Final: post-midterm material (estimators, CI, hypothesis testing)

Review:

I. before-midterm : probability

① common distributions:

discrete
$$\begin{cases} \text{Bernoulli}(p) : E[X]=p,\ var[X]=p(1-p) & X\in\{0,1\} \\ \text{Binomial}(n,p) : E[X]=np,\ var[X]=np(1-p) & X\in\{0,1,\cdots,n\} \\ \text{Poisson}(\lambda) : E[X]=var[X]=\lambda & X\in\{0,1,\cdots\} \end{cases}$$

continuous
$$\begin{cases} N(\mu,\sigma^2) : E[X]=\mu,\ var[X]=\sigma^2 \quad X\in R \\ \text{Gamma}(\alpha,\beta) \text{ w pdf } f(x)=\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} : E[X]=\alpha\beta,\ var[X]=\alpha\beta^2 \end{cases}$$

Exponential(λ) = Gamma$(1,\frac{1}{\lambda})$ $\quad X\geq 0$

χ^2_n = Gamma$(\frac{n}{2}, 2)$

② MGF: $M_X(t)=E[e^{tX}]\quad t\in\mathbb{R}$

$E[X]=M_X'(0),\ \cdots,\ E[X^k]=M_X^{(k)}(0)\quad k=1,2,\ldots$

$M_{aX+b}(t)=e^{bt}\cdot M_X(at)$

$X\perp Y \Rightarrow M_{X+Y}(t)=M_X(t)M_Y(t)$

$\Downarrow$

$X_1,\cdots,X_n \overset{iid}{\sim} \text{Poisson}(\lambda) \Rightarrow \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$

$X_1,\cdots,X_n \overset{iid}{\sim} N(\mu,\sigma^2) \Rightarrow \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$

Continuity Thm: $X_1,\cdots,X_n,\cdots \quad X$

$M_{X_n}(t)\to M_X(t)$ for all t

$\Rightarrow F_{X_n}(x)\to F_X(x)$ for all x at which F_X is cont at

allows us to prove the convergence of distribution based on MGFs

③ CLT: $X_1,\cdots,X_n$ iid w $E[X_i]=\mu$, $var[X_i]=\sigma^2<\infty$, then

$$\frac{S_n-n\mu}{\sqrt{n\sigma^2}}=\frac{\bar{X}_n-\mu}{\sqrt{\sigma^2/n}}\xrightarrow{d} N(0,1)$$

when n is large (>20), $\bar{X}_n \overset{approx}{\sim} N(\mu,\frac{\sigma^2}{n})$

④ Delta method (Taylor expansion) — approximation of E & Var after g, a non-linear transformation

$E[X] = \mu_X$, $var[X] = \sigma_X^2$

$$E[g(X)] \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 \cdot g''(\mu_X)$$

$$var[g(X)] \approx \sigma_X^2 \cdot [g'(\mu_X)]^2$$

③+④: $X_1, \cdots, X_n$ iid with $E[X_i] = \mu$, $var[X] = \sigma^2 < \infty$

approximate the dist of $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

Ans: $\frac{1}{n}\sum_{i=1}^{n} g(X_i) \overset{approx}{\sim} N\left(g(\mu) + \frac{1}{2}\sigma^2 g''(\mu), \frac{1}{n}\sigma^2[g'(\mu)]^2\right)$

⑤ Normal distribution derivatives

$X_1, \cdots, X_n \overset{iid}{\sim} N(0,1) \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$

~~$S_n^2 = \frac{1}{n}$~~

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\downarrow \sum_{i=1}^{n}(X_i - \bar{X}_n) = 0$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$E[S_n^2] = var[X_i]$$

$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \bar{X}_n \perp S_n^2 \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$

$X \sim N(0,1)$, $S^2 \sim \chi_{df}^2$, $X \perp S^2 \Rightarrow \frac{X}{\sqrt{S^2/df}} \sim t_{df}$

So
$$\frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_n^2/\sigma^2}{n-1}}} = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

$X \sim \chi_n^2$, $Y \sim \chi_m^2$, $X \perp Y \Rightarrow \frac{X/n}{Y/m} \sim F_{n,m}$

So $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$, $Y_1, \cdots, Y_m \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi_{n-1}^2 \qquad \frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi_{m-1}^2$$

So
$$\frac{\frac{(n-1)S_X^2}{\sigma_X^2}/(n-1)}{\frac{(m-1)S_Y^2}{\sigma_Y^2}/(m-1)} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$$

Ⅱ. post-midterm : statistics

① estimators

a random sample $X_1, \cdots, X_n \overset{iid}{\sim} F_\theta$

$\hat{\theta} = g(X_1, \cdots, X_n)$ is an estimator of θ

How to construct $\hat{\theta}$?

