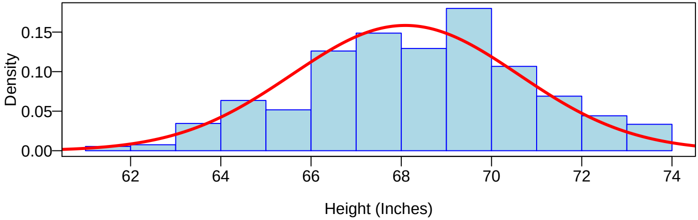
Biostatistics 682: Applied Bayesian Inference
Lecture 3: Normal Model

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Children's Height



- *In 1886, Francis Galton collected data in a table, showing a cross-tabulation of 928 adult children born to 205 fathers and mothers, by their height and their mid-parent's height. Let us first focus on the children's height data. The sample mean is 68.09 inches and the sample standard deviation is 2.62 inches*
- What is the sampling distribution of data?
- Why we can assume that?

Normal model

- The normal distribution is fundamental to most statistical modeling.
- The central limit theorem helps to justify using the normal likelihood in many statistical problems, as an approximation to a less analytically convenient actual likelihood.
- Even when the normal distribution does not itself provide a good model fit, it can be useful as a component of a more complicated model involving t or finite mixture distributions. We will see it in the future lectures.

Estimating a normal mean with known variance

- $y_i \mid \mu \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2), i = 1, \ldots, n$ with σ^2 known. Let $y = (y_1, \ldots, y_n)$.
- What is the likelihood?

$$\pi(y \mid \mu) = (2\pi\sigma^2)^{n/2} \exp\left\{-(2\sigma^2)^{-1} \sum_{i=1}^{n} (y_i - \mu)^2\right\}$$

- What prior models should we consider?

Conjugate prior distributions

If $\mathcal{F}$ is a class of sampling distributions $\pi(y \mid \theta)$, and $\mathcal{P}$ is a class of prior distributions for θ, then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ is

$$\pi(\theta \mid y) \in \mathcal{P}, \text{ for all } \pi(\cdot \mid \theta) \in \mathcal{F} \text{ and } \pi(\cdot) \in \mathcal{P}.$$

- Is beta distribution conjugate for binomial distribution?
- What distributions are conjugate for normal distribution?

The natural conjugate prior families: $\mathcal{P}$ is the set of all densities having the same functional form as the likelihood.

Natural conjugate prior for normal model

- We can write

$$\pi(y \mid u) \propto \exp\{a(y)\mu^2 + b(y)\mu + c(y)\}$$

$$\begin{aligned}
a(y) &= -n(2\sigma^2)^{-1} \\
b(y) &= \sigma^{-2}\sum_{i=1}^{n} y_i = n\sigma^{-2}\bar{y} \\
c(y) &= -(2\sigma^2)^{-1}\sum_{i=1}^{n} y_i^2
\end{aligned}$$

- To specify the natural conjugate prior, we set

$$\pi(\mu) \propto \exp\left\{\tilde{a}\mu^2 + \tilde{b}\mu + \tilde{c}\right\} = \exp\left\{-\frac{1}{2\tau^2}(\mu - \zeta)^2\right\} \Rightarrow \mu \sim N(\zeta, \tau^2).$$

- This further implies that

$$\pi(\mu \mid y) \sim \mathrm{N}(\tilde{\zeta}, \tilde{\tau}^2)$$

$$\begin{aligned}
\tilde{\zeta} = E(\mu \mid y) &= (n\sigma^{-2} + \tau^{-2})^{-1}(n\sigma^{-2}\bar{y} + \tau^{-2}\zeta) \\
\tilde{\tau}^2 = \mathrm{Var}(\mu \mid y) &= (n\sigma^{-2} + \tau^{-2})^{-1}
\end{aligned}$$

Limiting case

- Posterior mean is a weighted average of the prior mean and the sample mean

$$\tilde{\zeta} = \frac{n\sigma^{-2}}{n\sigma^{-2} + \tau^{-2}}\bar{y} + \frac{\tau^{-2}}{n\sigma^{-2} + \tau^{-2}}\zeta.$$

