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

Practice Problems Solutions

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) $Binom(n = 1, p = 0.25)$. This is the same thing as $Bern(0.25)$ **distribution**

(b) Ω **sample space/support**

(c) $x|y$ **event**

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

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

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

(g) 0.3145 **probability**

(h) I_j, where $I_j \sim Bern(0.5)$ **random variable**

(i) All x where $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ **sample space/support**

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

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

(l) P, where $P \in (0, 1)$ **random variable**

(m) $\varnothing$ **sample space/support**

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

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

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

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

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

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

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

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

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

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

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

(y) Your score on the midterm **random variable**

(z) You acing the midterm! **event**

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 “independent" version of Avalon where 7 freshmen are dealt 7 cards, with the pack containing the four special cards (Merlin, Percival, Mordrid, and Morgana) in addition to 2 generic good and 1 generic bad cards. We now assign all the cards to all players randomly such that each card has the equal probability of being assigned to any one of the student regardless of whether that particular student has been already dealt a card. This will mean it could be the case where one student possesses all of the cards and the other 6 students possesses none of the cards. Additionally, by the pigeonhole principle, this will mean there will be at least one freshman with at least 1 card. How many possible arrangements of cards are **played among the seven players**?

Solutions

(a) $\boxed{\binom{7}{4}}$ Order of players doesn't matter. Replacement does not exist (One student choosing a card will necessarily decrease the cards available for other students). We therefore use the form $\binom{n}{k}$.

(b) $\boxed{\frac{7!}{5!} * \binom{5}{3}}$ Order of players matter for who is chosen Merlin and Percival. In addition, replacement does not exist so we prioritize the first term to be of the form $\frac{n!}{(n-k)!}$. Finally, to choose the remaining goodies and baddies, we use the form $\binom{n}{k}$ since order doesn't matter and replacement does not exist. Finally, we apply multiplication rule for these two terms.

(c) $\boxed{\frac{7!}{3!} * \binom{3}{1}}$ Same reasoning as (b).

(d) $\boxed{2^7}$ We reframe the problem in this way: Each player is ordered to be good or bad. Every time we assign a good card to a player, the rest of the players still have the option to be good. Therefore, the good or bad cards are constantly **replaced** (we can think of replacement as a sample space that never changes in light of events). Finally, the ordering of the players matter, since we are interested in which specific players are good and which specific players are bad. We therefore use the form n^k.

(e) $\boxed{\binom{13}{7}}$ For each card, it is assigned to a student. The order of these students do not matter and the students are being replaced after each card assignment (we can think of replacement as a sample space that never changes in light of events). Therefore, we use $\binom{n+k-1}{k}$.

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 a supposed baddy on the team with a supposed goody. 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 and, similarly, that a mission failed 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 has voiced out all of the goodies who should be on the team since Merlin knows all the goodies. Therefore, all missions will pass. However, Merlin disclosing this information is not desirable since the baddies win the game if they guess correctly who Merlin is.

Let X be the event that Merlin discloses information. Assume the probability that Merlin discloses information is $P(X) = c$ and that G and X are independent. If Merlin is *not* on the team, then the missions are conditionally independent on G. Let $a_0 = P(M_j|G, X^c)$ and $b_0 = P(M_j|G^c, X^c)$. Find the posterior probability that the team consists of all goodies given that all n missions have passed.

Solution

Let $M = M_1 \cap M_2 \ldots \cap M_n$ be the event that all missions pass.

(a) Using Bayes Rule and the LOTP, we get:

$$\begin{aligned} P(G|M) &= \frac{P(M|G)P(G)}{P(M)} \\ &= \frac{P(M|G)P(G)}{P(M|G)P(G) + P(M|G^c)P(G^c)} \\ &= \boxed{\frac{pa^n}{pa^n + qb^n}} \end{aligned}$$

(b) Using the same formula as above, this time we have different values for $P(M|G)$ and $P(M|G^c)$ because we need to condition on whether Merlin discloses information.

$$\begin{aligned} P(M|G) &= P(M|X, G) \cdot P(X|G) + P(T|X^c, G)P(X^c|G) \\ &= P(M|X, G) \cdot P(X) + P(T|X^c, G)P(X^c) = c + (a_0^n)(1-c) \\ P(T|G^c) &= P(T|X, G^c) \cdot P(X|G^c) + P(T|X^c, G^c) \cdot P(X^c|G^c) \\ &= P(T|X, G^c) \cdot P(X) + P(T|X^c, G^c) \cdot P(X^c) = c + b_0^n(1-c) \end{aligned}$$

Therefore, the final answer is:

$$P(G|M) = \boxed{\frac{p(c + (a_0^n)(1-c))}{p(c + (a_0^n)(1-c)) + q(c + b_0^n(1-c))}}$$

Expected Values and Standard Deviation of Indicator Random Variables

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 $E(I_k) = \frac{300}{300^2} = \frac{1}{300}$ 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:

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

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:

$$P(X = x) = \frac{e^{-\lambda}\lambda^2}{2!} = \boxed{\frac{1}{2e}}$$

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 ($\leq, \geq, =$). Also, spot the category error.

