PROBABILITY: QUIZ 1

MARK DITSWORTH

1. PROBLEM 1

Indicate *True* or *False* for each claim.

1.1. Part A. If $\mathbf{P}(A) \leq \mathbf{P}(B)$, then $A \subseteq B$.
False. Let A and B be independent, but $\mathbf{P}(A) \leq \mathbf{P}(B)$.

1.2. Part B. If $\mathbf{P}(B) > 0$, $\mathbf{P}(A|B)$ is at least as large as $\mathbf{P}(A)$.
False. $\mathbf{P}(A)$ can decrease if it is dependent on B. For example, the probability of rain today ($\mathbf{P}(A)$) may historically be 35%. But the probability of rain today given clear skies ($\mathbf{P}(A|B)$) is lower than $\mathbf{P}(A)$.

1.3. Part C. If $\mathbf{E}[X] > \mathbf{E}[Y]$, then $\mathbf{E}[X^2] > \mathbf{E}[Y^2]$.
False. Let X take the values 0 and 1 with equal probability, and Y take the values -4 and -2 with equal probability. $\mathbf{E}[X] = 0.5 > \mathbf{E}[Y] = -3$. But $\mathbf{E}[X^2] = 0.5 < \mathbf{E}[Y^2] = 10$.

1.4. Part D. Suppose $\mathbf{P}(A) > 0$. Then $\mathbf{E}[X] = \mathbf{E}[X|A] + \mathbf{E}[X|A^C]$.
False. Assume X is independent of A. Then $\mathbf{E}[X] = \mathbf{E}[X|A] = \mathbf{E}[X|A^C]$, so $\mathbf{E}[X|A] + \mathbf{E}[X|A^C] = 2\mathbf{E}[X]$.

1.5. Part E. If X and Y are independent and $\mathbf{P}(C) > 0$, then $p_{X,Y|C}(x,y) = p_{X|C}(x)p_{Y|C}(y)$.
False. Independence between X and Y does not guarantee conditional independence.

1.6. Part F. If for some constant c, we have $\mathbf{P}(\{X > c\}) = \frac{1}{2}$, then $\mathbf{E}[X] > \frac{c}{2}$.
False. Let X take the values between -1 and 1, uniformly. $\mathbf{P}(\{X > 0\}) = \frac{1}{2}$, and $\mathbf{E}[X] = 0 \not> \frac{0}{2}$.

1.7. Part G. In a game involving the flip of a fair coin, you win a dollar for each head flipped. The maximum number of flips is 10, but the game ends when you flip a tail. The expected gain from this game is 1.
False. $\mathbf{E}[g(X)] = \sum g(x)p_X(x) = \sum_{k=0}^{9}\left(k(0.5)^{k+1}\right) + (0.5)^{10} = 0.99902$.

1.8. **Part H.** Let X be a uniformly distributed continuous random variable over some interval $[a, b]$. We can uniquely describe $f_X(x)$ from knowing its mean and variance.

True. For a uniform distribution, $\mathbf{E}[X] = \frac{a+b}{2}$ and $\sigma_X^2 = \frac{(b-a)^2}{12}$. Two equations and two unknowns result in a solvable system of equations for a and b.

1.9. **Part I.** Let X be an exponentially distributed random variable with a probability density function $f_X(x) = e^{-x}$. $P(\{0 \leq X \leq 3\} \cup \{2 \leq X \leq 4\}) = 1 - e^{-4}$.

True. $P(\{0 \leq X \leq 3\} \cup \{2 \leq X \leq 4\}) = \int_0^4 e^{-x} dx = 1 - e^{-4}$.

1.10. **Part J.** Let X be a normal random variable with mean 1 and variance 2. Let Y be a normal random variable with mean 1 and variance 1. $\mathbf{P}(X < 0) < \mathbf{P}(Y < 0)$.

False. Since X has a larger variance than Y its tails are thicker, thus the integral over the left tail will be greater for X and Y.

2. Problem 2

Marketing Team A has found that 1/4 of their customers are low-frequency buyers. But out of their low-frequency buyers, 1/3 are high spenders, whereas 1/10 of their high-frequency buyers are high spenders. Assume that each customer is either low-frequency or high frequency, and low spender of high spender.

2.1. **Part A.** Compute the probability that a randomly chosen customer is a high spender.
$\mathbf{P}(HS) = \mathbf{P}(HFB)\mathbf{P}(HS|HFB) + \mathbf{P}(LFB)\mathbf{P}(HS|LFB) = \frac{1}{4}\frac{1}{3} + \frac{3}{4}\frac{1}{10} = \frac{19}{120}$.

