Normal-Inverse Gamma Inference

STAT 251, Unit 7

Overview

Review

Inference with the Normal Distribution: σ^2 unknown, μ known

Inverse-Gamma (IG) Family of Distributions

Posterior Distribution of σ^2 (for normal likelihood, known μ)

Review

Review: Structure of Units 6–8

Suppose $y_i|(\mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2)$.

There are two parameters to consider. We will work through the following progression:

- Unit 6: Inference with μ unknown but σ^2 known.
- Unit 7: Inference with σ^2 unknown but μ known.
- Unit 8: Inference with μ and σ^2 unknown

While Unit 8 is the only realistic scenario among the three, that unit will rely on results that we establish in Units 6 and 7.

Posterior for μ (σ^2 known)

If $y_i|\{\mu,\sigma^2\} \overset{iid}{\sim} N(\mu,\sigma^2)$, and $\mu \sim N(m,v)$, and σ^2 is known, then

$$\mu|\{y_1,\ldots,y_n,\sigma^2\} \sim N(m^\star, v^\star)$$

with

$$m^\star = \frac{\frac{n\bar{y}}{\sigma^2} + \frac{m}{v}}{\frac{n}{\sigma^2} + \frac{1}{v}} = \frac{nv\bar{y} + \sigma^2 m}{nv + \sigma^2}$$

and

$$v^\star = \left(\frac{n}{\sigma^2} + \frac{1}{v}\right)^{-1} = \frac{v\sigma^2}{nv + \sigma^2}$$

The derivation of this result was part of Unit 6 class notes.

Inference with the Normal Distribution: σ^2 unknown, μ known

Prior for σ^2 (with known μ)

Consider the normal family of distributions—that is, the collection of all probability density functions $f(\cdot|\mu, \sigma^2)$ such that

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

and $\mu \in (-\infty, \infty), \sigma^2 > 0$. Note that any $\sigma^2 > 0$ could be used with a normal distribution.

Question: To allow σ^2 to potentially be any positive value (and to ensure it cannot be negative), which of the following distributions would work as a prior distribution for σ^2?

a. $\sigma^2 \sim \textit{Normal}$
b. $\sigma^2 \sim \textit{Beta}$
c. $\sigma^2 \sim \textit{Gamma}$
d. $\sigma^2 \sim \textit{Binomial}$
e. $\sigma^2 \sim \textit{Poisson}$

Question: To allow σ^2 to potentially be any positive value (and to ensure it cannot be negative), which of the following distributions would work as a prior distribution for σ^2?

a. $\sigma^2 \sim \textit{Normal}$
b. $\sigma^2 \sim \textit{Beta}$
c. $\sigma^2 \sim \textit{Gamma}$
d. $\sigma^2 \sim \textit{Binomial}$
e. $\sigma^2 \sim \textit{Poisson}$

Only (c.) gives the requested support of $(0, \infty)$, so it would work as a justifiable family of distributions for the prior.

The Gamma Distribution! (?)

Recall that if $\theta \sim Gamma(a, b)$, then

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{(\theta > 0)}$$

Although it has the correct support, $\sigma^2 \in (0, \infty)$, there is a good reason to prefer another family of distributions based on the gamma family of distributions:

A Gamma prior for σ^2 is NOT *conjugate* in this setting (normal likelihood, μ known)

Inverse-Gamma (IG) Family of Distributions

Inverse-Gamma (IG) Family of Distributions

θ has the *inverse-gamma(a,b)* distribution **if and only if** the inverse of θ, $1/\theta$, has the *gamma(a,b)* distribution.

$$\frac{1}{\sigma^2} \sim Gamma(a, b) \Leftrightarrow \sigma^2 \sim InverseGamma(a, b)$$

The probability density function of the inverse-gamma(a,b) distribution is

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-b/\theta} \mathbb{1}_{(\theta > 0)}$$

Comparing pdf's of Gamma(a,b), InverseGamma(a,b)

The pdf of the inverse-gamma(a,b) distribution is

$$\pi_{IG}(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-b/\theta} \mathbb{1}_{(\theta>0)}$$

