

HW Solution 4 — Not Due

Lecturer: Prapun Suksompong, Ph.D.

Problem 1. The random variable V has pmf

$$p_V(v) = \begin{cases} cv^2, & v = 1, 2, 3, 4, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the value of the constant c .
- (b) Find $P[V \in \{u^2 : u = 1, 2, 3, \dots\}]$.
- (c) Find the probability that V is an even number.
- (d) Find $P[V > 2]$.
- (e) Sketch $p_V(v)$.
- (f) Sketch $F_V(v)$.

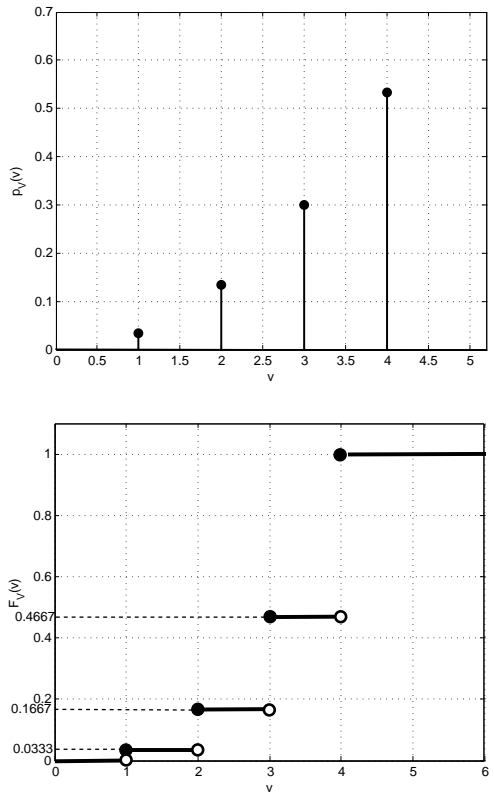
Solution: [Y&G, Q2.2.3]

- (a) We choose c so that the pmf sums to one:

$$\sum_v p_V(v) = c(1^2 + 2^2 + 3^2 + 4^2) = 30c = 1.$$

Hence, $c = \boxed{1/30}$.

- (b) $P[V \in \{u^2 : u = 1, 2, 3, \dots\}] = p_V(1) + p_V(4) = c(1^2 + 4^2) = \boxed{17/30}$.
- (c) $P[V \text{ even}] = p_V(2) + p_V(4) = c(2^2 + 4^2) = 20/30 = \boxed{2/3}$.
- (d) $P[V > 2] = p_V(3) + p_V(4) = c(3^2 + 4^2) = 25/30 = \boxed{5/6}$.
- (e) Sketch of $p_V(v)$:
- (f) Sketch of $F_V(v)$:



Problem 2. An optical inspection system is to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent.

- (a) Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X . [Montgomery and Runger, 2010, Q3-20]
- (b) Let the random variable Y denote the number of parts that are incorrectly classified. Determine the probability mass function of Y .

Solution:

- (a) X is a binomial random variable with $n = 3$ and $p = 0.98$. Hence,

$$p_X(x) = \begin{cases} \binom{3}{x} 0.98^x (0.02)^{3-x}, & x \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

In particular, $p_X(0) = 8 \times 10^{-6}$, $p_X(1) = 0.001176$, $p_X(2) = 0.057624$, and $p_X(3) = 0.941192$. Note that in MATLAB, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.98)`.

(b) Y is a binomial random variable with $n = 3$ and $p = 0.02$. Hence,

$$p_Y(y) = \begin{cases} \binom{3}{y} 0.02^y (0.98)^{3-y}, & y \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise} \end{cases} \quad (4.2)$$

In particular, $p_Y(0) = 0.941192$, $p_Y(1) = 0.057624$, $p_Y(2) = 0.001176$, and $p_Y(3) = 8 \times 10^{-6}$. Note that in **MATLAB**, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.02)`.

Alternatively, note that there are three parts. If X of them are classified correctly, then the number of incorrectly classified parts is $n - X$, which is what we defined as Y . Therefore, $Y = 3 - X$. Hence, $p_Y(y) = P[Y = y] = P[3 - X = y] = P[X = 3 - y] = p_X(3 - y)$.

