The Normal Model

Yingbo Li

Clemson University

MATH 9810

The Pygmalion Study

Do teachers' expectations impact academic development of children?

- Researchers gave IQ test to elementary school children
- They randomly picked six children and told teachers that the test predicts them to have high potential for accelerated growth
- They randomly picked six children and told teachers that the test predicts them to have no potential for growth
- At end of school year, they gave IQ test again to all students
- They recorded the change in IQ scores of each student

Data

Accelerated group (A): 20, 10, 19, 15, 9, 18

No growth group (N): 3, 2, 6, 10, 11, 5

Summary statistics

- $\bar{y}_A = 15.2$; $sd(Y_A) = 4.7$
- $\bar{y}_N = 6.2$; $sd(Y_N) = 3.6$

Impossible to tell family of pdf with only $n = 6$ observations in each group. But, IQ test scores are well known to be approximately normally distributed. So...

Model for Changes in Scores

- $y_i^{(A)} \sim N(\mu_A, \sigma_A^2)$
- $y_j^{(N)} \sim N(\mu_N, \sigma_N^2)$
- Want posterior distribution of $\mu_A - \mu_N$

Inference for $\mu_A - \mu_N$ is complicated in frequentist paradigm when $\sigma_A^2 \neq \sigma_N^2$.

But, it is trivial with Bayesian inference!

Normal Model

IID observations $Y = (y_1, y_2, \ldots y_n)$

$$y_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

unknown parameters μ and σ^2.

Some prefer to work with the *precision*, ϕ, where $\phi = 1/\sigma^2$.

Likelihood

$$\begin{aligned} L(\mu, \phi | Y) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \phi^{1/2} \exp\{-\frac{1}{2}\phi(y_i - \mu)^2\} \\ &\propto \phi^{n/2} \exp\{-\frac{1}{2}\phi \sum_i (y_i - \mu)^2\} \end{aligned}$$

Likelihood Factorization with ϕ

$$
\begin{aligned}
L(\mu,\phi|Y) &\propto \phi^{n/2}\exp\{-\frac{1}{2}\phi\sum_i (y_i-\mu)^2\} \\
&\propto \phi^{n/2}\exp\{-\frac{1}{2}\phi\sum_i [(y_i-\bar{y})-(\mu-\bar{y})]^2\} \\
&\propto \phi^{n/2}\exp\{-\frac{1}{2}\phi\left[\sum_i (y_i-\bar{y})^2+n(\mu-\bar{y})^2\right]\} \\
&\propto \phi^{n/2}\exp\{-\frac{1}{2}\phi s^2(n-1)\}\exp\{-\frac{1}{2}\phi n(\mu-\bar{y})^2\}
\end{aligned}
$$

$\bar{y}=\sum_{i=1}^{n} y_i$ is sample mean.
$s^2=\sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)$ is the sample variance.

Likelihood for Normal Model

$$L(\mu, \phi|Y) \quad \propto \quad \phi^{n/2} \exp\{-\frac{1}{2}\phi s^2(n-1)\} \exp\{-\frac{1}{2}\phi n(\mu - \bar{y})^2\}$$

Sufficient statistics:

- sample mean $\bar{y}$
- sample sum of squares SS $= s^2(n-1) = \sum_i (y_i - \bar{y})^2$

MLEs:

- $\hat{\mu} = \bar{y}$
- $\hat{\phi} = n/\text{SS}$, and $\hat{\sigma}^2 = \text{SS}/n$

Conjugate Prior Distribution

Recall the gamma distribution has a kernel of $\theta^{a-1}e^{-\theta b}$. The likelihood on the previous slide looks a lot like a gamma kernel times a normal kernel. So....

The conjugate prior for (μ, ϕ) is Normal-Gamma.

$$\begin{aligned} \mu|\phi &\sim \mathsf{N}(\mu_0, 1/(\kappa_0\phi)) \\ \phi &\sim \mathsf{G}(v_0/2, \mathsf{SS}_0/2) \end{aligned}$$

where $-\infty < \mu_0 < \infty, \kappa > 0, \mathsf{SS}_0 > 0, v_0 > 0$

$$p(\mu, \phi) \propto \phi^{v_0/2-1} \exp\{-\phi \frac{\mathsf{SS}_0}{2}\} \phi^{1/2} \exp\{-\phi \frac{\kappa_0}{2} (\mu - \mu_0)^2\}$$

Note: book uses $\sigma_0^2 = \mathsf{SS}_0/v_0$.

