Exam 1 Practice Questions I –solutions, 18.05, Spring 2022

Notes.

Not every possible problem can be covered in 13 problems. Look at the other review problems as well as the readings, psets and class problems.

Even the first 13 problems are much longer than the actual test will be,

Problem 1. *In a ballroom dancing class the students are divided into group A and group B. There are six people in group A and seven in group B. If four As and four Bs are chosen and paired off, how many pairings are possible?*

Solution: Build the pairings in stages and count the ways to build each stage:

Stage 1: Choose the 4 from group A: $\binom{6}{4}$.

Stage 2: Choose the 4 from group B: $\binom{7}{4}$

We need to be careful because we don't want to build the same 4 couples in multiple ways.
Line up the 4 A's A_1, A_2, A_3, A_4
Stage 3: Choose a partner from the 4 Bs for A_1: 4.
Stage 4: Choose a partner from the remaining 3 Bs for A_2: 3
Stage 5: Choose a partner from the remaining 2 Bs for A_3: 2
Stage 6: Pair the last B with A_4: 1

Number of possible pairings: $\binom{6}{4}\binom{7}{4}4!$.

Note: we could have done stages 3-6 in one go as: Stages 3-6: Arrange the 4 Bs opposite the 4 As: 4! ways.

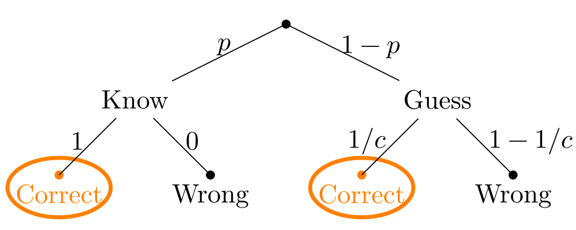
Problem 2. *Let A and B be two events. Suppose the probability that neither A or B occurs is 2/3. What is the probability that one or both occur?*

Solution: We are given $P(A^c \cap B^c) = 2/3$ and asked to find $P(A \cup B)$.

$A^c \cap B^c = (A \cup B)^c \;\Rightarrow\; P(A \cup B) = 1 - P(A^c \cap B^c) = \boxed{1/3.}$

Problem 3. *Suppose you are taking a multiple-choice test with c choices for each question. In answering a question on this test, the probability that you know the answer is p. If you don't know the answer, you choose one at random. What is the probability that you knew the answer to a question, given that you answered it correctly?*

Solution: The following tree shows the setting



Let C be the event that you answer the question correctly. Let K be the event that you actually know the answer. The left circled node shows $P(K \cap C) = p$. Both circled nodes together show $P(C) = p + (1-p)/c$. So,

$$P(K|C) \;=\; \frac{P(K\cap C)}{P(C)} \;=\; \frac{p}{p+(1-p)/c}$$

Or we could use the algebraic form of Bayes' theorem and the law of total probability: Let G stand for the event that you're guessing. Then we have,
$P(C|K) = 1$, $P(K) = p$, $P(C) = P(C|K)P(K) + P(C|G)P(G) = p + (1-p)/c$. So,

$$P(K|C) = \frac{P(C|K)P(K)}{P(C)} = \frac{p}{p+(1-p)/c}$$

Problem 4. *Two dice are rolled.*

A = 'sum of two dice equals 3'
B = 'sum of two dice equals 7'
C = 'at least one of the dice shows a 1'

(a) *What is $P(A|C)$?*

(b) *What is $P(B|C)$?*

(c) *Are A and C independent? What about B and C?*

Solution: Sample space =

$$\Omega = \{(1,1),\,(1,2),\,(1,3),\,\ldots,\,(6,6)\,\} = \{(i,j)\,|\,i,j = 1,\,2,\,3,\,4,\,5,\,6\,\}.$$

(Each outcome is equally likely, with probability 1/36.)

