"A human is a random variable, with a hope to find a function to let his life have some value."

Jinesh Parakh

11

Pairs of random variables II

Functions of two random variables

- for discrete random variables X, Y the random variable $W = g(X,Y)$ has the PMF

$$P_W(w) = \sum_{(x,y):\ g(x,y)=w} P_{X,Y}(x,y)$$

- for continuous random variables X and Y, the CDF of $W = g(X,Y)$ is

$$F_W(w) = P(W \le w) = \iint_{g(x,y)\le w} f_{X,Y}(x,y)\ dxdy$$

- in particular, the CDF of $W = max(X,Y)$ is

$$F_W(w) = \int_{-\infty}^{w}\int_{-\infty}^{w} f_{X,Y}(x,y)\ dxdy$$

- the expected valued of $W = g(X,Y)$ is

$$E(W) = \sum_{x\in S_X}\sum_{y\in S_Y} g(x,y)P_{X,Y}(x,y)$$

in the discrete case and

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

in the continuous case.

In general

$$E(X+Y) = E(X) + E(Y)$$

for any random variables X, Y, but

$$var(X+Y) = var(X) + var(Y) + 2 \cdot cov(X,Y)$$

where $cov(X,Y) = E(XY) - E(X)E(Y)$, with

$$var(X+Y) = var(X) + var(Y)$$

in the case when X, Y are independent

Conditioning by an event

• for any event B, a region of the X, Y plane with $P(B) > 0$ the conditional joint probability mass function of X, Y given B is:

$$P_{X,Y\big|B} = \begin{cases} \frac{P_{X,Y}(x,y)}{P(B)}, & \text{if } (x,y) \in B \\ 0, & \text{otherwise} \end{cases}$$

• the conditional joint probability density function of X and Y is:

$$f_{X,Y\big|B} = \begin{cases} \frac{f_{X,Y}(x,y)}{P(B)}, & \text{if } (x,y) \in B \\ 0, & \text{otherwise} \end{cases}$$

• the conditional expected value of $W = g(X,Y)$ given B is:

$$E(W\big|B) = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) P_{X,Y\big|B}(x,y)$$

in the discrete case and:

$$E(W\big|B) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y\big|B}(x,y)\ dxdy$$

in the continuous case.

• for the variance one has:

$$var(W\big|B) = E\left(W^2\big|B\right) - \left[E(W\big|B)\right]^2.$$

Conditioning by a random variable

- for any event $Y = y$ such that $P_Y(y) > 0$, the conditional PMF is

$$P_{X|Y}(x|y) = P(X = x\ |Y = y)$$

- one can observe that

$$P_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

thus for a pair of random variables X, Y one has the general rule

$$P(X = x, Y = y) = P_{X|Y}(x|y) \cdot P_Y(y) = P_{Y|X}(y|x) \cdot P_X(x)$$

- for a continuous random variable the conditioning PDF of X given $Y = y$, $f_Y(y) > 0$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- there are similar formulae for the conditional expected value:

$$E(X|Y = y) = \sum_{x \in S_X} x \cdot P_{X|Y}(x|y)$$

$$E(g(X, Y)|Y = y) = \sum_{x \in S_X} g(x, y) \cdot P_{X|Y}(x|y)$$

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\ dx$$

$$E(g(X, Y)|Y = y) = \int_{-\infty}^{\infty} g(x, y) \cdot f_{X|Y}(x|y)\ dx$$

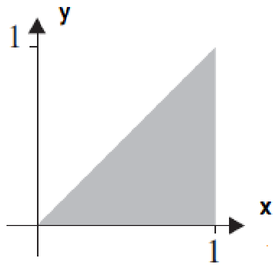
Solved problems

Problem 1. *Random variables X and Y have joint PDF*

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

Find the PDF $f_Y(y)$, the conditional PDF $f_{X|Y}(x|y)$, and the conditional expected value $E[X|Y = y]$.

