Quiz 2

P1. $F_X(t) = \mathbb{P}[X \le t] = \begin{cases} 1 - t^{\alpha} & t \ge 1 \\ 0 & t < 1 \end{cases}$

(a) $f_X(t) = F_X'(t) = \begin{cases} \alpha t^{-\alpha-1} & t \ge 1 \\ 0 & t < 1 \end{cases}$

(b) $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} t^2 f_X(t)\,dt = \int_1^{\infty} 3\, t^2 t^{-3-1}\,dt = 3\int_1^{\infty} t^{-2}\,dt$

$= 3(-t^{-1})\Big|_{t=1}^{\infty} = 3(\frac{-1}{\infty}) - 3(\frac{-1}{1}) = 3$

P2

$\mathbb{E}[X \mid Y=1] = \sum_x x\, \mathbb{P}[X=x \mid Y=1] = \sum_x x \frac{\mathbb{P}[X=x, Y=1]}{\mathbb{P}[Y=1]}$

Def. A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (a_n, b_n)$, where a_n, b_n are functions of $X_1, \dots, X_n$ s.t.

$$\mathbb{P}_\theta[\theta \in C_n] \ge 1 - \alpha \quad \forall \theta \in \Theta$$

C_n = random $\qquad \theta$ = fixed but unknown

Once we sample C_n, there are no more probability - either $\theta \in C_n(\omega)$ or $\theta \notin C_n(\omega)$

Ex. $Y = \begin{cases} \theta + 1 & \text{wp } \frac{1}{2} \\ \theta - 1 & \text{wp } \frac{1}{2} \end{cases}$ $\theta \in \mathbb{R}$ fixed, unknown parameter

Observe iid samples $Y_1, Y_2 \sim Y$

$$C_n = \left(\min(Y_1, Y_2, \dots, Y_n) - \frac{1}{4}, \max(Y_1, Y_2, \dots, Y_n) + \frac{1}{4}\right)$$

$$\theta \in C_n \iff \min(Y_1, Y_2, \dots, Y_n) \neq \max(Y_1, Y_2, \dots, Y_n)$$

$\mathbb{P}[\theta \in C_n] = 1 - \frac{1}{2^{n-1}} \Rightarrow C_n$ is a $1 - \frac{1}{2^{n-1}}$ CI for θ.

$\underbrace{1 - \frac{1}{2^{n-1}} \text{ the time}}_{\text{this means over samples of } C_n}$, $\theta \in C_n$

$n=2$: if $Y_1 = 15$, $Y_2 = 17$ then $C_2 = (14.75, 17.25)$ but we are 100% sure $\theta = 16$. $\mathbb{P}[\theta \in C_2 \mid Y_1 = 15, Y_2 = 17] = 1$

instance/sample of CI is not a probability statement about θ (θ is not random)

X t-step random walk with parameter p $\quad X = \sum_{i=1}^{t} Y_i \quad , Y_i \sim \begin{cases} -1 & \text{wp } p \\ +1 & \text{wp } 1-p \end{cases}$

$X_1, X_2, \dots, X_n \sim X$ iid samples

$\hat{p} = \frac{1}{2} - \frac{1}{2nt} \sum_{i=1}^{n} X_i$ a point estimate for p

$\mathbb{E}[\hat{p}] = p, \quad \mathbb{V}[\hat{p}] = \frac{p(1-p)}{nt}$

Fact: $\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathbb{V}[X]}{a^2}$ (Chebyshev)

Find $C_n = (a_n, b_n)$ st. C_n is a $1-\alpha$ confidence interval for p

$\mathbb{P}[|\hat{p} - p| < a] = \mathbb{P}[-a < p - \hat{p} < a] = \mathbb{P}[\hat{p} - a < p < \hat{p} + a]$

$C_n = (\hat{p} - a, \hat{p} + a)$

$\mathbb{P}[p \in C_n] = 1 - \mathbb{P}[|\hat{p} - p| \geq a] \geq 1 - \frac{\mathbb{V}[X]}{a^2} = 1 - \alpha$

if $\alpha = \frac{\mathbb{V}[X]}{a^2} \iff a = \sqrt{\mathbb{V}[X]/\alpha}$

$C_n = \left(\hat{p} - \sqrt{\frac{p(1-p)}{\alpha n t}}, \hat{p} + \sqrt{\frac{p(1-p)}{\alpha n t}} \right)$

note that p is <u>unknown</u> thus, to get a usable CI.

Note that $p(1-p) \leq \frac{1}{4} \; \forall p \in [0,1]$. Thus, we can use

$C_n = \left(\hat{p} - \frac{1}{2\sqrt{\alpha n t}}, \hat{p} + \frac{1}{2\sqrt{\alpha n t}} \right)$