Problem 3. The thickness of the wood paneling (in inches) that a customer orders is a random variable with the following cdf:

$$F_X(x) = \begin{cases} 0, & x < \frac{1}{8} \\ 0.2, & \frac{1}{8} \leq x < \frac{1}{4} \\ 0.9, & \frac{1}{4} \leq x < \frac{3}{8} \\ 1 & x \geq \frac{3}{8} \end{cases}$$

Determine the following probabilities:

- (a) $P[X \leq 1/18]$
- (b) $P[X \leq 1/4]$
- (c) $P[X \leq 5/16]$
- (d) $P[X > 1/4]$
- (e) $P[X \leq 1/2]$

[Montgomery and Runger, 2010, Q3-42]

Solution:

- (a) $P[X \leq 1/18] = F_X(1/18) = 0$ because $\frac{1}{18} < \frac{1}{8}$.
- (b) $P[X \leq 1/4] = F_X(1/4) = 0.9$
- (c) $P[X \leq 5/16] = F_X(5/16) = 0.9$ because $\frac{1}{4} < \frac{5}{16} < \frac{3}{8}$.
- (d) $P[X > 1/4] = 1 - P[X \leq 1/4] = 1 - F_X(1/4) = 1 - 0.9 = 0.1$.

(e) $P[X \leq 1/2] = F_X(1/2) = 1$ because $\frac{1}{2} > \frac{3}{8}$.

Alternatively, we can also derive the pmf first and then calculate the probabilities.

Problem 4. Plot the Poisson pmf for $\alpha = 10, 30,$ and 50 .

Solution: See Figure 4.1.

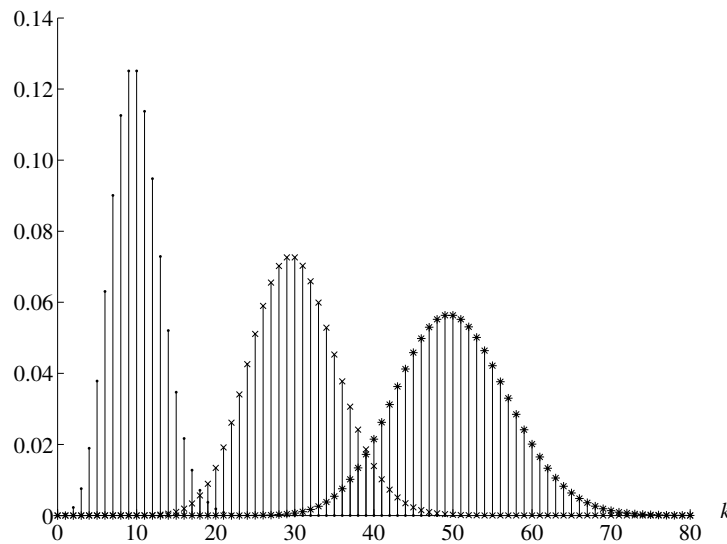


Figure 4.1: The Poisson pmf for $\alpha = 10, 30,$ and 50 from left to right, respectively. [Gubner, 2006, Figure 2.5]

Problem 5. Let $X \sim \mathcal{P}(\alpha)$.

- Evaluate $P[X > 1]$. Your answer should be in terms of α .
- Compute the numerical value of $P[X > 1]$ when $\alpha = 1$.

Solution:

(a) $P[X > 1] = 1 - P[X \leq 1] = 1 - (P[X = 0] + P[X = 1]) = \boxed{1 - e^{-\alpha}(1 + \alpha)}$.

(b) $\boxed{0.264}$.

Problem 6. When n is large, binomial distribution $\text{Binomial}(n, p)$ becomes difficult to compute directly because of the need to calculate factorial terms. In this question, we will consider an approximation when p is close to 0. In such case, the binomial can be approximated by the Poisson distribution with parameter $\alpha = np$.

More specifically, suppose X_n has a binomial distribution with parameters n and p_n . If $p_n \rightarrow 0$ and $np_n \rightarrow \alpha$ as $n \rightarrow \infty$, then

$$P[X_n = k] \rightarrow e^{-\alpha} \frac{\alpha^k}{k!}.$$

- Let $X \sim \text{Binomial}(12, 1/36)$. (For example, roll two dice 12 times and let X be the number of times a double 6 appears.) Evaluate $p_X(x)$ for $x = 0, 1, 2$.
- Compare your answers in the previous part with the Poisson approximation.
- Compare the plot of $p_X(x)$ and $\mathcal{P}(np)$.

