ECS 315: Probability and Random Processes **2012/1**

HW Solution 5 — Due: Sep 13

Lecturer: Prapun Suksompong, Ph.D.

Instructions

(a) ONE part of a question will be graded (5 pt). Of course, you do not know which part will be selected; so you should work on all of them.

(b) It is important that you try to solve all problems. (5 pt)

(c) Late submission will be heavily penalized.

(d) Write down all the steps that you have done to obtain your answers. You may not get full credit even when your answer is correct without showing how you get your answer.

Problem 1. When n is large, binomial distribution 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[1] by the Poisson distribution with parameter $\alpha = np$.

(a) Let $X \sim$ 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 $x = 0, 1, 2$.

(b) Compare your answers in the previous part with the Poisson approximation.

(c) Compare `MATLAB` plots of $p_X(x)$ and the pmf of $\mathcal{P}(np)$.

(d) 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:

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

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

(a) 0.7132, 0.2445, 0.0384.

(b) 0.7165, 0.2388, 0.0398.

(c) See Figure **??**.

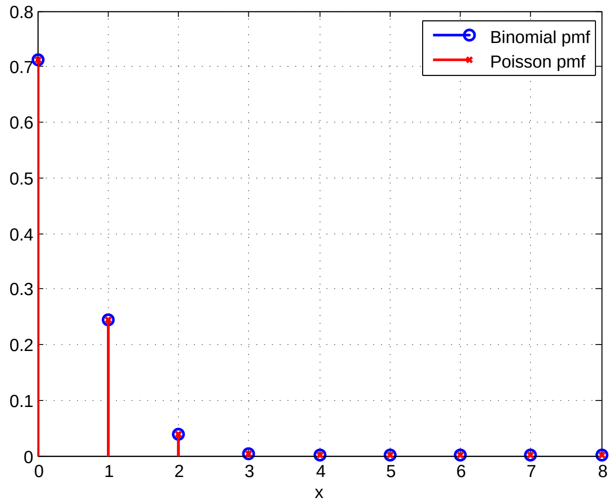

Figure 5.1: Poisson Approximation

urrett, 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 2. An article in Information Security Technical Report ["Malicious Software—Past, Present and Future" (2004, Vol. 9, pp. 618)] provided the data (shown in Figure 5.1) 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 detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology.

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

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

(a) What is the probability that at least one instance is "Klez"?

(b) What is the probability that three or more instances are "Klez"?

(c) 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, $N \sim$binomial(n,p) where $n = 20$ and $p = 0.6122$.

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

(b)

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

(c) $\mathbb{E}N = np = 20 \times 0.6122 = 12.244$.
$\sigma_N = \sqrt{\operatorname{Var} N} = \sqrt{np(1-p)} = \sqrt{20 \times 0.6122 \times 0.3878} \approx 2.179$.

Problem 3. 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 $\operatorname{Var} V$

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

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

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

Problem 4. Suppose X is a uniform discrete random variable on $\{-3, -2, -1, 0, 1, 2, 3, 4\}$. Find

(a) $\mathbb{E}X$

(b) $\mathbb{E}[X^2]$

(c) $\operatorname{Var} X$

(d) σ_X

Solution: All of the calculations in this question are simply plugging in numbers into appropriate formula. See the corresponding MATLAB file.

(a) $\mathbb{E}X = 0.5$

(b) $\mathbb{E}\left[X^2\right] = 5.5$

(c) $\operatorname{Var} X = 5.25$

(d) $\sigma_X = 2.2913$

Alternatively, we can find a formula for the general case of uniform random variable X on the sets of integers from a to b. Note that there are $n = b - a + 1$ values that the random variable can take. Hence, all of them has probability $\frac{1}{n}$.

(a) $\mathbb{E}X = \sum_{i=a}^{b} i\frac{1}{n} = \frac{1}{n}\sum_{i=a}^{b} i = \frac{1}{n} \times \frac{n(a+b)}{2} = \frac{a+b}{2}$.

(b) First, note that

$$\begin{aligned}\sum_{i=a}^{b} i\left(i-1\right) &= \sum_{i=a}^{b} i\left(i-1\right)\left(\frac{\left(i+1\right)-\left(i-2\right)}{3}\right)\\ &= \frac{1}{3}\left(\sum_{i=a}^{b}\left(i+1\right)i\left(i-1\right) - \sum_{i=a}^{b} i\left(i-1\right)\left(i-2\right)\right)\\ &= \frac{1}{3}\left(\left(b+1\right)b\left(b-1\right) - a\left(a-1\right)\left(a-2\right)\right)\end{aligned}$$

where the last equality comes from the fact that there are many terms in the first sum that is repeated in the second sum and hence many cancellations.

Now,

$$\begin{aligned}\sum_{i=a}^{b} i^2 &= \sum_{i=a}^{b}\left(i\left(i-1\right)+i\right) = \sum_{i=a}^{b} i\left(i-1\right) + \sum_{i=a}^{b} i\\ &= \frac{1}{3}\left(\left(b+1\right)b\left(b-1\right) - a\left(a-1\right)\left(a-2\right)\right) + \frac{n\left(a+b\right)}{2}\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{i=a}^{b} i^2\frac{1}{n} &= \frac{1}{3n}\left(\left(b+1\right)b\left(b-1\right) - a\left(a-1\right)\left(a-2\right)\right) + \frac{a+b}{2}\\ &= \frac{1}{3}a^2 - \frac{1}{6}a + \frac{1}{3}ab + \frac{1}{6}b + \frac{1}{3}b^2\end{aligned}$$

(c) $\operatorname{Var} X = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}X\right)^2 = \frac{1}{12}\left(b-a\right)\left(b-a+2\right) = \frac{1}{12}(n-1)(n+1) = \frac{n^2-1}{12}$.

