Probability and Statistics Homework #5

Ari Feiglin

Question 5.1:

Z is a random variable with the following probability function is:

$$P_Z(x) = \begin{cases} 0.5, & Z = 0 \\ 0.2 & Z = 0.1 \\ 0.1 & Z = 2 \\ 0.2 & Z = 3 \\ 0 & \text{else} \end{cases}$$

(1) Compute $P(Z < 2)$

(2) Compute $P(1 \leq Z < 3)$

(3) Compute $P(Z = 2 \mid Z < 3)$

Answer:

(1) We know that:

$$P(Z < 2) = \sum_{z<2} P(Z = x) = P(Z = 0) + P(Z = 1) = \boxed{0.7}$$

(2) Similarly:

$$P(1 \leq Z < 3) = \sum_{1 \leq z < 3} P(Z = z) = P(Z = 1) + P(Z = 2) = \boxed{0.3}$$

(3) We know that:

$$P(Z = 2 \mid Z < 3) = \frac{P(Z = 2, Z < 3)}{P(Z < 3)} = \frac{P(Z = 2)}{P(Z < 3)}$$

And we know:

$$P(Z < 3) = \sum_{z<3} P(Z = z) = P(Z = 0) + P(Z = 1) + P(Z = 2) = 0.8$$

So:

$$P(Z = 2 \mid Z < 3) = \frac{0.1}{0.8} = \boxed{\frac{1}{8}}$$

■

Question 5.2:

Two fair dice are rolled. Let X be the maximum result and Y be the minimum result. Find their joint probability distribution and compute

$$P\left(1 < X < 4\,,\, 2 \leq Y \leq 3\right)$$

Answer:

We know that if $y > x$ then:

$$P\left(X = x\,,\, Y = y\right) = 0$$

As $Y \leq X$.
If $x = y$ then $P\left(X = x\,,\, Y = y\right)$ is the probability that both dice rolled the same result, x. There is a $\frac{1}{36}$ probability of this happening as there are 36 possible results.
If $x > y$, then $P\left(X = x\,,\, Y = y\right)$ is the probability that one die rolled x and the other rolled y. There are 2 ways for this to happen (first die rolled an x or a y), and 36 pssible results, so:

$$P\left(X = x\,,\, Y = y\right) = \frac{2}{36} = \frac{1}{18}$$

So all in all:

$$\boxed{P\left(X = x\,,\, Y = y\right) = \begin{cases} 0 & y > x \\ \frac{1}{36} & y = x \\ \frac{1}{18} & y < x \end{cases}}$$

And to compute:

$$P\left(1 < X < 4\,,\, 2 \leq Y \leq 3\right) = \sum_{\substack{1<x<4 \\ 2\leq y\leq 3}} P\left(X = x\,,\, Y = y\right) =$$

$$= P\left(X = 2\,,\, Y = 2\right) + P\left(X = 2\,,\, Y = 3\right) + P\left(X = 3\,,\, Y = 2\right) + P\left(X = 3\,,\, Y = 3\right) =$$

$$= \frac{1}{36} + 0 + \frac{1}{18} + \frac{1}{36} = \boxed{\frac{1}{9}}$$

■

Question 5.3:

In a cup there are 4 black balls, 3 white balls, and 2 red balls. We remove 4 balls from the cup. Let X be the number of black balls removed and Y be the number of red balls removed.

(1) Find the probability distribution of X.

(2) Find the probability distribution of Y.

(3) Are X and Y dependent?

Answer:

Before I answer the question, I will first generalize.
Suppose there are n colors, and a_n balls of the nth color in the cup. We will remove k balls from the cup. Let X_i be the number of balls of the ith color removed.
Let $a \coloneqq \sum_{i=1}^{n} a_i$ be the total number of balls in the cup.
Notice that $P(X_i = x) = 0$ if $x > a_i$ as we can't choose more balls than the amount of balls in the cup. And by choosing x balls of color i, we must choose $k - x$ balls of the $a - a_i$ balls not colored i, so $a - a_i$ must be greater or equal to $k - x$, so if $a - a_i < k - x \iff x < a_i + k - a$, $P(X_i = x) = 0$.

