Probability Theory 1 – Proposed solution of moed aleph exam 2020b

1. **(a)** It readily follows from the definitions of X and Y that the support of both random variables is $\{0,1,2\}$. For every $0 \le i \le 2$ and $0 \le j \le 2$, the table below shows the value of $P(X=i, Y=j)$. We explain three of these calculations in greater detail:

   If $X=1$ and $Y=0$, then either the result of the first die roll was 1 and the result of the second die roll was 2,3,4 or 5, or the result of the second die roll was 1 and the result of the first die roll was 2,3,4 or 5. Since the die rolls are independent, we conclude that $\mathbb{P}(X=1, Y=0) = 2 \cdot 1/6 \cdot 4/6$.

   If $X=2$, then the result of both die rolls must be 1 and so $Y=0$. In particular, $\mathbb{P}(X=2, Y=1) = \mathbb{P}(X=2, Y=2) = 0$.

	$Y=0$	$Y=1$	$Y=2$
$X=0$	$(2/3)^2$	$2 \cdot 1/6 \cdot 2/3$	$(1/6)^2$
$X=1$	$2 \cdot 1/6 \cdot 2/3$	$2 \cdot (1/6)^2$	0
$X=2$	$(1/6)^2$	0	0

   (b) Using the table we calculated in part (a) we obtain

$$\begin{aligned}\mathbb{P}(X=0) &= \mathbb{P}(X=0, Y=0) + \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=0, Y=2) \\ &= (2/3)^2 + 2 \cdot 1/6 \cdot 2/3 + (1/6)^2 = \frac{25}{36},\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X=1) &= \mathbb{P}(X=1, Y=0) + \mathbb{P}(X=1, Y=1) + \mathbb{P}(X=1, Y=2) \\ &= 2 \cdot 1/6 \cdot 2/3 + 2 \cdot (1/6)^2 + 0 = \frac{10}{36},\end{aligned}$$

   and

$$\mathbb{P}(X=2) = \mathbb{P}(X=2, Y=0) + \mathbb{P}(X=2, Y=1) + \mathbb{P}(X=2, Y=2) = (1/6)^2 + 0 + 0 = \frac{1}{36}.$$

   Hence

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) + 2 \cdot \mathbb{P}(X=2) = \frac{10}{36} + \frac{2}{36} = \frac{1}{3}.$$

   Similarly

$$\mathbb{E}(X^2) = 0^2 \cdot \mathbb{P}(X=0) + 1^2 \cdot \mathbb{P}(X=1) + 2^2 \cdot \mathbb{P}(X=2) = \frac{10}{36} + \frac{4}{36} = \frac{7}{18}.$$

   We conclude that $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{7}{18} - \frac{1}{9} = \frac{5}{18}$.

   Note: One could also simply observe that $X \sim \text{Bin}(2, 1/6)$ and thus $\mathbb{E}(X) = 2 \cdot 1/6 = 1/3$ and $Var(X) = 2 \cdot 1/6 \cdot (1 - 1/6) = 5/18$.

(c) It is apparent from their definition or from the table we calculated in part (a) that X and Y have the same distribution. It thus follows by part (b) that $\mathbb{E}(Y) = \mathbb{E}(X) = 1/3$ and $Var(Y) = Var(X) = 5/18$. Using the same table, it is easy to calculate $\mathbb{E}(XY)$.

$$\mathbb{E}(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy\mathbb{P}(X = x, Y = y) = 1 \cdot 1 \cdot \mathbb{P}(X = 1, Y = 1) = 1/18.$$

Therefore

$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{18} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{18}.$$

We conclude that

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} = \frac{-1/18}{5/18} = -1/5.$$

(d) Using the various rules proved in the lectures regarding the effect of additive and multiplicative constants on the variance and covariance we obtain

$$\begin{aligned} Var(2X - Y + 1) &= Var(2X + (-Y)) = Var(2X) + Var(-Y) + 2Cov(2X, -Y) \\ &= 4Var(X) + Var(Y) - 4Cov(X, Y) = 4 \cdot 5/18 + 5/18 - 4 \cdot (-1/18) \\ &= 29/18. \end{aligned}$$

2. **(a)** Since the two sampled balls are returned to the urn, in each round of the experiment, the probability of "success", that is, the probability of sampling balls of two different colours, is $\frac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}} = \frac{3}{5}$. Since, moreover, the sampling in any given round is independent of all previous rounds and we continue sampling until the first success, we conclude that $X \sim \text{Geom}(3/5)$.

