COMP 233 Winter 2016

PROBABILITY AND STATISTICS FOR COMPUTER SCIENCE

Assignment 2 Solutions

1. (a) Suppose 1 in 1000 persons has a certain disease. A test detects the disease in 95% of diseased persons. The test also "detects" the disease in 1% of healthy persons. With what probability does a positive test diagnose the disease?

   (b) Machines M_1, M_2, M_3 produce these proportions of a article: $M_1 : 5\%$, $M_2 : 25\%$, $M_3 : 70\%$. The probability the machines produce defective articles is $M_1 : 5\%$, $M_2 : 4\%$, $M_3 : 2\%$. What is the probability a random article was made by machine M_1, given that it is defective?

   (c) A machine M consists of three independent parts, M_1, M_2, and M_3. Suppose that M_1 functions properly with probability $\frac{9}{10}$, M_2 functions properly with probability $\frac{9}{10}$, M_3 functions properly with probability $\frac{8}{10}$, and that the machine M functions if and only if its three parts function. What is the probability for the machine M to malfunction?

   SOLUTION:

   (a) Given $P(D) = 0.001$, $P(+|D) = 0.95$, $P(+|D^c) = 0.01$. By Bayes' formula:

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|D^c) \cdot P(D^c)} = \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.01 \cdot 0.999} = 8.68\%$$

   (b) We are given that $P(M_1) = 0.05$, $P(M_2) = 0.25$, $P(M_3) = 0.70$, and $P(D|M_1) = 0.05$, $P(D|M_2) = 0.04$, $P(D|M_3) = 0.02$. Thus

$$\begin{aligned} P(M_1|D) &= \frac{P(D|M_1) \cdot P(M_1)}{P(D|M_1) \cdot P(M_1) + P(D|M_2) \cdot P(M_2) + P(D|M_3) \cdot P(M_3)} \\ &= \frac{0.05 \cdot 0.05}{0.05 \cdot 0.05 + 0.04 \cdot 0.25 + 0.02 \cdot 0.70} = 9.4\%. \end{aligned}$$

   (c) The machine functions with probability $\frac{9}{10} \cdot \frac{9}{10} \cdot \frac{8}{10} = 64.8\%$, and hence malfunctions with probability $1 - 0.648 = 35.2\%$.

2. Three balls are selected at random from a bag containing 3 red, 3 green, and 4 blue balls. Define the random variables $R =$ the number of red balls drawn, and $G =$ the number of green balls drawn. List the values of

   (a) the joint probability mass function $p_{R,G}(r, g)$.

   (b) the marginal probability mass functions $p_R(r)$ and $p_G(g)$.

   (c) the joint distribution function $F_{R,G}(r, g)$.

   (d) the marginal distribution functions $F_R(r)$ and $F_G(g)$.

SOLUTION:

(a), (b)

$p_{R,G}(r,g)$

	$g=0$	$g=1$	$g=2$	$g=3$	$p_R(r)$
$r=0$	$\frac{4}{120}$	$\frac{18}{120}$	$\frac{12}{120}$	$\frac{1}{120}$	$\frac{35}{120}$
$r=1$	$\frac{18}{120}$	$\frac{36}{120}$	$\frac{9}{120}$	0	$\frac{63}{120}$
$r=2$	$\frac{12}{120}$	$\frac{9}{120}$	0	0	$\frac{21}{120}$
$r=3$	$\frac{1}{120}$	0	0	0	$\frac{1}{120}$
$p_G(g)$	$\frac{35}{120}$	$\frac{63}{120}$	$\frac{21}{120}$	$\frac{1}{120}$	1

(c), (d)

$F_{R,G}(r,g)$

	$g=0$	$g=1$	$g=2$	$g=3$	$F_R(r)$
$r=0$	$\frac{4}{120}$	$\frac{22}{120}$	$\frac{34}{120}$	$\frac{35}{120}$	$\frac{35}{120}$
$r=1$	$\frac{22}{120}$	$\frac{76}{120}$	$\frac{97}{120}$	$\frac{98}{120}$	$\frac{98}{120}$
$r=2$	$\frac{34}{120}$	$\frac{97}{120}$	$\frac{118}{120}$	$\frac{119}{120}$	$\frac{119}{120}$
$r=3$	$\frac{35}{120}$	$\frac{98}{120}$	$\frac{119}{120}$	1	1
$F_G(g)$	$\frac{35}{120}$	$\frac{98}{120}$	$\frac{119}{120}$	1	1

3. For the preceding problem, also determine

   (a) The conditional probability mass functions $p_{R|G}$ and $p_{G|R}$. Are R and G independent?

