IPDA1005 Introduction to Probability and Data Analysis

Worksheet 9

Solution

1. A student attempts 48 true-false questions in a test. For each question, independently, he has $\frac{3}{4}$ probability of answering correctly. Using an appropriate approximation, find the probability that he answers

   (a) at least 40 questions correctly

   (b) between 35 and 40 questions, inclusive

   (c) exactly 36 questions.

Solution: Here the condition that np and $n(1-p)$ are both at least 10 is satisfied, so the appropriate approximation is normal distribution with mean $\mu = np = 36$ and variance $np(1-p) = 9$.

(a)

$$\begin{aligned} P(X \geq 40) = P(X \geq 39.5) &\approx P\left(Z \geq \frac{39.5-36}{3}\right) \\ &= P(Z \geq 1.17) = 1 - .8790 = .1210 \end{aligned}$$

(b)

$$\begin{aligned} P(35 \leq X \leq 40) = P(34.5 \leq X \leq 40.5) &\approx P\left(\frac{34.5-36}{3} \leq Z \leq \frac{40.5-36}{3}\right) \\ &= P(-0.5 \leq Z \leq 1.5) = 0.6247 \end{aligned}$$

(c)

$$\begin{aligned} P(X = 36) = P(35.5 \leq X \leq 36.5) &\approx P\left(\frac{35.5-36}{3} \leq Z \leq \frac{36.5-36}{3}\right) \\ &= P(-0.1667 \leq Z \leq 0.1667) = 0.1324 \end{aligned}$$

2. Let $X \sim \text{Gamma}(\alpha, \lambda)$.

   (a) Show that $\mu_r' = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)\lambda^r}$ (Hint: Make use of the fact that the gamma density function integrates to 1.)

   (b) Using (a), show that $E(X) = \frac{\alpha}{\lambda}$ and $\text{Var}(X) = \frac{\alpha}{\lambda^2}$.

Solution:

(a) As the gamma density function integrates to 1, we have

$$\int_0^\infty x^{\alpha-1}e^{-\lambda x} = \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

Consequently,

$$\begin{aligned}
\mu'_r &= E(X^r) \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^r x^{\alpha-1}e^{-\lambda x} \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+r-1}e^{-\lambda x} \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+r)}{\lambda^{\alpha+r}} \\
&= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)\lambda^r}
\end{aligned}$$

(b) Substituting $r = 1$ in (a), we get $E(X) = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}$. From Property 2 of gamma function given in the lecture, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, so $E(X) = \frac{\alpha}{\lambda}$.

Substituting $r = 2$ in (a), we get $E(X^2) = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2}$. Applying the gamma function property twice, we get $\Gamma(\alpha+2) = \alpha(\alpha+1)\Gamma(\alpha)$, so $E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$.

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

3. Let $X \sim N(\mu, \sigma^2)$ and let $Y = e^X$. The by definition, $Y \sim \text{Lognormal}(\mu, \sigma^2)$. Derive the density function of Y and hence show that

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}$$

Solution: $Y = h(X)$ where $h(x) = e^x$. $g(y) = h^{-1}(y) = \ln y$, so $g'(y) = \frac{1}{y}$. Thus for $y > 0$,

$$f_Y(y) = f_X(g(y))|g'(y)| = \frac{1}{y}\phi(\ln y) = \frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}$$

4. The number of accidents on a highway is modelled by $\{N(t) : t \geq 0\}$, a Poisson process with rate $\lambda = 2$, where t represents time in days.

(a) Find the probability that $N(4) - N(2)$ is zero. Interpret this quantity.

(b) Express the number of accidents in 3 days in terms of N. Find the probability that there are no more than two accidents in 3 days.

(c) Find the expected time between accidents.

Solution:

(a) $N(4) - N(2)$ has Poisson(4) distribution, $P(N(4) - N(2) = 0) = e^{-4} = 0.0183$. This is the probability that no accidents occur during the period after Day 2 and before Day 4.

(b) The number of accidents in 3 days is $N(3)$ (or $N(t+3) - N(t)$ for any t.) As $N(3) \sim Pois(6)$

$$P(N(3) \leq 2) = e^{-6}\left(1 + 6 + \frac{6^2}{2!}\right) = 0.0620$$

(c) The inter-accident period has $Exp(2)$ distribution whose expected value is $\frac{1}{2}$.

