Statistics I: Chapter 5 - Expected values of functions of bi-dimensional random variables 23/04/2020

Carlos Oliveira

E-mail: carlosoliveira@iseg.ulisboa.pt

ISEG - Lisbon School of Economics and Management

Let (X, Y) be a two-dimensional random variable and $D_{(X,Y)}$ the set of points of discontinuity of the joint cumulative distribution function $F_{X,Y}(x, y)$.

Definition: Let $g(X, Y)$ be a function of the two-dimensional random variable (X, Y). Then, the expected value of $g(X, Y)$ is given by

- (X, Y) is a two-dimensional discrete random variable:

$$E[g(X, Y)] = \sum_{(x,y)\in D_{(X,Y)}} g(x, y) f_{X,Y}(x, y)$$

  provided that $\sum_{(x,y)\in D_{(X,Y)}} |g(x, y)| f_{X,Y}(x, y) < +\infty$.

- (X, Y) is a two-dimensional continuous random variable:

$$E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dxdy$$

  provided that $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |g(x, y)| f(x, y) dxdy < +\infty$.

Example: Let (X, Y) be a discrete bidimensional random variable such that

$$f_{X,Y}(x,y) = \begin{cases} x, & 0 < x < 1, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

Compute the expected value of $g(X, Y) = X + Y$.
Answer: Using the definition of expected value, one gets

$$\begin{aligned} E(X+Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y) f_{X,Y}(x,y) dxdy \\ &= \int_0^2 \int_0^1 x(x+y) dxdy \\ &= \int_0^2 \frac{1}{3} + \frac{y}{2} dy = \frac{5}{3} \end{aligned}$$

Theorem: Let (X, Y) be a discrete two-dimensional random variable with joint probability function $f_{X,Y}(x,y)$:

1. If $g(X,Y) = h(X)$ that is $g(X,Y)$ only depends on X , then

$$E(g(X,Y)) = E[h(X)] = \sum_{(x,y)\in D_{(X,Y)}} h(x) f_{X,Y}(x,y)$$
$$= \sum_{x\in D_X} h(x) \sum_{y\in D_Y} f_{X,Y}(x,y) = \sum_{x\in D_X} h(x) f_X(x)$$

provided that $\sum_{(x,y)\in D_{(X,Y)}} |h(x)| f_{X,Y}(x,y) < +\infty$.

2. If $g(X,Y) = v(Y)$ that is $g(X,Y)$ only depends on Y , then

$$E[v(Y)] = \sum_{(x,y)\in D_{(X,Y)}} v(y) f_{X,Y}(x,y)$$
$$= \sum_{y\in D_Y} v(y) \sum_{x\in D_X} f_{X,Y}(x,y) = \sum_{y\in D_Y} v(y) f_Y(y)$$

provided that $\sum_{(x,y)\in D_{(X,Y)}} |v(y)| f_{X,Y}(x,y) < +\infty$.

Example: Let (X, Y) be a two-dimensional random variable such that

$$f_{X,Y} = \begin{cases} \frac{1}{5}, & x = 1, 2,\ y = 0, 1, 2,\ y \leq x \\ 0, & \text{otherwise} \end{cases}.$$

Compute the expected value of X.

Solution:

(i) By using the joint probability function:

$$E(X) = \sum_{(x,y) \in D_{(X,Y)}} x f_{X,Y}(x, y) = \sum_{x=1}^{2} \sum_{y=0}^{x} \frac{1}{5} x$$
$$= \frac{8}{5}$$

Example: Let (X, Y) be a two-dimensional random variable such that

$$f_{X,Y} = \begin{cases} \frac{1}{5}, & x = 1, 2, \; y = 0, 1, 2, \; y \leq x \\ 0, & \text{otherwise} \end{cases}.$$

Compute the expected value of X.

Solution:

(ii) By using the marginal function:

$$f_X(x) = \sum_{y=0}^{x} f_{X,Y}(x, y) = \begin{cases} \frac{2}{5}, & x = 1 \\ \frac{3}{5}, & x = 2 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore,

$$E(X) = \sum_{x=1}^{2} x f_X(x) = 1 \times \frac{2}{5} + 2 \times \frac{3}{5} = \frac{8}{5}.$$

Theorem: Let (X, Y) be a continuous two-dimensional random variable with joint probability function $f_{X,Y}(x, y)$:

- If $g(X, Y) = h(X)$ that is $g(X, Y)$ only depends on X , then

$$E[h(X)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x) f_{X,Y}(x, y) dxdy$$
$$= \int_{-\infty}^{+\infty} h(x) \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy\right) dx = \int_{-\infty}^{+\infty} h(x) f_X(x)\, dx$$

provided that $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |h(x)|\, f_{X,Y}(x, y) dxdy < +\infty$.

