CSE 357 (Fall 2019)
Statistical Methods for Data Science

Practice mid-term 1
(6 questions, 30 points total)

I acknowledge that engaging in dishonest behavior during the exam will result in a score of 0. Dishonest behavior includes copying from other students, referring to any form of notes, conversing with other students without the permission of the instructor, etc.
By taking this exam, I agree to the above terms.

Please write your name on the line below.

_____

For instructor's use only.

Q1) 6 points:

Q2) 5 points

Q3) 7 points:

Q4) 5 points:

Q5) 3 points:

Q6) 4 points:

_____

Total (out of 30 points):

_____

Q1) THE LINEAR DISTRIBUTION **(Total 6 points)**

Consider a new continuous distribution called *Linear* that is defined on the range (1, 2) and has p.d.f. $f(x) = C \cdot x$ for $x \in (1, 2)$, where C is some constant, and f(x) = 0, otherwise.

(a) Find the value of C. (2 points)

(b) Draw a diagram to clearly show the p.d.f. of *Linear* in the range [0, 3]. (1 point)

(c) Draw a diagram to clearly show the c.d.f. of *Linear* in the range [0, 3]. (1 point)

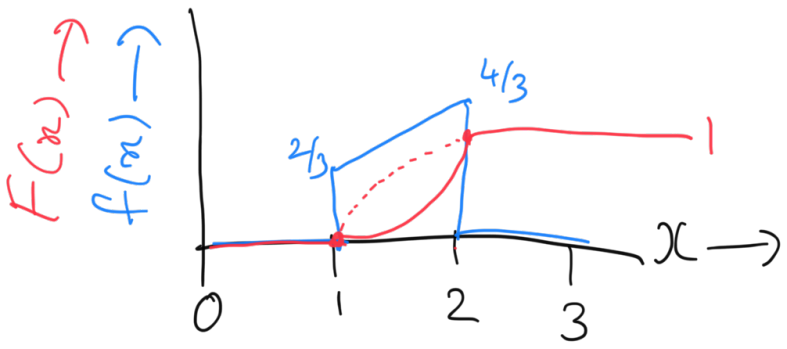
(d) Find the value of the expectation (or mean) of the *Linear* distribution. (2 points)

(a) $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{1}^{2} C \cdot x \cdot dx = 1 \Rightarrow C \cdot \frac{x^2}{2}\Big|_1^2 = 1$

$\Rightarrow C\left(\frac{4}{2} - \frac{1}{2}\right) = 1 \Rightarrow C = 2/3$

(b)

(C)



(d) $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

$= \int_{1}^{2} x \cdot \frac{2}{3} \cdot x \cdot dx = \frac{2}{3} \cdot \int_{1}^{2} x^2\,dx$

$= \frac{2}{3} \cdot \frac{x^3}{3}\Big|_1^2 = \frac{2}{3}\left(\frac{8}{3} - \frac{1}{3}\right)$

$= \frac{2}{3} \times \frac{7}{3} = \frac{14}{9}$

Q2) VARIANCE OF A UNIFORM **(Total 5 points)**

Let X be distributed as Uniform(a, b), for some b > a > 0.

(a) $E[X^2]$ in terms of a and b. (2 points)

(b) Given that $E[X] = (a+b)/2$, show that $Var[X] = (b-a)^2/12$. (3 points)

(a) $f(x) = \begin{cases} \frac{1}{b-a} & 0 \le x \le b \\ 0 & \text{o.w.} \end{cases}$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_a^b x^2 \cdot \frac{1}{(b-a)} \cdot dx$$

$$= \frac{1}{b-a} \cdot \int_a^b x^2 \cdot dx = \frac{b^3 - a^3}{3(b-a)}$$

$$= \frac{b^2 + ab + a^2}{3}$$

(b) $Var(X) = E[X^2] - (E[X])^2$

$$= \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{b^2+ab+a^2}{3} - \frac{(a^2+2ab+b^2)}{4}$$

$$= \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12}$$

$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$

Q3) CONSISTENCY OF eCDF **(Total 7 points)**

Let sample data $\{X_1, X_2, \ldots, X_n\}$ be i.i.d. RVs with true CDF *F*. Let the eCDF of the sample data be $\hat{F}$, as defined in class.

(a) Derive $E[\hat{F}]$ in terms of *F*. (2 points)

(b) Derive $se(\hat{F})$ in terms of *F*. (3 points)

(c) Show that $\hat{F}$ is a consistent estimator. (2 points)

(a) $\hat{F}(\alpha) = \frac{\sum I(X_i \le \alpha)}{n}$

$E[\hat{F}(\alpha)] = E\left[\frac{\sum I(X_i \le \alpha)}{n}\right]$

$= \frac{1}{n} E[\sum I(X_i \le \alpha)] \overset{LOE}{=} \frac{1}{n} \sum E[I(X_i \le \alpha)]$

$\overset{iid}{=} \frac{1}{n} \cdot n \cdot E[I(X_1 \le \alpha)] \overset{Indicator}{=} Pr(X_1 \le \alpha)$