(a)
$$E[Y|2X = x] \quad = \quad E[Y|X = x]$$

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

(c)
$$E[X|Y] \quad ? \quad E[X|Y = 2]$$

(d)
$$E[\frac{X^2}{X^2 + Y^2}] \quad \geq \quad \frac{1}{4}$$

(e)
$$Var(X^2) \quad \leq \quad E(X^4)$$

(f)
$$E[X^2] \quad \geq \quad E[XY]$$

(g)
$$E[2X + 2] \quad \leq \quad 2E[X] + 4$$

(h)
$$E[|X|] \quad \geq \quad |E[X]|$$

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

(j)
$$E[I_j(1 - I_j)] \quad \leq \quad E[1 - I_j]$$

(a) X and Y are independent. Therefore, the expected value of Y does not change whether we know the value of $2X$ or just X.

(b) $X + Y = 2$ covers more of the sample space than $X = 1, Y = 1$

(c) Category Error!

(d) Using symmetry: X and Y are independent and identical so that $E[\frac{X^2}{X^2+Y^2}] = E[\frac{Y^2}{X^2+Y^2}] = \frac{1}{2}$

(e) Using definition of variance: $Var(X^2) = E[X^4] - E[X^2]^2 \leq E[X^4]$

(f) Using independence: $E[XY] = E[X]E[Y] = E[X]^2 \leq E[X^2]$

(g) Using linearity: $E[2X + 2] = 2E[X] + 2$

(h) Consider the standard normal distribution or any distribution with positive values, negative values, and a mean centered at 0.

(i) Using Fundamental Bridge: $E[I_j] + P(I_j = 0) = P(I_j = 1) + P(I_j = 0) = 1$

(j) Indicator random variables follow the Bernoulli distribution, taking on the support of only 0 and 1. $E[I_j(1 - I_j)] = 0$ when $I_j = 0$ and when $I_j = 1$. Therefore, $E[I_j(1 - I_j)] = 0 \leq E[1 - I_j]$, as $E[1 - I_j]$ can take on a positive value as long as $p \neq 1$ for $I_j \sim Bern(p)$

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

Solution

(a) Note that $V + W = X + Y$ because adding the larger and smaller of two numbers is the same as adding both numbers). Therefore, by linearity of expectation, we get:

$$E(V) + E(W) = E(V + W) = E(X + Y) = E(X) + E(Y) = \boxed{\frac{2n+1}{2}}$$

(b) Note that $|X - Y| = W - V$ because the absolute difference between two numbers is equal to the larger minus the smaller. Therefore,

$$E|X - Y| = E(W - V) = E(W) - E(V)$$

(c) We can use the property of variances, which says that $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$. We get

$$\text{Var}(n - X) = \text{Var}(n) + \text{Var}(-X) = 0 + \text{Var}(X) = \boxed{n/4}$$

.

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 $^1/_2$.

(g) If X and Y share the same expected values, symmetry can be applied.

(h) It is impossible for X and Y to be dependent and share the same distribution without being identical.

Solution

(a) **True**. If X and Y have the same CDF, then they would have the same distribution, and hence the same expectation.

(b) **False**. Consider the expected values and CDFs of $\text{Bern}(1)$ and $\text{Bin}(2, 0.5)$.

(c) **False**. Having the same CDF (and equivalently distribution) does not indicate anything about the dependency of the two random variables.

(d) **False**. Same as above.

(e) **False**. Think of a clock with two hands, whereby one of the hands is always one unit faster than the other hand. They share the same distribution of probabilities, but one is always greater than the other by dependence.

(f) **True**. This means that X and Y are i.i.d., and hence by symmetry, $P(X < Y) = {^1/_2} \cdot (1 - P(X = Y))$ which is at most $\frac{1}{2}$.

(g) **False**. Same expectations does not imply same distribution, seen in (b). If two variables X and Y are independent and have the same distribution, then symmetry holds.

(h) **False**. Think of a clock with two hands, whereby one of the hands is always one unit faster than the other hand. They share the same distribution of probabilities, are dependent, and are not identical.