Recall that the pdf of the Gamma(a,b) distribution is

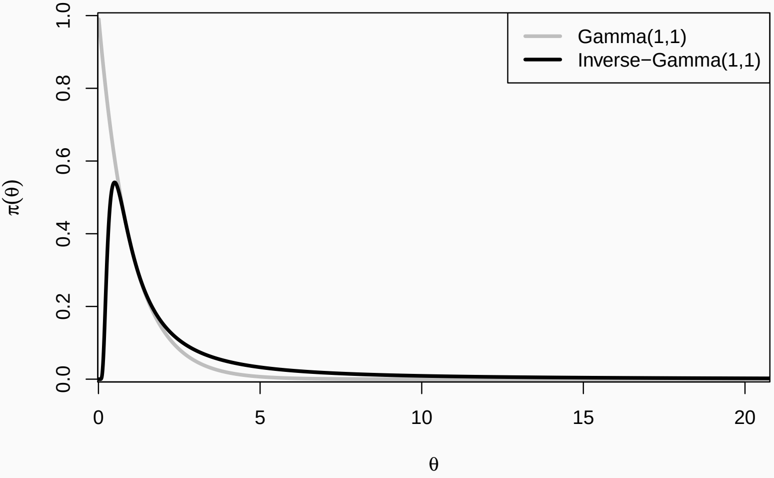
$$\pi_{G}(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{(\theta>0)}$$

$$\frac{1}{\sigma^2} \sim Gamma(a, b) \Leftrightarrow \sigma^2 \sim InverseGamma(a, b)$$

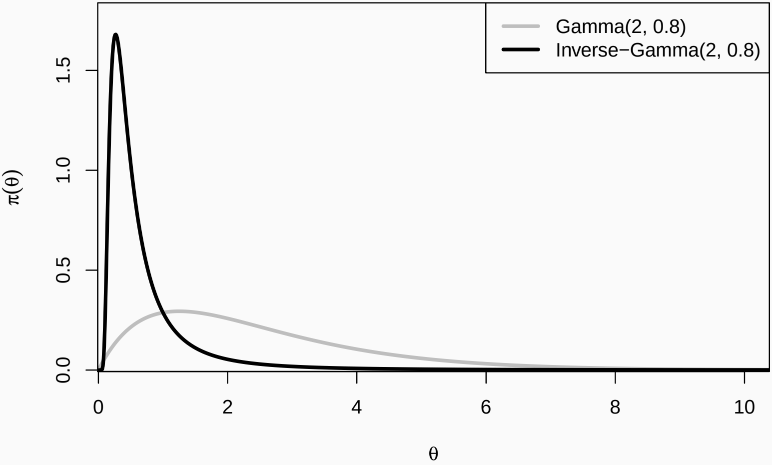
TRUE/FALSE: $\pi_{IG}(\theta) = \pi_G(\theta)$ for all values of θ.

TRUE/FALSE: $\pi_{IG}(\theta) = \pi_G(1/\theta)$ for all values of θ.

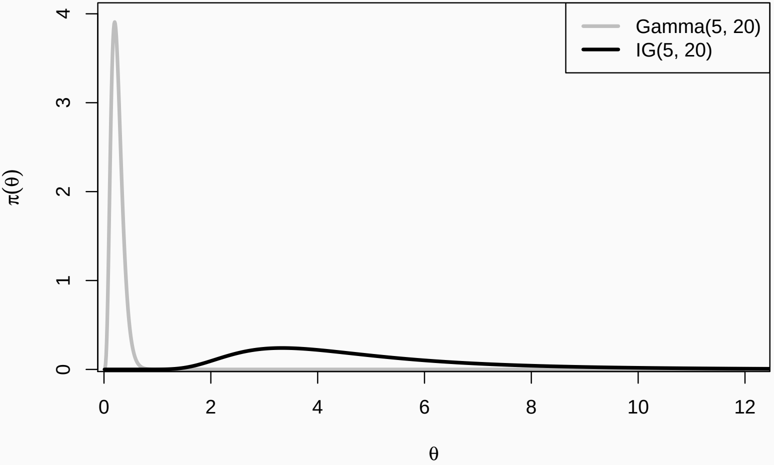
Comparison of Gamma, Inverse Gamma pdfs



Comparison of Gamma, Inverse Gamma pdfs



Comparison of Gamma, Inverse Gamma pdfs



Summaries of the Inverse Gamma distribution

Recall that for the $Gamma(a, b)$ distribution,

- mean$= \frac{a}{b}$
- variance$= \frac{a}{b^2}$
- mode$=\frac{a-1}{b}$, provided $a > 1$.

For the Inverse-Gamma(a,b) distribution,

- mean$=\frac{b}{a-1}$, provided $a > 1$
- variance$=\frac{b^2}{(a-1)^2(a-2)} = \frac{mean^2}{a-2}$, provided $a > 2$
- mode$=\frac{b}{a+1}$.

These summaries for the inverse gamma distribution are NOT the inverse of the summaries for the gamma distribution. For instance, $E\left(\frac{1}{\theta}\right) \neq \frac{1}{E(\theta)}$.

