu|y) \propto \exp\left[\frac{-(\mu-Vv)^2}{2V}\right]$$

$$p(\mu|y) \propto \exp\left[\frac{-(\mu - Vv)^2}{2\text{V}}\right]$$

Can be expanded to...

$$\exp\left[\frac{-\mu^2 - 2\mu V v + V^2 v^2}{2\text{V}}\right]$$

$$\exp\left[\frac{-1}{2\text{V}}\mu^2 - v\mu + \frac{1}{2}V v^2\right]$$

$$p(\mu|y) \propto \exp\left[-\frac{1}{2}\left(\frac{(y-\mu)^2}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\tau^2}\right)\right]$$

Can be expanded to...

$$\exp\left[-\frac{1}{2}\left(\frac{y^2-2y\mu+\mu^2}{\sigma^2}+\frac{\mu^2-2\mu\mu_0+\mu_0^2}{\tau^2}\right)\right]$$

$$\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2+\left(\frac{y}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)\mu-\frac{y^2}{2\sigma^2}-\frac{\mu_0^2}{2\tau^2}\right]$$

$$\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)\mu^2+\left(\frac{y}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)\mu-\frac{y^2}{2\sigma^2}-\frac{\mu_0^2}{2\tau^2}\right]$$

$$\exp\left[\frac{-1}{2V}\mu^2-\nu\mu+\frac{1}{2}V\nu^2\right]$$

$$\frac{1}{V}=\frac{1}{\sigma^2}+\frac{1}{\tau^2}$$

$$\nu=\frac{y}{\sigma^2}+\frac{\mu_0}{\tau^2}$$

$$p(\mu | y) = N(\mu | Vv, V)$$

$$= N\left(\mu \;\middle|\; \frac{\left(\frac{y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)} \;,\; \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)} \right)$$

Precision weighted average of data and prior

Precision = 1/variance

Second case: Multiple data points

$$p(\mu|y) \propto \prod_{i=1}^{n} N(y_i|\mu, \sigma^2) \cdot N(\mu|\mu_0, \tau^2)$$

$$= \exp\left[-\frac{1}{2}\left(\frac{\sum_{i=1}^{n}(y_i - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\tau^2}\right)\right]$$

$$= \exp\left[-\frac{1}{2} \left(\frac{\sum_{i=1}^{n} (y_i - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\tau^2} \right) \right]$$

$$\frac{1}{V} = \sum \frac{1}{\sigma^2} + \frac{1}{\tau^2} = \frac{n}{\sigma^2} + \frac{1}{\tau^2}$$

$$v = \sum \frac{y_i}{\sigma^2} + \frac{\mu_0}{\tau^2} = \frac{n}{\sigma^2} \bar{y} + \frac{\mu_0}{\tau^2}$$

$$p(\mu|y) = N(\mu|Vv, V)$$

$$\frac{1}{V} = \frac{n}{\sigma^2} + \frac{1}{\tau^2}$$

$$\lim_{n \to \infty} Vv = \bar{y}$$

$$v = \frac{n}{\sigma^2}\bar{y} + \frac{\mu_0}{\tau^2}$$

$$\lim_{n \to \infty} V = \frac{\sigma^2}{n}$$

Example

- Assume a prior that is $N(\mu | \mu_0 = 90, \tau^2 = 2)$
- Compare fits of two data sets that have the same sample mean $\bar{Y}$ = 100 and variance = 10 but vary in sample size: n = 5 vs n = 20

n = 5

$$\frac{1}{V} = \frac{n}{\sigma^2} + \frac{1}{\tau^2} = \frac{5}{10} + \frac{1}{2} = 1$$

$$v = \frac{n}{\sigma^2}\bar{y} + \frac{\mu_0}{\tau^2} = \frac{5}{10}\cdot 100 + \frac{1}{2}\cdot 90 = 95$$