Approach 1: method of moments (MOM)

$$\bar{X}_n = E[X_i] = h(\theta)$$

$$\Rightarrow \hat{\theta}_{MOM} = h^{-1}(\bar{X}_n)$$

If θ is 2-dimensional, e.g. $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$,

$$\begin{cases} \bar{X}_n = E[X_i] = h_1(\theta) \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = E[X_i^2] = h_2(\theta) \end{cases}$$

Approach 2: maximum likelihood estimation (MLE)

likelihood: $L(\theta) = \begin{cases} P(X_1, \cdots, X_n; \theta) = \prod_{i=1}^{n} P(X_i; \theta) \\ f(X_1, \cdots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta) \end{cases}$

e.g. $X_i = \begin{cases} 1 & \text{w.p. } \frac{\theta}{2} \\ 2 & \text{w.p. } \frac{\theta}{2} \\ 3 & \text{w.p. } 1-\theta \end{cases}$ $0<\theta<1$ $P(X_i; \theta) = \left(\frac{\theta}{2}\right)^{I(X_i = 1 \text{ or } 2)} \cdot (1-\theta)^{I(X_i=3)}$

$$\Rightarrow L(\theta) = \prod_{i=1}^{n} P(X_i; \theta) = \left(\frac{\theta}{2}\right)^{\sum_{i=1}^{n} I(X_i = 1 \text{ or } 2)} \cdot (1-\theta)^{\sum_{i=1}^{n} I(X_i=3)}$$

$$= \left(\frac{\theta}{2}\right)^{\#\text{ of 1's and 2's among } X_1, \cdots, X_n} \cdot (1-\theta)^{\#\text{ of 3's}}$$

log-likelihood: $\ell(\theta) = (\#\text{ 1's and 2's})(\log\theta - \log 2) + (\#\text{ of 3's}) \cdot \log(1-\theta)$

$$\ell'(\theta) = \frac{\#\text{ 1's and 2's}}{\theta} - \frac{\#\text{ of 3's}}{1-\theta}$$

Setting $\ell'(\theta) = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{\#\text{ of 1's and 2's}}{n} = \frac{\sum_{i=1}^{n} I(X_i = 1 \text{ or } 2)}{n}$

② estimator properties:

$$\left.\begin{aligned} \text{bias}(\hat{\theta}) &= E[\hat{\theta}] - \theta \\ \text{var}[\hat{\theta}] &= E[(\hat{\theta} - E[\hat{\theta}])^2] \end{aligned}\right\} \Rightarrow MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{bias}^2(\hat{\theta}) + \text{var}[\hat{\theta}]$$

Cramér-Rao Inequality: if $\hat{\theta}$ is unbiased $\Rightarrow var[\hat{\theta}] \geq \frac{1}{nI(\theta)}$.

$X \sim f_\theta$ pdf.

$$I(\theta) := E\left[\left(\frac{d}{d\theta}\log f_\theta(X)\right)^2\right] = \int\left(\frac{d}{d\theta}\log f_\theta(x)\right)^2 \cdot f_\theta(x)\,dx$$

$$= -E\left[\frac{d^2}{d\theta^2}\log f_\theta(X)\right]$$

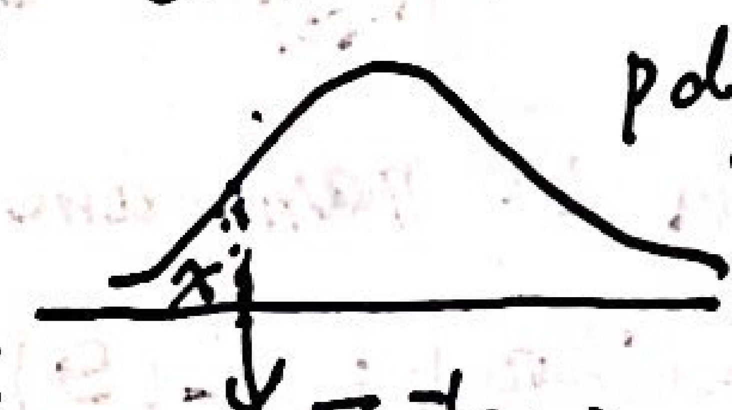
If $bias(\hat{\theta}) = 0$, $var[\hat{\theta}] = \frac{1}{nI(\theta)} \Rightarrow \hat{\theta}$ minimum variance unbiased estimator

MLE theory: $\hat{\theta}_{MLE} \overset{approx}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$ when n is large.

③ CI level: $(1-\alpha)$ $0<\alpha<1$ e.g. $\alpha = 0.05$, level $= 95\%$

$(1-\alpha)$ CI of θ: use the dist of $\hat{\theta}$

Say $\hat{\theta} \sim F_\theta$ CDF of $\hat{\theta}$



$$\mathbb{P}\left(F_\theta^{-1}\left(\tfrac{\alpha}{2}\right) \leq \hat{\theta} \leq F_\theta^{-1}\left(1-\tfrac{\alpha}{2}\right)\right) = 1-\alpha$$

$\Updownarrow$ reorganize the 2 inequalities

$$\cdots\hat{\theta}\cdots \leq \theta \leq \cdots\hat{\theta}\cdots$$

2 "random" boundaries of the $(1-\alpha)$ CI

④ Hypothesis testing:

$H_0, H_1, \alpha, T \Rightarrow$ rejection region

$\uparrow$ dist of T under H_0

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$

e.g. $H_0: \mu = 0$, $H_1: \mu > 0$

What is the rejection region if $T = \sum_{i=1}^{n} X_i$? or if $T = X_1$?

which is more powerful?

p-value $= \mathbb{P}_{H_0}(T$ is more extreme than its realization$)$

type I error $= \mathbb{P}_{H_0}($rejection region$)$

type II error $= \mathbb{P}_{H_1}($do not reject$)$

power $= 1 -$ type II error $= \mathbb{P}_{H_1}($rejection region$)$