Working with the inverse gamma distribution in R

Base R does not have the same suite of **d**dist, **p**dist, **q**dist, **r**dist functions for the inverse gamma distribution as it does for other common distributions (such as the binomial, Poisson, gamma, and normal distributions).

Get the *dinvgamma*, *pinvgamma*, *qinvgamma*, and *rinvgamma* functions by installing and loading the "invgamma" package.

```
install.packages("invgamma") # one-time installation
```

```
library("invgamma") #load package--repeat each session
```

Working with the inverse gamma distribution in R

```r
dinvgamma(.2, 5, 4)
```

```
## [1] 0.001374102
```

```r
pinvgamma(.2, 5, 4)
```

```
## [1] 1.694474e-05
```

```r
qinvgamma(.2, 5, 4)
```

```
## [1] 0.5951514
```

```r
rinvgamma(3, 5, 4)
```

```
## [1] 1.2539607 1.1346713 0.8274988
```

Posterior Distribution of σ^2 (for normal likelihood, known μ)

Setup for Deriving Posterior of σ^2 with μ known

If $\sigma^2 \sim \mathit{InverseGamma}(a, b)$, and $y_i | \{\sigma^2, \text{ known } \mu\} \overset{iid}{\sim} N(\mu, \sigma^2)$, what is posterior distribution of $\sigma^2 | \{\mu, y_1, \ldots, y_n\}$?

According to the prior distribution,

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp\left[-\frac{b}{\sigma^2}\right] \mathbb{1}_{(\sigma^2 > 0)}$$

Because the y_i's are conditionally iid and normally distributed,

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

Derivation of the posterior is left as a homework problem.

Key Result

If we assume that:

- $y_i|\{\mu, \sigma^2\} \overset{iid}{\sim} N(\mu, \sigma^2)$—that is, the data are conditionally *iid* and normally distributed; AND
- μ is known; AND
- $\sigma^2 \sim \mathit{InverseGamma}(a, b)$;

THEN

$$\sigma^2|\{\mu, y_1, \ldots, y_n\} \sim IG\left(a^\star = a + \frac{n}{2}, b^\star = b + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

The inverse gamma prior on σ^2 is *conjugate* when the likelihood is normal and μ is known.

An Example

Standardized tests are often designed so that they will be (approximately) normally distributed, and with a given mean and standard deviation. However, such tests are periodically modified in such a way that the distribution of the scores might be changed.

Scenario: Suppose that the current version of an IQ test is normally distributed with mean 100 and standard devation 15. A proposed modification of the test will leave the mean unchanged at 100, but the standard deviation might be a little different. The new version is administered to 20 adults in a pilot study, and their scores are determined.

Example: Continued

Setup: $y_i|\{\mu = 100, \sigma^2\} \overset{iid}{\sim} N(\mu = 100, \sigma^2),\ i = 1, \ldots, 20.$

Prior Distribution for σ^2?

- Variance of scores for the current test version is $15^2 = 225$.
- It is believed the standard deviation with the modified version "might be a little different."
- Inverse Gamma prior is a convenient choice—it is a conjugate prior for σ^2 if we assume a normal likelihood for the data and μ is known.

I'll assume $\sigma^2 \sim IG(a = 5, b = 900)$

Implications of the prior: $\sigma^2 \sim IG(5, 900)$

- $E(\sigma^2) =$

- $Pr(200 \leq \sigma^2 \leq 250) =$

- $Pr(150 \leq \sigma^2 \leq 300) =$

- $Pr(10 \leq \sigma \leq 20) =$

Data

In the pilot study, the 20 scores are given by the table below.

Individual	1	2	3	4	5	6	7	8	9	10
Score	108	88	97	129	91	80	114	118	87	89
Individual	11	12	13	14	15	16	17	18	19	20
Score	106	106	102	112	90	119	100	106	93	97

Table 1: Scores for 20 adults in modified IQ test (fictional data)

```
scores <- c(108, 88, 97, 129, 91, 80, 114, 118,
            87, 89, 106, 106, 102, 112, 90, 119,
            100, 106, 93, 97)
```

In-class practice

For the IQ example,

1. Determine the posterior distribution of σ^2.
2. What is the mean of the posterior distribution?
3. What is the mode?
4. What is the central 95% posterior credible interval?
5. Does it seem reasonable to believe the new test version has a variance of 225, as does the current version? Explain.
6. Plot the prior and posterior distribution on the same graph.

Hint 1: The R function *length* can be applied to a vector to determine the number of elements it contains.

Hint 2: In R, arithmetic functions of a vector are applied elementwise, so that if X is a vector of the values $x_1, x_2, \ldots, x_n$, then X**2 is a vector of the squared x_i values.

Hint 3: The R function *sum* can be applied to a vector.