STAT 135 Lecture 12

Henry Liev

22 July 2025

Remark 0.1

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$
Probability of observing what you observed, as a funciton of θ
MLE = pick $\hat{\theta}$ to be the θ for which the data has the highest probability
Probability of what you observed:
$f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta)$
$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$
$l(\theta) = \sum_{i=1}^{n} \log(f(x_i|\theta))$
$I(\theta) = -E(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta)))$
$I_n(\theta) = -E(\frac{\partial^2}{\partial\theta^2}\log(f(x_i, \ldots, x_n))) = nI(\theta)$

Example 0.2

$Poisson(\lambda)$
$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$

Remark 0.3 (Two-Sample Problem Continued)

$X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_x, \sigma^2)$
$Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_y, \sigma^2)$
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_x - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m}))$
$\bar{X} - \bar{Y} \pm 1.96SE(\bar{X} - \bar{Y})$
$\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim \mathcal{N}(0, 1^2)$

Example 0.4 (Same problem but don't know σ^2)

$\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$
$S_p^2 = \frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$
$\bar{X} - \bar{Y} \pm t_{n+m-2}(\frac{\alpha}{2})S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$

Remark 0.5 ("t-test requires normal data")

In practice, t-test "works" even when data not very normal "robust"
Simulation studies from data that is not quite exactly normal shows that the "t-test" still "works" for non normal data

Example 0.6 (Two large samples)

$X_1, \ldots, X_n \overset{iid}{\sim} E(X_i) = \mu_X, Var(X_i) = \sigma_X^2$
$Y_1, \ldots, Y_m \overset{iid}{\sim} E(Y_i) = \mu_Y, Var(Y_i) = \sigma_Y^2$
m and n large, inference about $\mu_X = \mu_Y$
$\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma_X^2}{n}), \bar{Y} \approx \mathcal{N}(\mu, \frac{\sigma_Y^2}{m})$
$\bar{X} - \bar{Y} \approx \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$
$\bar{X} - \bar{Y} \pm 1.96 SE(\bar{X} - \bar{Y})$ approximate interval

Remark 0.7 (Confidence Intervals)

Data Assumption	Sample Size	Variance Known	Variance Equal	Approach	Exact/ Approx
Normal	Any	Yes	Yes or No	Z-test	Exact
Normal	Any	No	Yes	t-test	Exact
Any	Large	No	No	Z-test	Approx
Normal	Any	No	No	t-test	Approx

Method 0.8 (Sample Size Calculations)

Planning to collect data
$X_1, \ldots, X_n$ mean μ_X, Variance σ^2
$Y_1, \ldots, Y_n$ mean μ_Y, Variance σ^2
$H_0 : \mu_X = \mu_Y$
$H_1 : \mu_X \neq \mu_Y$
Assume that n will be large enough for "large sample"

Solution. Significance level: $\alpha = 0.05$
Power: 80%
Pick Δ we would like to be able to detect
$H_0 : \mu_X = \mu_Y$
$H_1 : \mu_X - \mu_Y = \Delta$
5% level with 80% chance to detect this Δ
Under $H_0 : \bar{X} - \bar{Y} \approx \mathcal{N}(0, \frac{2\sigma^2}{n})$
$|\bar{X} - \bar{Y}| \geq 1.96\sigma\sqrt{\frac{2}{n}}$
Choose n such that 80% power when $\mu_x - \mu_y = \Delta$
$P(|\bar{X} - \bar{Y}| \geq 1.96\sigma\sqrt{\frac{2}{n}}) = 0.8$ under $H_0 : \mu_X - \mu_Y = \Delta$
$P(\bar{X} - \bar{Y} \leq -1.96\sigma\sqrt{\frac{2}{n}}) + P(\bar{X} - \bar{Y} \geq 1.96\sigma\sqrt{\frac{2}{n}})$
$P(\frac{\bar{X} - \bar{Y} - \Delta}{\sigma\sqrt{\frac{2}{n}}} < \frac{-1.96\sigma\sqrt{\frac{2}{n}} - \Delta}{\sigma\sqrt{\frac{2}{n}}})$
$1 - \Phi(1.96 - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}) + \Phi(-1.96 - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}) = 0.8$, for large n, $\Phi(-1.96 - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}) \approx 0$
$\Phi(1.96 - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}) = 0.2$ ■

Method 0.9 (Ingredients of Sample Size Calculation (Pick n))

1. Significance level
2. Δ (Minimum detectable effect) MDE
3. Desired power to detect Δ
4. Need σ^2 (We don't have this yet), prior data, intuition