Basic Math for Big Data

Solution for Final 2017

Question 1

(a) We define two symbols R, B_1 and B_2 which mean a red ball, and two blue balls, respectively.

1. $S = \{B_1B_2, B_2B_1, B_1R, B_2R, RB_1, RB_2\}$
2. $E = \{B_1R, B_2R, RB_1, RB_2\}$
3. $P(E) = |E|/|S| = 2/3$

(b) Let X be an event that the first flip came up heads, and Y be an event that exactly for heads appear. Then, $P(X) = 1/2$ and $P(X \cap Y) = 1/2^5 \cdot C(4,3) = 1/8$. Therefore, $P(Y|X) = 1/4$.

(c) Let S be the sample space, and E be the event space that the robot passes through the point (1, 2). Then, $|S|$ means the number of paths from (0, 0) to (3, 3). The robot can move through a diagonal up to 3 times. We obtain $|S|$ as follows:

$$|S| = \frac{6!}{3!\,3!} + \frac{5!}{2!\,2!} + \frac{4!}{2!} + 1 = 63$$

$|E|$ means the number of paths from (0, 0) to (3, 3) through the point (1, 2). By multiplying the number of paths from (0, 0) to (1, 2) and the number of paths from (1, 2) to (3, 3), we obtain $|E|$ as follows:

$$|\mathrm{E}| = \left(\frac{3!}{2!} + 2!\right)^2 = 25$$

Therefore, the probability that the robot passes through the point (1, 2) is $|E|/|S| = 25/63$.

Question 2

(a) 1.

x	$P(X = x)$	$x \cdot P(X = x)$
0	0.3	0
1	0.2	0.2
2	0.1	0.2
3	0.4	1.2

2. $E(X) = 1.6$

(b) Recall $E(X) = \int_p^q xf(x)dx$ where p and q are the lower and the upper boundary points in the domain. From the given $E(X) = 3/5$, we obtain

$$\begin{aligned} \mathrm{E(X)} &= \int_0^1 xf(x)dx = \int_0^1 x(a+bx^2)dx \\ &= \left[\frac{ax^2}{2} + \frac{bx^4}{4}\right]_0^1 \\ &= \frac{a}{2} + \frac{b}{4} \\ &= \frac{3}{5} \end{aligned}$$

Since $f(x)$ is a probability density function, $\int_p^q f(x)dx = 1$ where p and q are the lower and the upper boundary points in the domain. Therefore,

$$\begin{aligned} \int_0^1 f(x)dx &= \int_0^1 a + bx^2 \\ &= \left[ax + \frac{bx^3}{3}\right]_0^1 \\ &= a + \frac{b}{3} \\ &= 1 \end{aligned}$$

By solving the simultaneous equations, we obtain $a = 3/5$ and $b = 6/5$.

(c) Recall that we can convert $P(X > x) = P(Z > \frac{x-\mu}{\sigma})$.

1. $P(X > 16) = 1 - P(X \leq 16) = 1 - \left(0.5 + P\left(0 \leq Z \leq \frac{16-10}{6}\right)\right) = 1 - (0.5 + 0.34134) = 0.15866.$

2. $P(4 \leq X \leq 16) = P\left(\frac{4-10}{6} \leq Z \leq \frac{16-10}{6}\right) = P(-1 \leq Z \leq 1) = 2 \cdot P(0 \leq Z \leq 1) = 2 \cdot 0.34134 = 0.68268.$

Question 3

(a) 1. The distance y the driver can see

2. The age x of taxi drivers

3. $\mathrm{y = -3x + 576}$

4. No, as the age of driver decreases, the distance the driver can see increases.

(b) 1. The yield x_1

2. The level of two factors x_2 and x_3.

3. $\mathrm{x_1} = \alpha + \beta_2 x_2 + \beta_3 x_3$ where α is the value of intercept, β_2 and β_3 are the slopes for the two factors x_2 and x_3.

Question 4

(a) maximize (x+30-75) + (y+90-95) = (x+y-50), s.t.

50x + 24y <= 2400

30x + 33y <= 2100

x >= 75 – 30 = 45

y >= 95 – 90 = 5

(b) Solution: x = 45, y = 6.25, objective function value = 1.25