Bayesian Inference in Normal Models

Mark Andrews

Inference in normal distributions

- The univariate normal, or Gaussian, probability distribution with mean μ and variance σ^2 has the density

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- A random variable x that is normally distributed with mean μ and variance σ^2 can be written

$$x \sim N(\mu, \sigma^2).$$

Inference in normal distributions

- A common elementary data analysis problem is when we observe n iid samples from $N(\mu, \sigma^2)$, where μ and σ^2 are unknown, and we aim to infer μ.
- In other words, we observe

  $$D = x_1, x_2 \cdots x_n,$$

  where we assume

  $$x_i \sim N(\mu, \sigma^2), \quad \text{for } i \in 1, 2 \cdots n,$$

  where both μ and σ^2 are unknown, and we aim to infer the value of μ.
- In other words, we aim to determine the posterior distribution

  $$P(\mu|D).$$

Likelihood of μ and σ^2

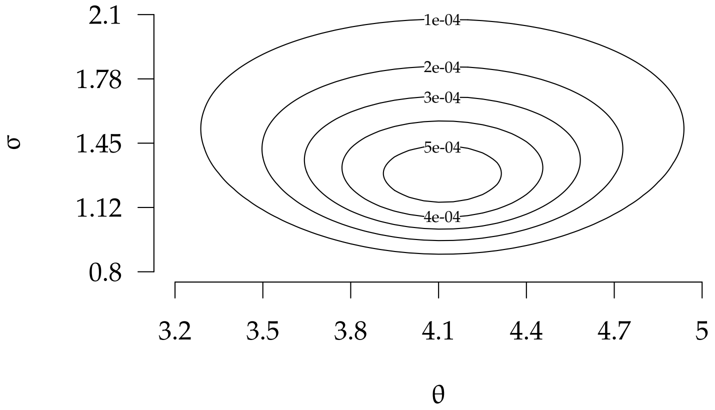
- The likelihood of μ and σ^2 given D is

$$
\begin{aligned}
P(x_1, x_2 \cdots x_n | \mu, \sigma^2) &= \prod_{i=1}^{n} P(x_i | \mu, \sigma^2), \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}, \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)}, \\
&\propto \sigma^{-n} e^{\left(-\frac{1}{2\sigma^2} [ns^2 + n(\bar{x} - \mu)^2] \right)},
\end{aligned}
$$

with

$$
\bar{x} \doteq \frac{1}{n} \sum_{i=1}^{n} x_i, \quad s^2 \doteq \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.
$$

Likelihood of μ and σ^2



The likelihood of μ and σ when

$$D = 2.41, 5.37, 5.28, 4.89, 4.40, 4.63, 4.67, 4.52, 1.10, 3.86$$

with $\bar{x} = 4.11$ and $s = 1.28$.

Conjugate prior for μ and σ

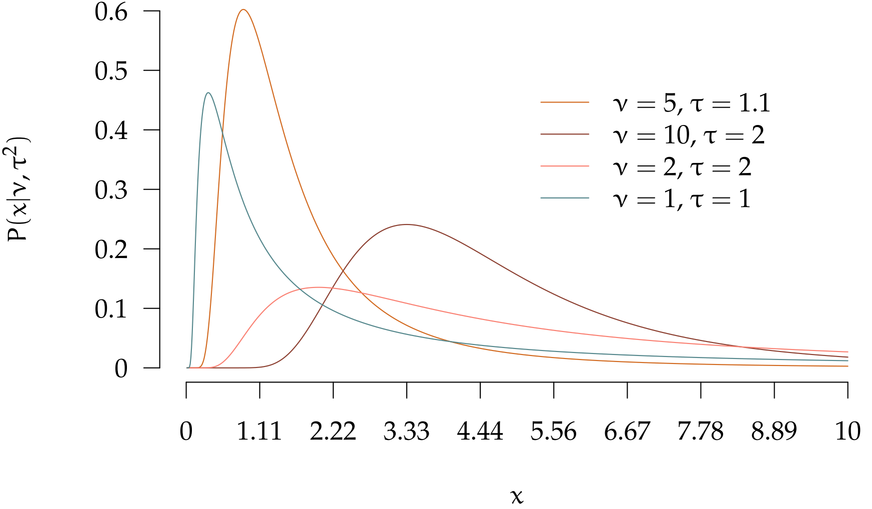
- A common choice of conjugate prior is the normal/inverse-gamma distribution, also known as the *normal* $\times$ *scaled inverse-*χ^2 distribution.
- Thus, our full model is

$$
\begin{aligned}
x_i &\sim N(\mu, \sigma^2), \quad \text{for } i \in 1, 2 \cdots n, \\
\mu &\sim N(\mu_0, \sigma^2/\kappa_0), \\
\sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2).
\end{aligned}
$$