(d) $\sigma_X = \sqrt{\operatorname{Var} X} = \sqrt{\frac{n^2-1}{12}}$.

Problem 5. (Expectation + pmf + Gambling + Effect of miscalculation of probability) In the eighteenth century, the famous French mathematician Jean Le Rond d'Alembert, author of several works on probability, analyzed the toss of two coins. He reasoned that because this experiment has THREE outcomes, (the number of heads that turns up in those two tosses can be 0, 1, or 2), the chances of each must be 1 in 3. In other words, if we let N be the number of heads that shows up, Alembert would say that

$$p_N(n) = 1/3 \quad \text{for} \quad N = 0, 1, 2.$$

[Mlodinow, 2008, p 50–51]

We know that Alembert's conclusion was ***wrong***. His three outcomes are not equally likely and hence classical probability formula can not be applied directly. The key is to realize that there are FOUR outcomes which are equally likely. We should not consider 0, 1, or 2 heads as the possible outcomes but rather the sequences (heads, heads), (heads, tails), (tails, heads), and (tails, tails). These are the 4 possibilities that make up the sample space. The actual pmf for N is

$$p_N(n) = \begin{cases} 1/4, & n = 0, 2, \\ 1/2 & n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose you travel back in time and meet Alembert. You could make the following bet with Alembert to gain some easy money. The bet is that if the result of a toss of two coins contains exactly one head, then he would pay you \$150. Otherwise, you would pay him \$100.

(a) Let R be the amount of money that Alembert gets from this bet. Then, $R = -150$ if you win and $R = +100$ otherwise. Use Alembert's *miscalculated* probabilities to determine the pmf of R (from Alembert's belief).

(b) Use Alembert's *miscalculated* probabilities (or the corresponding miscalculated pmf that you got from part (a)) to calculate $\mathbb{E}R$, the expected amount of money that he expects to gain.

Remark: You should find that $\mathbb{E}R > 0$ and hence Alembert will be very happy to accept your bet.

(c) Use the *actual* probabilities, to determine the pmf of R.

(d) Use the *actual* pmf, to determine $\mathbb{E}R$.

Remark: You should find that $\mathbb{E}R < 0$ and hence Alembert should not accept your bet if he calculates the probabilities correctly.

(e) Let Y be the return of this bet for you. Then, $Y = +150$ if you win and $Y = -100$ otherwise. Use the *actual* probabilities to determine the pmf of Y.

(f) Use the *actual* probabilities, to determine $\mathbb{E}Y$.

Remark: You should find that $\mathbb{E}Y > 0$. This is the amount of money that you expect to gain each time that you play with Alembert. Of course, Alembert, who still believes that his calculation is correct, will ask you to play this bet again and again believing that he will make profit in the long run.

By miscalculating probabilities, one can make wrong decisions (and lose a lot of money)!

Solution:

(a) $P[R = -150] = P[N = 1]$ and $P[R = +100] = P[N \neq 1] = P[N = 0] + P[N = 2]$. So,

$$\boxed{p_R(r) = \begin{cases} p_N(1), & r = -150, \\ p_N(0) + p_N(2), & r = +100, \\ 0, & \text{otherwise.} \end{cases}}$$

Using Alembert's *miscalculated* pmf,

$$p_R(r) = \begin{cases} 1/3, & r = -150, \\ 2/3, & r = +100, \\ 0, & \text{otherwise} \end{cases}$$

(b) From $p_R(r)$ in part (a), we have $\mathbb{E}R = \sum_r p_R(r) = \frac{1}{3} \times (-150) + \frac{2}{3} \times 100 = \boxed{\frac{50}{3}} \approx 16.67$

(c) Again,

$$p_R(r) = \begin{cases} p_N(1), & r = -150, \\ p_N(0) + p_N(2), & r = +100, \\ 0, & \text{otherwise} \end{cases}$$

Using the actual pmf,

$$p_R(r) = \begin{cases} \frac{1}{2}, & r = -150, \\ \frac{1}{4} + \frac{1}{4}, & r = +100, \\ 0, & \text{otherwise} \end{cases} = \boxed{\begin{cases} \frac{1}{2}, & r = -150 \text{ or } +100, \\ 0, & \text{otherwise.} \end{cases}}$$

(d) From $p_R(r)$ in part (c), we have $\mathbb{E}R = \sum_r p_R(r) = \frac{1}{2} \times (-150) + \frac{1}{2} \times 100 = \boxed{-25}$.

(e) Observe that $Y = -R$. Hence, using the answer from part (d), we have

$$p_R(r) = \boxed{\begin{cases} \frac{1}{2}, & r = +150 \text{ or } -100, \\ 0, & \text{otherwise.} \end{cases}}$$

(f) Observe that $Y = -R$. Hence, $\mathbb{E}Y = -\mathbb{E}R$. Using the actual probabilities, $\mathbb{E}R = -25$ from part (d). Hence, $\mathbb{E}Y = \boxed{+25}$.