Quiz Week 3 - Expectation - Solutions

FIT5197 teaching team

Question 1

X is a discrete random variable over $\chi \in \{1, 2, 3\}$ with probability mass function $P(X = 1) = 0.5,\ P(X = 2) = 0.4,\ P(X = 3) = 0.1$

(a) What is $E[\ln X]$?

(b) What is $E\left[\ln\left(\frac{1}{P(X)}\right)\right]$?

Answer 1

(a). $E[\ln X] = \sum_{x\in\{1,2,3\}} \ln x \cdot P(X = x)$

$= \ln 1 \cdot P(X = 1) + \ln 2 \cdot P(X = 2) + \ln 3 \cdot P(X = 3)$

$= \ln 1 \times 0.5 + \ln 2 \times 0.4 + \ln 3 \times 0.1 = 0.3871$

(b). $E\left[\ln\left(\frac{1}{P(X)}\right)\right] = \sum_{x\in\{1,2,3\}} \ln\left(\frac{1}{P(X)}\right) \cdot P(X = x)$

$= \ln\left(\frac{1}{P(X=1)}\right) \cdot P(X = 1) + \ln\left(\frac{1}{P(X=2)}\right) \cdot P(X = 2) + \ln\left(\frac{1}{P(X=3)}\right) \cdot P(X = 3)$

$= \ln\left(\frac{1}{0.5}\right) \times 0.5 + \ln\left(\frac{1}{0.4}\right) \times 0.4 + \ln\left(\frac{1}{0.1}\right) \times 0.1 = 0.94$

Question 2

Consider the triangular distribution $p(x) = 1 - |x|$ defined on the interval $x \in [-1, 1]$ (Tip: draw $p(x)$)

(a) What is $E[X]$?

(b) What is $E[X^2]$?

Answer 2

(a). $E[X] = \int x \cdot p(x)dx = \int_{-1}^{1} x \cdot (1 - |x|)dx = \int_{-1}^{0} x \cdot (1 + x)dx + \int_{0}^{1} x \cdot (1 - x)dx$

Rather than integrate using usual methods, i.e.

$\int_{a}^{b} x^n dx = \left[\frac{x^{n+1}}{n+1} \right]_{a}^{b}$ (1),

we can use substitution with $x^{'} = -x$. if $x = -1$, then $x^{'} = 1$, if $x = 0$, then $x^{'} = 0$, and $dx^{'} = -dx$.

Also, we have $\int_{a}^{b} f(x)dx = -\int_{b}^{a} f(x)dx$ (2).

Applying these, we get:

$$\begin{aligned} E[X] &= \int_{1}^{0} -x^{'} \cdot (1 - x^{'}) \cdot (-dx^{'}) + \int_{0}^{1} x \cdot (1 - x) \cdot dx \\ &= -\int_{0}^{1} x^{'} \cdot (1 - x^{'}) \cdot dx^{'} + \int_{0}^{1} x \cdot (1 - x) \cdot dx = 0 \end{aligned}$$

(b). $E[X^2] = \int x^2 \cdot p(x)dx = \int_{-1}^{1} x^2 \cdot (1 - |x|)dx = \int_{-1}^{0} x^2 \cdot (1 + x)dx + \int_{0}^{1} x^2 \cdot (1 - x)dx$.

Apply the substitution

$x^{'} = -x$, if $x = -1$, then $x^{'} = 1$, if $x = 0$, then $x^{'} = 0$, and $dx^{'} = -dx$, so:

$$
\begin{aligned}
E[X^2] &= \int_1^0 x'^2 \cdot (1-x') \cdot (-dx') + \int_0^1 x^2 \cdot (1-x)dx \\
&= -\int_1^0 x'^2 \cdot (1-x') \cdot dx' + \int_0^1 x^2 \cdot (1-x)dx \\
&= \int_0^1 x'^2 \cdot (1-x')dx' + \int_0^1 x^2 \cdot (1-x)dx \quad using\ (2) \\
&= 2 \times \int_0^1 x^2 \cdot (1-x)dx = 2 \times \int_0^1 (x^2 - x^3)dx \\
&= 2 \times \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 \qquad using\ (1) \\
&= 2 \times \left(\frac{1}{3} - \frac{1}{4}\right) - 2 \times (0) = \frac{1}{6}
\end{aligned}
$$

Question 3

$P(X=1) = 0.5$, $P(X=2) = 0.4$, $P(X=3) = 0.1$

What is $V[X]$?