- Posterior mean is prior mean adjusted toward the sample mean

$$\tilde{\zeta} \quad = \zeta + \frac{n\sigma^{-2}}{n\sigma^{-2} + \tau^{-2}}(\bar{y} - \zeta)$$

- Posterior mean is sample mean that are shrunk toward the prior mean

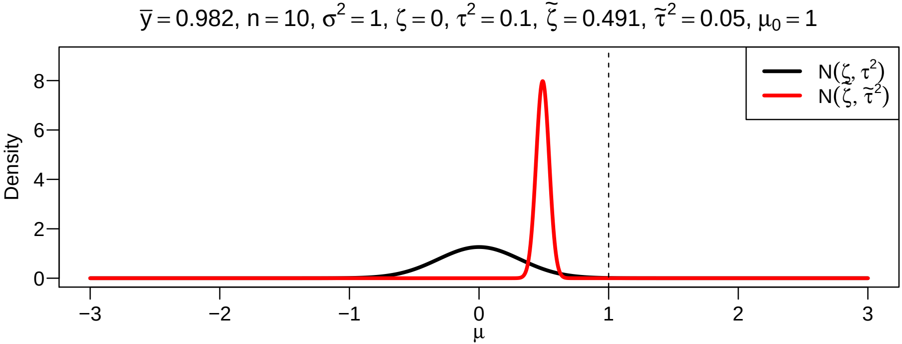
$$\tilde{\zeta} \quad = \bar{y} - \frac{\tau^{-2}}{n\sigma^{-2} + \tau^{-2}}(\bar{y} - \zeta)$$

- A non-informative prior when $\tau^2 \to \infty$. We write $\pi(\mu) \propto 1$.

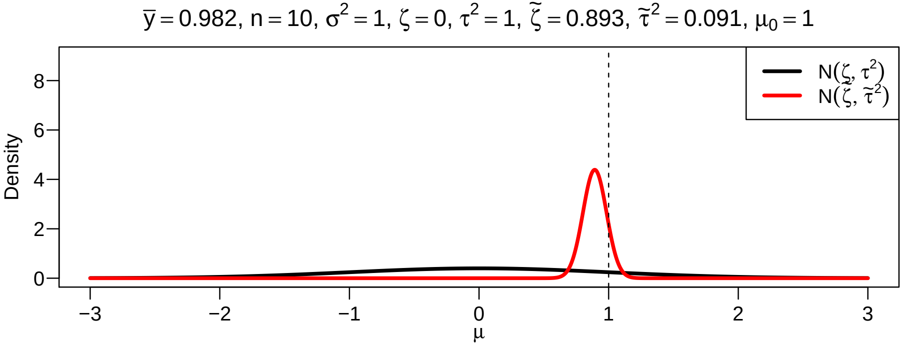
$$\lim_{\tau^2 \to \infty} \tilde{\zeta} = \bar{y}, \qquad \lim_{\tau^2 \to \infty} \tilde{\tau}^2 = \frac{\sigma^2}{n}$$

It is the MLE!

