Selected Problems Chapter 3
Introduction to Probability for Data Science
Stanley Chan

Mustaf Ahmed

November 27, 2022

Problem 3.2. Two dice are tossed. Let X be the absolute difference in the number of dots facing up.

a. Find and plot the PMF of X.
b. Find the probability $X \leq 2$.
c. Find $\mathbb{E}[X]$ and $Var[X]$.

Part a. The possible random variable states are $X(\Omega) = \{0, 1, 2, 3, 4, 5\}$.The probability mass function for this random variable is

$$p_X(0) = P(\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}) = \frac{1}{6}$$

$$p_X(1) = P(\{(1,2),(2,1),(2,3),(3,2),(3,4),(4,3),(4,5),(5,4),(5,6),(6,5)\}) = \frac{5}{18}$$

$$p_X(2) = P(\{(1,3),(3,1),(2,4),(4,2),(3,5),(5,3),(4,6),(6,4)\}) = \frac{2}{9}$$

$$p_X(3) = P(\{(1,4),(4,1),(2,5),(5,2),(3,6),(6,3)\}) = \frac{1}{6}$$

$$p_X(4) = P(\{(1,5),(5,1),(2,6),(6,2)\}) = \frac{1}{9}$$

$$p_X(5) = P(\{(1,6),(6,1)\}) = \frac{1}{18}$$

Part b.

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= \frac{2}{3} \end{aligned}$$

Part c.

$$\begin{aligned} \mathbb{E}[X] &= 0 * \frac{1}{6} + 1 * \frac{5}{8} + 2 * \frac{2}{9} + 3 * \frac{1}{6} + 4 * \frac{1}{9} + 5 * \frac{1}{18} \\ &= \frac{35}{18} \end{aligned}$$

$$\begin{aligned} Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X] \\ &= \mathbb{E}[X^2] - \frac{35}{18} \\ &= 0^2 * \frac{1}{6} + 1^2 * \frac{5}{8} + 2^2 * \frac{2}{9} + 3^2 * \frac{1}{6} + 4^2 * \frac{1}{9} + 5^2 * \frac{1}{18} - \frac{35}{18} \\ &= \frac{665}{324} \end{aligned}$$

Problem Theorem 3.4. Prove that the expectation of a random variable X has the following properties:

(a) Function. For any g,

$$\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p_X(x).$$

(b) Linearity. For any function g and h,

$$\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)].$$

(c)Scale. For any constant c,

$$\mathbb{E}[cX] = c\mathbb{E}[X].$$

(d) DC shift. For any constant c,

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c$$

Part a. Let g be a function.
By the definition of expectation, we have

$$\mathbb{E}[g(X)] = \sum_{s \in g(X(\Omega))} s p_{g(X)}(s).$$

Expanding the the term $p_{g(X)}(s)$, we get

$$\begin{aligned} p_{g(X)}(s) &= P(\{e \in \Omega \mid g(X(e)) = s\}) \\ &= \sum_{x \in g^{-1}(s)} p_X(x). \end{aligned}$$

We can now simplify the original equation:

$$\begin{aligned} \mathbb{E}[g(X)] &= \sum_{s \in g(X(\Omega))} s p_{g(X)}(s) \\ &= \sum_{s \in g(X(\Omega))} s \sum_{x \in g^{-1}(s)} p_X(x) \\ &= \sum_{s \in g(X(\Omega))} \sum_{x \in g^{-1}(s)} s p_X(x) \\ &= \sum_{s \in g(X(\Omega))} \sum_{x \in g^{-1}(s)} g(x) p_X(x) \\ &= \sum_{x \in X(\Omega)} g(x) p_X(x), \end{aligned}$$

where the last relation comes summing over each part of the partition of $X(\Omega)$.

Part b. Let g, h be functions. By 3.4(a), we have

$$\begin{aligned}
\mathbb{E}[g(X)+h(X)] &= \sum_{x\in X(\Omega)} (g(x)+h(x))p_X(x) \\
&= \sum_{x\in X(\Omega)} g(x)p_X(x) + \sum_{x\in X(\Omega)} h(x)p_X(x) \\
&= \mathbb{E}[g(X)] + \mathbb{E}[h(X)].
\end{aligned}$$

Part c. Let g be a function, and let c be a constant. Define $h(X) = cX$. By 3.4(a), we have

$$\begin{aligned}
\mathbb{E}[h(X)] &= \mathbb{E}[cX] \\
&= \sum_{x\in X(\Omega)} h(x)p_X(x) \\
&= \sum_{x\in X(\Omega)} cxp_X(x) \\
&= c\sum_{x\in X(\Omega)} xp_X(x) \\
&= c\mathbb{E}[X].
\end{aligned}$$