Statement 5.3.1:

The probability of picking x balls of a color out of k balls is independent of the order they are chosen. (Assuming we *can* pick k balls of the color.)

Proof:

Let $\{m_j\}_{j=1}^{x}$ be the ordered series of points when we pick a ball of color i, so the first time we pick a ball of color i is m_1, the second time is m_2, and so on. Furthermore, $m_j \le x$ as we only choose x balls. For simplicity, let $m_0 \coloneqq 0$, and $m_{x+1} = k + 1$.
So we need to prove that the probability of choosing x balls of color i out of k balls chosen, with the order specified by $\{m_j\}$ is independent of $\{m_j\}$.
Notice that at the tth step:
Case 1: If $t \in \{m_j\}$

Suppose $t = m_j$. Then $j - 1$ balls of color i have already been removed, and $t - 1$ balls of any color have already been removed. We want to compute the probability of removing a ball of color i.
There are $a_i - j + 1$ balls of color i remaining, and $a - t + 1$ balls all in all in the cup, so the probability of choosing a ball of color i is:

$$\frac{a_i - j + 1}{a - t + 1} = \frac{a_i - j + 1}{a - m_j + 1}$$

Case 2: If $t \notin \{m_j\}$

Let m_j be the maximum such that $m_j < t$. There are $a_i - j$ balls of color i remaining (as j of them have been removed), and $t - 1$ balls in total have been removed. We want to compute the probability of *not* removing a ball of color i.
There are $a - t + 1$ balls remaing in the cup, of which $a_i - j$ of them are of color i. So there are $a - a_i + j - t + 1$ balls that aren't of color i in the cup. So the probability is:

$$\frac{a - a_i + j - t + 1}{a - t + 1}$$

Let the probability of removing the correct ball at step t be denoted by p_t. So the probability we want to find (of removing x balls colored i out of k balls) is:

$$\prod_{t=1}^{k} p_t = \prod_{j=1}^{x} p_{m_j} \cdot \prod_{j=0}^{k} \prod_{t=m_j+1}^{m_{j+1}-1} p_t$$

The idea is that the first product is the product of p_t for all the steps where a ball of color i is removed. The second product is the product of all p_ts where one isn't, partitioned into the steps between the m_js.

We know that this is equal to:

$$\prod_{j=1}^{x} \frac{a_i - j + 1}{a - m_j + 1} \cdot \prod_{j=0}^{x} \prod_{t=m_j+1}^{m_{j+1}-1} \frac{a - a_i + j - t + 1}{a - t + 1}$$

We can reorder this to get a fraction whose denominator is:

$$\prod_{j=0}^{x-1} \left(\prod_{t=m_j+1}^{m_{j+1}-1} (a - t + 1) \cdot (a - m_{j+1} + 1) \right) \cdot \prod_{t=m_x+1}^{k} (a - t + 1)$$

And whose numerator is:

$$\prod_{j=1}^{x} (a_i - j + 1) \cdot \prod_{j=0}^{x} \prod_{t=m_j+1}^{m_{j+1}-1} (a - a_i + j - t + 1)$$

The denominator is equal to:

$$\prod_{j=0}^{x-1} \prod_{t=m_j+1}^{m_{j+1}} (a - t + 1) \cdot \prod_{t=m_x+1}^{k} (a - t + 1) = \prod_{t=m_0+1}^{m_x} (a - t + 1) \cdot \prod_{t=m_x+1}^{k} (a - t + 1) =$$

$$\prod_{t=1}^{k} (a - t + 1) = a \cdot (a-1) \cdots (a - k + 1) = \frac{a!}{(a-k)!}$$