- If $g(X, Y) = v(Y)$ that is $g(X, Y)$ only depends on Y , then

$$E[v(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} v(y) f_{X,Y}(x, y) dxdy$$
$$= \int_{-\infty}^{+\infty} v(y) \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx\right) dy = \int_{-\infty}^{+\infty} v(Y) f_Y(y)\, dy$$

provided that $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |v(y)|\, f_{X,Y}(x, y) dxdy < +\infty$.

Example: Let (X, Y) be a discrete bidimensional random variable such that

$$f_{X,Y}(x,y) = \begin{cases} x, & 0 < x < 1, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

Compute the expected value of $3X + 2$.

Answer:

(i) Using the joint density function.

Using the definition of marginal expected value, one gets

$$E(3X+2) = 3E(X) + 2 = 3\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{X,Y}(x,y)dxdy + 2$$

$$= 3\int_{0}^{2}\int_{0}^{1} x^2 dxdy + 2$$

$$= 3\int_{0}^{2} \frac{1}{3} dy + 2 = 4$$

Example: Let (X, Y) be a discrete bidimensional random variable such that

$$f_{X,Y}(x,y) = \begin{cases} x, & 0 < x < 1, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

Compute the expected value of $3X + 2$.

Answer:

(ii) Using the marginal density function.

The marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, $E(3X + 2) = 3E(X) + 2 = 4$, because

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \int_0^1 2x^2 dx = \frac{2}{3}$$

Properties:

1. $E\left[h(X)+v(Y)\right]=E\left[h(X)\right]+E\left[v(Y)\right]$ provided that $E\left[|h(X)|\right]<+\infty$, $E\left[|v(Y)|\right]<+\infty$
2. $E\left[\sum_{i=1}^{N}X_i\right]=\sum_{i=1}^{N}E\left[X_i\right]$, where N is a finite integer, provided that $E\left[|X_i|\right]<+\infty$ for $i=1,2,...,N$.

Example: Let (X,Y) be a discrete bidimensional random variable such that

$$f_{X,Y}(x,y)=\begin{cases} x, & 0<x<1, 0<y<2 \\ 0, & \text{otherwise} \end{cases}$$

Compute the expected value of Y.

Answer: We know that $E(X+Y)=E(X)+E(Y)=\frac{5}{3}$. Since $E(X)=\frac{2}{3}$, then we get that $E(Y)=1$.

Definition: The r th and s th moment of products about the origin of the random variables X and Y, denoted by $\mu'_{r,s}$ is the expected value of $X^r Y^s$, for $r = 1, 2, ...;\ s = 1, 2, ...$ which is given by

- if X and Y are discrete random variables:

$$\mu'_{r,s} = E[X^r Y^s] = \sum_{(x,y)\in D_{(X,Y)}} x^r y^s f_{X,Y}(x,y)$$

- if X and Y are continuous random variables:

$$\mu'_{r,s} = E[X^r Y^s] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^r y^s f(x,y) dxdy$$

Remarks:

- If $r = s = 1$, we have $\mu'_{1,1} = E[XY]$
- *Cauchy-Schwarz Inequality*: For any two random variables X and Y, we have $|E[XY]| \leq E\left[X^2\right]^{1/2} E\left[Y^2\right]^{1/2}$ provided that $E[|XY|]$ is finite.
- If X and Y are independent random variables, $E[h(X)v(Y)] = E(h(X))E(v(Y))$ for any two functions $h(X)$ and $v(Y)$.
  [**Warning:** The reverse is not true.]
- If $X_1, X_2, ..., X_n$ are independent random variables independent, $E[X_1X_2...X_n] = E(X_1)E(X_2)...E(X_n)$.
  [**Warning:** The reverse is not true.]

