Inference for an Unknown Mean and Variance of Normal Distribution

Edps 590BAY

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

Overview

- Likelihood
- Priors
- Mathematical result.
- Revisit anorexia data.
- Monto Carlo sampling.
- Revisit Getting what you paid for practice problem.

Depending on the book that you select for this course, read either Gelman et al. pp 63-69 or Kruschke Chapters pp 449-454. I also used Hoff.

Basics

In the previous set of notes, we conditioned the analysis on knowing σ^2, but now we want to simultaneously estimate both the mean and variance of a variable from normally distributed population.

Goal: Find the posterior distribution of θ and σ^2 where data $y_i \sim N(\theta, \sigma^2)\ i.i.d$; that is,

$$p(\theta, \sigma^2 | y_1, \ldots, y_n).$$

We need to break the posterior, $p(\theta, \sigma^2|y_1, \ldots, y_n)$, joint distribution of θ & σ^2 into likelihood and priors using Bayes Theorem.

$$\begin{aligned} p(\theta, \sigma^2 | y_i, \ldots, y_n) &= \frac{p(y_1, \ldots, y_n|\theta, \sigma^2) \times p(\theta, \sigma^2)}{p(y_1, \ldots, y_n)} \\ &\propto \underbrace{p(y_1, \ldots, y_n|\theta, \sigma^2)}_{\text{likelihood}} \times \underbrace{p(\theta, \sigma^2)}_{\text{prior}} \end{aligned}$$

Likelihood

This is the same as before; that is,...

If we have n independent observations (i.e., $(y_1, y_2, \ldots, y_n)$) where each y_i is from population $N(\theta, \sigma^2)$ or $y_i \sim N(\theta, \sigma^2)$ $i.i.d$, then the joint distribution of $(y_1, y_2, \ldots, y_n)$ is

$$\begin{aligned} p(y_1, y_2, \ldots, y_n|\theta, \sigma^2) &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\left(\frac{y_i - \theta}{\sigma}\right)^2\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \theta}{\sigma}\right)^2\right\} \end{aligned}$$

Breaking Problem Down Further

$$\begin{aligned} p(\theta,\sigma|y_i,\ldots,y_n) &= \frac{p(y_1,\ldots,y_n|\theta,\sigma^2)\times p(\theta,\sigma^2)}{p(y_1,\ldots,y_n)} \\ &\propto p(y_1,\ldots,y_n|\theta,\sigma^2)\times p(\theta,\sigma^2) \\ p(y_1,\ldots,y_n|\theta,\sigma^2)\times p(\theta,\sigma^2) &= p(y_1,\ldots,y_n|\theta,\sigma^2)\times p(\theta|\sigma^2)\times p(\sigma^2) \end{aligned}$$

We saw in previous set of notes that the if σ^2 is known then the conjugate prior for $p(\theta|\sigma^2, y_1, \ldots y_n)$ is normal $N(\mu_0, \tau_0^2)$, and the posterior is $N(\mu_n, \tau_n^2)$ where

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_o + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \qquad \text{and} \qquad \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

Conjugate Prior for $\theta|\sigma^2$

We start with the conjugate prior for θ given σ^2 and we got

$$\theta \sim N(\mu_n, \tau_n^2)$$

where

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_o + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \qquad \text{and} \qquad \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

$\mu_0 =$ mean from previous observations. When we estimate both σ^2 and μ, the equation for μ_n will differ slightly because the θ (i.e., the mean) now depends on σ^2.

$$p(\theta|\sigma^2) \times p(\sigma^2)$$

Conjugate Prior for σ^2: $p(\sigma^2)$

- σ^2 must be non-negative; that is, $0 \geq \sigma^2$.
- A family of distributions where this hold is the Gamma distribution, but the Gamma is not conjugate for normal likelihood.
- Thre Gamma is a conjugate for the precision, $1/\sigma^2$, in which case the prior for σ^2 is inverse-Gamma.
- Prior for precision and variance,

$$\begin{array}{lcl} \text{Precision:} & 1/\sigma^2 & \sim \text{Gamma}(a,b) \\ \text{Variance:} & \sigma^2 & \sim \text{Inverse Gamma}(a,b) \end{array}$$

- Sometimes you may see these referred to the inverse Gamma as, IGamma or InvGamma.