Now, for the numerator, let's focus on:

$$\prod_{j=0}^{x} \prod_{t=m_j+1}^{m_{j+1}-1} (a - a_i + j - t + 1)$$

We know that:

$$\prod_{t=m_j+1}^{m_{j+1}-1} (a - a_i + j - t + 1) = \prod_{t=1}^{m_{j+1}-m_j-1} (a - a_i + j - m_j + 1 - t) =$$

$$= (a - a_i + j - m_j) \cdots (a - a_i + j - m_{j+1} + 2) = \frac{(a - a_i + j - m_j)!}{(a - a_i + j + 1 - m_{j+1})!}$$

So the product is equal to:

$$\prod_{j=0}^{x} \frac{(a - a_i + j - m_j)!}{(a - a_i + j + 1 - m_{j+1})!} = \prod_{j=0}^{x} (a - a_i + j - m_j)! \cdot \prod_{j=0}^{x} \frac{1}{(a - a_i + j + 1 - m_{j+1})!} =$$

$$= \prod_{j=0}^{x} (a - a_i + j - m_j)! \cdot \prod_{j=1}^{x+1} \frac{1}{(a - a_i + j - m_j)!} = (a - a_i)! \cdot \frac{1}{(a - a_i + x - k)!} = \frac{(a - a_i)!}{(a - a_i + x - k)!}$$

And so all in all, the numerator is equal to:

$$\frac{a_i!}{(a_i - x)!} \cdot \frac{(a - a_i)!}{(a - a_i + x - k)!}$$

And the probability is:

$$\frac{a_i! \cdot (a - a_i)! \cdot (a - k)!}{a! \cdot (a_i - x)! \cdot (a - a_i + x - k)!}$$

■

For every x, there are $\binom{k}{x}$ series of $\{m_j\}$, as we just choose x numbers from $[k]$ and order them. Since the probability of choosing x i-colored balls with any $\{m_j\}$ is equal to the expression above, that means that the probability of choosing x i-colored bals in any order is:

$$\binom{k}{x} \cdot \frac{a_i! \cdot (a-a_i)! \cdot (a-k)!}{a! \cdot (a_i - x)! \cdot (a - a_i + x - k)!}$$

Which means that the probability density of X_i is:

$$P_{X_i}(x) = \begin{cases} \binom{k}{x} \cdot \frac{a_i! \cdot (a-a_i)! \cdot (a-k)!}{a! \cdot (a_i - x)! \cdot (a - a_i + x - k)!} & x \leq a_i, x \geq a_i + k - a, x \in \mathbb{N}_0 \\ 0 & \text{else} \end{cases}$$

(1) In this case $a_i = 4$, $a = 9$, and $k = 4$, so:

$$P_X(x) = \binom{4}{x} \cdot \frac{4! \cdot 5! \cdot 5!}{9! \cdot (4-x)! \cdot (1+x)!} = \binom{4}{x} \cdot \frac{20}{21 \cdot (4-x)! \cdot (x+1)!}$$

So:

$$\boxed{P_X(x) = \begin{cases} \binom{4}{x} \cdot \frac{20}{21 \cdot (4-x)! \cdot (x+1)!} & x \leq 4 \\ 0 & \text{else} \end{cases}}$$

(2) In this case, $a_i = 2$, $a = 9$, and $k = 4$, so:

$$P_Y(y) = \binom{4}{y} \cdot \frac{2! \cdot 7! \cdot 5!}{9! \cdot (2-y)! \cdot (3+y)!} = \binom{4}{y} \cdot \frac{10}{3 \cdot (2-y)! \cdot (y+3)!}$$

So:

$$\boxed{P_Y(y) = \begin{cases} \binom{4}{y} \cdot \frac{10}{3 \cdot (2-y)! \cdot (y+3)!} & y \leq 2 \\ 0 & \text{else} \end{cases}}$$

(3) It is obvious that X and Y are dependent. This is because we know:

$$P(X = 4\,, Y = 2) = 0$$

As we can't choose 4 black balls and 2 red balls if we're only choosing 4 balls in total.
But

$$P(X = 4)\,, P(Y = 2) \neq 0$$

So:

$$P(X = 4\,, Y = 2) \neq P(X = 4) \cdot P(Y = 2)$$

So X and Y are dependent.