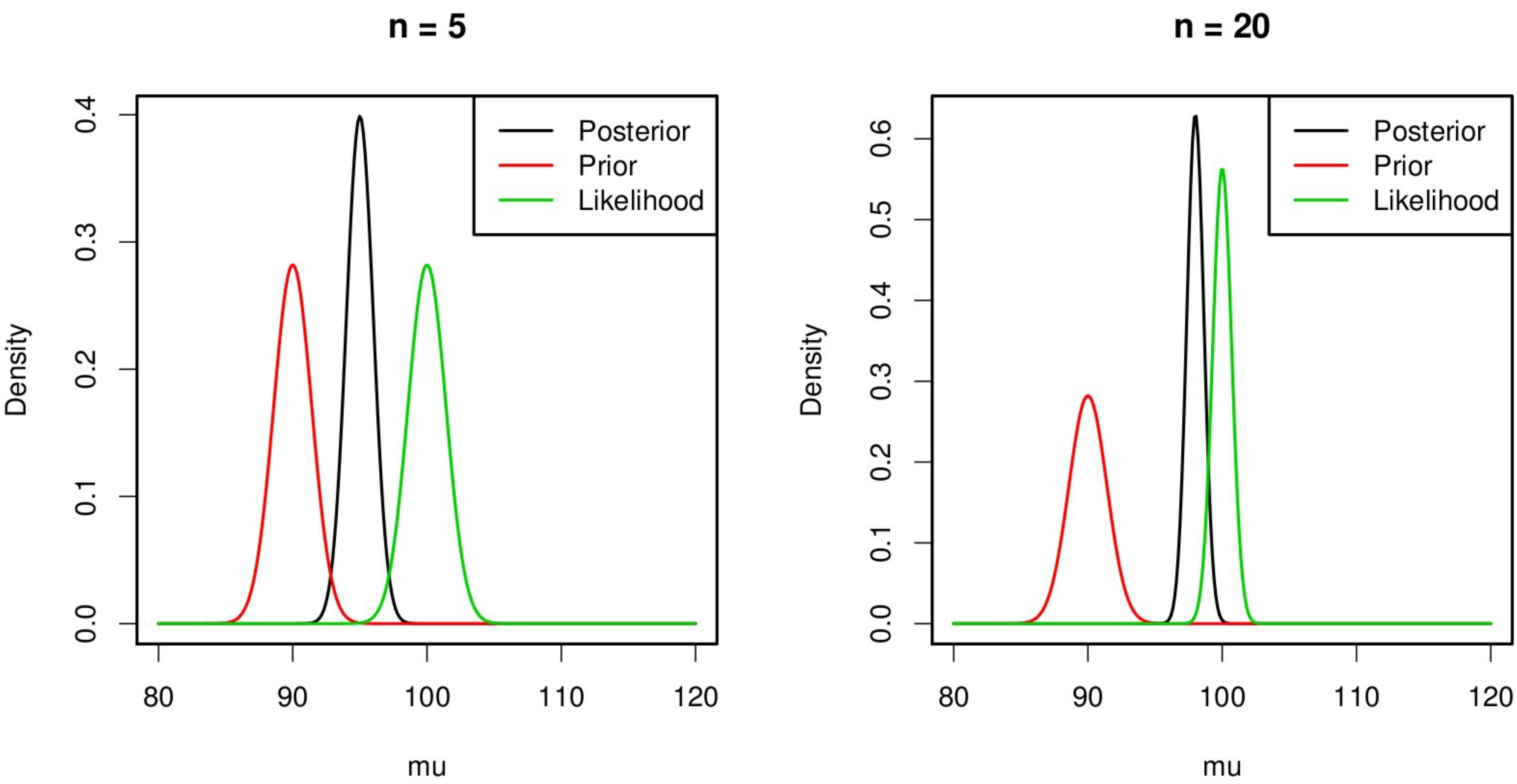
$$p(\mu|y) = N(\mu|95,1)$$

n = 20

$$\frac{1}{V} = \frac{n}{\sigma^2} + \frac{1}{\tau^2} = \frac{20}{10} + \frac{1}{2} = 2.5$$

$$v = \frac{n}{\sigma^2}\bar{y} + \frac{\mu_0}{\tau^2} = \frac{20}{10} \cdot 100 + \frac{1}{2} \cdot 90 = 245$$

$$p(\mu|y) = N(\mu|98, 0.4)$$


n = 5
Posterior
Prior
Likelihood
Density
0.0
0.1
0.2
0.3
0.4
80
90
100
110
120
mu
n = 20
Posterior
Prior
Likelihood
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
80
90
100
110
120
mu

Third case: Known mean, unknown variance

$$L = p(\vec{y}|\sigma^2) = N(\vec{y}|\mu,\sigma^2) \propto \prod_{i=1}^{n} (\sigma^2)^{\frac{-1}{2}} \exp\left[\frac{-(y_i-\mu)^2}{2\sigma^2}\right]$$

$$\text{prior} = p(\sigma^2) = IG(\sigma^2|\alpha,\beta) \propto (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right)$$

- Inverse Gamma commonly chosen because of conjugacy

$$p(\sigma^2|\vec{y}) \propto p(\vec{y}|\sigma^2)\, p(\sigma^2) = N(\vec{y}|\mu, \sigma^2) \cdot IG(\sigma^2|\alpha, \beta)$$

$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left[\frac{-\sum (y_i - \mu)^2}{2\sigma^2}\right] \cdot (\sigma^2)^{-(\alpha+1)} \exp\left[\frac{-\beta}{\sigma^2}\right]$$

$$= (\sigma^2)^{-(\frac{n}{2}+\alpha+1)} \exp\left[-\frac{1}{\sigma^2}\left(\beta + \frac{1}{2}\sum (y_i - \mu)^2\right)\right]$$

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

Looks like an Inverse Gamma PDF

$$IG(\sigma^2|a,b) \propto (\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right)$$