Solution:

- 0.7132, 0.2445, 0.0384.
- 0.7165, 0.2388, 0.0398.
- See Figure 4.2.

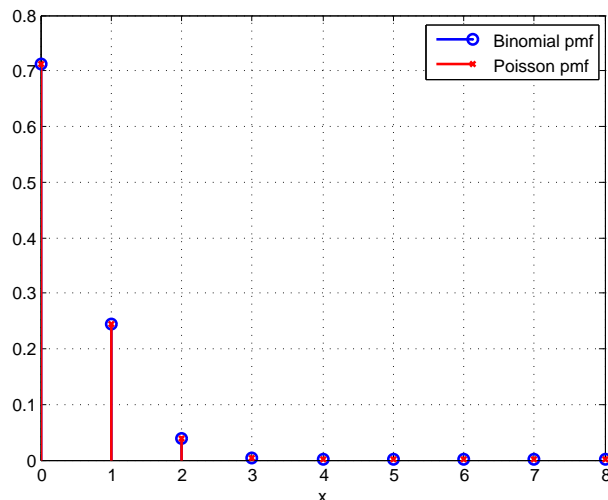


Figure 4.2: Poisson Approximation

Problem 7. In one of the New York state lottery games, a number is chosen at random between 0 and 999. Suppose you play this game 250 times. Use the Poisson approximation to estimate the probability that you will never win and compare this with the exact answer.

Solution: [Durrett, 2009, Q2.41] Let W be the number of wins. Then, $W \sim \text{Binomial}(250, p)$ where $p = 1/1000$. Hence,

$$P[W = 0] = \binom{250}{0} p^0 (1-p)^{250} \approx 0.7787.$$

If we approximate W by $\Lambda \sim \mathcal{P}(\alpha)$. Then we need to set

$$\alpha = np = \frac{250}{1000} = \frac{1}{4}.$$

In which case,

$$P[\Lambda = 0] = e^{-\alpha} \frac{\alpha^0}{0!} = e^{-\alpha} \approx 0.7788$$

which is very close to the answer from direct calculation.

Problem 8. Suppose X is a random variable whose pmf at $x = 0, 1, 2, 3, 4$ is given by $p_X(x) = \frac{2x+1}{25}$. Determine its expected value and variance. [Montgomery and Runger, 2010, Q3-51]

Solution:

$$\begin{aligned} \mathbb{E}X &= \sum_{x=0}^4 x p_X(x) = \sum_{x=0}^4 x \frac{2x+1}{25} = 0 + 1 \left(\frac{3}{25} \right) + 2 \left(\frac{5}{25} \right) + 3 \left(\frac{7}{25} \right) + 4 \left(\frac{9}{25} \right) \\ &= \frac{70}{25} = \frac{14}{5} = \boxed{2.8}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=0}^4 x^2 p_X(x) = \sum_{x=0}^4 x^2 \frac{2x+1}{25} = 0 + 1^2 \left(\frac{3}{25} \right) + 2^2 \left(\frac{5}{25} \right) + 3^2 \left(\frac{7}{25} \right) + 4^2 \left(\frac{9}{25} \right) \\ &= \frac{230}{25} = \frac{46}{5} = 9.2 \end{aligned}$$

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 9.2 - 2.8^2 = \boxed{1.36}.$$

Problem 9. An article in Information Security Technical Report [“Malicious Software—Past, Present and Future” (2004, Vol. 9, pp. 618)] provided the data (shown in Figure 4.3) on the top ten malicious software instances for 2002. The clear leader in the number of registered incidences for the year 2002 was the Internet worm “Klez”. This virus was first

Place	Name	% Instances
1	I-Worm.Klez	61.22%
2	I-Worm.Lentin	20.52%
3	I-Worm.Tanatos	2.09%
4	I-Worm.BadtransII	1.31%
5	Macro.Word97.Thus	1.19%
6	I-Worm.Hybris	0.60%
7	I-Worm.Bridex	0.32%
8	I-Worm.Magistr	0.30%
9	Win95.CIH	0.27%
10	I-Worm.Sircam	0.24%

Figure 4.3: The 10 most widespread malicious programs for 2002 (Source—Kaspersky Labs).

detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology.