Part d. Let g be a function, and let c be a constant. Define $h(X) = X + c$. By 3.4(a), we have

$$\begin{aligned}
\mathbb{E}[h(X)] &= \mathbb{E}[X+c] \\
&= \sum_{x\in X(\Omega)} h(x)p_X(x) \\
&= \sum_{x\in X(\Omega)} (x+c)p_X(x) \\
&= \sum_{x\in X(\Omega)} xp_X(x) + \sum_{x\in X(\Omega)} cp_X(x) \\
&= \sum_{x\in X(\Omega)} xp_X(x) + c\sum_{x\in X(\Omega)} p_X(x) \\
&= \sum_{x\in X(\Omega)} xp_X(x) + c \\
&= \mathbb{E}[X] + c.
\end{aligned}$$

Problem Theorem 3.5. The variance of a random variable X has the following properties:

(i) Moment.

$$Var[X] = E[X^2] - E[X]^2$$

(ii) Scale. For any constant c,

$$Var[cX] = c^2 Var[X]$$

(iii) DC Shift. For any constant c,

$$Var[X + c] = Var[X]$$

Proof. **Part (i).**

$$\begin{aligned} Var[X] &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Part (ii).

$$\begin{aligned} Var[cX] &= E[(cX - c\mu)^2] \\ &= E[c^2(X - \mu)^2] \\ &= c^2 E[(X - \mu)^2] \\ &= c^2 Var[X] \end{aligned}$$

Part (iii).

$$\begin{aligned} Var[X + c] &= E[(X + c - c - E[X])^2] \\ &= E[(X - E[X])^2] \\ &= Var[X] \end{aligned}$$

□

Problem Theorem 3.6. If $X \sim Bernoulli(p)$, then $E[X] = p$, $E[X^2] = p$ and $Var[X] = p(1-p)$.

Proof. **(i).**

$$\begin{aligned} E[X] &= 0 * (1-p) + 1 * p \\ &= p. \end{aligned}$$

(ii).

$$\begin{aligned} E[X^2] &= 0^2 * (1-p) + 1^2 * p \\ &= p. \end{aligned}$$

(iii).

$$\begin{aligned} Var[X] &= E[X^2] - E[X]^2 \\ &= p - p^2 \\ &= p(1-p) \end{aligned}$$

□

Problem 3.3. Let X be a random variable with PMF $p_k = c/2^k$ for $k = 1, 2, \ldots$

a. Determine the value of c.
b. Find $P(X > 4)$ and $P(6 \le X \le 8)$.
c. Find $E[X]$

Part a. We need to find the value c such that the sum of the probabilities of each state is 1.

$$\begin{aligned}1 &= \sum_{k=1}^{\infty} \frac{c}{2^k} \\ &= c\sum_{k=1}^{\infty} \frac{1}{2^k} \\ &= \frac{c}{2}\sum_{i=0}^{\infty} \frac{1}{2^i} \\ &= c.\end{aligned}$$

Hence, $c = 1$.

Part b.

$$\begin{aligned}P(X > 4) &= 1 - P(X <= 4) \\ &= 1 - P(X = 4) - P(X = 3) - P(X = 2) - P(X = 1) \\ &= 1 - \frac{1}{16} - \frac{1}{8} - \frac{1}{4} - \frac{1}{2} \\ &= \frac{1}{16}\end{aligned}$$

$$\begin{aligned}P(6 \le X \le 8) &= P(X = 6) + P(X = 7) + P(X = 8) \\ &= \frac{1}{64} + \frac{1}{128} + \frac{1}{256} \\ &= \frac{7}{256}\end{aligned}$$

Part c.

$$
\begin{aligned}
E[X] &= \sum_{k=1}^{\infty} \frac{k}{2^k} \\
&= \frac{1}{2} \sum_{k=1}^{\infty} \frac{k}{2^{k-1}} \\
&= \frac{1}{2} \frac{1}{(1 - \frac{1}{2})^2} \\
&= 2
\end{aligned}
$$

Problem 3.5. A modem transmits a +2 voltage signal into a channel. The channel adds to this signal a noise term that is drawn from the set $\{0, -1, -2, -3\}$ with respective probabilities $\{4/10, 3/10, 2/10, 1/10\}$.

(a). Find the PMF of the output Y of the channel.

(b). What is the probability that the channel's output is equal to the input of the channel?

(c). What is the probability that the channel's output is positive?