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

Sample size — Sum of Squares

Example

- Assume a prior that is $IG(\sigma^2|\alpha=1,\beta=250)$
- Compare fits of two data sets that have the same mean = 0 and same sample variance = 100 but vary in sample size: n = 5 vs n = 20
- This would give a sum of squares of 500 and 2000 respectively

n=5

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

$$= IG\left(\sigma^2 \middle| 1 + \frac{5}{2},\ 100 + \frac{1}{2}500\right)$$

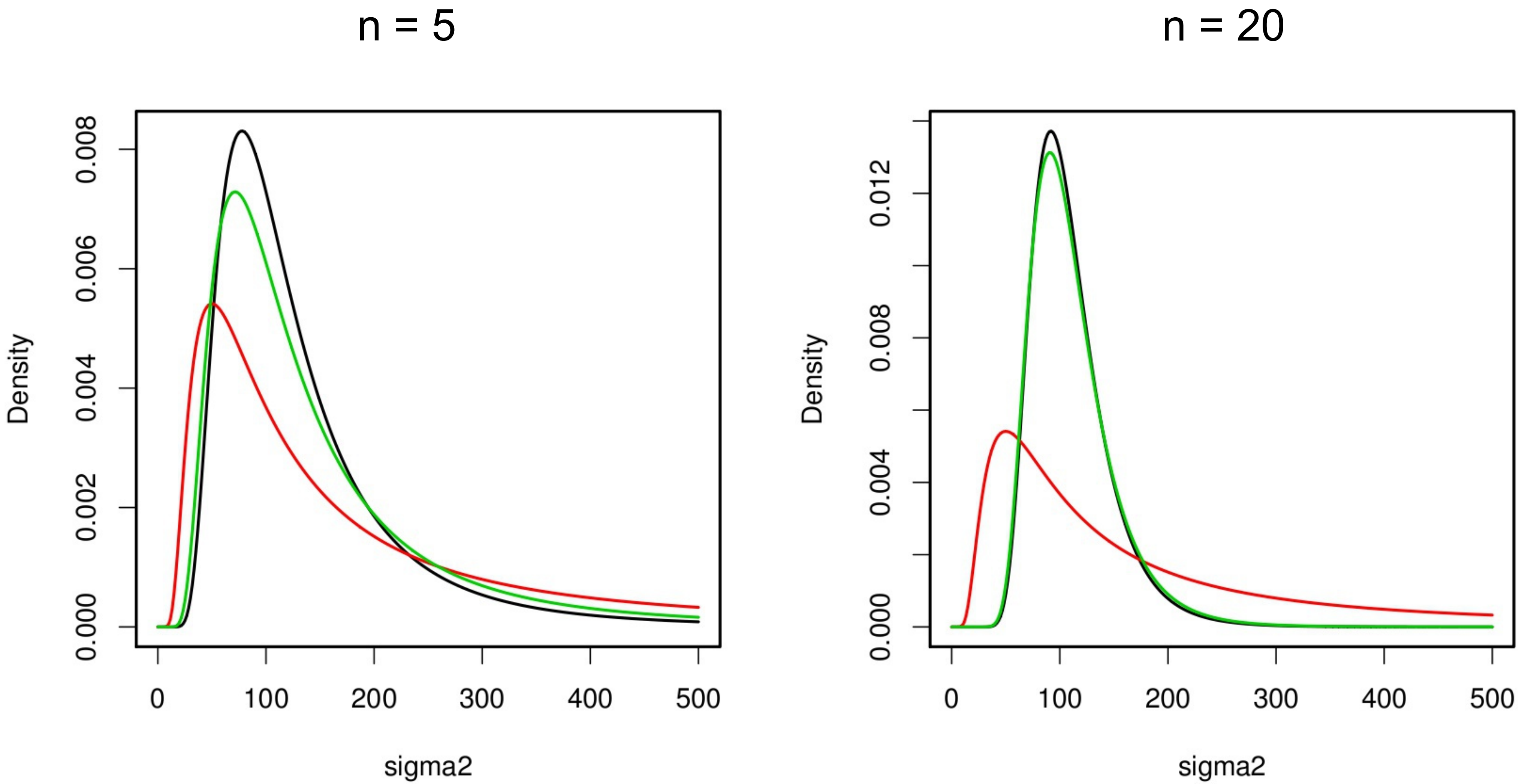
$$= IG\left(\sigma^2 \middle| 3.5,\ 350\right)$$

n=20

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

$$= IG\left(\sigma^2 \middle| 1 + \frac{20}{2},\ 100 + \frac{1}{2}2000\right)$$

$$= IG\left(\sigma^2 \middle| 11,\ 1100\right)$$



n = 5
Density
0.000
0.002
0.004
0.006
0.008
0
100
200
300
400
500
sigma2
n = 20
Density
0.000
0.004
0.008
0.012
0
100
200
300
400
500
sigma2

General pattern

- Write down likelihood
  - Data model
  - Process model
- Write down priors for the free parameters in the likelihood
  - Parameter model
- Solve for posterior distribution of model parameters