   (b) $E[R]$ and $E[G]$

   (c) $Var(R)$ and $Var(G)$

   (d) $cov(R,G)$

SOLUTION:

(a)

$p_{R|G}(r)$

	$g=0$	$g=1$	$g=2$	$g=3$
$r=0$	$\frac{4}{35}$	$\frac{18}{63}$	$\frac{12}{21}$	1
$r=1$	$\frac{18}{35}$	$\frac{36}{63}$	$\frac{9}{21}$	0
$r=2$	$\frac{12}{35}$	$\frac{9}{63}$	0	0
$r=3$	$\frac{1}{35}$	0	0	0

$p_{G|R}(r)$

	$g=0$	$g=1$	$g=2$	$g=3$
$r=0$	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$
$r=1$	$\frac{18}{63}$	$\frac{36}{63}$	$\frac{9}{63}$	0
$r=2$	$\frac{12}{21}$	$\frac{9}{21}$	0	0
$r=3$	1	0	0	0

R and G are not independent.

(b) $E[R]=E[G] \;=\; \frac{35}{120}\cdot 0 \;+\; \frac{63}{120}\cdot 1 \;+\; \frac{21}{120}\cdot 2 \;+\; \frac{1}{120}\cdot 3 \;=\; \frac{108}{120}$.

(c) $E[R^2]=E[G^2] \;=\; \frac{35}{120}\cdot 0 \;+\; \frac{63}{120}\cdot 1 \;+\; \frac{21}{120}\cdot 4 \;+\; \frac{1}{120}\cdot 9 \;=\; \frac{156}{120}$.

$Var(R)=E[R^2]-(E[R])^2=0.49 \;=\; Var(G)$.

(d) $E[RG]=\frac{36}{120}\cdot 1 \;+\; \frac{9}{120}\cdot 2 \;+\; \frac{9}{120}\cdot 2 \;=\; \frac{72}{120}$.

$cov(R,G)=E[RG]-E[R]\cdot E[G] \;=\; \frac{72}{120} \;-\; (\frac{108}{120})^2 \;=-0.21$.

4. (a) A trial consists of tossing two dice. The result is counted as successful if the sum of the outcomes is 12. What is the probability that the number of successes in 36 such trials is greater than one? What is this probability if we approximate it using the Poisson random variable?

   (b) Customers arrive at a counter at the rate of 20 per hour. Assume the arrivals have a Poisson distribution. What is the probability that more than two customers arrive in a period of 5 minutes?

SOLUTION:

(a) The probability of success is $p = 1/36$. Let the random variable X measure the number of successes in 36 trials. We have

$$P(X = 0) = \binom{36}{0} p^0 (1-p)^{36} = (35/36)^{36} \text{ and } P(X = 1) = \binom{36}{1} p^1 (1-p)^{35} = (35/36)^{35}.$$

Thus $P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - (35/36)^{36} - (35/36)^{35} \cong 26.421\%$

Using the Poisson approximation to the Binomial we have, with $\lambda = np = 36/36 = 1$,

$$P(X = 0) = e^{-1}1^0/0! = e^{-1} \qquad \text{and} \qquad P(X = 1) = e^{-1}1^1/1! = e^{-1},$$

so that $P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - 2e^{-1} \cong 26.424\%$

(b) Here $\lambda = \frac{5}{60} \cdot 20 \ = \ 1.667$, and

$$P(X > 2) \ = \ 1 - P(X = 0) - P(X = 1) - P(X = 2) \ = \ 1 - e^{-\lambda}\frac{\lambda^0}{0!} - e^{-\lambda}\frac{\lambda^1}{1!} - e^{-\lambda}\frac{\lambda^2}{2!} \ = \ 23.4\%.$$

5. Approximately 20,000 marriages took place in Québec last year. Assuming that each person's birthday is equally likely to be any of the 365 days of the year, estimate the probability that for one or more of these couples:

   (a) both partners were born on April 1;

   (b) both partners celebrated their birthday on the same day of the year.

(Hint: The Poisson random variable can be used.)

SOLUTION:

(a) The probability that a person was born on April 1 is $\frac{1}{365}$. Thus by the multiplication rule, the probability that both partners were born on April 1 is $1/365^2$. So, on average, the number of couples out of the 20,000 that were born on April 1 is $20,000/365^2 \cong 0.15$. The occurrence of couples being born on April 1 can be assumed to be a Poisson distribution with mean $\lambda = 0.15$. Thus

$$P\{N_\lambda \geq 1\} \ = \ 1 - P\{N_\lambda = 0\} \ = \ 1 - e^{-0.15} \ \cong \ 0.14.$$

(b) The probability that both partners were born on a specific day is again $1/365^2$. The probability that both partners were born on some day of the year is $365/365^2 = \frac{1}{365}$. Thus, on average, the number of couples out of the 20,000 that were born on the same day is $20,000/365 \cong 54.8$. The occurrence of couples being born on the same day can be assumed to be a Poisson distribution with mean of $\lambda \cong 54.8$. Thus

$$P\{N_\lambda \geq 1\} \ = \ 1 - P\{N_\lambda = 0\} \ = \ 1 - e^{-54.8} \ \cong \ 1.$$

6. For the random variable X with density function

$$f(x) = \begin{cases} 4x\ , & 0 < x \le \frac{1}{2} \\ 4-4x\ , & \frac{1}{2} < x \le 1 \\ 0\ , & \text{otherwise} \end{cases}$$

(a) Determine the distribution function $F(x)$, and draw the graphs $f(x)$ and of $F(x)$.