Gamma and Inverse Gamma

$$\phi \sim \mathsf{G}(v_0/2, \mathsf{SS}_0/2)$$

is equivalent to

$$1/\sigma^2 \sim \mathsf{G}(v_0/2, \mathsf{SS}_0/2)$$

We say that σ^2 has an <u>inverse Gamma</u> distribution

Updating the Posterior Parameters

Under the Normal-Gamma prior distribution:

$$\mu \mid \phi, Y \sim \mathsf{N}\left(\mu_n, \frac{1}{\kappa_n \phi}\right)$$
$$\phi \mid Y \sim \mathsf{G}\left(\frac{v_n}{2}, \frac{\mathsf{SS}_n}{2}\right)$$

where

$$\begin{aligned} \kappa_n &= \kappa_0 + n \\ \mu_n &= \frac{\phi n \bar{y} + \phi \kappa_0 \mu_0}{\phi \kappa_n} \\ v_n &= v_0 + n \\ \mathsf{SS}_n &= \mathsf{SS}_0 + \mathsf{SS} + \frac{n \kappa_0}{\kappa_n} (\bar{y} - \mu_0)^2 \end{aligned}$$

Interpretation

- κ_n: like sample size for estimating μ (precision $= \phi\kappa_n$)
- μ_n: expected value for μ is weighted average

$$\mu_n = \frac{n}{\kappa_n}\bar{y} + \frac{\kappa_0}{\kappa_n}\mu_0$$

- v_n: degrees of freedom for estimating ϕ

$$\phi \sim G(a/2, b/2) \Leftrightarrow \phi b \sim \chi^2_a \text{ with degrees of freedom } a$$

- $\mathsf{SS}_n = \mathsf{SS}_0 + \mathsf{SS} + \frac{n\kappa_0}{\kappa_n}(\bar{y} - \mu_0)^2$: total posterior variation
  - prior variation,
  - observed variation (sum of squares),
  - variation between prior mean and sample mean

Noninformative Prior

Let $p(\mu, \phi) = 1/\phi$, or equivalently $p(\mu, \sigma^2) = 1/\sigma^2$.
This is the limiting case where $\kappa_0 = 0$ and $\nu_0 = 0$.

This is an improper prior distribution. But, it leads to a proper posterior distribution that yields inferences similar to the frequentist ones. We have

$$\mu \mid \phi, Y \sim \mathsf{N}\left(\bar{y}, \frac{1}{n\phi}\right)$$

$$\phi \mid Y \sim \mathsf{G}(\frac{n}{2}, \frac{\mathsf{SS}}{2})$$

Pygmalion: Questions of Interest

- Is the average improvement for the accelerated group larger than that for the no growth group? We want $P(\mu_A > \mu_N | Y_A, Y_N)$
- Is the variance of improvement scores for the accelerated group larger than that for the no growth group? We want $P(\sigma_A^2 > \sigma_N^2 | Y_A, Y_N)$
- What are the posterior distributions of the coefficient of variation in each group? We want $P(\mu_A/\sigma_A | Y_A)$ and $P(\mu_N/\sigma_N | Y_N)$
- What is the probability that a randomly selected child assigned to the accelerated group will have larger improvement than a randomly selected child assigned to the no growth group? We want $P(Y_A^* > Y_N^* | Y_A, Y_N)$

Analysis with Noninformative Priors

Summary statistics ($n_A = n_B = 6$)

- $\bar{y}_A = 15.2$ and $\bar{y}_N = 6.2$. $SD_A = 4.7$ and $SD_N = 3.6$
- $SS_A = 4.7^2 * (6-1) = 110.45$. $SS_B = 3.6^2 * (6-1) = 64.8$

Posterior distributions:

$$\mu_A \mid \phi_A, Y_A \sim \mathsf{N}\left(15.2, \frac{1}{6\phi_A}\right)$$
$$\phi_A \mid Y_A \sim \mathsf{G}\left(\frac{6}{2}, \frac{110.45}{2}\right)$$

$$\mu_N \mid \phi_N, Y_N \sim \mathsf{N}\left(6.2, \frac{1}{6\phi_N}\right)$$
$$\phi_N \mid Y_N \sim \mathsf{G}\left(\frac{6}{2}, \frac{64.8}{2}\right)$$

Informative Priors for (μ_A, σ_A^2), (μ_B, σ_B^2)

- Suppose no predisposition if students should improve on average: then, prior medians equal zero. Set $\mu_{0A} = \mu_{0B} = 0$
- Suppose you don't have a lot of faith in this belief, and think it is the equivalent of having only 1 additional observation in each group. Set $\kappa_{0A} = \kappa_{0N} = 1$.
- Suppose you think SD of change scores should be around 10 in each group. And you don't have a lot of faith in this belief. Set $v_{0A} = v_{0A} = 1$. Set $\mathsf{SS}_{0A} = 10^2 v_{0N} = 100$ and $\mathsf{SS}_{0B} = 10^2 v_{0B} = 100$.

