Statistics

Def 1: A "(random) sample" of size n :

$(X_1, X_2, \ldots, X_n)$ where X_i's are independently

drawn from the same dist of r.v. X.

Ex: $(X_1, \ldots, X_{100})$ is a sample of size 100 from

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

Def 2: A "statistic" is a function of a sample.

E.g., $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$: sample mean

$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$: sample variance

Def 3: A "(point) estimate" of a paremeter θ is a single number that can be regarded as a sensible value of θ, e.g., a suitable statistic computed from a given sample. Use $\hat{\theta}$ (or rarely $\tilde{\theta}$) to customarily denote the estimate.

Ex: 1. $\hat{\mu} = \bar{X}$ is an estimate of μ using a sample $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

2. $\widehat{\sigma^2} = S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$ and

$\widetilde{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2$ are estimates of σ^2.

3. $\hat{p} = \frac{\sum_{i=1}^{N} X_i}{N}$ is an estimate of p from a sample $X_1, \ldots, X_N \overset{iid}{\sim} \text{Ber}(p)$

1. Method of Moments Estimator

① Write theoretical and sample moments

② Equate to solve for the params

Ex: Find $\hat{\mu}_{MM}$ for a sample $(X_1, \dots, X_n)$ drawn from dist w/ mean μ.

Ans: $m_1 = \mathbb{E}[X] = \mu$ $\qquad$ $S_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$

Set $m_1 = S_1$, we have

$$\left(\mu = \frac{1}{n}\sum X_i\right)\Big|_{\mu = \hat{\mu}_{MM}}$$

$$\hat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Ex: Find $\widehat{\sigma^2}_{MM}$ for a sample $(X_1, \dots, X_n)$ drawn from a dist w/ mean μ and variance σ^2.

Ans: $m_1 = E[X] = \mu$ $\qquad S_1 = \bar{X}$

$m_2 = E[X^2] = \mu^2 + \sigma^2$ $\qquad S_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

Set $m_i = S_i$, $i = 1, 2$, namely,

$$\begin{pmatrix} \mu = \bar{X} \\ \mu^2 + \sigma^2 = \frac{1}{n}\sum X_i^2 \end{pmatrix} \Bigg| \begin{matrix} \mu = \hat{\mu}_{MM} \\ \sigma^2 = \widehat{\sigma^2}_{MM} \end{matrix}$$

$$\Rightarrow \quad \hat{\mu}_{MM} = \bar{X}$$

$$\widehat{\sigma^2}_{MM} = \frac{1}{n}\sum X_i^2 - (\bar{X})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{(biased)}$$

Ex: Find $\hat{p}_{MM}$ for a sample of size $n \overset{iid}{\sim} Ber(p)$

Ans:

$$m_1 = \cancel{E}\,\mathbb{E}[X] = (1-p)\,0 + (p)\cdot 1 \qquad S_1 = \frac{1}{n}\sum_i X_i$$

$$= p$$

$$(m_1 = S_1)_{p=\hat{p}_{MM}}, \qquad \hat{p}_{MM} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

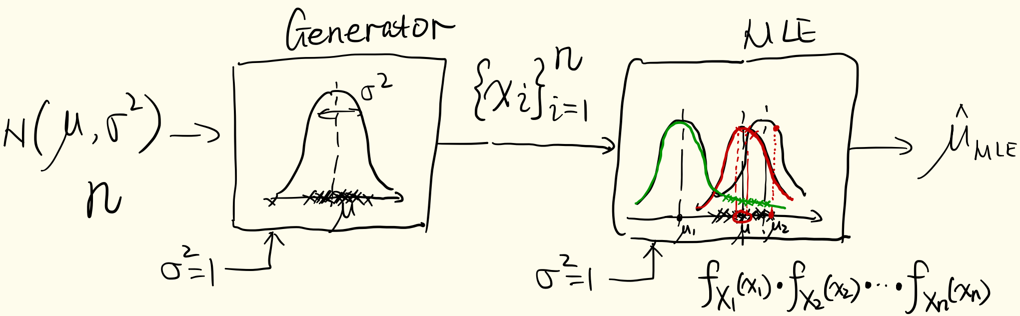
2. Maximum Likelihood Estimator (MLE)

① Write the likelihood of sample/data $(X_1, \dots, X_n)$, or the joint dist,

$$\mathcal{L}(\theta) = f_{X_1 X_2 \cdots X_n}(x_1, x_2, \dots, x_n)$$

$$= \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i)$$

② Find $\hat{\theta}$ that maximizes the (log) likelihood.