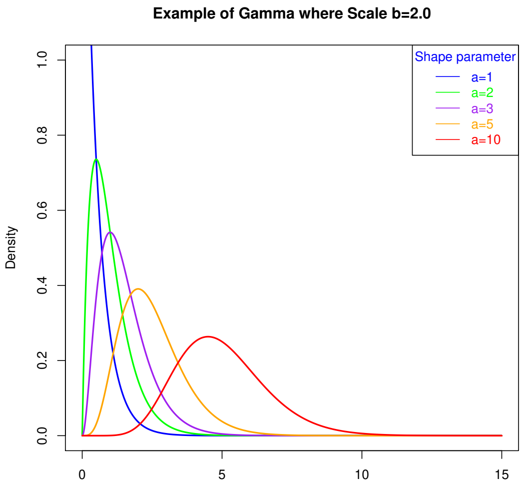
The Gamma Distribution

Gamma(a, b) has two parameters,

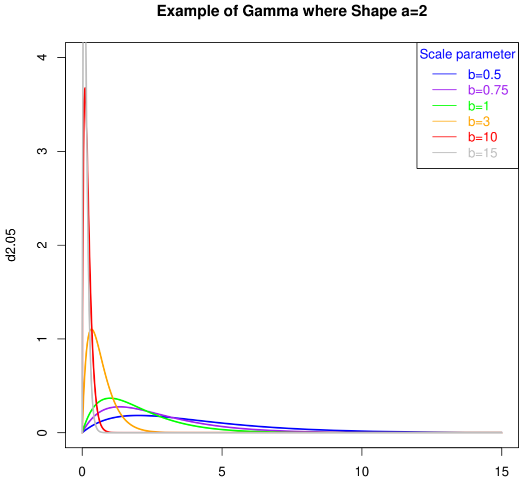
- Possible values for Gamma distributed random variable are Real numbers $0 \leq x < \infty$
- a is the "shape" parameter and $a > 0$.
- b is the "scale" parameter and $b > 0$.
- The mean $= ab$.
- No simple closed form for the median.
- The mode $= (a-1)b$.
- The variance $= ab^2$
- The chi-square distribution is a special case of Gamma when $\nu = 2b$ and $a = 2$ where ν is the degrees of freedom of the chi-square distribution.

Some examples. . .

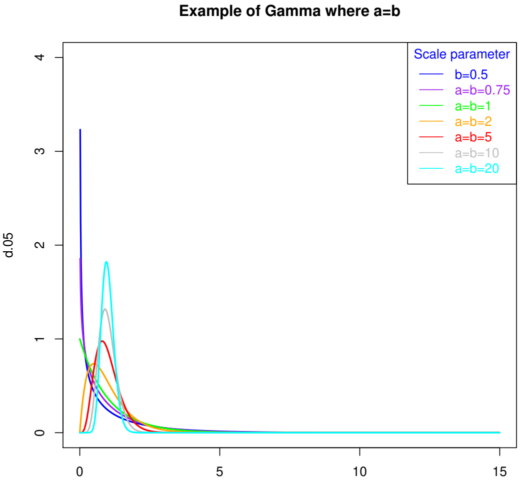
Examples of Gamma Distribution



Examples of Gamma Distribution



Examples of Gamma Distribution



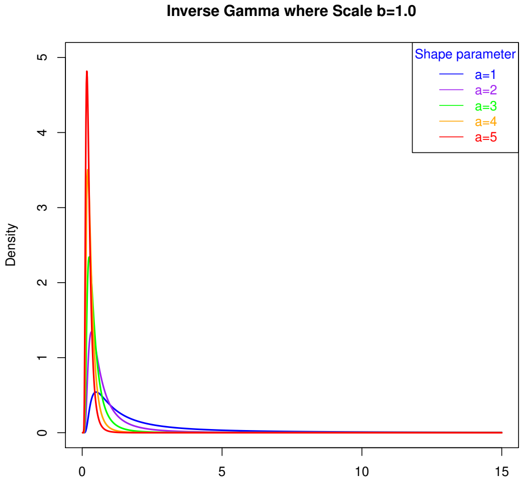
The Inverse Gamma Distribution

This is also a family 2 parameter distributions which is the distribution of reciprocals of gamma distributed variables.

