Lecture 6: The Normal Model

Professor Alexander Franks

2024-02-21

Announcements

- Reading: Section 5.3.3 and 5.3.4

The Normal Distribution

- One of the most utilized probability models in data analysis
- Central Limit Theorem $\to Z_1, \ldots Z_n$ iid, $\bar{Z} \sim$ normal
- Separate parameters for the mean and the variance (intuitive)

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

2 parameter model

Specify $P(\mu, \sigma^2)$ (prior)

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

Normal Distribution

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

$\mu \pm 1.96\sigma$

- Density:

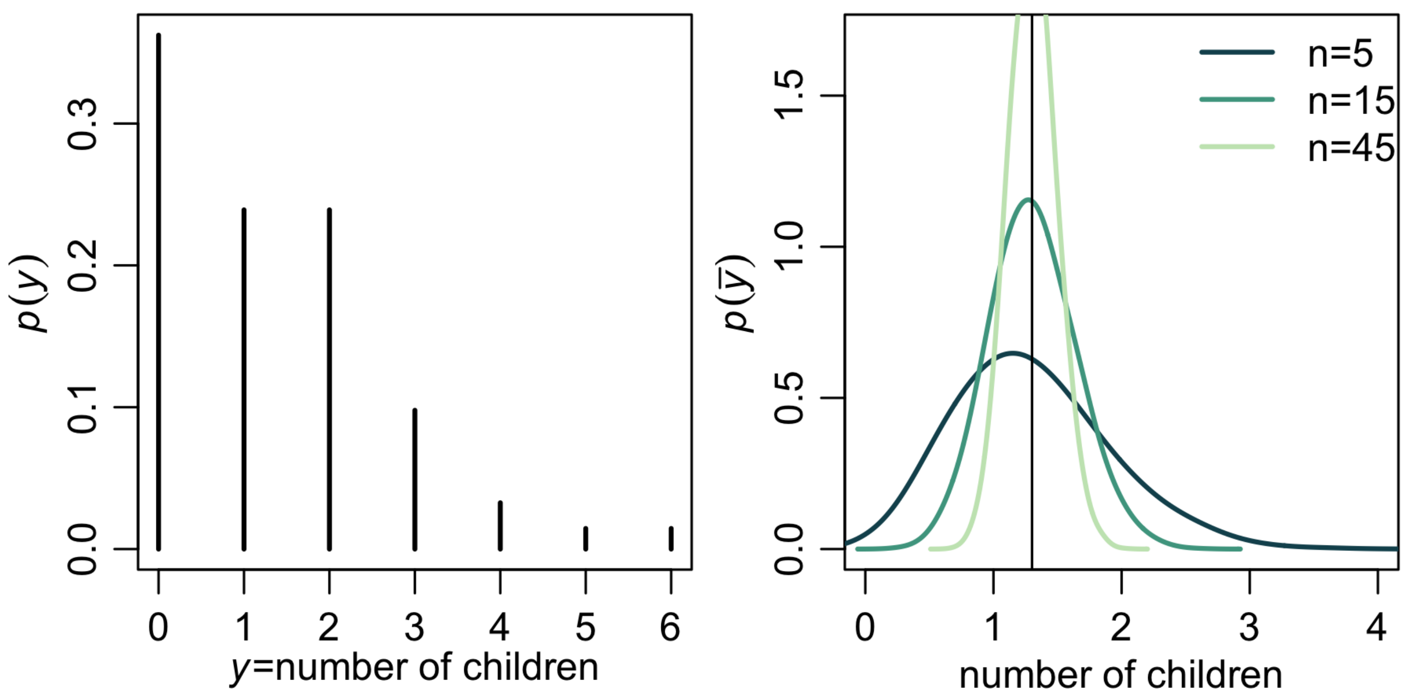
$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \quad -\infty < y < \infty$$

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

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

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

The Central Limit Theorem



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

Bayesian inference in the normal model

- Assume $y_1, \ldots y_n \sim N(\mu, \sigma^2)$ with σ^2 a known constant

  $=1$

- Lets start with a non-informative, improper prior: $p(\mu) \propto \text{const}$

  improper

- What is the posterior distribution $p(\mu \mid y_1, \ldots y_n, \sigma^2)$?

$$\int_{-\infty}^{\infty} p(\mu) = \infty$$

but we get a proper posterior.

