

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

Edps 590BAY

Carolyn J. Anderson

Department of Educational Psychology



©Board of Trustees, University of Illinois

I 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.

... Next set of notes will be on joint estimation of mean and variance.

I 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.

I 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

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

I The Likelihood for an Independent Sample

If we have n independent observations (i.e., (y_1, y_2, \dots, 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, \dots, y_n) is

$$p(y_1, y_2, \dots, y_n | \theta, \sigma^2) = \prod_{i=1}^n p(y_i | \theta, \sigma^2) \quad \text{due to independence}$$

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

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

I Further examination of normal

If we expand the term in the exponent, we find

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

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.

I 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:

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

$$\text{Prior} \quad \theta \sim N(\mu_o, \tau_o^2)$$

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

I 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.

I Examine Impact of $\mu_0, \tau_0^2, \bar{y}, \sigma^2$

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)

Note: Function works best if keep `mu0` and `m.y` within ± 10 or change line in function

```
y <- seq(from=-12, to=12, length.out=1000)
```

I Posterior

$$\begin{aligned} p(\theta|\sigma^2, y_1, \dots, y_n) &\propto p(\theta|\mu_o, \tau_o^2) \times p(y_1, \dots, y_n|\theta, \sigma) \\ &\propto \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left\{-\frac{1}{2}\left(\frac{\theta - \mu_o}{\tau_o}\right)^2\right\} \\ &\quad \times (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left(\frac{y_i - \theta}{\sigma}\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\tau_o^2}(\theta - \mu_o)^2\right)\right\} \times \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\left(\sum_{i=1}^n y_i - \theta\right)^2\right)\right\} \\ &\propto \exp\left\{\frac{1}{\tau_o^2}(\theta^2 - 2\theta\mu_o + \mu_o^2) + \frac{1}{\sigma^2}\left(\left(\sum_i y_i\right)^2 - 2\theta\sum_i y_i + n\theta^2\right)\right\} \end{aligned}$$

I Posterior (continued)

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

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

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

or

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

where

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

I Posterior (one last step)

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

I Results for Posterior

The model:

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

The posterior distribution of θ :

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

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

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

I Posterior Mean is Weighted Average

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

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

So...

$$\begin{aligned} \text{Mean: } \mu_n &= \frac{b}{a} = \frac{\frac{1}{\tau_o^2} \mu_o + \frac{n}{\sigma^2} \bar{y}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}} = \frac{\frac{1}{\tau_o^2}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}} \mu_o + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau_o^2} + \frac{n}{\sigma^2}} \bar{y} \\ &= \left(\frac{\tilde{\tau}_o^2}{\tilde{\tau}_o^2 + n\tilde{\sigma}^2} \right) \mu_o + \left(\frac{n\tilde{\sigma}^2}{\tilde{\tau}_o^2 + n\tilde{\sigma}^2} \right) \bar{y} \end{aligned}$$

What happens when

- Increase $1/\sigma^2 = \tilde{\sigma}^2$ (i.e., precision)?
- Increase $1/\tau_o^2 = \tilde{\tau}_o^2$?
- Increase sample size n ?

I 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).

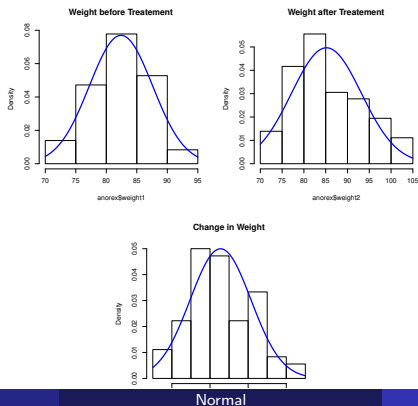
$$\tilde{\sigma}^2 = 1/\sigma^2 \quad \text{sampling precision}$$

$$\tilde{\tau}_o^2 = 1/\tau_o^2 \quad \text{prior precision}$$

$$\tilde{\tau}_n^2 = 1/\tau_n^2 = \frac{1}{\tau_o^2} + \frac{n}{\sigma^2} \quad \text{posterior precision}$$

I 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.



I 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 = weight₂ - weight₁.

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

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

Note that $\bar{y}_{57} = 2.98245$.

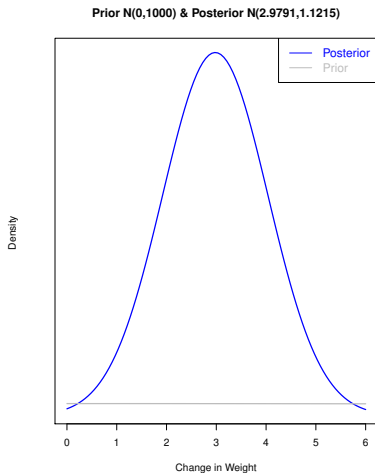
- Mean:

$$\mu_n = \frac{\frac{1}{1000}0 + \frac{57}{8^2}2.982456}{\frac{1}{1000} + \frac{57}{8^2}} = 2.979111$$

Variance:

$$\tau_n^2 = \frac{1}{\frac{1}{1000} + \frac{57}{8^2}} = 1.1215$$

I Posterior Dist: Mean of Anorexia Data



I Adding in the Hold Out Data

- For these $n = 15$ girls, their mean $\bar{y} = 1.9333$.
- We will use the posterior that we found for the $n = 57$ girls as our new prior so

$$\mu_o = 2.979111 \qquad \tau_o^2 = 1.1215$$

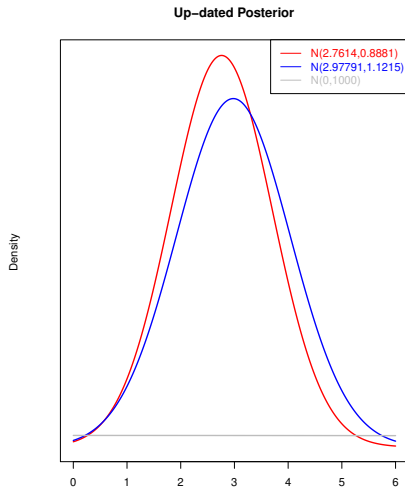
- The new posterior is normal with
 - mean:

$$\mu_n = \frac{\frac{1}{1.1215}2.979111 + \frac{15}{8^2}1.9333}{\frac{1}{1.1215} + \frac{15}{8^2}} = 2.7614 \qquad \text{note: } = \bar{y}_{72} = 2.7684$$

- Variance:

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

I Impact of Adding Data



I 95% Interval Estimates

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

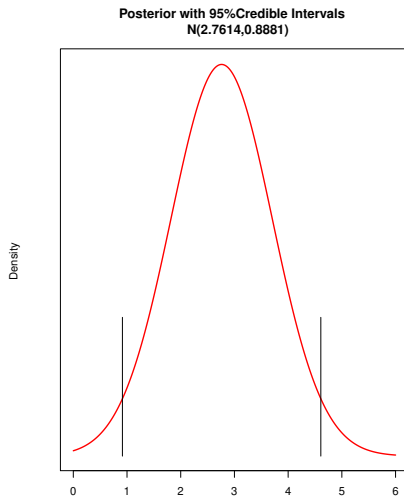
“The probability that the mean change of weight is between 0.91 and 4.61 equals .95.”

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?

I Adding 95% Intervals



I 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

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

The new prediction would be

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

I Prediction

... and the variance of new prediction:

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

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

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

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

I 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/teach 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 (use observed variance as “fixed”)
- 3 Find the 95% credible intervals and the high density intervals
- 4 Anything else you would like to do (e.g., change priors, plot salary by ave_tot)