$A = \{(1,2),\,(2,1)\}$,

$B = \{(1,6),\,(2,5),\,(3,4),\,(4,3),\,(5,2),\,(6,1)\}$

$C = \{(1,1),\,(1,2),\,(1,3),\,(1,4),\,(1,5),\,(1,6),\,(2,1),\,(3,1),\,(4,1),\,(5,1),\,(6,1)\}$

(a) $P(A|C) = \dfrac{P(A\cap C)}{P(C)} = \dfrac{2/36}{11/36} = \dfrac{2}{11}..$

(b) $P(B|C) = \dfrac{P(B\cap C)}{P(C)} = \dfrac{2/36}{11/36} = \dfrac{2}{11}..$

(c) $P(A) = 2/36 \neq P(A|C)$, so they are not independent. Similarly, $P(B) = 6/36 \neq P(B|C)$, so they are not independent.

Problem 5. *Suppose that $P(A) = 0.4, P(B) = 0.3$ and $P((A\cup B)^C) = 0.42$. Are A and B independent?*

Solution: We have $P(A\cup B) = 1 - 0.42 = 0.58$ and we know because of the inclusion-exclusion principle that

$$P(A\cup B) = P(A) + P(B) - P(A\cap B).$$

Thus,

$$P(A\cap B) = P(A) + P(B) - P(A\cup B) = 0.4 + 0.3 - 0.58 = 0.12 = (0.4)(0.3) = P(A)P(B).$$

So, A and B are independent.

Problem 6. *Suppose that X takes values between 0 and 1 and has probability density function $2x$. Compute $\mathrm{Var}(X)$ and $\mathrm{Var}(X^2)$.*

Solution: We will make use of the formula $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2$. First we compute

$$E[X] = \int_0^1 x \cdot 2x dx = \frac{2}{3}$$

$$E[X^2] = \int_0^1 x^2 \cdot 2x dx = \frac{1}{2}$$

$$E[X^4] = \int_0^1 x^4 \cdot 2x dx = \frac{1}{3}.$$

Thus,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

and

$$\mathrm{Var}(X^2) = E[X^4] - \left(E[X^2]\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Problem 7. *Suppose 100 people all toss a hat into a box and then proceed to randomly pick out of a hat. What is the expected number of people to get their own hat back.*

Hint: express the number of people who get their own hat as a sum of random variables whose expected value is easy to compute.

Solution: Let X be the number of people who get their own hat.

Following the hint: let X_j represent whether person j gets their own hat. That is, $X_j = 1$ if person j gets their hat and 0 if not.

We have, $X = \sum_{j=1}^{100} X_j$, so $E[X] = \sum_{j=1}^{100} E[X_j]$.

Since person j is equally likely to get any hat, we have $P(X_j = 1) = 1/100$. Thus, $X_j \sim \mathrm{Bernoulli}(1/100) \Rightarrow E[X_j] = 1/100 \Rightarrow \boxed{E[X] = 1.}$

Problem 8. *Let T be the waiting time for customers in a queue. Suppose that T is exponential with pdf $f(t) = 2\mathrm{e}^{-2t}$ on $[0, \infty)$.*

Find the pdf of the rate at which customers are served $R = 1/T$.

Solution: The CDF for T is

$$F_T(t) = P(T \le t) = \int_0^t 2\mathrm{e}^{-2u}\, du = -\mathrm{e}^{-2u}\Big|_0^t = 1 - \mathrm{e}^{-2t}.$$

Next, we find the CDF of R. R takes values in $(0, \infty)$.

For $0 < r$,

$$F_R(r) = P(R \le r) = P(1/T < r) = P(T > 1/r) = 1 - F_T(1/r) = \mathrm{e}^{-2/r}.$$

We differentiate to get $f_R(r) = \dfrac{d}{dr}\left(\mathrm{e}^{-2/r}\right) = \dfrac{2}{r^2}\mathrm{e}^{-2/r}$.

Problem 9. *Suppose that the cdf of X is given by:*

$$F(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{1}{5} & \text{for } 0 \le a < 2 \\ \frac{2}{5} & \text{for } 2 \le a < 4 \\ 1 & \text{for } a \ge 4. \end{cases}$$

Determine the pmf of X.

