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

Fall 2019

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
- Used throughout is anorexia data
- Practice: Getting what you paid for

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 and Gelman et al. for these notes.

. . . *Set of notes will be on joint estimation of mean and variance.*

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 θ 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

$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \qquad -\infty < y < \infty$$

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 $N(\theta, \sigma^2)$ or $y_i \sim N(\theta.\sigma^2)$ $i.i.d$ for short, then the joint distribution of $(y_1, y_2, \ldots, y_n)$ is

$$
\begin{aligned}
p(y_1, y_2, \ldots, y_n|\theta, \sigma^2) &= \prod_{i=1}^{n} p(y_i|\theta, \sigma^2) \qquad \text{due to independence} \\
\text{substitution} \qquad &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\theta}{\sigma}\right)^2\right\} \\
\text{re-arrange terms} \qquad &= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\left(\frac{y_i-\theta}{\sigma}\right)^2\right\} \\
\text{more algebra} \qquad &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2\right\}
\end{aligned}
$$

Further examination of normal

If we expand the term in the exponent, we find

$$\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2 = \frac{1}{\sigma^2}\underbrace{\sum_i y_i^2}_{data} - 2\frac{\theta}{\sigma^2}\underbrace{\sum_i y_i}_{data} + n\frac{\theta^2}{\sigma^2}$$

The terms based on the data are sufficient statistics; that is, $\sum_i y_i^2$ and $\sum_i y_i$. Also, these contain information to compute the variance and mean. In particular,

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

Therefore, $\bar{y}$ and s^2 are also sufficient statistics.

Prior

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:

$$\begin{array}{rrcl} \text{Likelihood} & y_i & \sim & N(\theta, \sigma^2)\ i.i.d \\ \text{Prior} & \theta & \sim & N(\mu_o, \tau_o^2) \\ & & & \\ \text{Posterior} & \theta|\sigma^2, y_1, \ldots, y_n & \sim & N(\mu_n, \tau_n^2) \end{array}$$

Prior

The prior distribution is just a uni-variate normal,

$$p(\theta|\mu_o, \tau_o^2) = \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left\{ \frac{-1}{2}\left(\frac{\theta - \mu_o}{\tau_o}\right)^2 \right\}$$

What to put in for μ_o and τ_o^2?

- If you have no knowledge use large value of τ_o, 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 μ_o.
- If you are up-dating the posterior, use values from the previous posterior.
- You can use educated guess, but be prepared to defend you choice.

Examine Impact of μ_o, τ_o^2, $\bar{y}$, σ^2

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

- File is on course web-site: "R_function_normal_prior_x_likelihood.txt"
- Input:
  - mu0 is mean of prior
  - tau0 is standard deviation of prior
  - m.y is mean of likelihood (i.e., $\bar{y}$, data based)
  - sigma is standard deviation of likelihood (s, data based)

Posterior

$$p(\theta|\sigma^2, y_1, \ldots, y_n) \propto p(\theta|\mu_o, \tau_o^2) \times p(y_1, \ldots, y_n|\theta, \sigma)$$

$$\begin{aligned}
\propto \quad & \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left\{\frac{-1}{2}\left(\frac{\theta-\mu_o}{\tau_o}\right)^2\right\} \\
\times \quad & (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2\right\} \\
\propto \quad & \exp\left\{\frac{-1}{2}\left(\frac{1}{\tau_o^2}(\theta-\mu_o)^2\right)\right\} \times \exp\left\{\frac{-1}{2}\left(\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} y_i - \theta\right)^2\right)\right\} \\
\propto \quad & \exp\left\{\frac{1}{\tau_o^2}(\theta^2 - 2\theta\mu_o + \mu_o^2) + \frac{1}{\sigma^2}\left((\sum_i y_i)^2 - 2\theta\sum_i y_i + n\theta^2\right)\right\}
\end{aligned}$$

Posterior (continued)

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

$$\propto \exp\left\{\frac{1}{\tau_o^2}(\theta^2 - 2\mu_o\theta + \mu_o^2) + \frac{1}{\sigma^2}\left((\sum_i y_i)^2 - 2(\sum_i y_i)\theta + n\theta^2\right)\right\}$$

$$\propto \exp\left\{\left(\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}\right)\theta^2 + \quad -2\left(\frac{\mu_o}{\tau_o^2} + \frac{\sum_i y_i}{\sigma^2}\right)\theta + \text{everything else}\right\}$$

or

$$a\theta^2 - 2b\theta + c \quad \longrightarrow \text{ Complete the square}$$

where

$$a = \frac{1}{\tau_o^2} + \frac{n}{\sigma^2} \qquad b = \frac{\mu_o}{\tau_o^2} + \frac{\sum_i y_i}{\sigma^2} = \frac{\mu_o}{\tau_o^2} + \frac{n}{\sigma^2}\bar{y}$$