Graph priors to see if they accord with your beliefs. Sampling new values of Y from the priors offers a good check.

Analysis with Informative Priors

Posterior hyperparameters for Accelerated group

$$
\begin{aligned}
\kappa_{nA} &= 1 + 6 = 7 \\
\mu_{nA} &= [(6)(15.2) + (1)(0)]/7 = 13.02 \\
v_{nA} &= 1 + 6 = 7 \\
\mathsf{SS}_{nA} &= 100 + 110.45 + (16.5 - 0)^2(6)(1)/7 = 443.81
\end{aligned}
$$

Using posterior on slide 10, we have

$$
\begin{aligned}
\mu_A \mid \phi_A, Y_A &\sim \mathsf{N}\left(13.02, \frac{1}{7\phi_A}\right) \\
\phi_A \mid Y_A &\sim \mathsf{G}\left(\frac{7}{2}, \frac{443.81}{2}\right)
\end{aligned}
$$

Samples from the Posterior

$$\mu_A \mid \phi_A, Y_A \sim \mathsf{N}\left(13.02, \frac{1}{7\phi_A}\right)$$

$$\phi_A \mid Y_A \sim \mathsf{G}\left(\frac{7}{2}, \frac{443.81}{2}\right)$$

Draw samples of $(\mu^{(i)}, \phi^{(i)}), i = 1, 2, \ldots, S$ from the posterior distribution

- Draw $\phi \mid Y$

```
phi = rgamma(10000, 7 / 2, rate = 443.81 / 2);
```

- Draw $\mu \mid \phi, Y$

```
mu = rnorm(10000, 13.02, 1 / sqrt(7 * phi));
```

Marginal Distribution for $\mu \mid Y$

$$
\begin{aligned}
p(\mu \mid Y) &= \int p(\mu, \phi \mid Y) d\phi = \int p(\mu \mid \phi, Y) p(\phi \mid Y) d\phi \\
&= \int (2\pi)^{-1/2} (\frac{1}{\kappa_n \phi})^{-1/2} \exp\left(-\kappa_n \phi (\mu - \mu_n)^2 / 2\right) \\
&\times \frac{(\mathsf{SS}_n/2)^{v_n/2}}{\Gamma(v_n/2)} \phi^{v_n/2-1} \exp\left(-\phi \mathsf{SS}_n / 2\right) d\phi \\
&\propto \int \phi^{\frac{v_n+1}{2}-1} \exp[-\phi \left\{ \frac{\mathsf{SS}_n + \kappa_n(\mu - \mu_n)^2}{2} \right\}] d\phi \\
&\propto \left[\frac{\mathsf{SS}_n + \kappa_n(\mu - \mu_n)^2}{2} \right]^{-(v_n+1)/2}
\end{aligned}
$$

Student t Distribution

X has a Student t distribution with location μ, scale s and degrees of freedom v if

$$p(x \mid v, \mu, s) \propto \left[1 + \frac{1}{v}\left(\frac{x-\mu}{s}\right)^2\right]^{-(v+1)/2}$$

Rearrange posterior distribution:

$$p(\mu \mid Y) \quad \propto \quad \left[\frac{\mathsf{SS}_n + \kappa_n(\mu - \mu_n)^2}{2}\right]^{-(v_n+1)/2}$$

Student $t_{v_n}(\mu_n, s_n)$ location μ_n, df $= v_n$, square of scale $s_n^2 = \frac{1}{\kappa_n}\frac{\mathsf{SS}_n}{v_n}$

Standard Student t

Standardize $X \sim t_v(\mu, s)$ by subtracting location and dividing by square root of the scale:

$$\frac{X - \mu}{s} \sim t_v(0, 1)$$

(new location 0 and scale 1)

$$\Rightarrow \frac{\mu - \mu_n}{s_n} \sim t_{v_n}(0, 1)$$

$$\mu \stackrel{D}{=} \mu_n + t_{v_n} s_n$$

Use `rt`, `qt`, `pt`, `dt` in `R`