Statistics 110: Introduction to Probability — Section Midterm Review, Fall 2018

Practice Problems

TF: Justin Zhu (justinzhu@college.harvard.edu)

Don't Make Category Errors!

Characterize the following as a probability, sample space/support, event, random variable, or distribution

(a) $Bern(n = 1, p = 0.25)$

(b) Ω

(c) $x|y$

(d) λ, where $\lambda \in S$ and $S = \{0, 1, 2, 3, \cdots, \infty\}$

(e) S, where $\lambda \in S$ and $S = \{0, 1, 2, 3, \cdots, \infty\}$

(f) p, where p = -$\frac{1}{2}$

(g) 0.3145

(h) I_j, where $I_j \sim Bern(0.5)$

(i) All x where $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

(j) All $P(X = x)$ where $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

(k) $Z = z$, where $z \in \Omega$

(l) P, where $P \in (0, 1)$

(m) $\varnothing$

(n) $P(Y) = 80\%$, where Y is true when $X = x$

(o) Y, where Y is 0 when $X = x$

(p) $\frac{P(A|B)P(B)}{P(A)}$

(q) $\sum_{k=0}^{n} \binom{n}{k} (\frac{1}{4})^k (\frac{3}{4})^{n-k}$, where $n > 0$

(r) $\lim_{n\to\infty}(1 - \frac{1}{n})^n$

(s) X, where $X = \sum_{k=1}^{N} \frac{1}{k}$, and $N > 1$

(t) $X = x|N = n$, where $X = \sum_{k=1}^{N} \frac{1}{k}$, and $N > 1$

(u) $N - X = x|N = n$, where $X = \sum_{k=1}^{N} \frac{1}{k}$, and $N > 1$

(v) $N - X|N = n$, where $N \sim HGeom(w = 3, b = 5, k = 2)$ and $X \sim HGeom(w = 3, b = 5, k = 2)$

(w) $Y^2|X$ where $Y|X \sim NBin(r, p)$

(x) Everything to the right of the "|" in $A|B, C, D, \cdots$

(y) Your score on the midterm

(z) You acing the midterm!

Counting

In the game of Avalon, 7 players are dealt four “good" characters and three “bad" characters. Consider the following:

(a) How many possible arrangements of unique good and bad players are there?

(b) We now make a distinction between generic “goodies" and two special good players, Merlin and Percival. How many more possible arrangements of unique good and bad players are there?

(c) In addition to the two special good characters, we now make a distinction between generic “baddies" and two special bad players, Mordrid and Morgana. How many more possible arrangements of unique good and bad players are there?

(d) A “probabilistic" version of Avalon is played where we still have 7 players but 14 cards are dealt, with the pack containing 7 “good" characters and 7 “bad" characters, no special characters. How many possible arrangements of unique good and bad players are there? (Hint: consider the extreme case where all players are “good" characters)

(e) Assume we play the normal “probabilistic" version of Avalon where 7 generic freshmen are dealt 14 cards, with the pack containing the 4 special characters of Merlin, Percival, Mordrid, and Morgana, and 5 generic good cards and 5 generic bad cards. Knowing we do not care who gets which card, how many possible arrangements of cards are played among the seven players?

Conditional Probability

In a game of Avalon, players go on missions, where goodies pass missions and baddies try to fail missions. After each mission, everybody discusses who they think are baddies, switching out the baddies on the team with the goodies. Therefore, it could be in a baddy's best interest to pass a mission in order to stay on the team so he can fail a mission in the future.

The game starts, and you must propose a mission. So far, n missions have passed, but there is a possibility that a baddie has been on the team this whole time and will fail the next mission.

We want to find whether this next mission consisting of you (a "goodie") and the existing team consists of all "goodies." Let G be the event that this team consists of all goodies, $p = P(G)$ be the probability that this team consists of all goodies, and $q = 1 - p$. Let M_j be the event that the j-th mission passes.

(a) Assume for this part that each mission passed is conditionally independent given that this team consists of all goodies. Let $a = P(M_j|G)$ and $b = P(M_j|G^c)$, where a and b don't depend on j. Find the posterior probability that the team contains all goodies given that all n of the n missions have passed.

(b) Suppose that the mission is passed on all n tests. However, it is likely that Merlin is on the team and he has selected all of the goodies on the team since Merlin knows all the goodies. Therefore, all missions will pass. However, having Merlin on the team is not desirable since the baddies win the game if they guess correctly who Merlin is.

Let X be the event that Merlin is on the team. Assume that $P(X) = c$ and that G and X are independent. If Merlin is *not* on the team, then the missions are conditionally independent given a team of all goodies. Let $a_0 = P(M_j|G, X^c)$ and $b_0 = P(M_j|G^c, X^c)$, where a_0 and b_0 don't depend on j. Find the posterior probability that the team consists of all goodies and no Merlin, given that all n missions have passed.