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

$$L(\mu)\, P(\mu) \propto$$

$$\prod_{i=1}^{n} \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}}\, e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} \times \cancel{\text{const}}$$

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

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

$$\propto \cancel{e^{-\sum y_i^2}}\, e^{-(n\mu^2 - 2\mu\sum y_i)}$$

Aside: Complete sq.

$$\boxed{\begin{aligned}(ax^2 - bx) &= a\left(x^2 - \tfrac{b}{a}x\right)\\ &= a\left(x - \tfrac{b}{2a}\right)^2 - \tfrac{b^2}{4a}\end{aligned}}$$

$$\propto e^{-\left(n\mu^2 - 2\mu\sum y_i\right)}$$

$$\propto e^{-\frac{n(\mu-\bar{y})^2}{2\sigma^2}} \quad \cancel{e^{\text{"}-\frac{b^2}{4a}\text{"}}}$$

$a = n$
$b = \Sigma y_i$
$x = \mu$

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

$$\Longrightarrow P(\mu \mid y_1, \ldots y_n)$$

$$\sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

What is the conjugate prior for μ?

Ans: normal

$$y_1, \ldots, y_n \sim N(\mu, \sigma^2)$$

$$p(\mu) \sim N\left(\mu_0, \frac{\sigma^2}{K}\right)$$

$$P(\mu \mid y_1, \ldots, y_n) \propto$$

$$L(\mu)\, P(\mu) \propto$$

$$\underbrace{e^{-\frac{(\mu-\bar{y})^2}{\sigma^2/n}}}_{L(\mu)} \; \underbrace{e^{-\frac{(\mu-\mu_0)^2}{\sigma^2/K}}}_{P(\mu)}$$

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

expand the square

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

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

$$\boxed{\begin{aligned}
&P(\mu \mid y_1, \ldots, y_n) \sim N(\mu_n, \sigma_n^2) \\
&\mu_n = \frac{n}{n+K}\bar{y} + \frac{K}{n+K}\mu_0 \\
&\sigma_n^2 = \frac{\sigma^2}{n+K}
\end{aligned}}$$

Bayesian inference in the normal model

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

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:

$$\tau^2 = \frac{\sigma^2}{K}$$

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

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

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

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?

The posterior mean and pseudo-counts

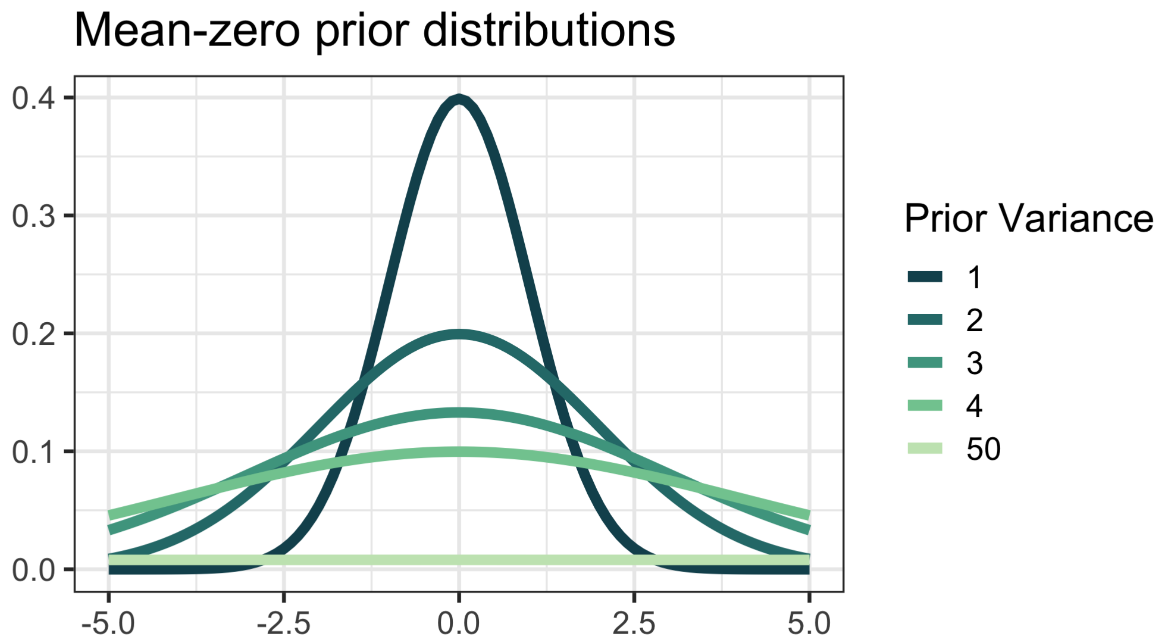
$$\mu_n = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\bar{y}$$
$$= (1 - w)\mu_0 + w\bar{y}$$