Solution: The jumps in the distribution function are at 0, 2, 4. The value of $p(a)$ at a jump is the height of the jump:

a	0	2	4
$p(a)$	1/5	1/5	3/5

Problem 10. Exponential Distribution
Suppose that buses arrive are scheduled to arrive at a bus stop at noon but are always X minutes late, where X is an exponential random variable with probability density function $f_X(x) = \lambda e^{-\lambda x}$. Suppose that you arrive at the bus stop precisely at noon.

(a) *Compute the probability that you have to wait for more than five minutes for the bus to arrive.*

Solution: We compute

$$P(X \ge 5) = 1 - P(X < 5) = 1 - \int_0^5 \lambda \mathrm{e}^{-\lambda x}\,dx = 1 - (1 - \mathrm{e}^{-5\lambda}) = \mathrm{e}^{-5\lambda}.$$

(b) *Suppose that you have already waiting for 10 minutes. Compute the probability that you have to wait an additional five minutes or more.*

Solution: We want $P(X \ge 15 | X \ge 10)$. First observe that $P(X \ge 15, X \ge 10) = P(X \ge 15)$. From similar computations in (a), we know

$$P(X \ge 15) = \mathrm{e}^{-15\lambda} \qquad\qquad P(X \ge 10) = \mathrm{e}^{-10\lambda}.$$

From the definition of conditional probability,

$$P(X \ge 15 | X \ge 10) = \frac{P(X \ge 15, X \ge 10)}{P(X \ge 10)} = \frac{P(X \ge 15)}{P(X \ge 10)} = \mathrm{e}^{-5\lambda}$$

Note: This is an illustration of the memorylessness property of the exponential distribution.

Problem 11. *Let X and Y be two continuous random variables with joint pdf*

$$f(x,y) = cx^2y(1+y) \quad \text{for } 0 \le x \le 3 \text{ and } 0 \le y \le 3,$$

and $f(x,y) = 0$ otherwise.

(a) *Find the value of c.*

(b) *Find the probability $P(1 \le X \le 2,\, 0 \le Y \le 1)$.*

(c) *Determine the joint cdf, $F(a,b)$, of X and Y for a and b between 0 and 3.*

(d) *Find marginal cdf $F_X(a)$ for a between 0 and 3.*

(e) *Find the marginal pdf $f_X(x)$ directly from $f(x,y)$ and check that it is the derivative of $F_X(x)$.*

(f) *Are X and Y independent?*

Solution: **(a)** Total probability must be 1, so

$$1 = \int_0^3 \int_0^3 f(x,y)\,dy\,dx = \int_0^3 \int_0^3 c(x^2y + x^2y^2)\,dy\,dx = c \cdot \frac{243}{2},$$

(Here we skipped showing the arithmetic of the integration) Therefore, $c = \dfrac{2}{243}$.

(b)

$$\begin{aligned} P(1 \le X \le 2,\, 0 \le Y \le 1) &= \int_1^2 \int_0^1 f(x,y)\,dy\,dx \\ &= \int_1^2 \int_0^1 c(x^2y + x^2y^2)\,dy\,dx \\ &= c \cdot \frac{35}{18} \\ &= \frac{70}{4374} \approx 0.016 \end{aligned}$$

(c) For $0 \le a \le 3$ and $0 \le b \le 3$. we have

$$F(a,b) = \int_0^a \int_0^b f(x,y)dy\,dx = c\left(\frac{a^3b^2}{6} + \frac{a^3b^3}{9}\right)$$

(d) Since $y = 3$ is the maximum value for Y, we have

$$F_X(a) = F(a,3) = c\left(\frac{9a^3}{6} + 3a^3\right) = \frac{9}{2}c\,a^3 = \frac{a^3}{27}$$

(e) For $0 \le x \le 3$, we have, by integrating over the entire range for y,

$$f_X(x) = \int_0^3 f(x,y)\,dy = cx^2\left(\frac{3^2}{2} + \frac{3^3}{3}\right) = c\frac{27}{2}x^2 = \frac{1}{9}x^2.$$

This is consistent with (c) because $\frac{d}{dx}(x^3/27) = x^2/9$.

(f) Since $f(x,y)$ separates into a product as a function of x times a function of y we know X and Y are independent.

Problem 12. (Table of normal probabilities)
Use the table of standard normal probabilities to compute the following. (Z is the standard normal.)