(b) Let Z be the random variable that counts the number of rounds in which two white balls were sampled. Since there is always a unique round in which we sample 1 white ball (namely, the last round) and in all other rounds we sample either 0 or 2 white balls, it follows that $Y = 2Z + 1$. Now, observe that $(Z|X = 3) \sim \text{Bin}(2, 1/2)$. Indeed, since $X = 3$, the experiment lasts 3 rounds, in the third round we sample 1 black ball and 1 white ball, and in the first two rounds we sample either 2 white balls or 2 black balls. Viewing the sampling of 2 white balls in a round as a success in that round, we see that there are two independent rounds and in each the probability of success is 1/2 (by symmetry, the probability of sampling 2 white balls is equal to the probability of sampling 2 black balls). Hence

$$\mathbb{E}(Z|X = 3) = 2 \cdot 1/2 = 1$$

and by the linearity of expectation it then follows that

$$\mathbb{E}(Y|X = 3) = \mathbb{E}(2Z + 1|X = 3) = 2\mathbb{E}(Z|X = 3) + 1 = 3.$$

(c) Following the same reasoning as in part (b) of this question, we see that $(Z|X = k) \sim \text{Bin}(k-1, 1/2)$ for every positive integer k. It then follows by the law of total probability that

$$\begin{aligned}\mathbb{P}(Y = 1) = \mathbb{P}(Z = 0) &= \sum_{k=1}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Z = 0|X = k) = \sum_{k=1}^{\infty} \left(\frac{2}{5}\right)^{k-1} \cdot \frac{3}{5} \cdot \binom{k-1}{0} \left(\frac{1}{2}\right)^{k-1} \\ &= \frac{3}{5} \sum_{k=1}^{\infty} \left(\frac{1}{5}\right)^{k-1} = \frac{3}{5} \cdot \frac{1}{1 - 1/5} = \frac{3}{4},\end{aligned}$$

where the third equality holds since $X \sim \text{Geom}(3/5)$ and $(Z|X = k) \sim \text{Bin}(k - 1, 1/2)$ and the penultimate equality holds by the formula for the infinite sum of a geometric series.

Note: One can also calculate $\mathbb{P}(Y = 1)$ without using conditional probability. Indeed

$$\mathbb{P}(Y = 1) = \sum_{k=1}^{\infty} \mathbb{P}(Y = 1, X = k) = \sum_{k=1}^{\infty} \left(\frac{1}{5}\right)^{k-1} \cdot \frac{3}{5} = \frac{3}{5} \cdot \frac{1}{1 - 1/5} = \frac{3}{4},$$

where the second equality holds since the event $\{Y = 1, X = k\}$ amounts to having $k - 1$ consecutive rounds in each of which we sampled 2 black balls, followed by one round in which we sampled 1 black ball and 1 white ball.

(d) X and Y are dependent. In order to see this, it suffices to find values a and b such that $\mathbb{P}(Y = a) \neq \mathbb{P}(Y = a|X = b)$ (and, in particular, $\mathbb{P}(X = b) > 0$). We have shown in part (c) of this question that $\mathbb{P}(Y = 1) = 3/4$. On the other hand, it is immediate from the definition of X and Y that $\mathbb{P}(Y = 1|X = 1) = 1$ (and $\mathbb{P}(X = 1) = 3/5 > 0$).