Definition: The r th and s th moment of products about the mean of the discrete random variables X and Y, denoted by $\mu_{r,s}$ is the expected value of $(X-\mu_X)^r(Y-\mu_Y)^s$, for $r=1,2,...;\ s=1,2,...$ which is given by

$$\begin{aligned}\mu_{r,s} &= E[(X-\mu_X)^r(Y-\mu_Y)^s] \\ &= \sum_{(x,y)\in D_{(X,Y)}} (x-\mu_X)^r(y-\mu_Y)^s f_{X,Y}(x,y)\end{aligned}$$

Definition: The r th and s th moment of products about the mean of the continuous random variables X and Y, denoted by $\mu_{r,s}$, for $r=1,2,...;\ s=1,2,...$ is given by

$$\begin{aligned}\mu_{r,s} &= E[(X-\mu_X)^r(Y-\mu_Y)^s] \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)^r(y-\mu_Y)^s f(x,y)dxdy\end{aligned}$$

The covariance is a measure of the joint variability of two random variables. Formally it is defined as

$$Cov(X, Y) = \sigma_{XY} = \mu_{1,1} = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

How can we interpret the covariance?

- When the variables tend to show similar behavior, the covariance is positive:
  - If high (small) values of one variable mainly correspond to high (small) values of the other variable;
- When the variables tend to show opposite behavior, the covariance is negative:
  - When high (small) values of one variable mainly correspond to low (high) values of the other;
- If there is no linear association, then the covariance will be zero.

Properties:

- $Cov\,(X,Y) = E(XY) - E(X)E(Y)$.
- If X and Y are independent $Cov\,(X,Y) = 0$.
- If $Y = bZ$, where b is constant,

$$Cov\,(X,Y) = bCov\,(X,Z)\,.$$

- If $Y = V + W$,

$$Cov\,(X,Y) = Cov\,(X,V) + Cov\,(X,W)\,.$$

- If $Y = b$, where b is constant,

$$Cov\,(X,Y) = 0.$$

- If follows from the Cauchy-Schwarz Inequality that $|Cov\,(X,Y)| \leq \sqrt{Var\,(X)\,Var\,(Y)}$.

The covariance has the inconvenient of depending on the scale of both random variables. For what values of the covariance can we say that there is a strong association between the two random variables? The correlation coefficient is a measure of the joint variability of two random variables that do not depend on the scale:

$$\rho_{X,Y} = \frac{Cov\,(X,Y)}{\sqrt{Var\,(X)\,Var\,(Y)}}.$$

Properties:

- If follows from the Cauchy-Schwarz Inequality that $-1 \leq \rho_{X,Y} \leq 1$.

If $Y = bX + a$, where b and a are constants

- $\rho_{X,Y} = 1$ if $b > 0$.
- $\rho_{X,Y} = -1$ if $b < 0$.
- If $b = 0$, it is not defined.

Summary of important results:

- If $Y = V \pm W$,

$$Var\left(Y\right) = Var\left(V\right) + Var\left(W\right) \pm 2Cov\left(V, W\right).$$

- If $X_1, ..., X_n$ are random variables and $a_1, ..., a_n$ are constants and $Y = \sum_{i=1}^{n} a_i X_i$, then

$$Var\left(Y\right) = \sum_{i=1}^{n} a_i^2 Var\left(X_i\right) + 2\underbrace{\sum_{i=1}^{n} \sum_{j=1, j<i}^{n} a_i a_j Cov(X_i, X_j)}_{=0,\text{ if } X_i, X_j \text{ are independent}}.$$

- If $X_1, ..., X_n$ are random variables, $a_1, ..., a_n$ are constants and $b_1, ..., b_n$ are constants, $Y_1 = \sum_{i=1}^{n} a_i X_i$, and $Y_2 = \sum_{i=1}^{n} b_i X_i$ then

$$Cov\left(Y_1, Y_2\right) = \sum_{i=1}^{n} a_i b_i Var\left(X_i\right) + \underbrace{\sum_{i=1}^{n} \sum_{j=1, j<i}^{n} \left(a_i b_j + a_j b_i\right) Cov(X_i, X_j)}_{=0,\text{ if } X_i, X_j \text{ are independent}}.$$

Definition: Let (X, Y) be a two dimensional random variable and $u(Y, X)$ a function of Y and X. Then, the conditional expectation of $u(Y, X)$ given $X = x$, is given by

- if X and Y are discrete random variables

$$E\left[u(Y,X)|X=x\right] = \sum_{y \in D_Y} u(y,x) f_{Y|X=x}(y)$$

  where D_Y is the set of discontinuity points of $F_Y(y)$ and $f_{Y|X=x}(y)$ is the value of the conditional probability function of Y given $X = x$ at y

- if X and Y are continuous random variables

$$E\left[u(Y,X)|X=x\right] = \int_{-\infty}^{+\infty} u(y,x) f_{Y|X=x}(y) dy$$

  where $f_{Y|X=x}(y)$ is the value of the conditional probability density function of Y given $X = x$ at y.

provided that the expected values exist and are finite.