(a) *(i)* $P(Z \leq 1.5)$ *(ii)* $P(-1.5 < Z < 1.5)$ $P(Z > -0.75)$.

(b) *Suppose* $X \sim N(2, (0.5)^2)$. *Find (i)* $P(X \leq 2)$ *(ii)* $P(1 < X \leq 1.75)$.

Solution: (a) (i) 0.9332 (ii) 0.9332 - 0.0668 = 0.8664

(iii) By symmetry $= P(Z < 0.75) = 0.7734$. (Or we could have used $1 - P(Z > -0.75)$.)

(b) (i) Since 2 is the mean of the normal distribution, $P(X \leq 2) = 0.5$.

(ii) Standardizing,

$$P(1 < X \leq 1.75) = P\left(\frac{1-2}{0.5} < Z \leq \frac{1.75-2}{0.5}\right) = P(-2 < Z < -0.5) = 0.3085 - 0.0228 = \boxed{0.2857}.$$

Problem 13. (**Central Limit Theorem**)
Let $X_1, X_2, \ldots, X_{81}$ *be i.i.d., each with expected value* $\mu = E[X_i] = 5$, *and variance* $\sigma^2 = \mathrm{Var}(X_i) = 4$. *Approximate* $P(X_1 + X_2 + \cdots X_{81} > 369)$, *using the central limit theorem.*

Solution: Let $T = X_1 + X_2 + \ldots + X_{81}$. The central limit theorem says that

$$T \approx \mathrm{N}(81 * 5, 81 * 4) = \mathrm{N}(405, 18^2)$$

Standardizing we have

$$\begin{aligned} P(T > 369) &= P\left(\frac{T-405}{18} > \frac{369-405}{18}\right) \\ &\approx P(Z > -2) \\ &\approx 0.975 \end{aligned}$$

The value of 0.975 comes from the rule-of-thumb that $P(|Z| < 2) \approx 0.95$. A more exact value (using R) is $P(Z > -2) \approx 0.9772$.

More problems

Problem 14. *There are 3 arrangements of the word DAD, namely DAD, ADD, and DDA. How many arrangements are there of the word PROBABILITY?*

Solution: Sort the letters: A BB II L O P R T Y. There are 11 letters in all. We build arrangements by starting with 11 'slots' and placing the letters in these slots, e.g

$$\underline{A}\ \underline{B}\ \underline{I}\ \underline{B}\ \underline{I}\ \underline{L}\ \underline{O}\ \underline{P}\ \underline{R}\ \underline{T}\ \underline{Y}$$

Create an arrangement in stages and count the number of possibilities at each stage:

Stage 1: Choose one of the 11 slots to put the A: $\binom{11}{1}$

Stage 2: Choose two of the remaining 10 slots to put the B's: $\binom{10}{2}$

Stage 3: Choose two of the remaining 8 slots to put the I's: $\binom{8}{2}$

Stage 4: Choose one of the remaining 6 slots to put the L: $\binom{6}{1}$

Stage 5: Choose one of the remaining 5 slots to put the O: $\binom{5}{1}$

Stage 6: Choose one of the remaining 4 slots to put the P: $\binom{4}{1}$

Stage 7: Choose one of the remaining 3 slots to put the R: $\binom{3}{1}$

Stage 8: Choose one of the remaining 2 slots to put the T: $\binom{2}{1}$

Stage 9: Use the last slot for the Y: $\binom{1}{1}$

Number of arrangements:

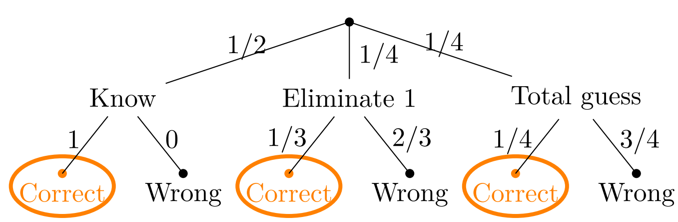
$$\binom{11}{1}\binom{10}{2}\binom{8}{2}\binom{6}{1}\binom{5}{1}\binom{4}{1}\binom{3}{1}\binom{2}{1}\binom{1}{1} = 11 \cdot \frac{10 \cdot 9}{2} \cdot \frac{8 \cdot 7}{2} \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 9979200$$

Note: choosing 11 out of 1 is so simple we could have immediately written 11 instead of belaboring the issue by writing $\binom{11}{1}$. We wrote it this way to show one systematic way to think about problems like this.

