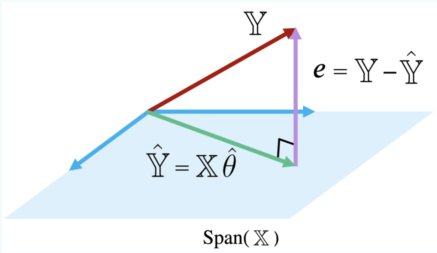
STAT 135 Lecture 19

Henry Liev

6 August 2025

Remark 0.1



$y_i \approx \beta_0 + x_{i1} + x_{i2} + \cdots + x_{ip}$
$\hat{\beta}$ minimizing $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2$
$d(\mathbb{Y}, \mathbb{X}) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_p - x_p)^2}$
$\hat{\mathbb{Y}} = x\hat{\beta}$
$\mathbb{X}^\intercal(\mathbb{Y} - \hat{\mathbb{Y}}) = 0$
$\mathbb{X}^\intercal\mathbb{Y} = (\mathbb{X}^\intercal\mathbb{X})\hat{\beta} \to \hat{\beta} = (\mathbb{X}^\intercal\mathbb{X})^{-1}\mathbb{X}\mathbb{Y} = \hat{\beta}$
Even if we do not have full rank the diagram will still look the same

Remark 0.2 (Reasons for being Bayesian)

Flip a "possibly biased" coin 5 times and get 0 heads. What is your best guess at p?
MLE $\to$ 0 is best guess
$\hat{p}$ is good in many ways: MLE, Unbiased, Minimum Variance
Frequentists approach ignores relevant information

- Symmetry of coin
- Historical evidence

Example 0.3

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ with known variance, interested in the mean
Precision $= \frac{1}{\text{Variance}}$
$\Theta \sim \mathcal{N}(\theta_0, \xi_{\text{prior}}^{-1})$
$\sigma^2 = \frac{1}{\xi_{\text{prior}}}$
$X = (x_1, \ldots, x_n)$

$$f_{\Theta|X}(\theta|X) \propto f_{X|\Theta}(x|\theta) f_\Theta(\theta)$$

$$\begin{aligned} f_{X|\Theta}(x|\theta) \cdot f_\Theta(\theta) &= \left(\frac{\xi_0}{2\pi}\right)^{\frac{n}{2}} \prod_{i=1}^{n} \exp\left(-\frac{\xi_0}{2}(x_i-\theta)^2\right) \cdot \left(\frac{\xi_{\text{prior}}}{2\pi}\right)^{\frac{1}{2}} \frac{1}{2} \exp\left(-\frac{\xi_{\text{prior}}}{2}(\theta-\theta_0)^2\right) \\ &\propto \exp\left[-\frac{1}{2}\left(\xi_0 \sum_{i=1}^{n}(x_i-\theta)^2 + \xi_{\text{prior}}(\theta-\theta_0)^2\right)\right] \\ &= \left[-\frac{1}{2}\left(\xi_0 \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\theta-x_0)^2 + \xi_{\text{prior}}(\theta-\theta_0)^2\right)\right] \end{aligned}$$

Get exp(something quadratic in θ) times some other stuff
$\exp\left(-\frac{\xi_{\text{post}}}{2}(x-\theta_{\text{post}})^2\right)$
Posterior is also normal
$\xi_{\text{post}} = n\xi_0 + \xi_{\text{prior}}$

$$\theta_{post} = \bar{x}\frac{n\xi_0}{n\xi_0 + \xi_{\text{prior}}} + \theta_0 \frac{\xi_{\text{prior}}}{n\xi_0 + \xi_{\text{prior}}}$$

- Posterior distribution $\Theta|X$ also normal
- Posterior mean is weighted combination of $\bar{X}$ & θ_0 with weight on $\bar{X} \to 1$ as $n \to \infty$
- Prior is flat $\to \xi_{\text{post}}^{-1} \approx \frac{\sigma^2}{n}$

With flat prior ($\xi_{\text{prior}} \approx 0$) or large data, result is similar to frequentist approach

Definition 0.4 (Conjugate Prior)

- Prior from family G
- Conditional on parameters of G, data has distribution H
- Posterior distribution also from family G
  "G is conjugate to H", "Beta conjugate to binomial, normal conjugate to normal"

Remark 0.5 (Some issues with priors)

- In previous example, what if we don't know $\hat{\mu}$?
- Could do Bayesian analysis (Θ, Σ), We'd need a prior on joint variables
- Tricky to specify prior joint distribution of all unknowns
  Sometimes we want a non-informative prior

Coin flipping $\to \theta \sim U(0,1)$
Knowing nothing about $\theta \Leftrightarrow \theta \sim U(0,1)$
θ^2 is not uniform on $U(0,1)$ if θ is
Same information about θ and θ^2
Try to model
non-informative = flat
Problem: Flat depends on parameterization
Not straightforward even to "model", "non-informative"

Remark 0.6 (Statistics pulls into two different directions)

1. Very strong assumptions and express questions & answers very clearly (Bayesian)
2. Weak assumptions but we have to accept indirect answers

Method 0.7 (Permutation Test)
$X_1, X_2, \dots, X_n, \sim F$
$Y_1, Y_2, \dots, Y_m \sim G$
Test F=G?
Mann-Whitney, Data $\to$ Ranks, $\binom{m+n}{n}$ assignments of the treatment groups are equally likely
Refer to the known distribution of Rank Sum (tables) or normal approximation
Look at all $\binom{n+m}{m}$ ways to assign m to treatment
Don't need to use all permutations, just take some B of them at random
Some true p-value P, we would get from all permutations, $T = \bar{Y} - \bar{X}, P(T > T_{obs})$
$\hat{p}$ = fraction of B permutations for which $T > T_{obs}$
$E(\hat{p}) = p, Var(\hat{p}) = \frac{p(1-p)}{B}$