Remarks:

1. If $u(Y,X) = Y$, then we have the *conditional mean* of Y, $E\left[u(Y,X)|X=x\right] = E\left[Y|X=x\right] = \mu_{Y|x}$ (notice that this is a function of x).
2. If $u(Y,X) = \left(Y - \mu_{Y|x}\right)^2$, then we have the *conditional variance* of Y

$$\begin{aligned} E\left[u(Y,X)|X=x\right] &= E\left[\left(Y-\mu_{Y|x}\right)^2|X=x\right] \\ &= E\left[\left(Y - E\left[u(Y)|X=x\right]\right)^2|X=x\right] \\ &= Var\left[Y|X=x\right] \end{aligned}$$

3. As usual, $Var\left[Y|X=x\right] = E\left[Y^2|X=x\right] - E\left[Y|X=x\right]^2$.
4. If Y and X are independent $E(Y|X=x) = E(Y)$.
5. Of course we can reverse the roles of Y and X, that is we can compute $E\left(u\left(X,Y\right)|Y=y\right)$, using definitions similar to those above.

Example: Let (X, Y) be two-dimensional random variable such that

$$f_{X,Y}(x,y) = \begin{cases} 1/2, & 0 < x < 2, 0 < y < x \\ 0, & \text{c.c.} \end{cases}.$$

Then the conditional density function of $Y|X = 1$ is given by

$$f_{Y|X=1}(y) = \begin{cases} \frac{f_{X,Y}(1,y)}{f_X(1)}, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases} = \begin{cases} \frac{1/2}{1/2}, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$
$$= \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$

where

$$f_X(x) = \begin{cases} \int_0^x f_{X,Y}(x,y)dy, & 0 < x < 2 \\ 0, & \text{c.c.} \end{cases} = \begin{cases} \frac{x}{2}, & 0 < x < 2 \\ 0, & \text{c.c.} \end{cases}.$$

Example: The conditional expected value can be computed as follows:

$$E(Y|X=1)=\int_0^1 y f_{Y|X=1}(y)dy=\int_0^1 y dy=\frac{1}{2}.$$

To compute the conditional variance, one may start by computing the following conditional expected value

$$E(Y^2|X=1)=\int_0^1 y^2 f_{Y|X=1}(y)dy=\int_0^1 y^2 dy=\frac{1}{3}.$$

Therefore

$$Var(Y|X=1)=E(Y^2|X=1)-(E(Y|X=1))^2$$
$$=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$$

Example: Let X and Y be two random variables such that

$$f_{X,Y}(x,y) = \frac{1}{9}, \text{ for } x = 1,2,3,\ y = 0,1,2,3,\ y \leq x$$

To compute the conditional expected value one has to compute the condition probability function:

$$f_{Y|X=1}(y) = \begin{cases} \frac{f_{X,Y}(1,Y)}{f_X(1)}, & y = 0,1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{2}, & y = 0,1 \\ 0, & \text{otherwise} \end{cases}$$

where

$$f_X(1) = \sum_{y=0}^{1} f_{X,Y}(1,y) = \sum_{y=0}^{1} \frac{1}{9} = \frac{2}{9}$$

Therefore,

$$E(Y|X=1) = \sum_{y \in D_Y} y f_{Y|X=1}(y) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}.$$

Notice that $g(y) = E(X|Y = y)$ is indeed a function of y. Therefore, $g(Y)$ is a random variable because Y can take different values according its distribution, i.e, if Y can take the value y, then $g(Y)$ can take $g(y)$ with probability $P(Y = y) > 0$.