- This corresponds to the joint prior density

$$
\begin{aligned}
P(\mu, \sigma^2) &= \text{N-Inv-}\chi^2(\mu, \sigma^2 | \mu_0, \kappa_0, \nu_0, \sigma_0^2), \\
&= N(\mu | \mu_0, \sigma^2/\kappa_0) \times \text{Inv-}\chi^2(\sigma^2 | \nu_0, \sigma_0^2), \\
&\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} e^{\left(-\frac{1}{2\sigma^2}[\nu_0 \sigma_0^2 + \kappa_0(\mu_0 - \mu)^2]\right)}
\end{aligned}
$$

The scaled inverse-χ^2 distribution



The mean is $\frac{\nu}{\nu-2}\sigma_0^2$, mode is $\frac{\nu}{\nu+2}\sigma_0^2$.

Joint posterior on μ and σ

- With the normal likelihood and normal/scaled inverse-χ^2 distribution, the joint posterior is:

$$\begin{aligned}
P(\mu, \sigma^2|D) &\propto P(D|\mu, \sigma^2)P(\mu, \sigma^2), \\
&\propto \sigma^{-n} e^{\left(-\frac{1}{2\sigma^2}[ns^2+n(\bar{x}-\mu)^2]\right)} \\
&\quad \times \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} e^{\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right)} \\
&= \text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_n, \kappa_n, \nu_n, \sigma_n^2)
\end{aligned}$$

where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_n}, \\
\kappa_n &= \kappa_0 + n, \\
\nu_n &= \nu_0 + n, \\
\sigma_n^2 &= \frac{1}{\nu_n}\left(\nu_0\sigma_0^2 + ns^2 + \frac{n\kappa_0}{\kappa_0+n}(\mu_0-\bar{x})^2\right).
\end{aligned}$$

Marginal posterior of σ^2

- The marginal posterior of σ^2 is obtained by integrating over μ:

$$
\begin{aligned}
P(\sigma^2|D) &= \int P(\sigma^2, \mu|D)\, d\mu, \\
&= \int \text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_n, \kappa_n, \nu_n, \sigma_n^2)\, d\mu, \\
&\propto (\sigma^2)^{-(\nu_n/2+1)} e^{\left(-\frac{1}{2\sigma^2}[\nu_n \sigma_n^2]\right)}, \\
&= \text{Inv-}\chi^2(\sigma^2|\nu_n, \sigma_n^2)
\end{aligned}
$$

Marginal posterior of μ

- The marginal posterior of μ is obtained by integrating over σ^2:

$$
\begin{aligned}
P(\mu|D) &= \int P(\sigma^2, \mu|D)\ d\sigma^2, \\
&= \int \text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_n, \kappa_n, \nu_n, \sigma_n^2) d\sigma^2, \\
&\propto \left[1 + \frac{\kappa_n}{\nu_n \sigma_n^2} (\mu - \mu_n)^2\right]^{-(\nu_n+1)/2}, \\
&= t_{\nu_n}\left(\mu|\mu_n, \sigma_n^2/\kappa_n\right).
\end{aligned}
$$

  which is a (non-standard) t-distribution with ν_n degrees of freedom, location ν_n and scale σ_n^2/κ_n.

Marginal posterior density of μ

- To infer the marginal probability of μ, we must integrate over the *nuisance* variable σ^2:

$$P(\mu|D) = \int P(\mu, \sigma^2|D)\, d\sigma^2,$$

which can be done by

$$= \int \overbrace{P(\mu|\sigma^2, D)}^{\substack{\text{conditional} \\ \text{posterior}}} \underbrace{P(\sigma^2|D)}_{\substack{\text{marginal} \\ \text{posterior}}}\, d\sigma^2.$$

- This demonstrates that the posterior distribution $P(\mu|D)$ is an infinite mixture of normal distributions.

Marginal posterior density of μ

- If we fix σ^2 at any value and, using the previous prior on μ, i.e. $\mu \sim N(\mu_0, \sigma^2/\kappa_0)$, then

$$\begin{aligned} P(\mu|D,\sigma^2) &\propto P(D|\mu,\sigma)P(\mu|\mu_0,\sigma^2/\kappa_0), \\ &= N(\mu_n, \sigma^2/\kappa_n), \end{aligned}$$

  where

$$\mu_n = \frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \quad \sigma^2/\kappa_n = \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}.$$

- Integrating over σ^2, when σ^2 has a scaled inverse χ^2 distribution with parameters ν_n, σ_n^2 leads to a non-standard t-distribution leads to $t_{\nu_n}\left(\mu|\mu_n, \sigma_n^2/\kappa_n\right)$ as before.

