Inference for a Mean of a Normal Distributed Variable when Variance is Known (or fixed)  
Edps 590BAY

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

©Board of Trustees, University of Illinois

Spring 2018
Overview

- The **Likelihood** for a Normally distributed variable
- The conjugate **Prior** of normal
- The **Posterior**
  - Estimation and inference of mean given variance
  - Prediction
  - Adding new information
- Practice analytic: anorexia data
- Joint inference for mean and variance
- Back to practice problem

Depending on the book that you select for this course, read either Gelman et al. p39-68 or Kruschke Chapters pp 450-459. I relied mostly on Hoff for this.
What we Are Working toward

The conjugate prior distribution of the normal likelihood is a normal distribution; that is, the product of a normal likelihood and a normal prior yields a posterior distribution for $\theta$ that is also normal.

Rather than just stating the result, we will walk through the math step-by-step.
The Normal Distribution

The normal is often a good approximation or model for data (i.e., likelihood) and parameters because Central Limit Theorem

- Variables or measures maybe sum of many things.
- Parameters are often sums or means of many values.
- Often a good approximation for those not normal (e.g., binomial with probability close to .5 and $n = 1000$).

Recall

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right), \quad -\infty < y < \infty$$
The Likelihood for an Independent Sample

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

$$p(y_1, y_2, \ldots, y_n | \theta, \sigma^2) = \prod_{i=1}^{n} p(y_i | \theta, \sigma^2)$$

due to independence.

Substitution

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{y_i - \theta}{\sigma} \right)^2 \right\}$$

Re-arrange terms

$$= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2} \left( \frac{y_i - \theta}{\sigma} \right)^2 \right\}$$

More algebra

$$= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{y_i - \theta}{\sigma} \right)^2 \right\}$$
Further examination of normal

If we expand the term in the exponent, we find

\[
\sum_{i=1}^{n} \left( \frac{y_i - \theta}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_i y_i^2 - 2\frac{\theta}{\sigma^2} \sum_i y_i + n\frac{\theta^2}{\sigma^2}
\]

The terms based on the data are sufficient statistics; that is \(\{\sum_i y_i^2, \sum_i y_i\}\) is a 2-dimensional sufficient for variance and mean. So are the sample mean and variances:

\[
\bar{y} = \sum_i y_i / n
\]

\[
s^2 = \sum_i (y_i - \bar{y})^2 / (n - 1)
\]

Therefore, \(\{\bar{y}, s^2\}\) are also sufficient statistics.
A normal prior is the conjugate of the normal likelihood; therefore, the prior should have a term in the exponent with a similar form as the likelihood.

To keep terms clear:

\[
\text{Likelihood} \quad y_i \sim N(\theta, \sigma^2) \quad i.i.d \\
\text{Prior} \quad \theta \sim N(\mu_o, \tau_o^2) \\
\text{Posterior} \quad \theta | \sigma^2, y_1, \ldots, y_n \sim N(\mu_n, \tau_n^2)
\]
Prior

The prior distribution is just a uni-variate normal,

\[
p(\theta|\mu_0, \tau_0) = \frac{1}{\sqrt{2\pi\tau^2_0}} \exp \left\{ \frac{-1}{2} \left( \frac{\theta - \mu_0}{\tau_0} \right)^2 \right\}
\]

What to put in for \(\mu_0\) and \(\tau_0\)?

- If you have no knowledge use large value of \(\tau_0\), which would give you a flat prior (similar to a uniform for the beta prior). This would also lessen the impact of whatever you put in as prior mean \(\mu_0\).
- If you are up-dating the posterior, use values from previous posterior.
- You can use educated guess, but be prepared to defend your choice.
Examine Impact of $\mu_0$, $\tau_0^2$, $\theta$, $\sigma^2$