- Discrete random variables

The random variable $Z = g(Y) = E(X|Y)$ takes the values $g(y) = E(X|Y = y)$. **Assume that all values of $g(y)$ are different.** Then,

$$Z \text{ takes the value } g(y) \text{ with probability } P(Y = y)$$

In general, the probability function of $Z = g(Y) = E(X|Y)$ can be computed in the following way

$$P(Z = z) = P(g(Y) = z) = P(Y \in \{y : g(y) = z\})$$

Example: Let (X, Y) be a discrete random variable such that $f_{X,Y}(x, y)$ is represented in the following table

X/Y	1	2	3
0	0.2	0.1	0.15
1	0.05	0.35	0.15

One may compute the following conditional probability functions:

$$f_{Y|X=0} = \begin{cases} 4/9, & y = 1 \\ 2/9, & y = 2 \\ 3/9, & y = 3 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad f_{Y|X=1} = \begin{cases} 1/11, & y = 1 \\ 7/11, & y = 2 \\ 3/11, & y = 3 \\ 0, & \text{otherwise} \end{cases} .$$

Consequently, $E(Y|X = 0) = 17/9$ and $E(Y|X = 1) = 24/11$. Therefore, the random variable $Z = E(Y|X)$ has the following probability function

$$P(Z = z) = \begin{cases} P(X = 0), & z = 17/9 \\ P(X = 1), & z = 24/11 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 0.45, & z = 17/9 \\ 0.55, & z = 24/11 \\ 0, & \text{otherwise} \end{cases} .$$

- Continuous random variables

The cumulative distribution function of $Z = g(Y) = E(X|Y)$ is, indeed

$$F_Z(z) = P(Z \leq z) = P(g(Y) \leq z) = P(Y \in \{y : g(y) \leq z\})$$

When g is an injective function, we get that

$$F_Z(z) = F_Y(g^{-1}(z)) \text{ or } F_Z(g(y)) = F_Y(y).$$

Therefore, we can calculate all the quantities that we know (the expected value, variance, ...) for $E(X|Y)$ or $E(Y|X)$

Theorem (*Law of iterated Expectations*) Let (X, Y) be a two dimensional random variable. Then, $E(Y) = E(E[Y|X])$ provided that $E(|Y|)$ is finite and $E(X) = E(E[X|Y])$ provided that $E(X)$ is finite.

Remark: This theorem shows that there are two ways to compute $E(Y)$ (resp., $E(X)$). The first is the direct way. The second way is to consider the following steps:

1. compute $E[Y|X = x]$ and notice that this is a function solely of x that is we can write $g(x) = E[Y|X = x]$,
2. according to the theorem replacing $g(x)$ by $g(X)$ and taking the mean we obtain $E[g(X)] = E[Y]$ for this specific form of $g(X)$.
3. This theorem is useful in practice in the calculation of $E(Y)$ if we know $f_{Y|X=x}(y)$ or $E[X|X = x]$ and $f_X(x)$ (or some moments of X), but not $f_{X,Y}(x, y)$.

Remarks: The results presented can be generalized for functions of X and Y, i.e., $E(u(X, Y)) = E(E(u(X, Y)|X))$, if $E(u(X, Y))$ exists.

Example: Let (X, Y) be a bi-dimensional continuous random variable such that

$$E(X|Y=y) = \frac{3y-1}{3} \quad \text{and} \quad f_Y(y) = \begin{cases} 1/2, & 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

Taking into account the previous theorem,

$$E(X) = E(E(X|Y)) = E\left(\frac{3Y-1}{3}\right) = \int_0^2 \frac{3y-1}{6} dy = 2/3.$$

Theorem: Assuming that $E(Y^2)$ exists then

$$Var(Y) = Var[E(Y|X)] + E[Var[Y|X]].$$

Theorem: Let X and Y be two random variables then

$$Cov(X, Y) = Cov(X, E(Y|X))$$

Example: Let (X, Y) be a bidimensional random variable such that

$$f_{X|Y=y}(x) = \frac{1}{y}, \quad 0 < x < y \quad \text{(for a fixed} y > 1)$$
$$f_Y(y) = 3y^{-4}, \quad y > 1$$

Compute $Var(X)$ using the previous theorem.

Exam question: Let X and Y be two random variables such that

$$E(X|Y = y) = y,$$

for all y such that $f_Y(y) > 0$. Prove that $Cov(X, Y) = Var(Y)$. Are the random variables independent? Justify your answer.