Ex: Find $\hat{\mu}_{MLE}$ for a sample $X_1, \dots, X_n \sim N(\underset{\sim}{\mu}, \sigma^2)$

Ans: $\underset{\mu}{\underset{\shortparallel}{L(\theta)}} = f_{X_1 \cdots X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]}$$

$$\ell(\mu) = \ln L(\mu) = -n \ln(\sqrt{2\pi}\,\sigma) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\partial \ln L(\mu) / \partial \mu = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i-\mu) \cdot (-1) \overset{\text{set}}{=} 0 \Big|_{\mu = \hat{\mu}_{MLE}}$$

$$\sum x_i = n \hat{\mu}_{MLE} \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \bar{x}$$

Ex: Find $\hat{p}_{MLE}$ for $X_1, \dots, X_n \sim Ber(p)$

Ans: $p_X(0) = 1-p$, $p_X(1) = p$

$$\Longleftrightarrow p_X(x) = p^x \cdot (1-p)^{1-x}$$

$$L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} (1-x_i)}$$

$$\ln L(p) = \left(\sum x_i\right) \ln p + \sum (1-x_i) \ln (1-p)$$

$$\frac{\partial}{\partial p} \ln L(p) = \left(\sum x_i\right) \frac{1}{p} + \left[\sum (1-x_i)\right] \frac{1}{p-1} \overset{\text{set}}{=} 0 \Big|_{p = \hat{p}_{MLE}}$$

$$\left(\sum x_i\right)(1-p) = \left[\sum (1-x_i)\right] \cdot \hat{p}$$

$$\sum x_i = \hat{p} \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} (1-x_i) \right) \Rightarrow \hat{p}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

Invariance Principle of MLE

Let $\hat{\theta}_1, \dots, \hat{\theta}_n$ be the MLEs for $\theta_1, \dots, \theta_n$, then

$$\underbrace{h(\hat{\theta}_1, \dots, \hat{\theta}_n)}_{=\,\hat{\nu}} \text{ is the MLE for } h(\theta_1, \dots, \theta_n) =: \nu .$$

Ex: $\widehat{\sigma^2_{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, then

$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

$\theta_1 = \sigma^2$

$\nu = \sqrt{\theta_1} = h(\theta_1)$

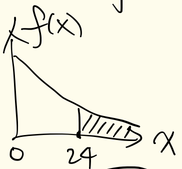
$\hat{\nu} = h(\hat{\theta}_1)$

$= \sqrt{\frac{1}{n}\sum(\quad)^2}$

Ex: At $t=0$, 20 identical components are tested. Life time $X_i \sim \exp(\lambda)$. After 24 hrs, 15 are still in operation. Derive MLE of λ.

Ans: $\hat{p}_{MLE} = 15/20$ "In operation flag", $W_i \sim Ber(p)$

$i = 1, \dots, 20$.



Exp:

$f(x) = \lambda e^{-\lambda x}, \ x \geq 0$

$F(x) = 1 - e^{-\lambda x}$

$\mathbb{E}[X] = \frac{1}{\lambda}$

$Var(X) = \frac{1}{\lambda^2}$

$$p = \mathbb{P}[X_i \geq 24] = 1 - \underbrace{\mathbb{P}[X_i < 24]}$$

$$= 1 - (1 - e^{-\lambda \cdot 24}) = e^{-\lambda \cdot 24}$$

$$\hat{p}_{MLE} = e^{-\hat{\lambda}_{MLE} \cdot 24}$$

$$\Rightarrow \hat{\lambda}_{MLE} = \ln(15/20) / (-24) = \cdots$$

Def 4. An estimate $\hat{\theta}$ is an "unbiased" estimator of θ if $\mathbb{E}[\hat{\theta}] = \theta, \ \forall \theta$.

$$\text{bias}(\hat{\theta}) \triangleq \mathbb{E}[\hat{\theta}] - \theta$$

Ex: $\mathbb{E}[\hat{\mu}_{MM}] = \mathbb{E}[\bar{X}] = \mathbb{E}[\frac{1}{n}\sum X_i] = \frac{1}{n}\sum \mathbb{E}X_i = \mu$ unbiased

$\mathbb{E}[\hat{\sigma}^2_{MM}] = (HW)$ biased

$\mathbb{E}[\hat{p}_{MM}] = \mathbb{E}[\frac{\sum X_i}{n}] = \frac{\sum \mathbb{E}X_i}{n} = \frac{\sum_{i=1}^{n} p}{n} = p$ unbiased

$\mathbb{E}[\hat{\mu}_{MLE}] = \mathbb{E}[\bar{X}] = \mu$ unbiased

$\mathbb{E}[\hat{\sigma}^2_{MLE}] = (HW)$ biased

$\mathbb{E}[\hat{p}_{MLE}] = p$ unbiased

Principles: When choosing among several estimators of θ, we usually prefer the unbiased one.

