

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)

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

Let X be distributed as $\text{Uniform}(a, b)$, 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 $\text{Var}[X] = (b-a)^2/12$.

(3 points)

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

Let sample data $\{X_1, X_2, \dots, X_n\}$ 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)

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 $\text{Var}(I_E)$ in terms of $\Pr(E)$ from scratch (via the definition of variance). (2 points)
- (b) Using part (a), show that $\text{Var}(\text{Binomial}(n, p)) = n \cdot p \cdot (1-p)$. You must express the Binomial as a sequence of individual coin flips, as in class. (3 points)

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

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

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.

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

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