Run function `normalPriorLike(mu0, tau0, m.y, sigma)`

- File is on course web-site:
  “R_function_normal_prior_x_likelihood.txt”
- Input:
  - $\mu_0$ is mean of prior
  - $\tau_0$ is standard deviation of prior
  - $\theta$ is mean of likelihood
  - $\sigma$ is standard deviation of likelihood
Posterior

\[ p(\theta | \sigma^2, y_1, \ldots, y_n) \propto p(\theta | \mu_0, \tau_0) \times p(y_1, \ldots, y_n | \theta, \sigma) \]

\[ \propto \frac{1}{\sqrt{2\pi \tau_0^2}} \exp \left\{ -\frac{1}{2} \left( \frac{\theta - \mu_0}{\tau_0} \right)^2 \right\} \]

\[ \times (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{y_i - \theta}{\sigma} \right)^2 \right\} \]

\[ \propto \exp \left\{ -\frac{1}{2} \left( \frac{1}{\tau_0^2} (\theta - \mu_0)^2 \right) \right\} \times \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} \left( \sum_{i=1}^{n} y_i - \theta \right)^2 \right) \right\} \]

\[ \propto \exp \left\{ \frac{1}{\tau_0^2} (\theta^2 - 2\theta \mu_0 + \mu_0^2) + \frac{1}{\sigma^2} \left( (\sum_{i} y_i)^2 - 2\theta \sum_{i} y_i + n\theta^2 \right) \right\} \]
Posterior (continued)

\[ p(\theta|\sigma^2, y_1, \ldots, y_n) \]

\[ \propto \exp \left\{ \frac{1}{\tau_0^2} (\theta^2 - 2\mu_0 \theta + \mu_0^2) + \frac{1}{\sigma^2} \left( (\sum_i y_i)^2 - 2(\sum_i y_i)\theta + n\theta^2 \right) \right\} \]

\[ \propto \exp \left\{ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \theta^2 + -2\left( \frac{\mu_0}{\tau_0^2} + \frac{\sum_i y_i}{\sigma^2} \right)\theta + \text{everything else} \right\} \]

or

\[ a\theta^2 - 2b\theta + c \rightarrow \text{Complete the square} \]

where

\[ a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \quad b = \frac{\mu_0}{\tau_0^2} + \frac{\sum_i y_i}{\sigma^2} = \frac{1}{\tau_n^2} + \frac{n}{\sigma^2} \bar{y} \]
Posterior (one last step)

\[
p(\theta | \sigma^2, y_1, \ldots, y_n) \propto \exp \left\{ -\frac{1}{2} (a\theta^2 - 2b\theta) \right\}
\]

take out \(a\)

\[
= \exp \left\{ -\frac{1}{2} a(\theta^2 - 2b\theta/a) \right\}
\]

\(+/- b^2/a\)

\[
= \exp \left\{ -\frac{1}{2} a(\theta^2 - 2b\theta/a + b^2/a) - b^2/a \right\}
\]

almost done

\[
\propto \exp \left\{ -\frac{1}{2} a(\theta - b/a)^2 \right\}
\]

\[
= \exp \left\{ -\frac{1}{2} \frac{(\theta - b/a)^2}{1/a} \right\}
\]
Results for Posterior

The model:

Likelihood \[ y_i \sim N(\theta, \sigma^2) \text{ i.i.d} \]

Prior \[ \theta \sim N(\mu_o, \tau_o^2) \]

The posterior distribution of $\theta$:

Posterior: \[ p(\theta|\sigma^2, y_1, \ldots, y_n) \sim N(\mu_n, \tau_n^2) \]

Mean: \[ \mu_n = \frac{1}{\tau_0^2} \mu_o + \frac{n}{\sigma^2} \bar{y} \]

Variance: \[ \tau_n^2 = \frac{1}{\tau_0^2 + \frac{n}{\sigma^2}} \]
Precision

It can be more useful to express dispersion as Precision, which is inversely related to variance, because

- It makes expressions for mean and variance of the posterior a bit simpler.
- It will be used when estimating variances (has nicer distribution).

\[
\tilde{\sigma}^2 = \frac{1}{\sigma^2} \quad \text{sampling precision}
\]

\[
\tilde{\tau}_0^2 = \frac{1}{\tau_0^2} \quad \text{prior precision}
\]

\[
\tilde{\tau}_n^2 = \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \quad \text{posterior precision}
\]
Example: Anorexia

The data are the weights of $n = 72$ girls before and after treatment for anorexia. The girls were in one of three treatment groups, but for now we will only look at the change in weight.
Example: Anorexia

For our example, I took a random sample of $n = 57$ girls and set $\sigma^2 = 8$, which is close to the variance of the full sample. We will model change = weight$_2$ – weight$_1$.

Question: Did girls' weights on average change (increase) after treatment for anorexia?

Prior: $\mu_0 = 0$

Likelihood: $\sigma^2 = 8$

Note that $\bar{y}_{57} = 2.6859$.