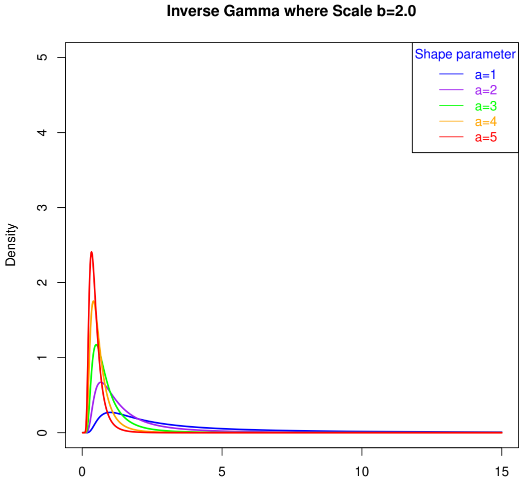
- A variable distribution at Inverse Gamma is $0 < x < \infty$
- $\alpha > 0$ (shape)
- $\beta > 0$ (scale)
- Values can be non-negative reals; that is, $0 < x$.
- The mean $= \beta/(\alpha - 1)$ for $\alpha > 1$
- The mode $= \beta/(\alpha + 1)$
- Variance $= \beta^2/(\alpha - 1)^2(\alpha - 2)$ for $\alpha > 2$

Some examples,. . .

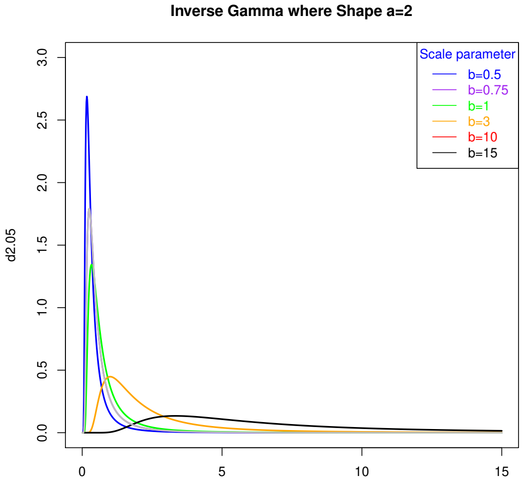
Examples of Inverse Gamma Distribution



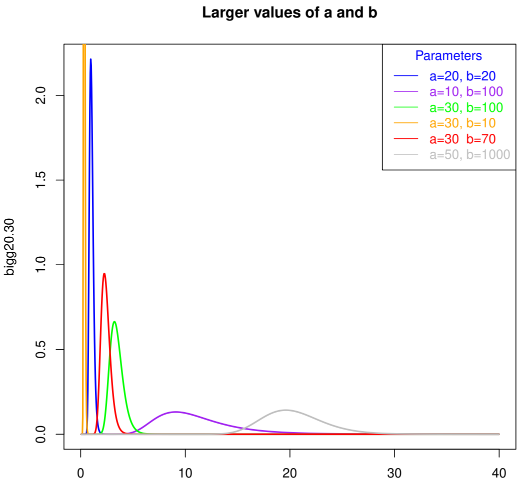
Examples of Inverse Gamma Distribution



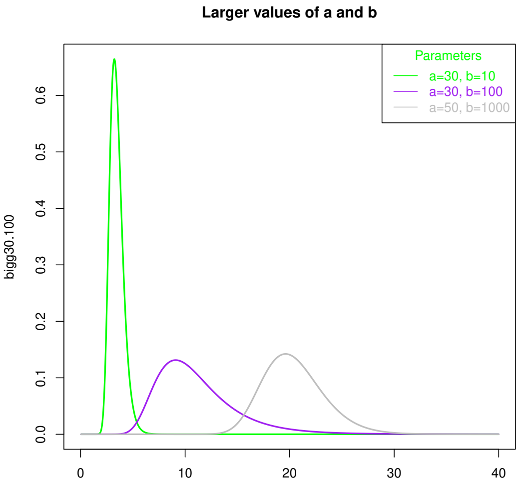
Examples of Inverse Gamma Distribution



Inverse Gamma Distribution: Bigger a and b



Inverse Gamma Distribution: Bigger a and b



The Parameterizations of the Priors

For the mean θ,

- μ_0 is the prior mean
- Then prior for θ is $N(\mu_0, \tau_0^2)$
- κ_0 number of prior measurements (min value=1)

