Lecture 24 - CLT

Tuesday, December 7, 2021 9:24 AM

Indicator Functions

$$\mathbb{1}(\text{statements}) = \begin{cases} 0 & \text{statement false} \\ 1 & \text{statement true} \end{cases}$$

Ex. $\mathbb{1}(x \in A) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$

PDF of a RV $X \sim Exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

$$\boxed{f(x) = \lambda e^{-\lambda x} \, \mathbb{1}(x > 0)}$$

More generally:

$$f(x) = \sim\!\sim \, \mathbb{1}(x \in \text{Support})$$

Fact: $\mathbb{1}(A \text{ and } B) = \mathbb{1}(A)\,\mathbb{1}(B)$

Independence:

$$X \perp\!\!\!\perp Y \iff f(x,y) = f(x)f(y)$$

Ex. $f(x,y) = c e^{-x} e^{-y}$ for $\underbrace{x>0 \text{ and } y>0}$

$$= c e^{-x} e^{-y}\, \mathbb{1}(x>0 \text{ and } y>0)$$

$$= c e^{-x} e^{-y}\, \mathbb{1}(x>0)\, \mathbb{1}(y>0)$$

$$= \underbrace{\left[c e^{-x}\, \mathbb{1}(x>0)\right]}_{\text{fn of } x} \underbrace{\left[e^{-y}\, \mathbb{1}(y>0)\right]}_{\text{fn of } y}$$

Defn: Random Sample

If $X_1, X_2, X_3, \dots, X_n \overset{iid}{\sim} f$

we call this a random sample of size n from f.

Fact:

$$\underset{\sim}{X} = (X_1, \ldots, X_n)$$
$$\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

$$f(\underset{\sim}{x}) = f(x_1, \ldots, x_n)$$
$$= f(x_1) f(x_2) \cdots f(x_n)$$
$$= \prod_{i=1}^{n} f(x_i)$$

Defn: Statistic

If $(X_i)_{i=1}^n$ are a rand. sample and

$$T : \mathbb{R}^n \to \mathbb{R}^d$$

then $T(\underset{\sim}{X})$ is called a statistic.

Ex. ① arithmetic mean

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n)$$

② Sample variance

$$S^2 = \frac{1}{} \sum_{i}^{n} (X_i - \overline{X})^2$$

$n-1 \; i=1$

③ Order Statistics

$X_{(1)}$ = minimum of X_i

$X_{(n)}$ = maximum of X_i

$X_{(r)}$ = r^{th} smallest of X_i

④ range: $R = X_{(n)} - X_{(1)}$.

Defn: Sampling Distribution of a Statistic

The sampling dist. of $T(\underset{\sim}{X})$ is simply the dist. of $T(\underset{\sim}{X})$.

Theorem: Central Limit Theorem

Intro stats: $\bar{X} \approx N(\mu, \sigma^2/n)$ for large n

451 statement:

If X_i are a random sample

w/ mean $EX_i = \mu$ and $Var X_i = \sigma^2 < \infty$

then

$$Y = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$$

$$\bar{X} \approx N(\mu, \sigma^2/n) \Rightarrow \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \approx N(0,1)$$

CDF/MGF of left
Converges to CDF/MGF of Right

Ex. CLT

$$X_i \overset{iid}{\sim} Bernoulli(p)$$

$\mu = EX_i = p$ and $Var X_i = p(1-p) = \sigma^2$

CLT says:

$$\rightarrow \quad \sqrt{n}\left(\frac{\bar{X} - p}{\sqrt{p(1-p)}}\right) \xrightarrow{d} N(0,1)$$

practically: $\boxed{\bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)}$

practically: $\boxed{\bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)}$

$$\bar{X} - p \approx N\left(0, \frac{p(1-p)}{n}\right)$$

$Var(aX) = a^2 Var X$

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

intro stats: 95% CI: $\bar{X} \pm 2\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$

$$Y = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right)$$

$$= \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma}\right)$$

$$= \frac{\left(\frac{\sqrt{n}}{n}\sum_{i=1}^{n} X_i - \sqrt{n}\mu\right)}{\sigma}$$

$$= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \frac{n}{\sqrt{n}}\mu}{\sigma}$$

$\sum_{i=1}^{n} \mu$

$$= \frac{\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}(X_i - \mu)\right)}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)$$

; $\mathbb{E}X_i = \mu$

$Var X_i = \sigma^2$

$Y_i = \frac{X_i - \mu}{\sigma}$

then $\mathbb{E}Y_i = 0$

$Var Y_i = 1$

$$\boxed{= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = Y}$$

Going to show that MGF of $Y \to$ MGF of $N(0,1)$

$$M_{\Sigma X_i}(t) = M_{X_1}(t) M_{X_2}(t) \cdots$$

$$M_{aX}(t) = M_X(at)$$

$$M_Y(t) = M_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i}(t) = M(t/\sqrt{n})^n$$

Taylor Series

$g : \mathbb{R} \to \mathbb{R}$ that is k-times diffentiable

then

$$T_k(x) = \sum_{r=0}^{k} \frac{g^{(r)}(0)}{r!} x^r$$

then $R = g(x) - T_k(x) \to 0$ as $x \to 0$

Approx M w/ a taylor series

$$M(t) = (1)t^0 \quad M^{(1)}(0)t^1 \quad M^{(2)}(0)t^2$$

$$M(t) = \frac{(1)\,t^0}{0!} + \frac{M^{(1)}(0)\,t^1}{1!} + \frac{M^{(2)}(0)\,t^2}{2!} + \cdots$$

$$= 1 + \underbrace{\cancel{E[Y_i]}}_{0}\,t + \frac{\underbrace{\cancel{E[Y_i^2]}}_{1}\,t^2}{2} + \cdots$$

$$= 1 + t^2/2 + \cdots$$

$$\boxed{\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^n = e^c}$$

$$M_Y(t) = M(t/\sqrt{n})^n = \left(1 + \frac{t^2/2}{n}\right)^n$$

$$\xrightarrow{n} e^{t^2/2}$$

MGF of $N(0,1)$