(b) Determine $P(\frac{1}{3} < X \le \frac{1}{2})$.

(c) Determine $E[X]$.

(d) Determine $Var(X)$, and $\sigma(X)$.

SOLUTION:

(a) For $x \in [0, \frac{1}{2}]$: $F(x) = \int_0^x 4x\ dx = 2x^2|_0^x = 2x^2$.
For $x \in (\frac{1}{2}, 1]$: $F(x) = \frac{1}{2} + \int_{\frac{1}{2}}^x 4 - 4x\ dx = \frac{1}{2} + (4x - 2x^2)|_{\frac{1}{2}}^x = \frac{1}{2} + (4x - 2x^2)\ -\ (2 - \frac{1}{2}) = 4x - 2x^2 - 1$.

(b) $P(\frac{1}{3} < X \le \frac{1}{2}) = \frac{5}{18}$.

(c) $E[X] = \frac{1}{2}$.

(d) $Var(X) = \frac{1}{24}, \quad \sigma(X) = \frac{\sqrt{6}}{12}$.

7. For the random variables X, Y with joint density function

$$f(x,y) = \begin{cases} cxy^2(1-x)(1-y^2), & 0 \le x \le 1\ ,\ 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) For what value of c is this a joint density function?

(b) Using this value of c, compute the density functions of X and Y.

(c) What is the value of $Cov(X,Y)$?

(d) Determine $P\{X > Y\}$.

SOLUTION:

(a)

$$\begin{aligned} \int_0^1 \int_0^1 f(x,y)dy\ dx &= \int_0^1 \int_0^1 cxy^2(1-x)(1-y^2)dy\ dx \\ &= c \cdot \int_0^1 \left\{ x(1-x) \int_0^1 y^2(1-y^2)dy \right\} dx \\ &= c \cdot \int_0^1 x(1-x)dx \cdot \int_0^1 y^2(1-y^2)dy \\ &= c \cdot (\frac{x^2}{2} - \frac{x^3}{3})|_0^1 \cdot (\frac{y^3}{3} - \frac{y^5}{5})|_0^1 \\ &= c \cdot \frac{1}{6} \cdot \frac{2}{15}\ =\ \frac{c}{45}. \end{aligned}$$

Hence $c = 45$.

(b)

$$f_X(x) = \int_0^1 f(x,y)dy = \int_0^1 45xy^2(1-x)(1-y^2)dy$$
$$= 45x(1-x)\int_0^1 y^2(1-y^2)dy$$
$$= 45x(1-x)\cdot\frac{2}{15} = 6x(1-x).$$

(c)

$$f_Y(y) = \int_0^1 f(x,y)dx = \int_0^1 45xy^2(1-x)(1-y^2)dx$$
$$= 45y^2(1-y^2)\int_0^1 x(1-x)dx$$
$$= 45y^2(1-y^2)\cdot\frac{1}{6} = \frac{15}{2}y^2(1-y^2).$$

X and Y are independent since $f(x,y) = f_X(x)f_Y(y)$. Thus $Cov(X,Y) = 0$.

(d)

$$P\{X > Y\} = \int_0^1\int_0^x f(x,y)dy\ dx = 45\int_0^1\int_0^x xy^2(1-x)(1-y^2)dy\ dx = 45\int_0^1 x(1-x)\int_0^x y^2 - y^4 dy\ dx$$
$$= 45\int_0^1 x(1-x)(\frac{y^3}{3} - \frac{y^5}{5})|_0^x dx = 45\int_0^1 x(1-x)(\frac{x^3}{3} - \frac{x^5}{5})dx = 45\int_0^1 (\frac{x^4}{3} - \frac{x^6}{5} - \frac{x^5}{3} + \frac{x^7}{5})dx$$
$$= 45(\frac{x^5}{15} - \frac{x^7}{35} - \frac{x^6}{18} + \frac{x^8}{40})|_0^1 = 45(\frac{1}{15} - \frac{1}{35} - \frac{1}{18} + \frac{1}{40}) = \frac{19}{56} \cong .34$$

8. The side measurement of a die manufactured by a company is a random number X that is uniformly distributed between 1 and 1.25 cm. (You may assume the die is a perfect cube.)