■

Question 5.4:

X is a discrete random variable. Find k given the probability distribution of X:

(1) $P_X(i) \coloneqq \frac{k-i}{3k}$ for $0 \le i \le 4 \in \mathbb{N}_0$.

(2) $P_X(i) \coloneqq k \cdot p \cdot (1-p)^{i+1}$ for $i \in \mathbb{N}_0$ and $p < 1$.

Answer:

For each subquestion, we need to find the k which satisfies:

$$\sum_{x \in \mathbb{R}} P(X = x) = 1$$

Since P is a probability function.

(1) We know:

$$\sum_{x \in \mathbb{R}} P(X = x) = \sum_{x \in \mathbb{R}} \frac{k-x}{3k} = \sum_{i=0}^{4} \frac{k-i}{3k} = \frac{5}{3} - \frac{1}{3k} \cdot \sum_{i=0}^{4} i = \frac{5}{3} - \frac{1}{3k} \cdot \frac{5}{2} \cdot 4 = \frac{5}{3} - \frac{10}{3k}$$

So:

$$\frac{5}{3} - \frac{10}{3k} = 1 \implies \frac{2}{3} = \frac{10}{3k} \implies \boxed{k = 5}$$

(2) We know:

$$\sum_{x \in \mathbb{R}} P(X = x) = \sum_{i=0}^{\infty} kp \cdot (1-p)^{i+1} = kp \cdot \sum_{i=1}^{\infty} (1-p)^i = kp \cdot \frac{(1-p)}{p} = k(1-p)$$

This is equal to 1, so:

$$k = \boxed{\frac{1}{1-p}}$$

∎

Question 5.5:

We choose randomly two numbers in $[n]$. Let X be the first number, and Y be the second number. Let $M \coloneqq \max\{X,Y\}$.

(1) Find the joint probability distribution of X,Y.

(2) Find the probability distribution of M.

Answer:

(1) We know that:

$$P_{X,Y}((x,y)) = P(X = x\,,\, Y = y)$$

And since the numbers are chosen randomly (with returns), so X and Y are independent. Meaning that:

$$P(X = x\,,\, Y = y) = P(X = x) \cdot P(Y = y)$$

Since the numbers are chosen uniformly over $[n]$, this is equal to:

$$\frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}$$

This is of course assuming that $1 \leq x, y \leq n$, so:

$$\boxed{P_{X,Y}((x,y)) = \begin{cases} \frac{1}{n^2} & x,y \in [n] \\ 0 & \text{else} \end{cases}}$$

(2) We know that $M = m$ if and only if $X = m$ and $Y \leq m$, or $X \leq m$ and $Y = m$, which is equivalent to $(X = m\,,\, Y < m) \vee (X < m\,,\, Y = m) \vee (X = Y = m)$. These are all disjoint so:

$$P_M(m) = P(X = m\,,\, Y < m) + P(X < m\,,\, Y = m) + P(X = Y = m)$$

Since X and Y are independent:

$$P(X = m\,,\, Y < m) = P(X = m) \cdot P(Y < m)$$

And we know that:

$$P(Y < m) = \frac{m-1}{n}$$

Since there are $m-1$ possible options for Y and the probability is uniform, so:

$$P(X = m\,,\, Y < m) = \frac{m-1}{n^2}$$

Similar for $X < m\,,\, Y = m$. So:

$$P_M(m) = \frac{2m-2}{n^2} + \frac{1}{n^2} = \frac{2m-1}{n^2}$$

This is assuming that $m \in [n]$, otherwise the probability is 0 as X and Y and thus M must be in $[n]$. So:

$$\boxed{P_M(m) = \begin{cases} \frac{2m-1}{n^2} & m \in [n] \\ 0 & \text{else} \end{cases}}$$

$\blacksquare$

Question 5.6:

X and Y are random variables. Prove that the following are equivalent:

(a) $X \perp\!\!\!\perp Y$

(b) The distribution of $X \mid Y = a$ is the same distribution for every a.