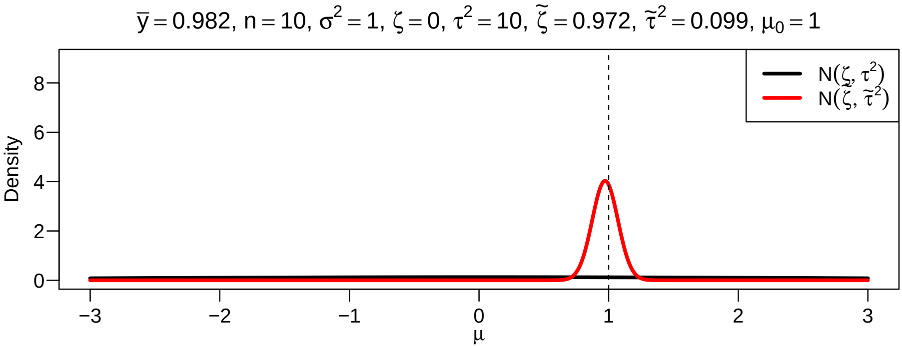
Limiting case: $\tau^2 \to \infty$



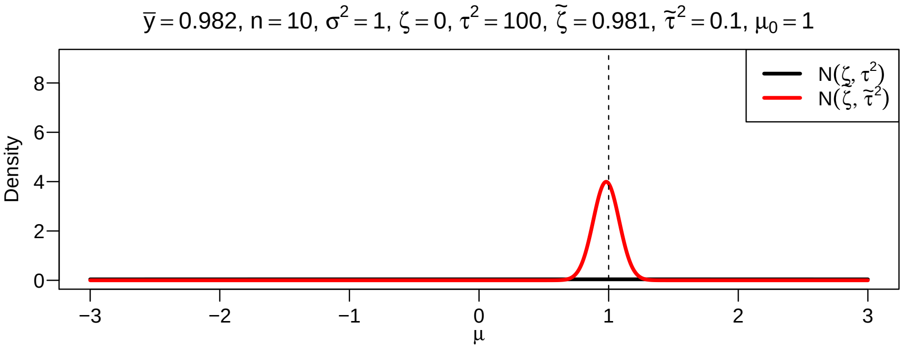
Limiting case: $\tau^2 \to \infty$



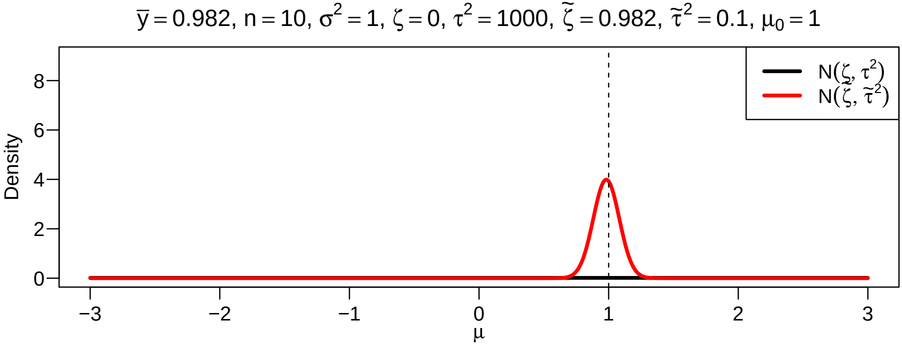
Limiting case: $\tau^2 \to \infty$



Limiting case: $\tau^2 \to \infty$



Limiting case: $\tau^2 \to \infty$



Predictions

- Prior predictive distribution:

$$\pi(\tilde{y}) = \int_{-\infty}^{\infty} \pi(\tilde{y} \mid \mu)\pi(\mu)d\mu = \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\left\{-\frac{(\tilde{y} - \zeta)^2}{2(\sigma^2 + \tau^2)}\right\}$$

- Posterior predictive distribution:

$$\pi(\tilde{y} \mid y) = \int_{-\infty}^{\infty} \pi(\tilde{y} \mid \mu)\pi(\mu \mid y)dy = \frac{1}{\sqrt{2\pi(\sigma^2 + \tilde{\tau}^2)}} \exp\left\{-\frac{(\tilde{y} - \tilde{\zeta})^2}{2(\sigma^2 + \tilde{\tau}^2)}\right\},$$

  where

- Do we have a simpler way to figure out those distributions?
  - The density of $(\tilde{y}, \mu)^{\mathrm{T}}$ must be bivariate normal and so the marginal density of $\tilde{y} \mid y$ must be normal.
  - The normal distribution is completely specified by its mean and variance.