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

- Let's reparameterize: $\tau^2 = \frac{\sigma^2}{\kappa_0}$
- Then: the posterior variance is $\frac{\sigma^2}{\kappa_0+n}$
- And: $(1 - w) = \frac{\kappa_0}{\kappa_0+n}$
- κ_0 are the prior counts and μ_0 is the prior sample average.

Conjugate prior with increasing variance

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning wa
## generated.
```



Estimators: Bayes / Frequentist Unification

- Bayesian inference provides a straightforward procedure for producing estimators given your prior beliefs.

    1. Compute posterior distribution

    2. Summarize the posterior distribution with a point estimator (e.g. posterior mean or posterior mode) and a probability interval

- Frequentists provide tools for evaluating the sampling properties of an estimator.

    - Bias, variance and MSE of an estimator
    - Well-calibrated probability intervals

- Both are useful!

The Bias-Variance Tradeoff

Reminder: an estimator is a random variable, an estimate is a constant

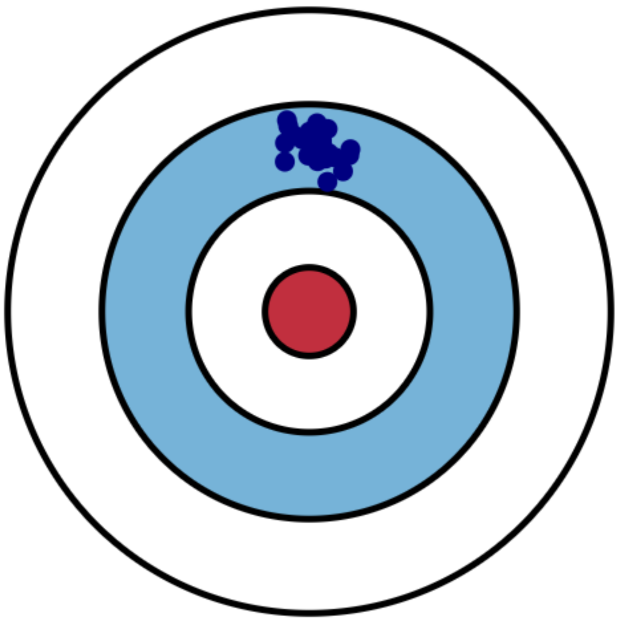
- *Bias*: systematic sampling error of the estimator
- *Variance*: variance of the estimator (from sampling & measurement error)
- Often we evaluate an estimator in terms of mean square error: $\text{MSE}(\hat{\theta}) = E_Y(\hat{\theta} - \theta)^2$
- The Bias-Variance tradeoff: $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$

The Bias-Variance Tradeoff

- Variance of an estimator comes sampling from a population