Answer:

(a) $\implies$ ***(b)*** We know that:

$$P(X = x \mid Y = a) = \frac{P(X = x\,,\, Y = a)}{P(Y = a)}$$

Since $X \perp\!\!\!\perp B$, we know this is equal to:

$$\frac{P(X = x) \cdot P(Y = a)}{P(Y = a)} = P(X = x)$$

So that means:

$$P_{X \mid Y = a} = P_X$$

So the distribution is independent of a, as required.

(b) $\implies$ ***(a)*** Let $x \in \mathbb{R}$ and $a \in \mathbb{R}$. We know that for every $y \in \mathbb{R}$:

$$P(X = x \mid Y = y) = P(X = x \mid Y = a)$$

And we know:

$$P(X = x) = \sum_{y \in \mathbb{R}} P(X = x \mid Y = y) \cdot P(Y = y) =$$

$$= \sum_{y \in \mathbb{R}} P(X = x \mid Y = a) \cdot P(Y = y) = P(X = x \mid Y = a) \cdot \sum_{y \in \mathbb{R}} P(Y = y)$$

And we know:

$$\sum_{y \in \mathbb{R}} P(Y = y) = 1$$

So:

$$P(X = x) = P(X = x \mid Y = a)$$

For every a. Now, notice that:

$$P(X = x, Y = y) = P(X = X \mid Y = y) \cdot P(Y = y) = P(X = x) \cdot P(Y = y)$$

So X and Y are independent, as required.

$\blacksquare$

Question 5.7:

Prove that the sum of two fair dice rolls modulo 6 is uniform in $\{0, 1, 2, 3, 4, 5\}$

Answer:

I will prove that given $X_1, \ldots, X_n \in \mathrm{Unif}(\{1, \ldots, \mathrm{k}\})$, if we define $Y \coloneqq X_1 + \cdots + X_n$ and $M \coloneqq Y \bmod k$, then

$$M \sim \mathrm{Unif}(\{0, \ldots, \mathrm{k} - 1\})$$

Through induction on n.

Base case: $n = 1$

In this case, $Y = X_1 \implies M = X_1 \bmod k$. This means that $M = m \iff X_1 \bmod k = m$. This means that $X_1 = m$ if $m > 0$, and $X_1 = k$ if $m = 0$. So:

$$P(M = m) = \begin{cases} P(X_1 = m) & m > 0 \\ P(X_1 = k) & m = 0 \end{cases} = \begin{cases} \frac{1}{k} & m > 0 \\ \frac{1}{k} & m = 0 \end{cases} = \frac{1}{k}$$

Which means that M distributes uniformly over $\{0, \ldots, k - 1\}$, as required.

Base case: $n = 2$

In this case, notice that $M = m$ if and only if $Y \bmod k = m$, which is if and only if there exists a q such that $Y = qk + m$. So:

$$P(M = m) = \mathbb{P}_Y(\{qk + m \mid q \in \mathbb{N}_0\})$$

And we know that $\{qk + m \mid q \in \mathbb{N}_0\}$ are disjoint as $qk + m = q'k + m \iff qk = q'k \iff q = q'$. So:

$$P(M = m) = \sum_{q \in \mathbb{N}_0} P(Y = qk + m)$$

And we know that $Y = X_1 + X_2 \leq 2k$, so:

$$P(M = m) = P(Y = m) + P(Y = k + m)$$

If $m > 0$, and:

$$P(M = 0) = P(Y = 0) + P(Y = k) + P(Y = 2k) = P(Y = k) + P(Y = 2k)$$

And in general:

$$P(Y = y) = \sum_{x \in \mathbb{N}_0} P(Y = y \mid X_1 = x) \cdot P(X_1 = x)$$

And we know that

$$P(Y = y \mid X_1 = x) = P(X_1 + X_2 = y \mid X_1 = x) = P(X_2 = y - x \mid X_1 = x)$$