For the variance σ^2,

- σ_0^2 is the prior variance.
- ν_0 be the sample size upon which σ_0^2 is based on (i.e., degrees of freedom).
- Then precision is $(1/\sigma^2) \sim$ Gamma$(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2)$
- or $\sigma^2 = 1/$precision

The JOINT Posterior Distribution

The Model

$$\begin{array}{rcl} \text{likelihood:} & y \sim & N(\theta, \sigma^2) \ i.i.d. \\ \text{prior for } \theta: & \theta|\sigma^2 \sim & N(\mu_o, \sigma_o^2/\kappa_o) \\ \text{prior for } \mathit{precision}: & 1/\sigma^2 \sim & \text{Gamma}(\nu_o/2, (\nu_o/2)\sigma_o^2) \end{array}$$

The posterior is a bi-variate distribution for μ and σ^2:

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n} \qquad \tau_n^2 = \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}$$

$$\kappa_n = \kappa_0 + n$$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = \frac{1}{\nu_0 + n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{(\kappa_0 + n)}(\bar{y} - \mu_0)^2\right]$$

Closer look at σ_n^2

$$\sigma_n^2 = \frac{1}{\nu_0 + n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{(\kappa_0 + n)}(\bar{y} - \mu_0)^2\right]$$

$$(\nu_0 + n)\sigma_n^2 = \left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{(\kappa_0 + n)}(\bar{y} - \mu_0)^2\right]$$

where

- $(\nu_o + n)\sigma_n^2$ posterior sum of squares
- $\nu_0\sigma_0^2$ prior sum of squares
- $(n-1)s^2$ sample sum of squares
- $\frac{\kappa_0 n}{(\kappa_0+n)}(\bar{y} - \mu_0)^2$ additional uncertainty due to difference between observed sample mean and the prior mean

The CONDITIONAL Posterior Distribution of θ

The conditional distribution of μ given σ, which is held constant (i.e., starting with the joint and then conditioning on σ^2) and data is $p(\mu|\sigma^2, y_1, \ldots, n_n)$; that is,

$$\begin{aligned} \mu|\sigma^2, y_1, \ldots, y_n &\sim N(\mu_n, \sigma^2/\kappa_n) \\ &= N\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right) \end{aligned}$$

Note: $\sigma^2/\kappa_n \approx \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}} = \tau_n^2$
For small κ_n (and large n), $\tau_n^2 \approx \sigma^2/n$.

The Marginal Posterior Distribution of σ^2

The marginal distribution of σ^2 is inverse Gamma (i.e., a scale inverse chi-squared) distribution; that is,

$$\sigma^2|y_1,\ldots,y_n \sim \textsf{InvGamma}((\nu_n/2),(\nu_n/2)\sigma_n^2)$$

Re-analysis of Anorexia Data

The sample statistics are

- $n = 72$ girls
- $\bar{y} = 2.7638$
- $s^2 = 63.7378$

Set prior values

- $\mu_0 = 0$
- $\kappa_0 = 1$
- $\sigma_0^2 = 500$
- $\nu_0 = 1$

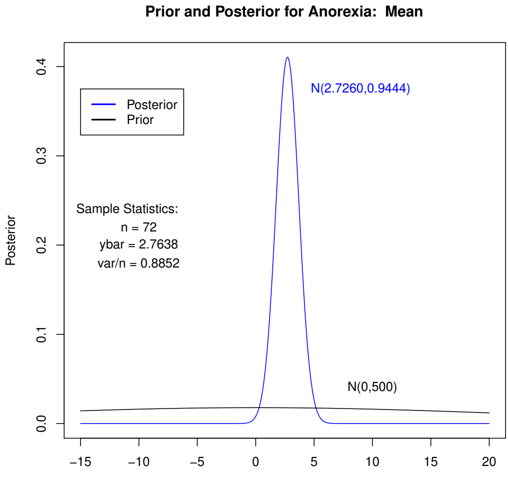
The Statistics for Posterior

$$\mu_n = \frac{\kappa_o\mu_o + n\bar{y}}{\kappa_o + n} = \frac{1(0) + 72(2.7638)}{1 + 72} = 2.7261$$