Estimating a normal variance with known mean

- Let
$$y_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2),$$
where μ is known and σ^2 is unknown.
- The likelihood function
$$\pi(y \mid \sigma^2) = (2\pi\sigma^2)^{n/2} \exp\left\{-(2\sigma^2)^{-1} \sum_{i=1}^{n} (y_i - \mu)^2\right\}$$
- For a natural conjugate prior, we seek a prior that has the same functional form"
$$\pi(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right),$$
which is the kernel of an inverse gamma distribution: $G^{-1}(\alpha, \beta)$. Thus the natural conjugate prior in this case is
$$\sigma^2 \sim G^{-1}(\alpha, \beta).$$
what is the normalizing constant in $\pi(\sigma^2)$? $\frac{\beta^\alpha}{\Gamma(\alpha)}$.
- Posterior distribution
$$\pi(\sigma^2 \mid y) \propto \quad (\sigma^2)^{-(n/2+\alpha+1)} \exp\left(-\frac{\nu/2 + \beta}{\sigma^2}\right),$$
where $\nu = n^{-1} \sum_i (y_i - \mu)^2$.

An improper prior

- An improper prior results in the limit as $\alpha \downarrow 0$ and $\beta \downarrow 0$. In this case,

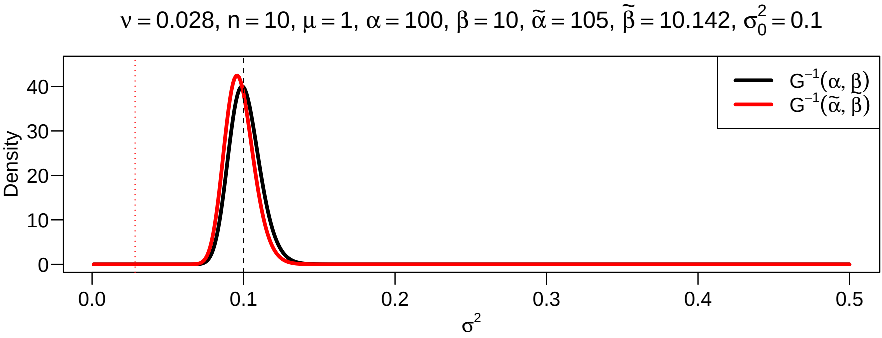
$$\lim_{\alpha \to 0, \beta \to 0} \pi(\sigma^2; \alpha, \beta) = \sigma^{-2}$$

- What is the posterior distribution? Is it proper or not?

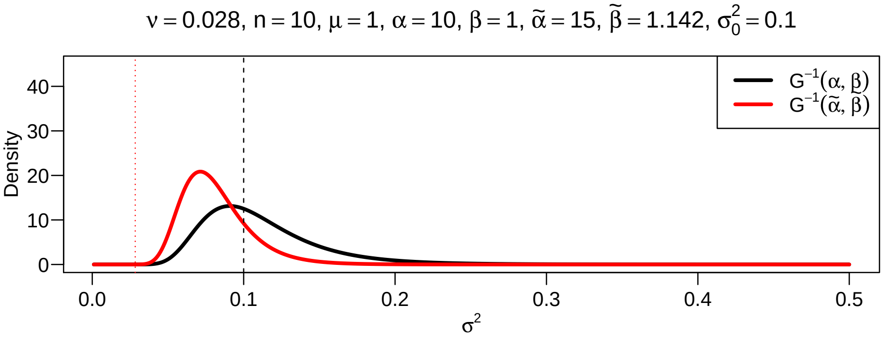
$$\sigma^2 \mid y \sim \mathrm{G}^{-1}(n/2, \nu/2)$$

It is proper.

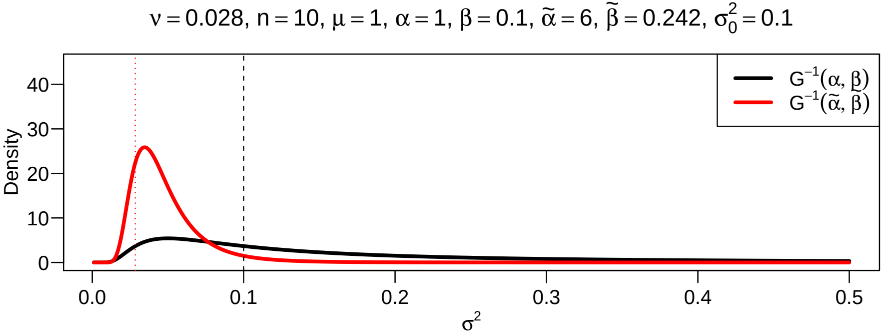
Limiting case: $\alpha \downarrow 0$ and $\beta \downarrow 0$

