Lecture 6: The Normal Model

Professor Alexander Franks

2023-02-07

Announcements

- Reading: Section 5.3.3 and 5.3.4 (Normal model, start after exam).

+ Quiz 2, out, due 2pm Wed.

+ No Thurs Section.

+ Lauren: Wed 2-4, Thurs 8:30-9:30. (SH 542)

Doris: Zoom Wed 5-6.

- Likelihood, simplify, drop normalizing constant.
- Conjugate Priors, meaning of the parameters (matching units)
- Prior & Posterior Mean for Beta / Gamma.

  $\frac{\alpha}{\alpha+\beta}$ / α/β
- write posterior mean as $w\hat{\theta}_{MLE} + (1-w)\theta_{prior}$
  - + How we derived that.
  - + what w is
- Basic idea behind M.C.
  - + Inversion / Rejection sampling.
- posterior Predictive Distn.

  $$P(\tilde{y}|y) = \int P(\tilde{y}|\theta)P(\theta|y)\,d\theta$$

- MC: sample $\theta^s \sim p(\theta|y)$
sample $\tilde{y} \sim p(\tilde{y}|\theta^s)$

- Uncertainty Quant.
  - Quantile and HPD strategies

$$p(\tilde{y}|\cancel{y}) = \int p(\tilde{y}|\theta)\,p(\theta|\cancel{y})\,d\theta$$

Prior Predictive.

$y \sim Bin(n, \theta)$

$Beta(a, b) \longrightarrow (\ )\,\theta^{a-1}(1-\theta)^{b-1}$

$$\theta^{y}(1-\theta)^{n-y} \times \theta^{a-1}(1-\theta)^{b-1}$$

$$(\ )\,\theta^{y+a-1}(1-\theta)^{b-1}$$

The Normal Distribution

- One of the most utilized probability models in data analysis
- Central Limit Theorem
- Separate parameters for the mean and the variance (intuitive)

mean

variance

$$Y \sim N(\mu, \sigma^2)$$

$Z_1, \ldots Z_n \sim$ iid

- 2 parameter model

$\bar{Z} \approx$ normal.

Need: $P(\mu, \sigma^2)$

Get: $P(\mu, \sigma^2 \mid y_1, \ldots y_n)$

Normal Distribution

- Symmetric with mode = median = mean = μ
- Approximately 95% of the population lies within two standard deviations of the mean

$\bar{y} \pm 1.96\hat{\sigma}$ (95% CI)

- Density:

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

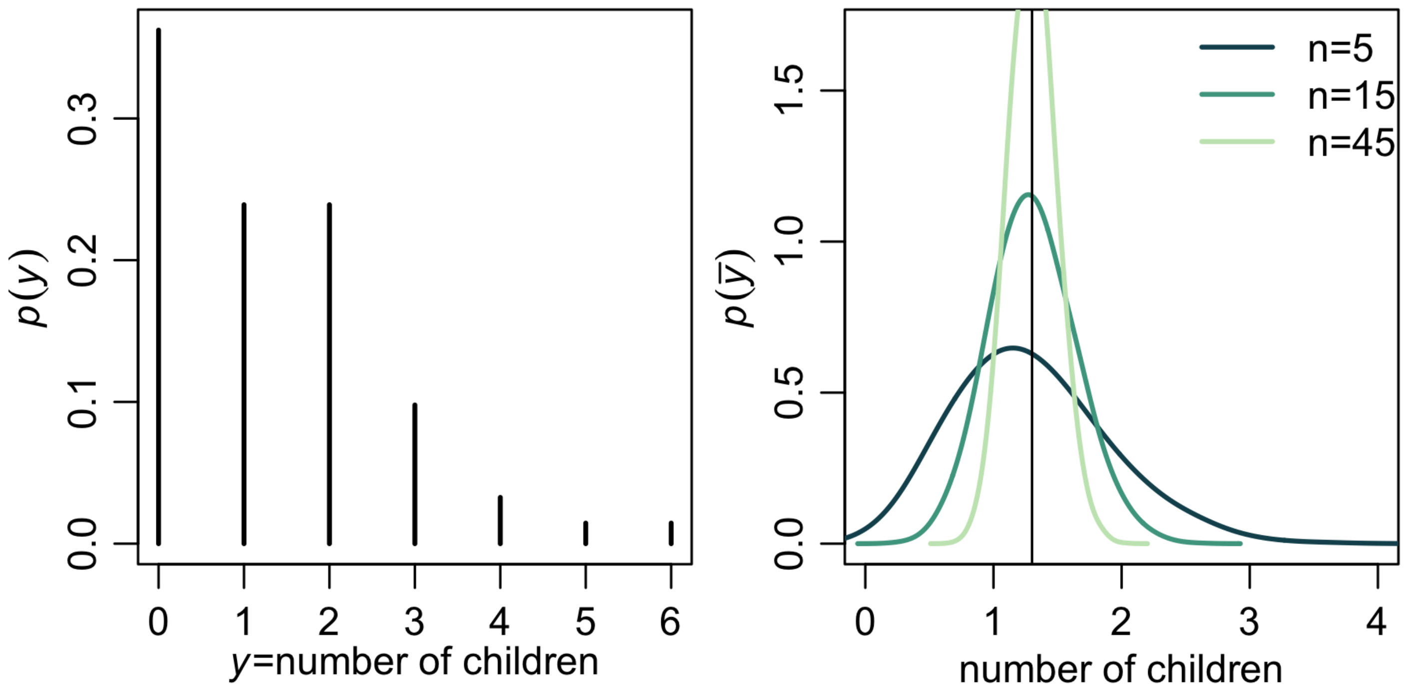
- $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ with X and Y independent then