And since X_1 and X_2 are independent, this is equal to:

$$P(X_2 = y - x)$$

So we know that $1 \leq x \leq k$ and $1 \leq y - x \leq k$, so $1, y - k \leq x \leq k, y - 1$, therefore:

$$P(Y = y) = \sum_{\max\{1, y-k\}}^{\min\{k, y-1\}} \frac{1}{k^2} = \frac{\min\{k, y - 1\} - \max\{1, y - k\} + 1}{k^2}$$

So for $y = m < k$, $\min\{k, y - 1\} = y - 1 = m - 1$ and $\max\{1, y - k\} = 1$, so:

$$P(Y = m) = \frac{m - 1}{k^2}$$

And for $y = m + k > k$, $\min\{k, y - 1\} = k$ and $\max\{1, y - k\} = y - k = m$, so:

$$P(Y = m + k) = \frac{k - m + 1}{k^2}$$

So:

$$P\left(M=m\right)=\frac{k}{k^2}=\frac{1}{k}$$

And:

$$P\left(M=0\right)=\frac{k-1}{k^2}+\frac{1}{k^2}=\frac{k}{k^2}=\frac{1}{k}$$

So for every $m\in\{0,\ldots,k-1\}$:

$$P\left(M=m\right)=\frac{1}{k}$$

Which means that $M\sim \text{Unif}\left(\{0,\ldots,\text{k}-1\}\right)$, as required.
(This is actually sufficient to answer the original question.)

Inductive step:

Suppose this is true for n, we will prove this for $n+1$. We know that $X_1,\ldots,X_n,X_{n+1}\in \text{Unif}\left(\{1,\ldots,\text{k}\}\right)$, then:

$$M=\left(X_1+\cdots+X_n+X_{n-1}\right) \bmod k=\left(\left(X_1+\cdots+X_n\right) \bmod k+X_{n+1} \bmod k\right) \bmod k$$

By our inductive assumption:

$$\left(X_1+\cdots+X_n\right) \bmod k\in \text{Unif}\left(\{0,\ldots,\text{k}-1\}\right)$$

As this is just the M of n.
And we also know from our first base case that $X_{n+1} \bmod k\in \text{Unif}\left(\{0,\ldots,\text{k}-1\}\right)$ as well.
And we know by our second base case that given two random variables which are uniform over the set, then their union modulo k is uniform over the set modulo k. Which means that:

$$M\in \text{Unif}\left(\{0,\ldots,\text{k}-1\}\right)$$

As required.

Now, since the dice rolls are $X_1\in \text{Unif}\left(\{1,\ldots,6\}\right)$ an $X_2\in \text{Unif}\left(\{1,\ldots,6\}\right)$ that means that (since $k=6$) $\left(X_1+X_2\right) \bmod 6\in \text{Unif}\left(\{0,\ldots,5\}\right)$ as required.

$\blacksquare$

Question 5.8:

X and Y are two random variables whose support is in $\mathbb{Z}$ and are independent. Let $Z \coloneqq X + Y$. Prove that for every $n \in \mathbb{Z}$:

$$P(Z = n) = \sum_{i \in \mathbb{Z}} P(X = i) \cdot P(Y = n - i)$$

Answer:

We know that:

$$P(Z = n) = \sum_{x \in \mathbb{R}} P(Z = n \mid X = x) \cdot P(X = x)$$

And since X's support is in $\mathbb{Z}$, we can sum over only $\mathbb{Z}$:

$$P(Z = n) = \sum_{i \in \mathbb{Z}} P(Z = n \mid X = i) \cdot P(X = i)$$

We know that:

$$P(Z = n \mid X = i) = P(X + Y = n \mid X = i) = P(Y = n - i \mid X = i)$$

And since X and Y are independent, this is equal to $P(Y = n - i)$. So:

$$P(Z = n) = \sum_{i \in \mathbb{Z}} P(X = i) \cdot P(Y = n - i)$$

As required.

$\blacksquare$