Posterior (one last step)

$$
\begin{array}{rcl}
p(\theta|\sigma^2, y_1, \ldots, y_n) & \propto & \exp\left\{(a\theta^2 - 2b\theta)\right\} \\
\text{take out } a \text{ of "( )"} & = & \exp\left\{a(\theta^2 - 2b\theta/a)\right\} \\
+/\text{- b}^2/a & = & \exp\left\{a(\theta^2 - 2b\theta/a + b^2/a) - b^2/a\right\} \\
\text{almost done} & \propto & \exp\left\{a(\theta - b/a)^2\right\} \\
 & = & \exp\left\{\dfrac{(\theta - b/a)^2}{1/a}\right\}
\end{array}
$$

Results for Posterior

The model:

$$\begin{array}{rrcl} \text{Likelihood} & y_i & \sim & N(\theta, \sigma^2)\ i.i.d \\ \text{Prior} & \theta & \sim & N(\mu_o, \tau_o^2) \end{array}$$

The posterior distribution of θ:

$$\text{Posterior:} \quad p(\theta|\sigma^2, y_1, \ldots, y_n) \quad \sim N(\mu_n, \tau_n^2)$$

$$\text{Mean:} \qquad \mu_n \qquad = \frac{b}{a} = \frac{\frac{1}{\tau_o^2}\mu_o + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}}$$

$$\text{Variance:} \qquad \tau_n^2 \qquad = \frac{1}{a} = \frac{1}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}}$$

Posterior Mean is Weighted Average

The posterior distribution of θ: Often, precision is used instead of variance,

$$\tilde{\tau}_o^2 = 1/\tau_o^2 \qquad \text{and} \qquad \tilde{\sigma}^2 = 1/\sigma^2$$

So...

$$\text{Mean:} \quad \mu_n = \frac{b}{a} = \frac{\frac{1}{\tau_o^2}\mu_o + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}} = \frac{\frac{1}{\tau_o^2}\mu_o}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}} + \frac{\frac{1}{\tau_o^2}\bar{y}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}}$$

$$= \left(\frac{\tilde{\tau}_o^2}{\tilde{\tau}_o^2 + n\tilde{\sigma}^2}\right)\mu_o + \left(\frac{n\tilde{\sigma}^2}{\tilde{\tau}_o^2 + n\tilde{\sigma}^2}\right)\bar{y}$$

What happens when

- Increase prior precision relative to precision of data?
- Increase prior precision relative to precision of data?
- Increase sample size versus small sample size?

Precision

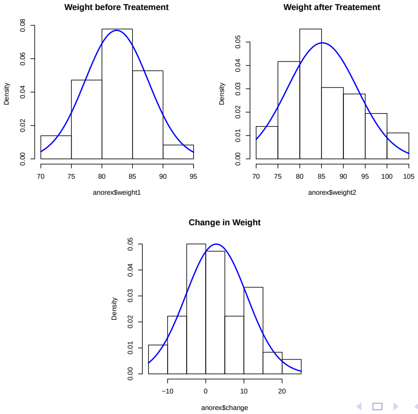
It can be more useful to express dispersion as Precision, which is inversely related to variance, because

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

$$
\begin{aligned}
\tilde{\sigma}^2 &= 1/\sigma^2 && \text{sampling precision} \\
\tilde{\tau}_o^2 &= 1/\tau_o^2 && \text{prior precision} \\
\tilde{\tau}_n^2 &= 1/\tau_n^2 = \frac{1}{\tau_o^2} + \frac{n}{\sigma^2} && \text{posterior precision}
\end{aligned}
$$

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 $\sigma = 8$, which is close to the standard deviation of the full sample.
We will model change $=$ weight$_2$ $-$ weight$_1$.
Question: Did girls' weights on average change (increase) after treatment for anorexia?

$$\begin{array}{llll} \text{Prior:} & \mu_o & = 0 & \\ & \tau_o^2 & = 1000 & \text{non-informative, like uniform} \\ \text{Likelihood:} & \sigma^2 & = 8^2 & \end{array}$$

Note that $\bar{y}_{57} = 2.6860$.

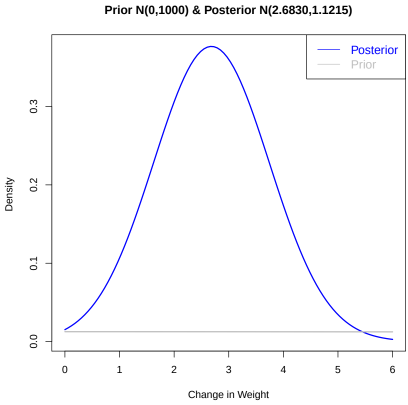
- Mean:

$$\mu_n = \frac{\frac{1}{1000}0 + \frac{57}{8^2}2.6860}{\frac{1}{1000} + \frac{57}{8^2}} = 2.6830$$