Problem 15. *A multiple choice exam has 4 choices for each question. A student has studied enough so that the probability they will know the answer to a question is 0.5, the probability that they will be able to eliminate one choice is 0.25, otherwise all 4 choices seem equally plausible. If they know the answer they will get the question right. If not they have to guess from the 3 or 4 choices.*

As the teacher you want the test to measure what the student knows. If the student answers a question correctly what's the probability they knew the answer?

Solution: We show the probabilities in a tree:



For a given problem let C be the event the student gets the problem correct and K the event the student knows the answer.

The question asks for $P(K|C)$.

We'll compute this using Bayes' rule:

$$P(K|C) = \frac{P(C|K)\,P(K)}{P(C)} = \frac{1 \cdot 1/2}{1/2 + 1/12 + 1/16} = \frac{24}{31} \approx 0.774 = 77.4\%$$

Problem 16. *Compute the expectation and variance of a Bernoulli(p) random variable.*

Solution: Make a table:

X:	0	1
prob:	(1-p)	p
X^2	0	1.

From the table, $\boxed{E[X] = 0 \cdot (1-p) + 1 \cdot p = p.}$

Since X and X^2 have the same table $E[X^2] = E[X] = p$.

Therefore, $\boxed{\text{Var}(X) = p - p^2 = p(1-p).}$

Problem 17. Transforming Normal Distributions
Suppose $Z \sim$ N(0,1) and $Y = \mathrm{e}^Z$.

(a) *Find the cdf $F_Y(a)$ and pdf $f_Y(y)$ for Y. (For the CDF, the best you can do is write it in terms of Φ the standard normal cdf.)*

Solution: Note, Y follows what is called a *log-normal distribution.*

$F_Y(a) = P(Y \le a) = P(\mathrm{e}^Z \le a) = P(Z \le \ln(a)) = \boxed{\Phi(\ln(a)).}$

Differentiating using the chain rule:

$$f_y(a) = \frac{d}{da} F_Y(a) = \frac{d}{da}\Phi(\ln(a)) = \frac{1}{a}\phi(\ln(a)) = \boxed{\frac{1}{\sqrt{2\pi}\,a}\mathrm{e}^{-(\ln(a))^2/2}.}$$

(b) *We don't have a formula for $\Phi(z)$ so we don't have a formula for quantiles. So we have to write quantiles* in terms *of Φ^{-1}.*

(i) Write the 0.33 quantile of Z in terms of Φ^{-1}

(ii) Write the 0.9 quantile of Y in terms of Φ^{-1}.

(iii) Find the median of Y.

Solution: (i) The 0.33 quantile for Z is the value $q_{0.33}$ such that $P(Z \le q_{0.33}) = 0.33$. That is, we want

$$\Phi(q_{0.33}) = 0.33 \;\Leftrightarrow\; \boxed{q_{0.33} = \Phi^{-1}(0.33)}.$$

(ii) We want to find $q_{0.9}$ where

$$F_Y(q_{0.9}) = 0.9 \;\Leftrightarrow\; \Phi(\ln(q_{0.9})) = 0.9 \;\Leftrightarrow\; \boxed{q_{0.9} = \mathrm{e}^{\Phi^{-1}(0.9)}}.$$

(iii) As in (ii) $q_{0.5} = \mathrm{e}^{\Phi^{-1}(0.5)} = \mathrm{e}^0 = \boxed{1}$.

Problem 18. *Suppose that $X \sim Bin(n, 0.5)$. Find the probability mass function of $Y = 2X$.*

Solution: For $y = 0, 2, 4, \ldots, 2n$,

$$P(Y = y) = P(X = \frac{y}{2}) = \binom{n}{y/2}\left(\frac{1}{2}\right)^n.$$