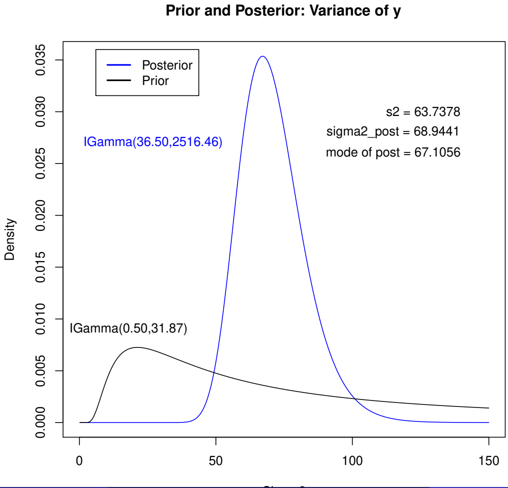
$$\begin{aligned}\sigma_n^2 &= \frac{1}{\nu_o + n}[\nu_o\sigma_o^2 + (n-1)s^2 + \frac{\kappa_o n}{(\kappa_o + n)}(\bar{y} - \mu_o)^2] \\ &= \frac{1}{1 + 72}[1(500) + (72-1)63.7378 + \frac{1(72)}{(1 + 72)}(2.7638 - 0)^2] \\ &= 68.9441\end{aligned}$$

and $\tau_n^2 = \sigma_n^2/\kappa_n = 0.94444$ or $\tau_n^2 = 1/(\kappa_0/\sigma_n^2 + n/\sigma_n^2) = 0.9438$.
For comparison, when we set $\sigma^2 = 64.00$ and we obtained $\mu_n = 2.6856$, and using just the data, $\bar{y} = 2.7639$ and var$(y) = 63.7378$

Pictures of the Prior & Posterior: Mean



Pictures of the Prior & Posterior: Variance



Bayesian Central Limit Theorem

Notice that Bayesian parameter estimates are close to sample statistics, which are all estimates of population values.

Note that

$$\begin{aligned}\lim_{n\to\infty} \hat{\mu}_n &= \lim_{n\to\infty} \frac{\kappa\mu + n\bar{y}}{\kappa + n} \\ &= \lim_{n\to\infty} \left(\frac{\kappa\mu}{\kappa + n} + \frac{n\bar{y}}{\kappa + n} \right) \\ &= \bar{y}\end{aligned}$$

As the sample size goes to infinity, the Bayesian posterior estimate of the mean $\hat{\mu}$ converges to the maximum likelihood estimate $\bar{y}$.

Bayesian Central Limit Theorem

And for precision of the mean... Using equation for precision with $1/\tau^2$ and $1/\sigma^2$ as the prior precision and data precision, respectively,

$$\begin{aligned}\lim_{n\to\infty} \hat{\sigma}^2_{\mu} &= \lim_{n\to\infty} \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \\ &= \lim_{n\to\infty} \frac{\sigma^2}{\frac{\sigma^2}{\tau^2} + n} \\ &= \frac{\sigma^2}{n}\end{aligned}$$

As the sample size goes to infinity, the Bayesian posterior estimate of the variance of the mean $\hat{\mu}$ converges to the maximum likelihood estimate.

Monte Carlo Sampling

If we are primarily interested in the mean, we could use Monte Carlo Sampling to get mean of $p(\theta|y_1,\ldots,y_n)$, or standard deviation, or any other function (e.g., $Pr(\theta_1 < \theta_2|y_{11},\ldots,y_{n_2 2})$).

Recall that $p(\theta,\sigma) = p(\theta|\sigma^2) \times p(\sigma^2)$, which suggests a Monte Carlo sampling method to get parameters. Take S random draws:

$$\begin{array}{rcl} \sigma^{2(1)} \sim \text{InvGamma}(\nu_n/2, \sigma_n^2\nu_n/2) & \rightarrow & \theta^{(1)} \sim N(\theta_n, \tau_n^{2(1)}) \\ \sigma^{2(2)} \sim \text{InvGamma}(\nu_n/2, \sigma_n^2\nu_n/2) & \rightarrow & \theta^{(2)} \sim N(\theta_n, \tau_n^{2(2)}) \\ \vdots & & \vdots \\ \sigma^{2(S)} \sim \text{InvGamma}(\nu_n/2, \sigma_n^2\nu_n/2) & \rightarrow & \theta^{(S)} \sim N(\theta_n, \tau_n^{2(S)}) \end{array}$$

Note $\tau_n^{2(s)} = \sigma_n^{2(2)}/(\kappa_0 + n)$.