Answer 3

$$V[X] = E[(X - E[X])^2] = \sum_{x \in \chi} (x - E[X])^2 \cdot P(X=x) = \sum_{x \in \{1,2,3\}} (x - E[X])^2 \cdot P(X=x)$$

First we need to find $E[X]$

$$E[X] = \sum_x x \cdot P(X=x) = 1 \times 0.5 + 2 \times 0.4 + 3 \times 0.1 = 1.6,$$

so:

$$V[X] = (1-1.6)^2 \times 0.5 + (2-1.6)^2 \times 0.4 + (3-1.6)^2 \times 0.1 = 0.44$$

Question 4

Show $V[X] = E[X^2] - E[X]^2$

Answer 4

Assume A,B are random variables, c is a constant, we have known:

$$\begin{aligned}
&E[A+B] = E[A] + E[B] \\
&E[c \cdot A] = c \cdot E[A] \\
&E[A \cdot E[B]] = E[A] \cdot E[B], \\
&E[X]^2 \textit{ is a constant and } E[c] = c
\end{aligned}$$

Start with definition:

$$\begin{aligned}
V[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2X \cdot E[X] + E[X]^2] \\
&= E[X^2] + E[-2X \cdot E[X]] + E[E[X]^2] \\
&= E[X^2] - 2 \cdot E[X \cdot E[X]] + E[E[X]^2] \\
&= E[X^2] - 2 \cdot E[X] \cdot E[X] + E[E[X]^2] \\
&= E[X^2] - 2 \cdot E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

Question 5

$P(X = x,\ Y = y)$ is defined as

	X = 1	X = 2	X = 3
Y = 1	0.05	0.15	0.1
Y = 2	0.25	0.15	0.3

What is $cov(X, Y)$?

Answer 5

$cov(X, Y) = E[XY] - E[X] \cdot E[Y]$

First we need $E[X]$ and $E[Y]$. To get this we need the marginal probabilities $P(X = x), P(Y = y)$.

From the table above we have $P(X = 1) = 0.3$, $P(X = 2) = 0.3$, $P(X = 3) = 0.4$, and $P(Y = 1) = 0.3, P(Y = 2) = 0.7$.

So $E[X] = \sum_x x \cdot P(X = x) = 1 \times 0.3 + 2 \times 0.3 + 3 \times 0.4 = 2.1$

and $E[Y] = \sum_y y \cdot P(Y = y) = 1 \times 0.3 + 2 \times 0.7 = 1.7$.

Now we can find:

$$\begin{aligned}
cov(X, Y) = E[XY] - E[X] \cdot E[Y] \\
&= \sum_x \sum_y xy \cdot P(X = x, Y = y) - (2.1 \times 1.7) \\
&= 1 \times 1 \cdot P(X = 1, Y = 1) + 2 \times 1 \cdot P(X = 2, Y = 1) + 3 \times 1 \cdot P(X = 3, Y = 1) \\
&\quad + 1 \times 2 \cdot P(X = 1, Y = 2) + 2 \times 2 \cdot P(X = 2, Y = 2) + 3 \times 2 \cdot P(X = 3, Y = 2) - (2.1 \times 1.7) \\
&= 1 \times 0.05 + 2 \times 0.15 + 3 \times 0.1 + 2 \times 0.25 + 4 \times 0.15 + 6 \times 0.3 - (2.1 \times 1.7) \\
&= -0.02
\end{aligned}$$

Question 6

The wealth of an individual is random variable with probability density function:

$$f(x) = \frac{C}{x^{\alpha+1}}, \; x \; \in [2, \infty), \; \alpha > 1$$

Furthermore you are given the following integral:

$$\int_2^\infty \frac{1}{x^n} = \begin{cases} \infty, & if \; n \leq 1 \\ \frac{2^{1-n}}{n-1}, & if \; n > 1 \end{cases}$$

(a) What is the value of C to make the distribution normalise to 1?