Mean:

$$\mu_n = \frac{1}{1000} 0 + \frac{57}{8} 2.6859 = \frac{1}{1000 + \frac{57}{8}} = 2.6856$$

Variance:

$$\tau_n^2 = \frac{1}{1000 + \frac{57}{8}} = 0.1403$$
Posterior for Anorexia Data

Prior N(0,1000) & Posterior N(2.6856,0.1403)
Adding in the Hold Out Data

- For these \( n = 15 \) girls, their mean \( \bar{y} = 3.06 \).
- We will use the posterior that we found for the \( n = 57 \) girls as our new prior so
  \[
  \mu_0 = 2.6856 \quad \text{and} \quad \tau_0^2 = 0.1403
  \]
- The new posterior is normal with
  - mean:
    \[
    \mu_n = \frac{1}{0.1403} \frac{2.6856}{1} + \frac{15}{8} \cdot 3.06 = 2.7636 = \bar{y}_{72}
    \]
  - Variance:
    \[
    \tau_n^2 = \frac{1}{0.1403} + \frac{15}{8} = 0.1111
    \]
Impact of Adding Data

Up-dated Posterior: $N(2.7636, 0.1111)$

![Graph showing the impact of adding data to a posterior distribution. The graph compares the updated posterior $N(2.7636, 0.1111)$ with two prior distributions: $N(2.6856, 0.1403)$ and $N(0, 1000)$, illustrating how the posterior changes with the addition of data.]
95% Interval Estimates

Finding the .025 and .975 quantiles of $N(2.11, 3.42)$ gives us our credible intervals:

“The probability that the mean change of weight is between (2.11, 3.42) equals .95.”

The 95% high density interval:

“The probability that the mean change of weight is between (2.11, 3.42) equals .95.”

Why are they the same?
Comparing Data and Posterior

Data and Posterior Overlaid

Change in Weight

Density

0.00 0.02 0.04 0.06

−15 −10 −5 0 5 10 15 20

C.J. Anderson (Illinois)
Comparing Data and Random Draws from Posterior

Data and Random Samples from Posterior Overlaid

Change in Weight

Density

Data
Random

-15 -10 -5 0 5 10 15 20

0.00 0.02 0.04 0.06

C.J. Anderson (Illinois)
Prediction

Problem: We would like to make a prediction of a new $\tilde{y}$.

Solution: We need the predictive distribution. We could write out integrals, but instead we’ll use the fact that $\tilde{Y} \sim \mathcal{N}(\theta, \sigma^2)$

Using our posterior predictive distribution, the new prediction would be

$$E(\tilde{Y}|\sigma^2, y_1, \ldots, y_n) = E[E(\theta)|\sigma^2, y_1, \ldots, y_n]$$

$$= E(\theta|\sigma^2, y_1, \ldots, y_n)$$

$$= \mu_n$$
Prediction

... and the variance of new prediction:

\[
\text{var}(\tilde{Y}|\sigma^2, y_1, \ldots, y_n) = E[\text{var})(\tilde{Y}|\theta, y_1, \ldots, y_n)|y_1, \ldots, y_n] \\
+ \text{var}(E[\tilde{y}|\theta, y_1, \ldots, y_n]|y_1, \ldots, y_n) \\
= E[\sigma^2|y_1, \ldots, y_n] + (\theta|y_1, \ldots, y_n) \\
= \sigma^2 + \tau_n^2
\]

Note that there are 2 sources of uncertainty of new observation:

- \(\tau_n\): uncertainty of the value of \(\theta\)
- \(\sigma^2\): uncertainty due to sampling.

To summarize:

\[
\tilde{Y} \sim N(\mu_n, (\tau_n + \sigma^2))
\]
Getting what you pay for

The data are on the course web-site and consist of the following variables, which are state averages:

- state = name of state
- exp_pp = expenditure per pupil
- ave_pp = pupil/teacher ratio
- salary = average teacher salary
- taking = percent taking SAT
- ave_v = verbal scores
- ave_m = math scores
- ave_tot = total score
- region = of country
- state_abv = state abbreviation

Using the state average total scores, (a) examine the distribution of total scores and overlay normal; (b) use a diffuse prior, and find the posterior mean and variance; (c) find the 95% credible intervals; and (d) anything else you would like to do.
Mean and Variance

Everything that we’ve done so far has been conditional on our assumption of $\sigma$, but not we are going to estimate both.