2.2. **Part B.** Compute the probability that a randomly chosen customer is a high-frequency buyer given that they are a low spender.
$\mathbf{P}(HFB|LS) = \frac{\mathbf{P}(HFB)\mathbf{P}(LS|HFB)}{\mathbf{P}(LS)} = \frac{\frac{3}{4}\frac{9}{10}}{1 - \frac{19}{120}} = \frac{81}{101}$

You are told that the only products sold are books (B), CDs (CD), and DVDs (DVD). Marketing team B has found three customer groups C_1, C_2, C_3. These groups are collectively exhaustive and mutually exclusive. Each customer is equally likely to be in any group, and customers are *i.i.d.*. Each customer only buys one item per day. Marketing team B has determined the following purhcasing conditional probabilities:

$$
\begin{aligned}
\mathbf{P}(B|C_1) &= \tfrac{1}{2} \\
\mathbf{P}(CD|C_1) &= \tfrac{1}{4} \\
\mathbf{P}(DVD|C_1) &= \tfrac{1}{4} \\
\mathbf{P}(B|C_2) &= \tfrac{1}{2} \\
\mathbf{P}(CD|C_2) &= 0 \\
\mathbf{P}(DVD|C_2) &= \tfrac{1}{2} \\
\mathbf{P}(B|C_3) &= \tfrac{1}{3} \\
\mathbf{P}(CD|C_3) &= \tfrac{1}{3} \\
\mathbf{P}(DVD|C_3) &= \tfrac{1}{3}
\end{aligned}
$$

2.3. **Part C.** Compute the probability that a customer purchases a book or CD.
Since C_1, C_2, and C_3 are mutually exclusive and customers are *i.i.d.*,
$\mathbf{P}(B \cup CD) = \mathbf{P}(B) + \mathbf{P}(CD)$
$\mathbf{P}(B) = \sum_{i=1}^{3} \mathbf{P}(C_i)\mathbf{P}(B|C_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{9} = \frac{4}{9}$
$\mathbf{P}(CD) = \sum_{i=1}^{3} \mathbf{P}(C_i)\mathbf{P}(CD|C_i) = \frac{1}{12} + 0 + \frac{1}{9} = \frac{7}{36}$
$\mathbf{P}(B \cup CD) = \frac{4}{9} + \frac{7}{36} = \frac{23}{36} \approx 0.6389$

2.4. **Part D.** Compute the probability that a customer is in group C_2 or C_3 given they purchased a book.
$\mathbf{P}(C_2 \cup C_3|B) = \frac{\mathbf{P}(C_2 \cup C_3)\mathbf{P}(B|C_2 \cup C_3)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B \cap (C_2 \cup C_3))}{\mathbf{P}(B)} = \frac{\mathbf{P}(B \cap C_2) + \mathbf{P}(B \cap C_3)}{\mathbf{P}(B)} = \frac{5}{8}$

Each book costs \$15, each CD costs \$10, and each DVD costs \$15.

2.5. **Part E.** Compute the PDF, expected value, and variance of the revenue (R) collected by a single customer's purchase.

$\mathbf{P}(R = \$10) = \mathbf{P}(CD) = \frac{7}{36}$

$\mathbf{P}(R = \$15) = \mathbf{P}(B) + \mathbf{P}(DVD) = \frac{4}{9} + \frac{13}{36}$

$$p_R(r) = \begin{cases} \frac{7}{36} & r = \$10 \\ \frac{29}{36} & r = \$15 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbf{E}[R] = \sum r \cdot p_R(r) = 10(\frac{7}{36}) + 15(\frac{29}{36}) = \$14.03$

$\text{Var}(R) = \sum (r - \mathbf{E}[R])^2 \cdot p_R(r) \approx 3.9159$

2.6. **Part F.** If the store gets n customers in a day, what is the total expected revenue and variance?

Total Revenue $TR = nR$

$\mathbf{E}[TR] = \mathbf{E}[nR] = n\mathbf{R} = n \cdot \$14.03$

$\text{Var}(nR) = n^2\text{Var}(R) = n^2 \cdot \$3.9159$

2.7. **Part G.** Skip is an abnormal customer. When he goes to the store, he flips a fair coin until he gets his seconds tails. The number of heads that have been flipped is the number of DVDs he buys. What is the expected dollar amount he spends and the variance?

This process follows the kth arrival time process, modeled by the pascal distribution. Each time Skip flips his coin, he has a 50% probability of getting tails. We want to know the expected number of flips it takes to get his second tails. The number of heads is 2 less than this number since 2 of them are the 2 arrivals of tails.

For pascal, $\mathbf{E}[Y_k] = \frac{k}{p}$

$\mathbf{E}[Y_2] = \frac{2}{0.5} = 4$

Thus, the expected number of heads is 2, resulting in an expected revenue of $30.

For pascal, $\text{Var}[Y_k] = \frac{k(1-p)}{p^2}$

$\text{Var}(Y_2) = \frac{2(1-0.5)}{0.5^2} = 4$

$\text{Var}(R) = \text{Var}(\$15 \cdot Y_2) = \$15^2\text{Var}(Y_2) = \$900$

3. Problem 3

We have s urns and n balls $(n \geq s)$. Each ball is placed in an urn at random. Each ball has equal probability of being placed in any urn, and it's placement is independent of the other ball placements. Each urn can fit any number of balls. Define the following random variables:

Let X_i be the number of balls in urn i.
Let Y_k be the number of urns that have exactly k balls.