Variance:

$$\tau_n^2 = \frac{1}{\frac{1}{1000} + \frac{57}{8^2}} = 0.1215$$

Posterior Dist: Mean of Anorexia Data



Adding in the Hold Out Data

- For these $n = 15$ girls, their mean $\bar{y} = 3.0600$.
- We will use the posterior that we found for the $n = 57$ girls as our new prior so

$$\mu_o = 2.6830 \qquad \tau_o^2 = 1.1215$$

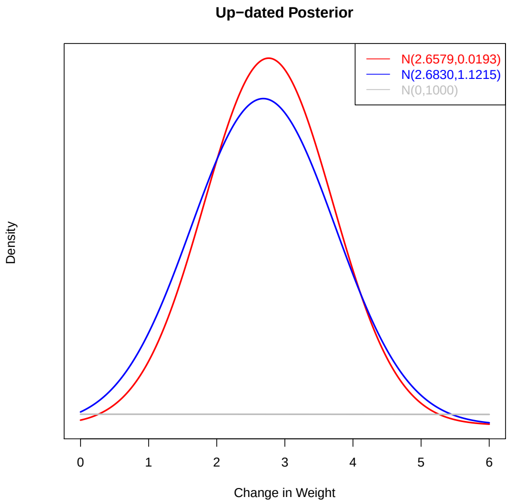
- The new posterior is normal with
  - mean:

$$\mu_n = \frac{\frac{1}{1.1215}2.6830 + \frac{15}{8^2}3.0600}{\frac{1}{1.1215} + \frac{15}{8^2}} = 2.7614 \qquad \text{note: } = \bar{y}_{72} = 2.6579$$

  - Variance:

$$\tau_n^2 = \frac{1}{\frac{1}{1.1215} + \frac{15}{8^2}} = 0.8881 \qquad \text{note: } = 8^2/72 = 0.8889$$

Impact of Adding Data



95% Interval Estimates

Finding the .025 and .975 quantiles of $N(2.7614, 0.8881)$ gives us out credible intervals:

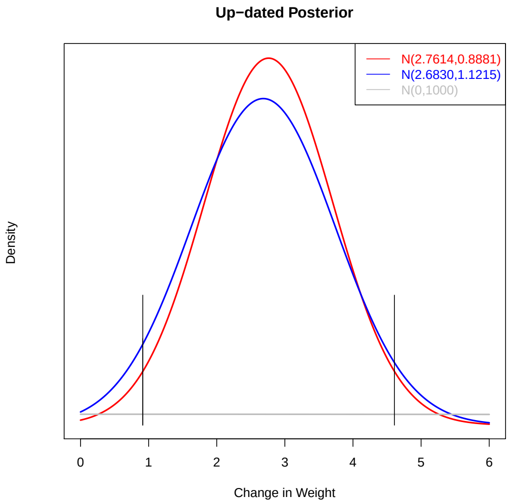
"The probability that the mean change of weight is between 0.91 and 4.61 equals .95."

The 95% high density interval:

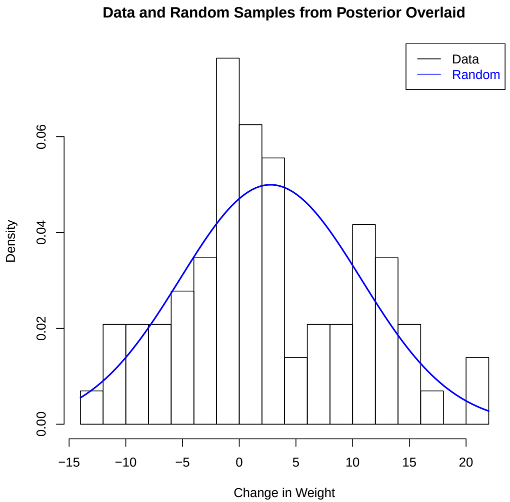
"The probability that the mean change of weight is between 0.91 and 4.61 equals .95."

Why are they the same?

Adding 95% Intervals



Comparing Data and Posterior



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 $\tilde{Y} \sim N(\theta, \sigma^2)$

First note that

$$\left\{\tilde{Y}|\sigma^2, y_1, \ldots y_n\right\} \sim N(\theta, \sigma^2) \Leftrightarrow \tilde{Y} = \theta + \epsilon \quad \text{where } \epsilon \sim N(0, \sigma^2)$$

The new prediction would be

$$\begin{aligned} E(\tilde{Y}|\sigma^2, y_1, \ldots, y_n) &= E(\theta + \epsilon|\sigma^2, y_1, \ldots, y_n) \\ &= E(\theta|\sigma^2, y_1, \ldots, y_n) + E(\epsilon|\sigma^2, y_1, \ldots, y_n) \\ &= \mu_n \end{aligned}$$

Prediction

...and the variance of new prediction:

$$\begin{aligned} \text{var}(\tilde{Y}+\epsilon|\sigma^2, y_1,\ldots,y_n) &= \text{var}(\theta|\sigma^2, y_1,\ldots,y_n|y_1,\ldots,y_n) \\ &\quad +\text{var}(\epsilon|\sigma^2, y_1,\ldots,y_n|y_1,\ldots,y_n) \\ &= \tau_n^2+\sigma^2 \end{aligned}$$

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

- τ_n: uncertainty of the value of θ
- σ^2: uncertainty due to sampling.

To summarize prediction of new: $\tilde{Y} \sim N(\mu_n, (\tau_n+\sigma^2))$

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_pt= pupil/treach 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 SAT scores,

1. Examine the distribution of total scores and overlay normal;
2. Use a diffuse prior and find the posterior mean and variance
3. Find the 95% credible intervals and the high density intervals
4. Anything else you would like to do.