$= F(\alpha)$

(b) $Var(\hat{F}) = Var\left(\frac{1}{n} \sum I(X_i \le \alpha)\right)$

$= \frac{1}{n^2} Var(\sum I(\cdot)) \overset{LOV}{\underset{iid}{=}} \frac{1}{n^2} \sum Var(I(X_i \le \alpha))$

$\overset{iid}{=} \frac{1}{n^2} \cdot n \cdot Var(I(X_1 \le \alpha)) = \frac{1}{n} Var(I(X_1 \le \alpha))$

$= \frac{1}{n} \cdot Pr(X_1 \le \alpha) \cdot (1 - Pr(X_1 \le \alpha)) = \frac{F(1-F)}{n}$

$se(\hat{F}) = \sqrt{Var(\hat{F})} = \sqrt{\frac{F(1-F)}{n}}$

(C) $bias(\hat{F}) = E[\hat{F}] - F = 0$

$se(\hat{F}) = \sqrt{\frac{F(1-F)}{n}} \longrightarrow 0$ as $n \to \infty$ } $\therefore$ consistent

Q4) INDICATORS TO THE RESCUE **(Total 5 points)**

(a) Let I_E be an indicator RV for event E. If Pr(E) is the probability that event E occurs, then derive Var(I_E) in terms of Pr(E) from scratch (via the definition of variance). (2 points)

(b) Using part (a), show that Var(Binomial(n, p)) = n·p·(1-p). You must express the Binomial as a sequence of individual coin flips, as in class. (3 points)

(a) $\mathrm{Var}(I_E) = E[I_E^2] - (E[I_E])^2$

$E[I_E] = \sum_{i \in \Omega} i \cdot \Pr(i)$ $\qquad I_E = \begin{cases} 1 & \text{w.p. } \Pr(E) \\ 0 & \text{w.p. } 1-\Pr(E) \end{cases}$

$= 1 \cdot \Pr(E) + 0 \cdot (1-\Pr(E)) = \Pr(E)$

$E[I_E^2] = \sum_i i^2 \Pr(i) = 1^2 \cdot \Pr(E) + 0^2 \cdot (\quad)^0 = \Pr(E)$

$\therefore \mathrm{Var}(\quad) = \Pr(E) - (\Pr(E))^2 = \Pr(E)(1-\Pr(E))$

(b) $\mathrm{Bin}(n, p) = I_{F_1} + I_{F_2} + I_{F_3} + \cdots + I_{F_n}$

I_{F_i} : Indicator RV for flip i to be successful

$\mathrm{Var}(\mathrm{Bin}(n,p)) = \mathrm{Var}(I_{F_1} + I_{F_2} + \cdots + I_{F_n})$

$\overset{\text{LOV}}{=} \sum_{i=1}^{n} \mathrm{Var}(I_{F_i})$

$\overset{\text{iid}}{=} n \cdot \mathrm{Var}(I_{F_1}) \overset{\text{part(a)}}{=} n \cdot \Pr(F_1)(1-\Pr(F_1))$

$= n \cdot p \cdot (1-p)$

Q5) IT'S SO NORMAL **(Total 3 points)**

Let X be distributed as Normal(μ, σ^2). Derive the 95th percentile of X in terms of ϕ^{-1}() and μ and σ, where ϕ is the CDF of the standard normal, Normal(0,1). Note that the 95th percentile of X is the point α such that Pr(X < α) = 0.95. This question is asking you to find the value of α. Clearly show how you transform X in terms of the standard normal.

$$Pr(X < \alpha) = 0.95 \qquad \frac{X-\mu}{\sigma} = Z$$

$$\Rightarrow Pr\left(\frac{X-\mu}{\sigma} < \frac{\alpha-\mu}{\sigma}\right) = 0.95$$

$$\Rightarrow Pr\left(Z < \frac{\alpha-\mu}{\sigma}\right) = 0.95$$

$$\Rightarrow \Phi\left(\frac{\alpha-\mu}{\sigma}\right) = 0.95$$

$$\Rightarrow \frac{\alpha-\mu}{\sigma} = \Phi^{-1}(0.95)$$

$$\Rightarrow \alpha = \mu + \sigma \cdot \Phi^{-1}(0.95)$$

Q6) RETURN OF THE PLUG-IN ESTIMATORS **(Total 4 points)**

Let D = {-1, -1, 0, 2} be a set of i.i.d. data samples drawn from some distribution X with true mean μ. Find the plug-in estimator for $E[X^3] + e^{\mu}$. Your final answer should be numerical.

$$E[X^3] + e^{\mu} = \sum i^3 \cdot p_X(i) + e^{\sum i \cdot p_X(i)}$$

$$\therefore \text{plug-in is } \sum X_i^3 \cdot \frac{1}{n} + e^{\sum X_i \cdot \frac{1}{n}}$$

$$D = \{-1, -1, 0, 2\} \Rightarrow n = 4. \quad \sum X_i = 0$$

$$\frac{1}{4}\sum X_i^3 + e^{0/4} = \frac{1}{4}(-1^3 + (-1)^3 + 0^3 + 2^3) + 1$$

$$= \frac{1}{4}(-1 - 1 + 8) + 1 = \frac{6}{4} + 1 = \frac{5}{2}$$

(page intentionally left blank for extra space or rough work)

(page intentionally left blank for extra space or rough work)