$$aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$$

$E[aX+bY] = aE[x] + bE[y]$

- In R: `dnorm, rnorm, pnorm, qnorm`.
  - Warning: the argument to the `norm` functions `R` is σ not σ^2!

The Central Limit Theorem



$Var\left(\frac{\sum Y_i}{n}\right) = \frac{1}{n^2} n \, Var(Y_i) = \frac{\lambda}{n}$

CLT: $\bar{y} \approx N(E[Y], \text{Var}[Y]/n)$

$Y_i \sim Pois(\lambda)$; $\bar{Y} = \frac{\sum Y_i}{n} \approx N\left(\lambda, \frac{\lambda}{n}\right)$

Bayesian inference in the normal model

- Assume $y_1, \dots y_n \sim N(\mu, \sigma^2)$ with σ^2 a known constant
- Lets start with a non-informative, improper prior: $p(\mu) \propto$ const "flat"
- What is the posterior distribution $p(\mu \mid y_1, \dots y_n, \sigma^2)$?

$$P(\mu \mid \sigma^2, y_1, \dots y_n) \propto L(\mu) \cancel{\times \text{const}}$$

$$P(\mu \mid \cdots) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

$$\propto \cancel{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n} e^{-\frac{\sum^n (y_i-\mu)^2}{2\sigma^2}}$$

$$\propto e^{-\frac{(\sum y_i^2 - 2\mu\sum y_i + n\mu^2)}{2\sigma^2}}$$

(expand the sq)

$$\propto \cancel{e^{-\frac{\sum y_i^2}{2\sigma^2}}} \; e^{\frac{2\mu\sum^n y_i - n\mu^2}{2\sigma^2}}$$

(complete the sq)

Aside

$$(ax^2 - bx) = a\left(x^2 - \frac{b}{a}x\right)$$

$$= a\left(x - \frac{b}{2a}\right)^2 - \frac{b^2}{4a}$$

$a = n$
$b = 2\sum y_i$
$x = \mu$

$$\propto e^{\frac{-n\mu^2 - 2\mu\sum y_i}{2\sigma^2}}$$

$$\propto e^{\frac{-n\left(\mu - \frac{\sum y_i}{n}\right)^2}{2\sigma^2}} \; \cancel{e^{\frac{-(2\sum y_i)^2}{4n}}}$$

$$\propto e^{\frac{-(\mu - \bar{y}_\cdot)^2}{2\sigma^2/n}}$$

$$P(\mu \mid y_1, \ldots y_n) \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

$$E[\mu / y_1, \ldots y_n] = \bar{y} = \hat{\mu}_{MLE}$$

$$Var(\mu / y_1, \ldots y_n) = \sigma^2/n$$

$$\int_{-\infty}^{\infty} e^{-\frac{(\mu - \bar{y})^2}{2\sigma^2/n}} d\mu < \infty$$

Bayesian inference in the normal model

- Assume $y_1, \ldots y_n \sim N(\mu, \sigma^2)$ with σ^2 a known constant
- The normal prior distribution is conjugate for μ in the normal sampling model
- Sampling distribution, prior distribution and posterior distribution are all normal.