Posterior predictive distribution

- Given our n observed values

$$D = 2.41, 5.37, 5.28, 4.89, 4.40, 4.63, 4.67, 4.52, 1.10, 3.86,$$

  and our assumed probabilistic generative model, what is *next* value that we predict?

- In other words, what is

$$P(x_{n+1}|D) = \int\int P(x_{n+1}|\mu, \sigma^2) P(\mu, \sigma^2|D) d\mu d\sigma^2$$

- This is again a t-distribution:

$$t_{\nu_n}\left(\mu|\mu_n, \left(1 + \frac{1}{\kappa_n}\right)\sigma_n^2\right)$$

Inferences concerning differences of normal means

- Let us assume that we have observed two sets of data as follows:

$$x_1, x_2 \cdots x_n \overset{\text{iid}}{\sim} N(\mu_x, \sigma^2), \quad y_1, y_2 \cdots y_n \overset{\text{iid}}{\sim} N(\mu_y, \sigma^2).$$

- In other words, both $x_1, x_2 \cdots x_n$ and $y_1, y_2 \cdots$ are independently and identically normally distributed, with different means μ_x, μ_y but with a common variance σ^2.
- In general, we know neither μ_x, μ_y nor σ^2.
- In such a situation, a common problem we are faced with is to infer the difference

$$\mu_x - \mu_y,$$

given $D = \{x_1, x_2 \cdots x_n, y_1, y_2 \cdots y_n\}$, while integrating over the nuisance variable σ^2, i.e.,

$$P(\mu_x - \mu_y | D).$$

Posterior inference of differences of means

- Inference of the differences of means is a special kind of posterior predictive inference:

$$\int P(\mu_x - \mu_y | \mu_x, \mu_y, \sigma^2) P(\mu_x, \mu_y, \sigma^2 | D) \, d\mu_x \, d\mu_y \, d\sigma^2.$$

- We can approach this problem using the methods we have used to deal with inference in the univariate normal distribution.
- We begin by assuming, as before,

$$\begin{aligned} \mu_x &\sim N(\mu_0, \sigma^2/\kappa_0), \\ \mu_y &\sim N(\mu_0, \sigma^2/\kappa_0), \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2). \end{aligned}$$

Posterior inference of differences of means

- Given that

$$P(\mu_x, \mu_y, \sigma^2 | D) = P(\mu_x, \mu_y, | \sigma^2, D) P(\sigma^2 | D)$$

- From previous results, we see that

$$\begin{aligned} P(\mu_x, \mu_y, | \sigma^2, D) &= P(\mu_x | \sigma^2, D) P(\mu_y | \sigma^2, D), \\ &= N(\mu_n^x, \sigma^2 / \kappa_n^x) N(\mu_n^y, \sigma^2 / \kappa_n^y) \end{aligned}$$

where μ_n^x, μ_n^y, κ_n^x and κ_n^y are defined analogously with previous results.

Posterior inference of differences of means

- As such,

$$\int P(\mu_x - \mu_y | \mu_x, \mu_y, \sigma^2) P(\mu_x, \mu_y | \sigma^2, D) d\mu_x d\mu_y$$
$$= N(\mu_n^x - \mu_n^y, \sigma^2/\kappa_n^x + \sigma^2/\kappa_n^y).$$

For simplicity, let's assume $\kappa_0 \doteq 0$. In that case,

$$= N(\bar{x} - \bar{y}, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}).$$

- Integrating over σ^2 as before leads to

$$\int N(\bar{x} - \bar{y}, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}) P(\sigma^2 | D)$$
$$= t_{\nu_n}\left(\bar{x} - \bar{y}, \sigma_n^2 \left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

Posterior inference of differences of means

- In the case of the non-informative prior case, i.e.,

$$P(\mu_x, \mu_y, \sigma^2) \propto \frac{1}{\sigma^2},$$

  the marginal posterior over $\mu_x - \mu_y$ is

$$P(\mu_x - \mu_y | D) = t_\nu \left(\bar{x} - \bar{y}, s^2 \left(\frac{1}{n} + \frac{1}{m} \right) \right)$$

  where

$$\nu = n + m - 2, \quad s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}.$$

- This is an identical distribution to the sampling distribution of the difference of the means.