3.1. Part A. Are the X_i's independent? Why or why not?
They are not independent. Without loss of generality, if $X_1 = j$, then X_2 through X_s can be at most $n - j$.

3.2. Part B. Find the PMF, mean, and variance of X_i.

If there are k balls in urn i, there must have been k occurrences of successful placement, and $n - k$ occurrences of unsuccessful placement: $p^k(1-p)^{(n-k)}$. Since the order in which the k balls are placed in urn i does not matter, there are $\binom{n}{k}$ possible combinations of this placement. Since urn placement is equally distributed, the probability of placement in urn i is $\frac{1}{s}$. Thus, the PMF is:

$$p_{X_i}(k) = \left(\frac{1}{s}\right)^k \left(1 - \frac{1}{s}\right)^{(n-k)} \binom{n}{k}$$

This PMF is the binomial distribution. Thus, the mean is:

$$\mathbf{E}(X_i) = np = \frac{n}{s}$$

and the variance is:

$$\text{Var}(X_i) = np(1-p) = \frac{n}{s}\left(1 - \frac{1}{s}\right)$$

3.3. Part C. Assume $n = 10$ and $s = 3$. Find the probability that the first urn has 3 balls, the second has 2, and the third has 5.

The number of combinations of dividing the 10 balls into partitions of 3, 2, and 5 is $\binom{10}{3,2,5}$. The total number of ways to divide 10 balls among the 3 urns is 3^{10}. Thus, the probability is $\frac{\binom{10}{3,2,5}}{3^{10}} \approx 0.0427$.

3.4. **Part D.** Compute $\mathbf{E}[Y_k]$.

Recall that Y_k is the number of urns that have exactly k balls. Define:

$$I_i = \begin{cases} 1 & \text{urn } i \text{ has } k \text{ balls} \\ 0 & \text{otherwise} \end{cases}$$

Now we can write $Y_k = \sum_{i=1}^{s} I_i$. Due to linearity, $\mathbf{E}[Y_k] = \mathbf{E}\left[\sum_{i=1}^{s}\right] = \sum_{i=1}^{s} \mathbf{E}[I_i]$.

$\mathbf{E}[I_i]$ is just the probability that urn i has k balls, $p_{X_i}(k)$, which is known from Part B. Thus,

$$\mathbf{E}[Y_k] = \sum_{i=1}^{s} \left(\frac{1}{s}\right)^k \left(1 - \frac{1}{s}\right)^{(n-k)} \binom{n}{k} = s \left(\frac{1}{s}\right)^k \left(1 - \frac{1}{s}\right)^{(n-k)} \binom{n}{k}$$

3.5. **Part E.** Compute $\text{Var}(Y_k)$.

$\text{Var}(Y_k) = \mathbf{E}[Y_k^2] - \mathbf{E}[Y_k]^2$. The second term is known from Part D. Thus, we need to find $\mathbf{E}[Y_k^2]$.

$$\mathbf{E}[Y_k^2] = \mathbf{E}\left[\left(\sum_{i=1}^{s} I_i\right)^2\right] = \mathbf{E}\left[\sum_{i=1}^{s} I_i^2\right] + 2\mathbf{E}\left[\sum_{i=1}^{s-1}\sum_{j=i+1}^{s} I_i I_j\right]$$

Since I_i is either 0 or 1, $I_i^2 = I_i$. Thus, $\mathbf{E}\left[\sum_{i=1}^{s} I_i^2\right] = \mathbf{E}\left[\sum_{i=1}^{s} I_i\right] = \mathbf{E}[Y_k]$, which is known from Part D. We now need to find $\mathbf{E}\left[\sum_{i=1}^{s-1}\sum_{j=i+1}^{s} I_i I_j\right]$ to be done.

$\mathbf{E}\left[\sum_{i=1}^{s-1}\sum_{j=i+1}^{s} I_i I_j\right] = \sum_{i=1}^{s-1}\sum_{j=i+1}^{s} \mathbf{E}[I_i I_j]$

$\mathbf{E}[I_i I_j]$ is the probability that X_i and X_j both equal k. This is $2k$ successful placements and $n - 2k$ unsuccessful placements. The number of combinations of these placements is the number of ways to partition the n balls into 2 sets of size k and a third set of size $n-2k$. Thus this probability is $\binom{n}{k,k,n-2k} \left(\frac{1}{s}\right)^{2k} \left(1 - \frac{1}{s}\right)^{n-2k}$.

Putting it all together:

$$\text{Var}(Y_k) = s \left(\frac{1}{s}\right)^k \left(1 - \frac{1}{s}\right)^{(n-k)} \binom{n}{k} + (s)(s-1)\binom{n}{k,k,n-2k} \left(\frac{1}{s}\right)^{2k} \left(1 - \frac{1}{s}\right)^{n-2k}$$

$$- \left(s \left(\frac{1}{s}\right)^k \left(1 - \frac{1}{s}\right)^{(n-k)} \binom{n}{k}\right)^2$$