(b) What is the mean of x?

Answer 6

(a). $C = \alpha 2^\alpha$ when using $n = \alpha + 1$ in the integral above (true for $\alpha > 0$) although we state $\alpha > 1$ above.

(b). Using the integral again for $n = \alpha$ (since we multiply $f(x)$ by x) and the value determined for C gives the mean of $\frac{2\alpha}{\alpha - 1}$ when $\alpha > 1$ (note it's ∞ otherwise but we say α >1 above).

Question 7

Let $E[Z] = 1$ and $E[Z^2] = 6$, $E[Y] = -2$ and $E[Y^2] = 5$, and Z and Y are independent, then what is $V[3Z + 2Y]$?

Answer 7

Compute as $E[(3Z + 2Y)^2] - E[(3Z + 2Y)]^2$.

$$\begin{aligned} E[(3Z + 2Y)^2] &= E[9Z^2 + 12ZY + 4Y^2] \\ &= 9E[Z^2] + 12E[Z]E[Y] + 4E[Y^2] \\ &= 9 \times 6 + 12 \times 1 \times (-2) + 4 \times 5 = 50 \end{aligned}$$

$$\begin{aligned} &E[(3Z + 2Y)] = 3E[Z] + 2E[Y] = 3 \times 1 + 2 \times (-2) = -1 \\ &V[3Z + 2Y] = 50 - 1 = 49 \end{aligned}$$

Question 8

In a lottery a four-digit number is chosen at random from the range 0000-9999. A lottery ticket costs \$2. You win \$50 if your ticket matches the last two digits but not the last three, \$500 if your ticket matches the last three digits but not last four, and \$5,000 if your ticket matches all four digits. What is the expected payoff on a lottery ticket? How much money does the lotto make on average per ticket sold?

Answer 8

Let the random variable X be the payoff (expressed in dollars) on a lottery ticket. The random variable X takes on the values 0, 50, 500 and 5,000 with respective probabilities.

$$P(X = 0) = \frac{1 \times 9 \times 10 \times 10 + 9 \times 10 \times 10 \times 10}{10,000} = \frac{99}{100},$$
$$P(X = 50) = \frac{1 \times 1 \times 9 \times 10}{10,000} = \frac{9}{1,000},$$
$$P(X = 500) = \frac{1 \times 1 \times 1 \times 9}{10,000} = \frac{9}{10,000},$$
$$P(X = 5,000) = \frac{1}{10,000}.$$

This gives:

$$E(X) = 0 \times \frac{99}{100} + 50 \times \frac{9}{1,000} + 500 \times \frac{9}{10,000} + 5,000 \times \frac{1}{10,000} = 1.4$$

So lotto makes $\$2 - \$1.4 = \$0.6$ per ticket.

R hackers mini power punch challenge

Using numerical integration in R with the built-in ‘integrate’ function, calculate the expected value of the mean and variance for a normal distribution with mean 2 and standard deviation 2.

In [1]:
```
#integrate takes first argument as function to integrate, followed by the limits for the integration
#... signifies arbitrary number of arguments to the function. This gives flexibility to calculate mean and varian
#for any probability distribution function. Try running this method for different PDFs by providing the relevant 
find_mean_variance <- function(f, ...){
    #mean is expected value of x which is integration of x*f(x) for -infinity to infinity
    exp_x <- round(integrate(function(x) x*f(x, ...), -Inf, Inf)$value)
    #variance is expected value of x^2 - square of expected value of x.
    exp_x2 <- round(integrate(function(x) x^2*f(x, ...), -Inf, Inf)$value)
    return(c(mean=exp_x, variance=(exp_x2 - exp_x^2)))
}
```

In [2]:
```
#dnorm gives PDF for normal distribution with mean 2 and sd 2 (passed as arguments)
find_mean_variance(dnorm, mean=2, sd=2)
```

mean
2
variance
4