5. Suppose we have three random numbers 0.588, 0.222 and 0.906 from $U(0, 1)$ distribution. Use these to generate three random numbers from

(a) exponential distribution with rate parameter 0.1

(b) geometric distribution with $p = 0.2$

(c) binomial(4, .6) distribution.

Solution:

(a) As $F(x) = 1 - e^{-.1x}$, we get that $F^{-1}(u) = -10\ln(1-u)$. Applying this function to the given values, we get

$$x_1 = F^{-1}(.588) = 8.867, \quad x_2 = F^{-1}(.222) = 2.510, \quad x_3 = F^{-1}(.906) = 23.645$$

(b) From the formula $F(k) = P(X \leq k) = 1 - q^{k+1} = 1 - .8^{k+1}$, we get

$$F(0) = 0.2, F(1) = 0.36, F(2) = 0.488,$$

$$F(3) = 0.5904, F(9) = 0.8926, F(10) = 0.9141.$$

Since $F(2) < .588 \leq F(3)$, the first simulated value is 3.
Since $F(0) < .222 \leq F(1)$, the second simulated value is 1.
Since $F(9) < .906 \leq F(10)$, the third simulated value is 10.

(c) From the cumulative tables of binomial or from `R`, we get

$$F(0) = 0.0256, F(1) = 0.1792, F(2) = 0.5248, F(3) = 0.8704, F(4) = 1.$$

Since $F(2) < .588 \leq F(3)$, the first simulated value is 3.
Since $F(1) < .222 \leq F(2)$, the second simulated value is 2.
Since $F(3) < .906 \leq F(4)$, the third simulated value is 4.

6. When an automobile is stopped by a roving safety patrol, each tyre is checked for tyre wear, and each headlight is checked to see whether it is properly aimed. Let X denote the number of headlights that need adjustment, and let Y denote the number of defective tyres.
   (a) If X and Y are independent with $p_X(0) = 0.5$, $p_X(1) = 0.3$, $p_X(2) = 0.2$, and $p_Y(0) = 0.6$, $p_Y(1) = 0.1$, $p_Y(2) = p_Y(3) = 0.05$, $p_Y(4) = 0.2$, display the joint PMF $p(x, y)$ in a joint probability table.
   (b) Calculate $P(X \leq 1, Y \leq 1)$ from the joint PMF, and verify that it equals the product $P(X \leq 1) \cdot P(Y \leq 1)$.
   (c) How would you express the probability of no violations? What is its value?
   (d) Calculate $P(X + Y \leq 1)$.
   (e) Display the joint CDF. There are two ways that you might go about doing this: (a) by adding up the appropriate probabilities from the joint PMF, or (b) using the marginal CDFs and the fact that X and Y are independent. Can you prove this latter result?

Solution:

(a) Because X and Y are independent, the joint PMF can be easily calculate as the outer product of their marginal PMFs, as follows:

		Y					
		0	**1**	**2**	**3**	**4**	$p_X(x)$
X	**0**	0.3	0.05	0.025	0.025	0.1	*0.5*
	1	0.18	0.03	0.015	0.015	0.06	*0.3*
	2	0.12	0.02	0.01	0.01	0.04	*0.2*
	$p_Y(y)$	*0.6*	*0.1*	*0.05*	*0.05*	*0.2*	

(b) $P(X \leq 1, Y \leq 1) = p(0, 0) + p(0, 1) + p(1, 0) + p(1, 1) = 0.56$. Furthermore,

$$P(X \leq 1) \cdot P(Y \leq 1) = (0.5 + 0.3) \cdot (0.6 + 0.1) = 0.56.$$

(c) The probability of no violations can be expressed as

$$P(X + Y = 0) = P(X = 0, Y = 0) = 0.30.$$

(d) $P(X+Y \le 1) = p(0,0) + p(0,1) + p(1,0) = 0.53$

(e) Because X and Y are independent, it's far easier to calculate the joint CDF by calculating the outer product of the marginal CDFs.

