Probability Theory
2024/25 Period IIb
Instructor: Gilles Bonnet
Exam
20/6/2024
Duration: 2 hours

Solutions

1. A traditional fair die is rolled 4 times. Give the probability of the following events.
   Remark 1: The justification of your answers can be limited to a brief computation (no need to write many words).
   Remark 2: Give your answer in the form of a fraction of integer numbers.
   (a) (5 points) A six turns up exactly twice.
   (b) (5 points) All numbers are odd.
   (c) (5 points) The sum of the four rolls is 6.

Solution: (a)

$$\begin{aligned}\mathbb{P}(\text{exactly two sixes}) &= \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 \\ &= 6\cdot\frac{1}{36}\cdot\frac{25}{36} \\ &= \frac{1}{6}\cdot\frac{25}{36} \\ &= \boxed{\frac{25}{216}}\end{aligned}$$

(without the simplification $6\times(1/36)=1/6$, this is $150/1296$)

(b)

$$\mathbb{P}(\text{all odd}) = \left(\frac{3}{6}\right)^4 = \left(\frac{1}{2}\right)^4 = \boxed{\frac{1}{16}}$$

(c) To get a sum of 6 we either have 3 ones and 1 three, or 2 twos and 2 ones. Thus we have a total of $\binom{4}{1}+\binom{4}{2}=4+6=10$ ways to get a sum of 6. Therefore

$$\begin{aligned}\mathbb{P}(\text{sum is 6}) &= \frac{\text{number of ways to get sum 6}}{\text{total outcomes}} \\ &= \frac{10}{6^4} = \frac{10}{1296} = \boxed{\frac{5}{648}}\end{aligned}$$

2. (15 points) Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with expectation $\lambda > 0$. Show that the sum $S_n = X_1 + X_2 + \ldots + X_n$ has a gamma distribution with parameters n and λ^{-1}.
Remark (to avoid confusion with the parametrization): the gamma distribution with parameters n and λ^{-1} has pdf proportional to $x^{n-1}e^{-x/\lambda}$ for $x > 0$.
Hint: You might want to do an induction on n.

Solution: We prove by induction on n that the sum $S_n = X_1 + \cdots + X_n$ has density

$$f_{S_n}(x) = \mathbf{1}(x > 0)\frac{1}{\lambda^n (n-1)!}x^{n-1}e^{-x/\lambda},$$

which is the Gamma(n, λ^{-1}) distribution.

Base case: For $n = 1$, we have $S_1 = X_1$ and

$$f_{S_1}(x) = \mathbf{1}(x > 0)\frac{1}{\lambda}e^{-x/\lambda},$$

which matches the gamma distribution with parameters $n = 1$, λ^{-1}.

Induction step: Assume the result holds for $n = k$, i.e.

$$f_{S_k}(x) = \mathbf{1}(x > 0)\frac{1}{\lambda^k (k-1)!}x^{k-1}e^{-x/\lambda}.$$

Now consider $S_{k+1} = S_k + X_{k+1}$, where $X_{k+1} \sim \text{Exp}(\lambda)$ independently. Then:

$$\begin{aligned}
f_{S_{k+1}}(x) &= \int_{-\infty}^{\infty} f_{S_k}(y) f_{X_{k+1}}(x-y)\, dy \\
&= \int_0^x \left(\frac{1}{\lambda^k (k-1)!}y^{k-1}e^{-y/\lambda}\right)\left(\frac{1}{\lambda}e^{-(x-y)/\lambda}\right) dy \\
&= \frac{e^{-x/\lambda}}{\lambda^{k+1}(k-1)!}\int_0^x y^{k-1}\, dy \\
&= \frac{e^{-x/\lambda}}{\lambda^{k+1}(k-1)!} \cdot \frac{x^k}{k} \\
&= \frac{1}{\lambda^{k+1}k!}x^k e^{-x/\lambda},
\end{aligned}$$

which is the gamma density with parameters $n = k + 1$, λ^{-1}.

Hence, by induction, the result holds for all $n \geq 1$.

3. Let (X, Y) be a continuous random vector with joint density function given (up to a constant) by:

$$f_{X,Y}(x, y) = c(x + y)e^{-(x+y)}, \quad \text{for } x > 0,\ y > 0,$$

and 0 elsewhere.

(a) (15 points) Compute the marginal densities $f_X(x)$ and $f_Y(y)$ and the constant c.
Hint: You might want to start by computing the marginals up to the constant c and then use this to compute c.

Solution: We start by computing $f_X(x)$:

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,Y}(x,y)dy = \int_0^\infty c(x+y)e^{-(x+y)}dy \\ &= ce^{-x}\int_0^\infty (x+y)e^{-y}dy \\ &= ce^{-x}\left[x\int_0^\infty e^{-y}dy + \int_0^\infty ye^{-y}dy\right] \\ &= ce^{-x}(x+1), \quad x>0. \end{aligned}$$

$$\boxed{f_X(x) = c(x+1)e^{-x}, \quad x>0}$$

Similarly, by symmetry,

$$\boxed{f_Y(y) = c(y+1)e^{-y}, \quad y>0}$$

Integrating the first marginal, we get:

$$\begin{aligned} \int_0^\infty f_X(x)dx &= \int_0^\infty c(x+1)e^{-x}dx \\ &= c\left[\int_0^\infty xe^{-x}dx + \int_0^\infty e^{-x}dx\right] \\ &= c(1+1) = 2c. \end{aligned}$$

So

$$\boxed{c = \frac{1}{2}}$$

(b) (5 points) Are X and Y independent? Justify your answer.