   (a) Determine the distribution function of the volume of the die.

   (b) What is the probability that the volume of a randomly selected die manufactured by this company is greater than 1.424?

   SOLUTION:

   (a) Let X denote the side-length of the die. We know that X is a number chosen randomly from the interval $[1, 1.25]$. The distribution function of the volume $Y = X^3$ is $F(y) = P(Y \leq y)$. Thus we have to determine $F(y) = P(X^3 \leq y)$ for $y \in \mathbb{R}$. We know that $X^3 \leq y$ if and only if $X \leq y^{1/3}$, so $F(y) = P(X \leq y^{1/3})$. Since X is randomly chosen from $[1, 1.25]$, if $y < 1$, then $P(X \leq y^{1/3}) = 0$. For y such that $1 \leq y^{1/3} \leq 1.25$ we have

$$P(X \leq y^{1/3}) = P(X \in [1, y^{1/3}]) = \frac{y^{1/3} - 1}{1.25 - 1} = 4(y^{1/3} - 1).$$

   Finally, if $y^{1/3} > 1.25$, then $P(X \leq y^{1/3}) = 1$. Hence we obtain

$$F(y) = \begin{cases} 0 & : \quad y < 1 \\ 4(y^{1/3} - 1) & : \quad 1 \leq y \leq (1.25)^3 \\ 1 & : \quad y > (1.25)^3. \end{cases}$$

(b) The probability $P(X > 1.424)$ can be calculated as

$$P(Y > 1.424) \;=\; 1 - P(Y \leq 1.424) \;=\; 1 - F(1.424) \;=\; 1 - 4((1.414)^{1/3} - 1) \;\cong\; 0.5.$$

9. From past experience, a professor knows that the test score of students taking a final examination is a random variable with mean 65.

   (a) Give an upper bound on the probability that a student's test score will exceed 75.

   (b) Suppose in addition the professor knows that the variance of a student's test score is equal to 30. What can be said about the probability that a student will score between 55 and 75?

   (c) How many students would have to take the examination so as to ensure, with probability at least 0.8, that the class average would be within 5 of 65?

SOLUTION:

(a) An upper bound to the probability that a student's test score will exceed 85 is 65/75, namely, by Markov's inequality:

$$P\{X \geq 75\} \leq E[X]/75 = 65/75.$$

(b) The probability that a student will score between 55 and 75 is greater than or equal to 0.70 by Chebyshev's inequality:

$$P\{|X - 65| \geq 10\} \leq \frac{\sigma^2}{10^2} = 30/100,$$

or

$$P\{|X - 65| < 10\} > 1 - 0.30 = 0.70.$$

(c) With $\overline{X} = \frac{X_1+X_2+\ldots+X_n}{n}$ and knowing that

$$Var\left(\tfrac{X_1+X_2+\ldots+X_n}{n}\right) = Var\left(\tfrac{X_1}{n}\right) + Var\left(\tfrac{X_2}{n}\right) + \ldots + Var\left(\tfrac{X_n}{n}\right)$$
$$= \tfrac{1}{n^2}Var(X_1) + \tfrac{1}{n^2}Var(X_2) + \ldots + \tfrac{1}{n^2}Var(X_n) = \tfrac{30n}{n^2} = \tfrac{30}{n}.$$

then from Chebyshev's inequality,

$$P\{|\overline{X} - 65| \geq 5\} \leq \frac{Var(\overline{X})}{5^2} = \frac{30/n}{25} = \frac{6}{5n},$$

which is equal to 0.2 when $n = 6$. So $n = 6$ would suffice.

10. A stick of length 1 is split at a randomly selected point X, i.e., X is uniformly distributed in the interval $[0, 1]$. Determine the expected length of the piece that contains the point 1/3.

SOLUTION:

Let the function $L(X)$ denote the length of the piece that contains the point 1/3 :

$$L(x) = \begin{cases} 1 - x & \text{if } x < 1/3 \;, \\ x & \text{if } x \geq 1/3 \;. \end{cases}$$

Since the density function of X is $f(x) = 1$ if $x \in [0, 1]$ and $f(x) = 0$ otherwise, we have

$$E[L(X)] \;=\; \int_0^1 L(x)f(x)\,dx \;=\; \int_0^1 L(x)\,dx \;=\; \int_0^{1/3}(1 - x)\,dx + \int_{1/3}^1 x\,dx$$
$$=\; (x - \frac{x^2}{2})\Big|_0^{\frac{1}{3}} \;+\; \frac{x^2}{2}\Big|_{\frac{1}{3}}^1 \;=\; \frac{13}{18}\;.$$