Example from Monte Carlo Sampling

Simple case: Draw random sample to get estimates of σ^2 and then use that estimate to sample from normal to get estimates of μ.

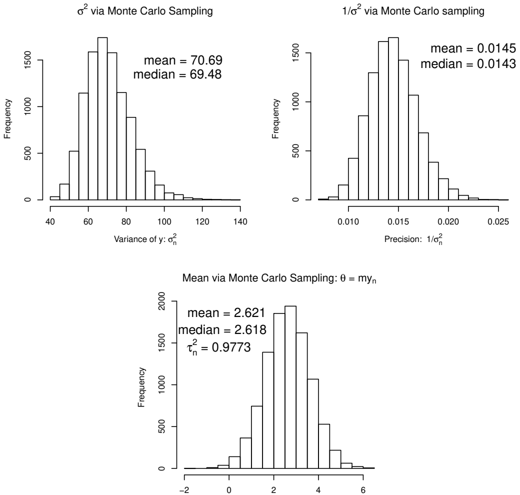
r

```r
sigma2.mc <- 1/rgamma(10000, nu.n/2, nu.n*sigma2.n/2)
tau2.mc <- sigma2.mc/(kappa.o + n)
theta.mc <- rnorm(10000, mu.n, sqrt(tau2.mc))
```

Note:

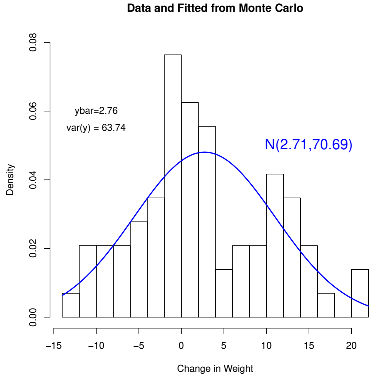
$$\begin{aligned}\tau^2 &= \frac{1}{\kappa_o/\sigma^2 + n/\sigma^2} \\ &= \frac{\sigma^2}{\kappa_o + n}\end{aligned}$$

Monte Carlo of Anorexia Data



Compare with Data

Plausible values for μ, 95% credible interval (-13.77, 19.19)



Some Comparisons

We've used the Anorexia data a number of times, so here's what we have obtained so far:

Sample statistics	Priors & Posteriors	Bayesian: Mean w/ fixed $\sigma^2 = 64$	Bayesian: Mean w/ fixed $\sigma^2 = 64$	Bayesian: Joint Estimation Analytic	Bayesian: Joint Estimation Monte Carlo
$n = 72$	n	57	$+15$	72	$S = 1000$
	μ_0	0	2.683	0	2.72
	τ_0^2	1000	1.123		
	σ_o^2	fixed	fixed	500	68.94
	κ_o			1	
	ν_o			1	73
$\bar{y} = 2.7639$	μ_n	2.683	2.761	2.726	2.62
$\sigma^2 = 63.7378$				68.944	70.69
$\sigma^2/72 = 0.8852$	τ_n^2	1.123	0.889	0.944	.87

Re-Analysis of Getting What you Paid For

Using the data set that we used before, but now we'll estimate both the mean and variance of the posterior distribution of $\mu, \sigma^2|\mathbf{y}$ $y_i =$ state total SAT.

- load data
- compute sample statistics
- set values for prior
- Find μ_n using $\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}$
- Find σ_n^2 using $\frac{1}{\nu_0+n}[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{(\kappa_0+n)}(\bar{y}-\mu_0)^2]$
- Compute τ_n^2 using $\sigma_n/(\kappa_0+n)$.
- Run a Monte Carlo simulation to get distributions of σ_n^2, $1/\sigma_n^2$, and μ_n.
- Compare analytic, monte carlo and sample descriptive statistics.

Summary

- Introduced the Gamma and Inverse Gamma Distributions
- σ^2 is Inverse Gamma distributed
- $1/\sigma^2$ (precision) is Gamma distributed
- Found posterior distributions for μ, σ^2, and τ^2.
- Computed summary statistics for posterior to get point estimates of μ, σ and τ^2.
- Monte Carlo sampling was used to "look" at the posterior joint distribution of θ and σ^2.
- Practiced using "getting what you paid for" — priors matter!