Solution: X and Y are $\boxed{\text{not independent}}$ since there are values of x and y for which

$$f_X(x)f_Y(y) = \frac{1}{4}(x+1)(y+1)e^{-(x+y)} \neq \frac{1}{2}(x+y)e^{-(x+y)} = f_{X,Y}(x,y).$$

For instance

$$f_X(2)f_Y(2) = \frac{1}{4}(2+1)(2+1)e^{-(2+2)} = \frac{9}{4}e^{-4} \neq \frac{1}{2}e^{-4} = \frac{1}{2}(2+2)e^{-(2+2)} = f_{X,Y}(2,2).$$

4. Consider an experiment where a fair coin is tossed $n = 10\,000$ times. Let X denote the number of heads observed.

   (a) (5 points) Give the expectation and variance of X.
   Remark: You do not need to justify your answer.

Solution: Since the coin is fair and each toss is independent, $X \sim \text{Bin}(n, 1/2)$. For a binomial random variable with parameters n and p, we have:

$$\mathbb{E}[X] = np = \frac{n}{2}, \quad \text{Var}(X) = np(1-p).$$

Thus, with $n = 10000$ and $p = \frac{1}{2}$, we get:

$$\boxed{\mathbb{E}[X] = 5000}, \quad \boxed{\text{Var}(X) = 2500}.$$

(b) (10 points) Show that $\mathbb{P}(|X - \mathbb{E}[X]| \geq 500) \leq 0.01$.

Solution: We apply Chebyshev's inequality:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq 500) \leq \frac{\text{Var}(X)}{500^2} = \frac{2500}{250000} = 0.01.$$

5. A deck contains exactly 10 cards, of which 4 are red and 6 are black. The cards are shuffled and drawn one by one, without replacement. Let n be the number of cards drawn, with $n \leq 10$. Define the random variables $X_1, \ldots, X_n$ such that $X_i = 1$ if the i-th card drawn is red, and 0 otherwise. Let $S_n = X_1 + \cdots + X_n$.

(a) (2 points) What is the variance of S_{10}?
Remark: You don't need to justify your answer.

Solution: Since all 10 cards are drawn, and exactly 4 are red, the total number of red cards drawn is always 4. Thus $S_{10} = 4$ with probability 1, and therefore

$$\boxed{\text{Var}(S_{10}) = 0}.$$

(b) (5 points) What are the expectation and variance of X_i?
Remark: You do not need to justify your answer.

Solution: X_i is a Bernoulli random variable with success probability equal to $p = 4/10$, since there are 4 red cards out of 10 total cards. Thus:

$$\boxed{\mathbb{E}[X_i] = p = 0.4} \quad \text{and} \quad \text{Var}(X_i) = p(1-p) = \frac{4}{10}\left(1 - \frac{4}{10}\right) = \frac{4}{10} \cdot \frac{6}{10} = \boxed{0.24}.$$

(c) (5 points) Compute $\mathbb{E}[X_i X_j]$ for $i \neq j$.

Solution: Note that $X_i X_j$ equals 1 if both draws i and j are red, and 0 otherwise. Thus:

$$\mathbb{E}[X_i X_j] = 1 \cdot \mathbb{P}(\text{both } i \text{ and } j \text{ are red}) + 0 \cdot \mathbb{P}(\text{otherwise}) = \mathbb{P}(\text{both } i \text{ and } j \text{ are red}).$$

Since the cards are drawn without replacement:

$$\boxed{\mathbb{E}[X_iX_j] = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}}.$$

(d) (5 points) Compute $\mathrm{Cov}(X_i, X_j)$ for $i \neq j$.

Solution: We have:

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]^2 = \frac{2}{15} - \left(\frac{4}{10}\right)^2 = \frac{2}{15} - \frac{16}{100} = -\frac{2}{75}.$$

$$\boxed{\mathrm{Cov}(X_i, X_j) = -\frac{2}{75}}$$

(e) (8 points) Compute the variance of S_n.
Sanity check: You might want to check compare with your answer of the first question of this exercise.

Solution: We compute:

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \\
&= n \cdot \mathrm{Var}(X_1) + n(n-1) \cdot \mathrm{Cov}(X_1, X_2) \\
&= \boxed{n \cdot 0.24 - \frac{2}{75} n(n-1)}.
\end{aligned}$$

Note that this gives

$$\mathrm{Var}(S_{10}) = 10 \cdot 0.24 + 10 \cdot 9 \cdot \left(-\frac{2}{75}\right) = 2.4 - \frac{180}{75} = 2.4 - 2.4 = 0,$$

which matches our answer to the first question of this exercise.