Suppose that 20 malicious software instances are reported. Assume that the malicious sources can be assumed to be independent.

- What is the probability that at least one instance is “Klez”?
- What is the probability that three or more instances are “Klez”?
- What are the expected value and standard deviation of the number of “Klez” instances among the 20 reported?

Solution: Let N be the number of instances (among the 20) that are “Klez”. Then, $N \sim \text{binomial}(n, p)$ where $n = 20$ and $p = 0.6122$.

$$(a) \quad P[N \geq 1] = 1 - P[N < 1] = 1 - P[N = 0] = 1 - p_N(0) = 1 - \binom{20}{0} \times 0.6122^0 \times 0.3878^{20} \approx 0.9999999941 \approx 1.$$

(b)

$$\begin{aligned} P[N \geq 3] &= 1 - P[N < 3] = 1 - (P[N = 0] + P[N = 1] + P[N = 2]) \\ &= 1 - \sum_{k=0}^2 \binom{20}{k} (0.6122)^k (0.3878)^{20-k} \approx 0.999997 \end{aligned}$$

$$(c) \quad \mathbb{E}N = np = 20 \times 0.6122 = 12.244.$$

$$\sigma_N = \sqrt{\text{Var } N} = \sqrt{np(1-p)} = \sqrt{20 \times 0.6122 \times 0.3878} \approx 2.179.$$

Problem 10. The random variable V has pmf

$$p_V(v) = \begin{cases} \frac{1}{v^2} + c, & v \in \{-2, 2, 3\} \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the value of the constant c .
- (b) Find $P[V > 3]$.
- (c) Find $P[V < 3]$.
- (d) Find $P[V^2 > 1]$.
- (e) Let $W = V^2 - V + 1$. Find the pmf of W .
- (f) Find $\mathbb{E}V$
- (g) Find $\mathbb{E}[V^2]$
- (h) Find $\text{Var } V$
- (i) Find σ_V
- (j) Find $\mathbb{E}W$

Solution:

- (a) The pmf must sum to 1. Hence,

$$\frac{1}{(-2)^2} + c + \frac{1}{(2)^2} + c + \frac{1}{(3)^2} + c = 1.$$

The value of c must be

$$c = \frac{1}{3} \left(1 - \frac{1}{4} - \frac{1}{4} - \frac{1}{9} \right) = \boxed{\frac{7}{54}} \approx 0.1296$$

Note that this gives

$$p_V(-2) = p_V(2) = \frac{41}{108} \approx 0.38 \quad \text{and} \quad p_V(3) = \frac{13}{54} \approx 0.241.$$

- (b) $P[V > 3] = \boxed{0}$ because all elements in the support of V are ≤ 3 .
- (c) $P[V < 3] = 1 - p_V(3) = \frac{41}{54} \approx 0.759$.

(d) $P[V^2 > 1] = \boxed{1}$ because the square of any element in the support of V is > 1 .

(e) $W = V^2 - V + 1$. So, when $V = -2, 2, 3$, we have $W = 7, 3, 7$, respectively. Hence, W takes only two values, 7 and 3. the corresponding probabilities are

$$P[W = 7] = p_V(-2) + p_V(3) = \frac{67}{108} \approx 0.62.$$

and

$$P[W = 3] = p_V(2) = \frac{41}{108} \approx 0.38.$$

Hence, the pmf of W is given by

$$p_W(w) = \begin{cases} \frac{41}{108}, & w = 3, \\ \frac{67}{108}, & w = 7, \\ 0, & \text{otherwise.} \end{cases} \approx \boxed{\begin{cases} 0.38, & w = 3, \\ 0.62, & w = 7, \\ 0, & \text{otherwise.} \end{cases}}$$

(f) $\mathbb{E}V = \frac{13}{18} \approx 0.7222$

(g) $\mathbb{E}V^2 = \frac{281}{54} \approx 5.2037$

(h) $\text{Var } V = \mathbb{E}V^2 - (\mathbb{E}V)^2 = \frac{1517}{324} \approx 4.682.$

(i) $\sigma_V = \sqrt{\text{Var } V} \approx 2.1638$

(j) $\mathbb{E}W = 5.4815$