Expected Number and Variance of Matches

Suppose 300 people are in Stat110, each completed a test. We make 300 copies of each person's test, grade these 300^2 copies and mix the tests, handing them out randomly to each person.

(a) What is the expected number of people who get their own test?

(b) What is the variance of the number of people who get their own test?

(c) Can we use a Poisson to approximate the probability of there being two people who receive their test back?

Solution

(a) We have $E(I_k) = \frac{1}{300}$ since the it is equally probable for the kth person to get any of the 300 tests. Hence,

$$\begin{aligned} E(X) &= E\left(\sum_{k=1}^{300} I_k\right) \\ &= \sum_{k=1}^{300} E(I_k) \\ &= 300 \cdot \frac{1}{300} \\ &= \boxed{1} \end{aligned}$$

(b) By the definition of variance,

$$\text{Var}(X) = E(X^2) - (E(X))^2 = E(X^2) - 1$$

using the result in (a). To find $E(X^2)$, we can expand the indicator r.v. representation and use properties of indicators:

$$\begin{aligned} E(X^2) &= E\left[(I_1 + I_2 + \ldots + I_{300})^2\right] \\ &= E(I_1^2 + I_2^2 + \ldots + I_{300}^2 + 2\sum_{i<j} I_i I_j) \\ &= E(I_1 + I_2 + \ldots + I_{300} + 2\sum_{i<j} I_i I_j) \\ &= 300E(I_1) + 2\binom{300}{2} E(I_1 I_2) \end{aligned}$$

Since $I_1 I_2$ is the indicator that the 1st and 2nd people both get their own tests, $E(I_1 I_2) = \frac{300}{300^2} \cdot \frac{300}{300^2-1}$. Thus,

$$E(X^2) = 300 \cdot \frac{1}{300} + 2\binom{300}{2} \cdot \frac{300}{300^2} \cdot \frac{300}{300^2 - 1} = 1 + \frac{300 \cdot 299}{(300+1)(300-1)} = \frac{601}{301}$$

and

$$\text{Var}(X) = E(X^2) - 1 = \frac{601}{301} - 1 = \boxed{\frac{300}{301}}$$

(c) Yes, the expected value 1 is approximately equal to the variance $\frac{300}{301}$, which is characteristic of a Poisson distribution. The probability would therefore be equal to the PDF at $X = x$ where $X \sim Pois(\lambda)$ and $\lambda = 1$, the expected value:

$$P(X = x) = \frac{e^{-\lambda}\lambda^2}{2!}$$

Inequality Practice

Let X and Y be independent and identical (i.i.d) random variables. I_j is an indicator variable.

Fill out the inequalities ($\leq, \geq, =$).

$$E[Y|2X] \quad \_\_\_\_\_\_ \quad E[Y|X]$$

$$P(X+Y=2) \quad \_\_\_\_\_\_ \quad P(X=1)P(Y=1)$$

$$E[X|Y] \quad \_\_\_\_\_\_ \quad E[X|Y=2]$$

$$E[\frac{X^2}{X^2+Y^2}] \quad \_\_\_\_\_\_ \quad \frac{1}{4}$$

$$Var(X^2) \quad \_\_\_\_\_\_ \quad E(X^4)$$

$$E[X^2] \quad \_\_\_\_\_\_ \quad E[XY]$$

$$E[2X+c] \quad \_\_\_\_\_\_ \quad 2E[X]+2c$$

$$E[|X|] \quad \_\_\_\_\_\_ \quad |E[X]|$$

$$E[I_j]+P(I_j=0) \quad \_\_\_\_\_\_ \quad 1$$

$$E[I_j(1-I_j)] \quad \_\_\_\_\_\_ \quad E[1-I_j]$$

Min and Max

Let $X \sim \text{Bin}\left(n, \frac{1}{2}\right)$ and $Y \sim \text{Bin}\left(n+1, \frac{1}{2}\right)$ independently.

(a) Let V = $\min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$. So if X crystalizes to x and Y crystalizes to y, then V crystalizes to $\min(x, y)$ and W crystalizes to $\max(x, y)$. Find $E(V) + E(W)$.

(b) Show that $E|X - Y| = E(W) - E(V)$, with notation as in (a).

(c) Compute $\text{Var}(n - X)$.

True and False

(a) If X and Y have the same CDF, they have the same expectation.

(b) If X and Y have the same expectation, they have the same CDF.

(c) If X and Y have the same CDF, they must be dependent.

(d) If X and Y have the same CDF, they must be independent.

(e) If X and Y have the same distribution, $P(X < Y)$ is at most $^1/_2$.

(f) If X and Y are independent and have the same distribution, $P(X < Y)$ is at most $^1/_2$.