3. **(a)** The event $\{X_1 = 0\}$ indicates that the result of the first die roll is 2 or 3. The event $\{Y_1 = 0|X_1 = 0\}$ indicates that the result of the second die roll was in $\{1, 4, 5, 6\}$ or that the result of the third die roll was in $\{1, 4, 5, 6\}$. Therefore

$$\begin{aligned}\mathbb{P}(X_1 = 0, Y_1 = 0) = \mathbb{P}(X_1 = 0)\mathbb{P}(Y_1 = 0|X_1 = 0) &= 1/3 \cdot [1 - \mathbb{P}(Y_1 = 1|X_1 = 0)] \\ &= 1/3 \cdot \left[1 - (1/3)^2\right] = 8/27.\end{aligned}$$

(b) For every $1 \leq i \leq n - 2$, the random variable Y_i is an indicator, implying that

$$\begin{aligned}\mathbb{E}(Y_i) &= \mathbb{P}(Y_i = 1) \\ &= \mathbb{P}(X_i = X_{i+1} = X_{i+2} = -1) + \mathbb{P}(X_i = X_{i+1} = X_{i+2} = 0) + \mathbb{P}(X_i = X_{i+1} = X_{i+2} = 1) \\ &= (1/6)^3 + (1/3)^3 + (1/2)^3 = \frac{1 + 8 + 27}{6^3} = 1/6.\end{aligned}$$

It then follows by the linearity of expectation that

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^{n-2} Y_i\right) = \sum_{i=1}^{n-2} \mathbb{E}(Y_i) = (n - 2)/6.$$

(c) We aim to use the formula

$$Var(Y) = Var\left(\sum_{i=1}^{n-2} Y_i\right) = \sum_{i=1}^{n-2} Var(Y_i) + 2 \sum_{1 \le i < j \le n-2} Cov(Y_i, Y_j).$$

For every $1 \le i \le n-2$, since Y_i is an indicator, it follows that

$$Var(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2 = \mathbb{E}(Y_i) - (\mathbb{E}(Y_i))^2 = 1/6 - (1/6)^2 < 1/6.$$

Since the die rolls are independent, it follows that whenever $j > i + 2$, the random variables Y_i and Y_j are independent and thus $Cov(Y_i, Y_j) = 0$. For every $1 \le i \le n-3$ we have

$$\begin{aligned}Cov(Y_i, Y_{i+1}) &= \mathbb{E}(Y_i Y_{i+1}) - \mathbb{E}(Y_i)\mathbb{E}(Y_{i+1}) = \mathbb{P}(Y_i = 1, Y_{i+1} = 1) - 1/36 \\ &< \mathbb{P}(X_i = X_{i+1} = X_{i+2} = X_{i+3}) = (1/6)^4 + (1/3)^4 + (1/2)^4 < 3 \cdot (1/2)^4 < 1/5.\end{aligned}$$

Similarly, for every $1 \le i \le n-4$ we have

$$\begin{aligned}Cov(Y_i, Y_{i+2}) &= \mathbb{E}(Y_i Y_{i+2}) - \mathbb{E}(Y_i)\mathbb{E}(Y_{i+2}) = \mathbb{P}(Y_i = 1, Y_{i+2} = 1) - 1/36 \\ &< \mathbb{P}(X_i = X_{i+1} = X_{i+2} = X_{i+3} = X_{i+4}) = (1/6)^5 + (1/3)^5 + (1/2)^5 \\ &< 3 \cdot (1/2)^5 < 1/5.\end{aligned}$$

We conclude that

$$\begin{aligned}Var(Y) &= \sum_{i=1}^{n-2} Var(Y_i) + 2 \sum_{1 \le i < j \le n-2} Cov(Y_i, Y_j) \\ &= \sum_{i=1}^{n-2} Var(Y_i) + 2 \sum_{i=1}^{n-3} Cov(Y_i, Y_{i+1}) + 2 \sum_{i=1}^{n-4} Cov(Y_i, Y_{i+2}) \\ &< (n-2)/6 + 2(n-3)/5 + 2(n-4)/5 < n.\end{aligned}$$

(d) Using Chebyshev's inequality we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(Y \ge n/5) &= \lim_{n\to\infty} \mathbb{P}(Y - \mathbb{E}(Y) \ge n/5 - (n-2)/6) \le \lim_{n\to\infty} \mathbb{P}(|Y - \mathbb{E}(Y)| \ge n/100) \\ &\le \lim_{n\to\infty} \frac{Var(Y)}{(n/100)^2} \le \lim_{n\to\infty} \frac{100^2}{n} = 0,\end{aligned}$$

where in the last inequality we used the inequality $Var(Y) \le n$ proved in part (c) of this question.