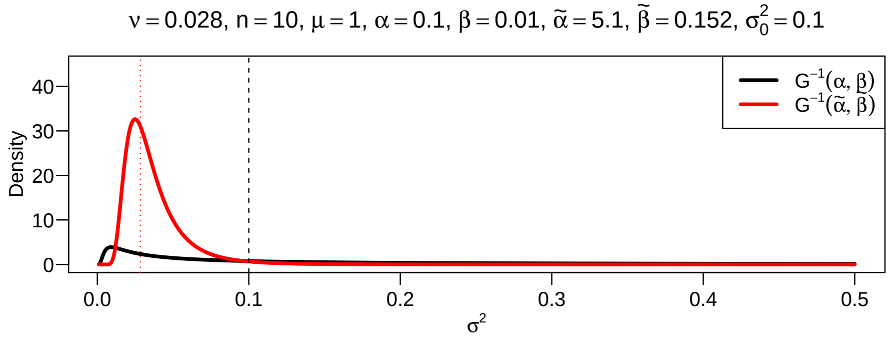


Limiting case: $\alpha \downarrow 0$ and $\beta \downarrow 0$

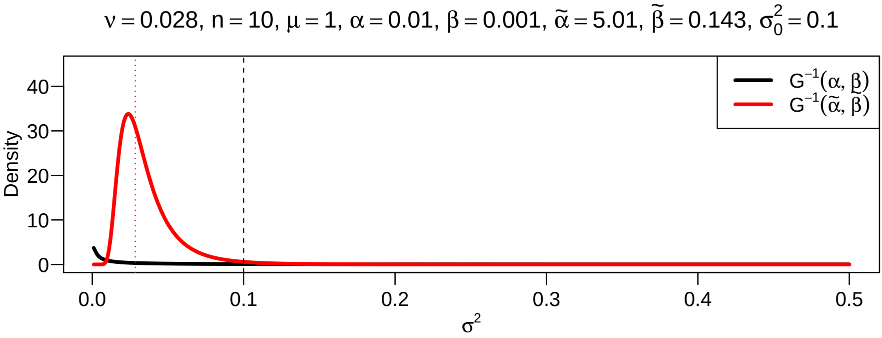


Limiting case: $\alpha \downarrow 0$ and $\beta \downarrow 0$




$\nu = 0.028$, $n = 10$, $\mu = 1$, $\alpha = 0.1$, $\beta = 0.01$, $\tilde{\alpha} = 5.1$, $\tilde{\beta} = 0.152$, $\sigma_0^2 = 0.1$
$G^{-1}(\alpha, \beta)$
$G^{-1}(\tilde{\alpha}, \tilde{\beta})$
Density
0
10
20
30
40
0.0
0.1
0.2
0.3
0.4
0.5
σ^2

Limiting case: $\alpha \downarrow 0$ and $\beta \downarrow 0$



Precision

- Bayesians often work directly with the precision, σ^{-2}, as opposed to the variance.
- The likelihood can be written as

$$\pi(y \mid \sigma^{-2}) \propto (\sigma^{-2})^{n/2} \exp\left\{-0.5\sigma^{-2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

- Using the fact that if $x \sim G^{-1}(\alpha,\beta)$, then $x^{-1} \sim G(\alpha,\beta)$, the natural conjugate prior is

$$\sigma^{-2} \sim G(\alpha,\beta).$$

  Therefore,

$$\pi(\sigma^{-2} \mid y) \propto (\sigma^{-2})^{\alpha+n/2-1} \exp\left\{-(\beta+\nu/2)\sigma^{-2}\right\}$$

  and

$$\sigma^{-2} \mid y \sim G(\alpha+n/2, \beta+\nu/2)$$

- An improper prior results in the limit as $\alpha \downarrow 0$ and $\beta \downarrow 0$.

$$\pi(\sigma^{-2}) = \sigma^{2}$$