Solution: The joint PDF is nonzero only in the shaded region illustrated below



In order to apply the formula of the marginal PDF f_Y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx$$

we have to investigate a few possible cases

For $0 \leq y \leq 1$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx = \int_{y}^{1} 2\ dx = 2(1 - y)$$

For $y > 1$ or $y < 0$, one can easily observe that $f_Y(y) = 0$, thus the marginal PDF is given by

$$f_Y(y) = \begin{cases} 2(1-y), & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Further, the conditional PDF of X given Y

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{1}{1-y}, & y \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The conditional expectation of X given $Y = y$ can be calculated as

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\ dx = \int_{y}^{1} x \cdot \frac{1}{1-y}\ dx = \left.\frac{x^2}{2(1-y)}\right|_{y}^{1} = \frac{1+y}{2}$$

Problem 2. *Calls arriving at a telephone switch are either voice calls v or data calls d. Each call is a voice call with probability p, independent of any other call. Observe calls at a telephone switch until you see two voice calls. Let M equal the number of calls up to and including the first voice call. Let N equal the number of calls observed up to and including the second voice call. Find the conditional PMFs $P_{M|N}(m|n)$ and $P_{N|M}(n|m)$. Interpret your results.*

Solution: First of all we have to observe that $N \geq M$. We will treat only the case $n > m$, since the case $n = m$ is much simpler. The main idea is to find the joint probability mass function

$$P_{M,N}(m, n) = P(M = m, N = n)$$

By the way M and N are defined, it involves estimating the probability of having $m-1$ calls until the first voice call and another sequence of $n-m-1$ calls until the second voice call

$$\underbrace{d\,d\,\ldots\,d}_{m-1\ \text{calls}}\ v\ \underbrace{d\,d\,\ldots\,d}_{n-m-1\ \text{calls}}\ v$$

In this way we have the event $N = n$, which means n calls up to and including the second voice call, will occur. Since v has the probability p and d will have the probability $1 - p$, one has

$$P(M = m, N = n) = (1 - p)^{m-1}p(1 - p)^{n-m-1}p = (1 - p)^{n-2}p^2$$

and thus

$$P_{M,N}(m, n) = \begin{cases} (1 - p)^{n-2}p^2, & m = 1, 2, \ldots, n - 1; \quad n = m + 1, m + 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

by the way M and N are defined, e.g. m can not be greater than $n - 1$.

For the conditional PMF's one has to use formulae like

$$P_{N|M}(n|m) = \frac{P_{M,N}(m, n)}{P_M(m)}$$

and thus we need to determine the marginal PMF's $P_N(n)$ and $P_M(m)$ For the first one:

$$P_N(n) = \sum_{m=1}^{n-1} P_{M,N}(m, n) = \sum_{m=1}^{n-1} (1 - p)^{n-2}p^2 = (n - 1)(1 - p)^{n-2}p^2$$

and eventually

$$P_N(n) = \begin{cases} (n - 1)(1 - p)^{n-2}p^2, & n = 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

The other PMF's is obtained adding the values $P_{M,N}(m, n)$ for $n \in S_N$, have a look in the previous handout

$$\begin{aligned} P_M(m) &= \sum_{n=m+1}^{\infty} P_{M,N}(m, n) = \sum_{n=m+1}^{\infty} (1 - p)^{n-2}p^2 = p^2 \left[(1 - p)^{m-1} + (1 - p)^m + \ldots)\right] \\ &= p^2(1 - p)^{m-1}\frac{1}{1 - (1 - p)} = p(1 - p)^{m-1} \end{aligned}$$

The complete expression will be

$$P_M(m) = \begin{cases} p(1 - p)^{m-1}, & m = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is now easy to find the conditional PMFs:

$$P_{N|M}(n|m) = \frac{P_{M,N}(m, n)}{P_M(m)} = \begin{cases} (1 - p)^{n-m-1}p, & n = m + 1, m + 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

and

$$P_{M|N}(m|n) = \frac{P_{M,N}(m,n)}{P_N(n)} = \begin{cases} \frac{1}{n-1}, & m = 1, 2, \ldots, n-1 \\ 0, & \text{otherwise} \end{cases}$$

In order to interpret these results, we have to observe that M has a geometric distribution, since it is the number of trials up to and including the first success and N has negative binomial distribution, since it is the number of trials required to see 2 successes.