Among all unbiased estimators of θ, choose the one that has the minimum variance. The resulting $\hat{\theta}$ is called the minimum variance unbiased estimator (MVUE) of θ.

Comment: If the MVUE has a larger variance than a biased one, we may choose either estimator, depending on the use case.

$$MSE = E\left[(\hat{\theta} - \theta)^2\right] = bias^2(\hat{\theta}) + Var(\hat{\theta})$$

Def 5. An estimator $\hat{\theta}_n$ is "consistent" if for every $\theta \in \Theta$, $\underbrace{\hat{\theta}_n \xrightarrow{P} \theta \text{ as } n \to \infty}$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left[|\hat{\theta}_n - \theta| > \varepsilon \right] = 0, \text{ for every } \varepsilon.$$

Note that MLE is consistent.

Claim: $\hat{\mu} = \bar{X}$ for $(X_1, \dots, X_n) \sim \mathcal{N}(\mu, \sigma^2)$ is consistent.

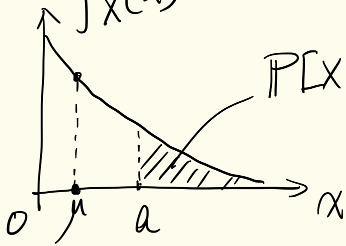
$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\theta = \mu$.

$$\mathbb{P}\Big[\underbrace{|\hat{\theta}_n - \theta| > \varepsilon}_{|\hat{\theta}_n - \theta|^2 > \varepsilon^2} \Big] \le \frac{\mathbb{E}\big[|\hat{\theta}_n - \overset{\mathbb{E}\hat{\theta}_n}{\theta}|^2 \big]}{\varepsilon^2} = \frac{\mathrm{Var}(\hat{\theta}_n)}{\varepsilon^2}$$

$$\mathrm{Var}(\hat{\theta}_n) = \mathrm{Var}\left(\tfrac{1}{n}\sum X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\,\mathrm{Var}(X)}{n^2} = \sigma^2/n$$

Hence, $\mathbb{P}[|\hat{\theta}_n - \theta| > \varepsilon] \le \frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$. $\square$

Markov's Inequality.



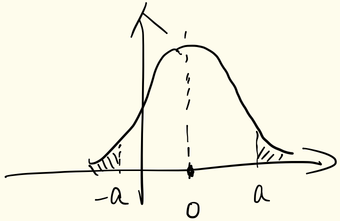
"Concentration Inequalities"

$$\mathbb{P}[X>a] \le \frac{\mathbb{E}[X]}{a}$$

$X \geq 0$, $a>0$

Chebyshev's Inequality

$$\mathbb{P}[|X-\mathbb{E}X|>a] \le \frac{\mathbb{E}[|X-\mathbb{E}X|^2]}{a^2} = \frac{Var(X)}{a^2}$$



$$V\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left[\left(\sum X_i - \mathbb{E}\sum X_i\right)^2\right]$$

$$= \mathbb{E}\left[\left[\sum_{i=1}^{n} \underbrace{(X_i - \mathbb{E}X_i)}_{u_i}\right]^2\right]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{n} u_i\right)^2\right]$$

$(u_1+u_2+u_3)(u_1+u_2+u_3)$

	u_1	u_2	u_3
u_1	✓	o	o
u_2	o	✓	o
u_3	o	o	✓

$$= \mathbb{E}\left[\sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} u_i u_j\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} u_i^2 + 2\sum_{i=1}^{n}\sum_{j<i} u_i u_j\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mathbb{E}X_i)^2\right] + 2\sum\sum \mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right]$$

$$= \sum_{i=1}^{n} V(X_i) + 2\sum_{i=1}^{n}\sum_{j<i} \operatorname{cov}(X_i, X_j)$$

If $\perp\!\!\!\perp_i X_i$, $\quad V(\sum X_i) = \sum V(X_i)$.