		Y					
		0	**1**	**2**	**3**	**4**	$F_X(x)$
	0	0.3	0.35	0.375	0.4	0.5	*0.5*
X	**1**	0.48	0.56	0.6	0.64	0.8	*0.8*
	2	0.6	0.7	0.75	0.8	1	*1*
	$F_Y(y)$	*0.6*	*0.7*	*0.75*	*0.8*	*1*	

From the definition of the joint CDF,
$F(x,y) = P(X \le x, Y \le y) = \sum_{u \le x} \sum_{v \le y} p(u,v)$, but because of independence, we can write $F(x,y) = \sum_{u \le x} \sum_{v \le y} p_X(u) \cdot p_Y(v)$, and hence
$F(x,y) = \sum_{u \le x} p_X(u) \sum_{v \le y} p_Y(v) = F_X(x) \cdot F_Y(y)$.

7. The joint probability distribution of the number X of cars and the number Y of buses per signal cycle at a proposed right-turn lane is displayed in the accompanying joint probability table.

			y	
	$p(x, y)$	0	1	2
	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
x	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

(a) What is the probability that there is exactly one car and exactly one bus during a cycle?

(b) What is the probability that there is at most one car and at most one bus during a cycle?

(c) What is the probability that there is exactly one car during a cycle? Exactly one bus?

(d) Suppose the right-turn lane is to have a capacity of five cars, and one bus is equivalent to three cars. What is the probability of an overflow during a cycle? Calculate this quantity by adding up the appropriate probabilities in the joint PMF. If you were to express the probability of overflow as $P(aX + bY > 5)$, what would the constants a and b be?

(e) Are X and Y independent random variables? Explain.

Solution:

(a) $p(1,1) = 0.030$.

(b) $P(X \leq 1, Y \leq 1) = p(0,0) + p(0,1) + p(1,0) + p(1,1) = .120$

(c) $P(X = 1) = p(1,0)+p(1,1)+p(1,2) = .100$; $P(Y = 1) = p(0,1)+\ldots+p(5,1) = .300$.

(d) $P(\text{overflow}) = P(X + 3Y > 5) = 0.380$, which is obtained by adding the probabilities in the red rectangles from the joint PMF below:

	$p(x, y)$	y = 0	1	2
	0	.025	.015	$\boxed{.010}$
	1	.050	.030	$\boxed{.020}$
	2	.125	.075	$\boxed{.050}$
x	3	.150	$\boxed{.090}$	$\boxed{.060}$
	4	.100	$\boxed{.060}$	$\boxed{.040}$
	5	.050	$\boxed{.030}$	$\boxed{.020}$

(e) The marginal probabilities for X (row sums) are $p_X(0) = 0.05, p_X(1) = 0.10, p_X(2) = 0.25, p_X(3) = 0.30, p_X(4) = 0.20, p_X(5) = 0.10$. For Y, the marginal probabilities (column sums) are $p_Y(0) = 0.5, p_Y(1) = 0.3, p_Y(2) = 0.2$. It is easy to verify that for every (x, y), $p(x, y) = p_X(x) \cdot p_Y(y)$, and hence that X and Y are independent.

8. An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X be the number of marked earned on the first part and Y be the number of points earned on the second part. Suppose that the joint PMF of X and Y is given in the accompanying table.

	$p(x, y)$	y = 0	5	10	15
	0	.02	.06	.02	.10
x	5	.04	.15	.20	.10
	10	.01	.15	.14	.01

(a) If the score recorded for each student is the total number of marks earned on the two parts, what is the expected recorded score $E(X + Y)$?

(b) Calculate the covariance between X and Y.

(c) Calculate the correlation coefficient ρ between X and Y.