Prior guess about μ (mean)

How certain

- Assume the prior is $p(\mu) \sim N(\mu_0, \tau^2)$
- What are the parameters of the posterior $p(\mu \mid y_1, \ldots y_n, \sigma^2)$?

Normal

$$P(\mu|y_1,\dots y_n,\sigma^2) \propto L(\mu)P(\mu)$$

$$\propto \underbrace{e^{-\frac{(\mu-\bar y)^2}{2\sigma^2/n}}}_{L(\mu)} \cancel{\frac{1}{\sqrt{2\pi\tau^2}}} \underbrace{e^{-\frac{(\mu-\mu_0)^2}{2\tau^2}}}_{P(\mu)}$$

$$\propto \exp\left[-\left(\frac{\cancel{\bar y^2} - 2\bar y\mu + \mu^2}{2\sigma^2/n} + \frac{\mu^2 - 2\mu\mu_0 + \cancel{\mu_0^2}}{2\tau^2}\right)\right]$$

$$\propto \exp\left[-\frac12\left(\frac{n}{\sigma^2}\mu^2 - \frac{2n}{\sigma^2}\bar y\mu + \frac{1}{\tau^2}\mu^2 - \frac{2\mu\mu_1}{\tau^2}\right)\right]$$

$$\propto \exp\left[-\frac12\left(\underbrace{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}_{a}\mu^2 - 2\underbrace{\left(\frac{n\bar y}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)}_{b}\mu\right)\right]$$

$$\propto \exp\left[-\frac12 \underbrace{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}_{1/\text{posterior var.}} \left(\mu - \underbrace{\frac{\frac{n\bar y}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}}_{\text{Post. Mean}}\right)^2\right]$$

(Post mean)

MLE

Prior Mean.

$$E[\mu|y_1,\dots y_n] = \frac{\frac{n}{\sigma^2}\bar y}{n/\sigma^2+1/\tau^2} + \frac{\frac{1}{\tau^2}\mu_0}{n/\sigma^2+1/\tau^2}$$

$w \qquad\qquad 1-w$

$$= w\bar{y} + (1-w)\mu_0$$

$$w = \frac{n/\sigma^2}{n/\sigma^2 + \frac{1}{\tau^2}}$$

$$Var(\mu \mid y_1, \ldots, y_n) = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

$\frac{n}{\sigma^2}$ is precision of $\bar{y}$

$1/\tau^2$ is prior precision for μ.

$$p(\mu \mid \ldots) \sim N(\mu_n, \sigma_n^2)$$

$$\mu_n = w\bar{y} + (1-w)\mu_0$$

$$w = \frac{n/\sigma^2}{n/\sigma^2 + 1/\tau^2}$$

$$\sigma_n^2 = w\frac{\sigma^2}{n}$$

A conjugate prior for the normal likelihood

- The normal distribution is conjugate for the normal likelihood
    - Often called the "normal-normal model"
- $Y_i \sim N(\mu, \sigma^2)$ and $\mu \sim N(\mu_0, \tau^2)$ implies that the posterior distribution $p(\mu \mid y)$ is also normally distributed:

$$\mu \mid Y \sim N(\mu_n, \tau_n^2)$$

where $\mu_n = \frac{\frac{1}{\tau^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$ and $\tau_n^2 = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$

The posterior mean and pseudo-counts

$$\mu_n = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\overline{y}$$
$$= (1 - w)\mu_0 + w\bar{y}$$

where $w = \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}$

Can we think about the normal prior parameters in terms of pseudo-counts?