(d). Find $E[Y]$ and $Var[Y]$.

(a). The states of the random variable Y are $Y(\Omega) = \{2, 1, 0, -1\}$. We can calculate the probabilities for each state now:

$$p_y(2) = P(\{0\}) = \frac{4}{10}$$

$$p_y(1) = P(\{-1\}) = \frac{3}{10}$$

$$p_y(0) = P(\{-2\}) = \frac{2}{10}$$

$$p_y(-1) = P(\{-3\}) = \frac{1}{10}$$

(b). The input of the chanel is +2, so we want to find the probability that $Y = 2$. From part (a), $p_y(2) = \frac{4}{10}$.

(c).

$$\begin{aligned} P(Y > 0) &= P(Y = 1) + P(Y = 2) \\ &= \frac{7}{10} \end{aligned}$$

(d).

$$\begin{aligned} E[Y] &= \frac{-1}{10} + 0 + \frac{3}{10} + \frac{8}{10} \\ &= 1 \end{aligned}$$

$$\begin{aligned} Var[Y] &= E[X^2] - E[X]^2 \\ &= E[X^2] - 1 \\ &= (\frac{1}{10} + 0 + \frac{3}{10} + \frac{16}{10}) - 1 \\ &= 2 - 1 \\ &= 1 \end{aligned}$$

Problem 3.7. Let

$$g(X) = \begin{cases} 1 & \text{if } x > 10 \\ 0 & \text{otherwise} \end{cases},$$

and

$$h(X) = \begin{cases} X - 10 & \text{if } X - 10 > 0 \\ 0 & \text{otherwise} \end{cases},$$

(a). Find $E[g(X)]$ for X as in Problem 1(a) with $X(\Omega) = \{1, \ldots, 15\}$.
(b). Find $E[H(X)]$ for X as in Problem 1(a) with $X(\Omega) = \{1, \ldots, 15\}$.

We need to first find the probability distribution over the states of X. We have

$$\begin{aligned} 1 &= \sum_{k=1}^{15} \frac{p}{k} \\ &= p \sum_{k=1}^{15} \frac{1}{k} \\ &= p * 3.32, \end{aligned}$$

so $p = .3012$.

(a).

$$\begin{aligned} E[g(X)] &= \sum_{k=1}^{15} g(k) \frac{p}{k} \\ &= \sum_{k=11}^{15} g(k) \frac{p}{k} \\ &= \sum_{k=11}^{15} \frac{p}{k} \\ &\approx 0.1172 \end{aligned}$$

(b).

$$\begin{aligned} E[h(X)] &= \sum_{k=1}^{15} h(k) \frac{p}{k} \\ &= \sum_{k=11}^{15} (k - 10) \frac{p}{k} \\ &\approx 0.333 \end{aligned}$$

Problem 3.8. A voltage X is uniformly distributed in the set $\{-3, \ldots, 3, 4\}$

(a). Find mean and variance of X.

(b). Find mean and variance of $Y = -2X^2 + 3$.

(c). Find mean and variance of $W = cos(\pi X/8)$.

(d). Find mean and variance of $Z = cos^2(\pi X/8)$.

(a).

Since the distribution of X is uniformly distributed, $P(X = x) = \frac{1}{8}$ for each $x \in X(\Omega)$. We can now calculate the mean and variance of X.

$$\begin{aligned} E[X] &= \sum_{x=-3}^{4} \frac{x}{8} \\ &= 1/2 \end{aligned}$$

$$\begin{aligned} Var[X] &= E[X^2] - E[X]^2 \\ &= E[X^2] - \frac{1}{4} \\ &= (\sum_{x=-3}^{4} \frac{x^2}{8}) - \frac{1}{4} \\ &= \frac{21}{4} \end{aligned}$$

(b).

$$\begin{aligned} E[Y] &= E[-2X^2 + 3] \\ &= E[-2X^2] + 3 \\ &= -2E[X^2] + 3 \\ &= -8 \end{aligned}$$

$$
\begin{aligned}
Var[Y] &= E[Y^2] - E[Y]^2 \\
&= E[Y^2] - 64 \\
&= E[(-2X^2+3)^2] + 8 \\
&= (\sum_{k=-3}^{4} (-2k^2+3)^2 P_x(k)) - 64 \\
&= (\frac{1}{8} \sum_{k=-3}^{4} (-2k^2+3)^2) - 64 \\
&= 105
\end{aligned}
$$

(c).