Solution:

(a) $E(X+Y) = \sum_x \sum_y (x+y)p(x,y) = (0+0)(0.02) + (0+5)(0.06) + \ldots + (10+15)(0.01) = 14.1$

(b) The expected values of X and Y can be calculated from their respected marginals, i.e., $E(X) = (0)(0.20) + (5)(0.49) + (10)(0.31) = 5.55$, and similarly, $E(Y) = 8.55$. We calculate $E(XY)$ as $\sum_x \sum_y xy \cdot p(x,y) = (0)(0)(0.02) + (0)(5)(0.06) + \ldots + (15)(10)(0.01) = 44.25$, and hence $\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = 44.25–(5.55)(8.55) = -3.20$.
Note: It is always true that $E(X+Y) = E(X)+E(Y)$, so the answer to (a) can be obtained more easily from the fact that $E(X) + E(Y) = 5.55 + 8.55 = 14.1$

(c) It is straightforward to show, after calculating the marginal distributions of X and Y, that $\sigma_X^2 = 12.45$ and $\sigma_Y^2 = 19.15$, and hence that

$$\rho_{XY} = \frac{-3.20}{\sqrt{(12.45)(19.15)}} = -0.207$$

9. (a) Use the rules of expected value to show that $\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$.
   (b) Use part (a) along with the rules of variance and standard deviation to show that $\text{Corr}(aX+b, cY+d) = \text{Corr}(X,Y)$ when a and c have the same sign.
   (c) What happens if a and c have opposite signs?

Solution:

(a)

$$\begin{aligned}
\text{Cov}(aX+b, cY+d) &= E[(aX+b)(cY+d)]–E(aX+b) \cdot E(cY+d) \\
&= E[acXY + adX + bcY + bd]–(aE(X)+b)(cE(Y)+d) \\
&= acE(XY)–acE(X)E(Y) \\
&= ac\text{Cov}(X,Y)
\end{aligned}$$

(b)

$$\begin{aligned}
\text{Corr}(aX+b, cY+d) &= \frac{\text{Cov}(aX+b, cY+d)}{\sqrt{\text{Var}(aX+b)}\sqrt{\text{Var}(cY+d)}} \\
&= \frac{ac\text{Cov}(X,Y)}{|a| \cdot |c|\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \\
&= \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}
\end{aligned}$$

because when a and c have the same sign, $|a| \cdot |c| = ac$.

(c) When a and c have different signs, $\operatorname{Corr}(aX + b, cY + d) = -\operatorname{Corr}(X, Y)$.

10. (a) Show that if X and Y are random variables with finite variances, and a and b are constants, then $\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)$.
    (b) What is $\operatorname{Var}(X - Y)$?
    (c) What is the expression for $\operatorname{Var}(aX + bY)$ when X and Y are independent?
    (d) What is $\operatorname{Var}(X - Y)$ when X and Y are independent?

Solution:

(a) For convenience, we denote $E(X)$ by μ_X and $E(Y)$ by μ_Y. Then, we can write $E(aX + bY) = a\mu_X + b\mu_Y$, and

$$\begin{aligned}
\operatorname{Var}(aX + bY) &= E[(aX + bY)^2] - (a\mu_X + b\mu_Y)^2 \\
&= E(a^2X^2 + b^2Y^2 + 2abXY) - (a^2\mu_X^2 + b^2\mu_Y^2 + 2ab\mu_X\mu_Y) \\
&= [E(a^2X^2) - a^2\mu_X^2] + [E(b^2Y^2) - b^2\mu_Y^2] \\
&\quad +[2abE(XY) - 2ab\mu_X\mu_Y] \\
&= a^2[E(X^2) - \mu_X^2] + b^2[E(Y^2) - \mu_Y^2] + 2ab[E(XY) - \mu_X\mu_Y] \\
&= a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)
\end{aligned}$$

(b) From (a), by taking $a = 1$ and $b = -1$, we get $\operatorname{Var}(X - Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X, Y)$.

(c) When X and Y are independent, $\operatorname{Cov}(X, Y) = 0$ and hence $\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y)$.

(d) From (c), by taking $a = 1$ and $b = -1$, we get $\operatorname{Var}(X - Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$ when X and Y are independent.

(*Sources: All questions adapted from Devore and Berk (2012) and Carlton and Devore (2017).*)

Bibliography

1. Carlton, M.A. and Devore, J.L. (2017) *Probability with Applications in Engineering, Science, and Technology*, 2nd ed. Springer: New York.

2. Devore, J.L. and Berk, K.N. (2012) *Modern Mathematical Statistics with Applications.* Springer: New York.