  - If you were to repeatedly draw new samples of the same size how much would your estimates vary?
  - e.g. if $y_i \sim N(\mu, \sigma^2)$ then $\mathrm{Var}(\bar{Y}) = \sigma^2/n$

Bias

The expected difference between the estimate and the response

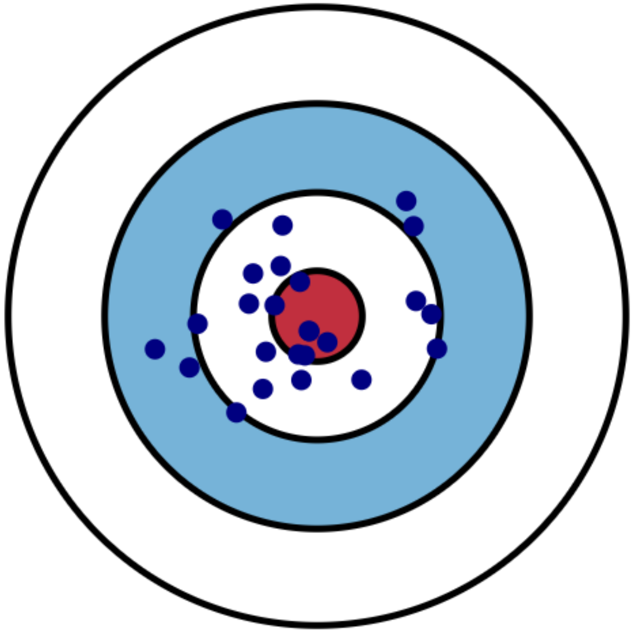


Statistical definition of bias:

$$E_Y[\hat{\theta} - \theta]$$

Variance

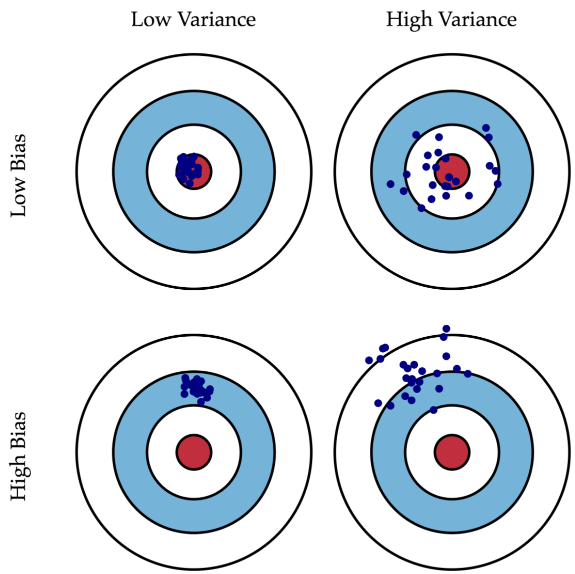
How variable is the prediction about its mean?



Statistical definition of variance:

$$E_Y[\hat{\theta} - E_Y[\hat{\theta}]]^2$$

Bias and Variance



$$\underbrace{\mathrm{MSE}(\hat{\theta})}_{\text{accuracy}} = \underbrace{\mathrm{Var}(\hat{\theta})}_{\text{variance}} + \underbrace{\Big(E[\hat{\theta} - \theta]\Big)^2}_{\text{bias squared}}$$

The Bias-Variance Tradeoff

- The prior distribution (usually) makes your estimator biased...
- But the prior distribution also (usually) reduces the variance!
- Example: compute the frequentist mean and variance of the posterior mean.

$$Y_1, \ldots Y_n \sim N(\mu, \sigma^2)$$

$$\hat{\mu}_{MLE} = \bar{Y}, \quad E[\bar{Y}] = \mu \text{ (unbiased)}$$

$$Var(\bar{Y}) = \frac{\sigma^2}{n} = MSE$$

$$\hat{\mu}_{\substack{post. \\ mean}} = w\bar{Y} + (1-w)\mu_0$$

Bias: $E[\hat{\mu} - \mu] = E[w\bar{Y} + (1-w)\mu_0 - \mu]$

$$= wE[\bar{Y}] + (1-w)\mu_0 - \mu$$

$$= w\mu + (1-w)\mu_0 - \mu$$

$$= (1-w).(\mu_0 - \mu)$$

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

$$w^2 \operatorname{Var}(\bar{Y}) = \frac{w^2 \sigma^2}{n}$$

$$\leq \sigma^2/n$$

$$MSE: \frac{w^2\sigma^2}{n} + (1-w)^2(\mu_0 - \mu)^2$$