$$
\begin{aligned}
E[W] &= E[cos(\pi X/8)] \\
&= \sum_{k=-3}^{4} cos(\pi k/8) p_x(k) \\
&= \frac{1}{8} \sum_{k=-3}^{4} cos(\pi k/8) \\
&\approx \frac{5}{8}
\end{aligned}
$$

$$
\begin{aligned}
Var[W] &= E[W^2] - E[W]^2 \\
&= E[W^2] - 0.390625 \\
&= (\sum_{k=-3}^{4} cos(\pi k/8)^2 p_x(k)) - 0.390625 \\
&= (\frac{1}{8} \sum_{k=-3}^{4} cos(\pi k/8)^2) - 0.390625 \\
&\approx 0.10938
\end{aligned}
$$

(d).

$$
E[Z] = E[W^2] = \frac{1}{2}
$$

$$
\begin{aligned}
Var[Z] &= Var[W^2] \\
&= Var[W^4] - Var[W^2]^2 \\
&= Var[W^4] - \frac{1}{4} \\
&= (\frac{1}{8}\sum_{k=-3}^{4} cos(\pi k/8)^4) - \frac{1}{4} && =\approx 0.125
\end{aligned}
$$

Problem 3.9. a. If X is $Poisson(\lambda)$, compute $E[1/(X+1)]$.

b. If X is $Bernoulli(p)$ and Y is $Bernoulli(q)$, compute $E[(X+Y)^3]$ if X and Y are independent.

c. Let X be a random variable with mean μ and variance σ^2. Let $\Delta(\theta) = E[(X-\theta)^2]$. Find θ that minimizes the error $\Delta(\theta)$.

d. Suppose that $X_1, \ldots, X_n$ are indpendent uniform random variables in $\{0, 1 \ldots 100\}$. Evaluate $P[min(X_1, \ldots, X_n) > \ell]$ for any $\ell \in \{0, 1, \ldots, 100\}$.

Part a.

$$\begin{aligned}
E[1/(X+1)] &= \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{\lambda^k e^- \lambda}{k!} \\
&= \frac{1}{\lambda} \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{\lambda^{k+1} e^{-\lambda}}{k!} \\
&= \frac{1}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1} e^{-\lambda}}{(k+1)!} \\
&= \frac{1}{\lambda}((\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}) - e^{-\lambda}) \\
&= \frac{1}{\lambda}(1 - e^{-\lambda})
\end{aligned}$$

Part b.

E[(X+Y)^3] $= E[X^3 + 3X^2Y + 3XY^2 + Y^3]$
$= E[X^3] + 3E[X^2Y] + 3E[XY^2] + E[Y^3]$
$=$

Part c. Simplying $\Delta(\theta)$, we have

$$\begin{aligned}
\Delta(\theta) &= E[(X-\theta)^2] \\
&= E[X^2 - 2\theta X + \theta^2] \\
&= E[X^2] - 2\theta E[X] + E[\theta^2] \\
&= \sigma^2 + \mu^2 + -2\theta\mu + \theta^2.
\end{aligned}$$

It suffices to find the minimum of the final expression above. We can find where the first derivative is zero :

$$\Delta(\theta)' = -2\mu + 2\theta = 0,$$

so $\theta = \mu$. We can now verify that this value of θ is a minimum.

$$\begin{aligned}\Delta(\mu)'' &= 2 \\ &> 0,\end{aligned}$$

so $\theta = \mu$ is the value that minimizes $\Delta(\theta)$.

Part d.

Let $\ell \in \{0, 1, \ldots, 100\}$.

$$\begin{aligned}P(min(X_1, \ldots, X_n) > \ell) &= P((X_1 > \ell) \cap (X_2 > \ell) \cap \cdots \cap (X_n > \ell)) \\ &= P(X_1 > \ell)P(X_2 > \ell) \ldots P(X_n > \ell) \\ &= P(X_1 > \ell)^n \\ &= (\frac{101 - l - 1}{101})^n\end{aligned}$$

Problem 3.10. a. Consider the binomial probability mass function $p_X(k) = \binom{n}{k}p^k(1-p)^{n-k}$. Show that the mean is np.

b.

c.

d.

Part a. Let X be a $Binomial(n, p)$. The sum of n independent Bernoulli random variables with parameter p equals X

$$X = X_1 + X_2 + \ldots X_n.$$

Taking the expectation, we have

$$\begin{aligned}
E[X] &= E[X_1 + X_2 + \cdots + X_n] \\
&= E[X_1] + E[X_2] + \cdots + E[X_n] \\
&= p + p + \cdots + p \\
&= np.
\end{aligned}$$

Part b.

Part c.

Part d.