Problem 3. *A company receives shipments from two factories. Depending on the size of the order, a shipment can be in:*
1 box for a small order,
2 boxes for a medium order,
3 boxes for a large order.
The company has two different suppliers: Factory Q is 60 *miles from the company. Factory R is* 180 *miles from the company. An experiment consists of monitoring a shipment and observing B, the number of boxes, and M, the number of miles the shipment travels. The following probability model describes the experiment*

	Factory Q	Factory R
small order	0.3	0.2
medium order	0.1	0.2
large order	0.1	0.1

a) Find $P_{B,M}(b,m)$, the joint PMF of the number of boxes and the distance. (You may present your answer as a matrix if you like.)

b) What is $E[B]$, the expected number of boxes?

c) What is $P_{M|B}(m|2)$, the conditional PMF of the distance when the order requires two boxes?

d) Find $E[M|B=2]$, the expected distance given that the order requires 2 boxes.

e) Are the random variables B and M independent?

f) The price per mile of sending each box is one cent per mile the box travels. C cents is the price of one shipment. What is $E[C]$, the expected price of one shipment?

Solution: First of all we have to observe two random variables M and B that model mathematically this problem. One has to observe that $M = 60$ corresponds to Factory Q and $M = 180$ corresponds to Factory R. Further $B = 1$ means small order, $B = 2$ a medium order and $B = 3$ a large order. We can rearrange the given information in the following way

$P_{B,M}(b,m)$	$m=60$	$m=180$
$b=1$	0.3	0.2
$b=2$	0.1	0.2
$b=3$	0.1	0.1

and the answer to a) is already given in this table.

For b) one needs the marginal PMFs of B and M, which can be computed using the strategies from our previous handout (adding the entries form every column or row)

$P_{B,M}(b,m)$	$m=60$	$m=180$	$P_B(b)$
$b=1$	0.3	0.2	0.5
$b=2$	0.1	0.2	0.3
$b=3$	0.1	0.1	0.2
$P_M(m)$	0.5	0.5	

Thus

$$E(B) = \sum_{b=1}^{3} b \cdot P_B(b) = 1 \cdot 0.5 + 2 \cdot 0.3 + 3 \cdot 0.2 = 1.7$$

For c) the conditional PMF of the distance when the order requires two boxes, means the conditional PMF of M given $B = 2$, which by its very formula is

$$P_{M|B}(m|2) = \frac{P_{M,B}(m,b)}{P_B(2)} = \begin{cases} 1/3, & m = 60 \\ 2/3, & m = 180 \\ 0, \text{ otherwise} \end{cases}$$

d) The conditional expectation of M given $B = 2$ is

$$E(M|B=2) = \sum_m m \cdot P_{M|B}(m,2) = 60 \cdot \frac{1}{3} + 180 \cdot \frac{2}{3} = 140$$

e) From the information we have in the previous table, we can conclude that M and B are not independent, since for example

$$P_{B,M}(1,60) = 0.3 \neq 0.5 \cdot 0.5 = P_B(1) \cdot P_M(m)$$

f) In terms of M and B, the cost (in cents) of sending a shipment is the random variable $C = B \cdot M$. The expected value of C is

$$E(C) = \sum_{b,m} bm \cdot P_{B,M}(b,m) = 1 \cdot 60 \cdot 0.3 + 2 \cdot 60 \cdot 0.1 + 3 \cdot 60 \cdot 0.1 +$$
$$+ 1 \cdot 180 \cdot 0.2 + 2 \cdot 180 \cdot 0.2 + 3 \cdot 180 \cdot 0.1 = 210$$

Proposed problems

Problem 1. *Let the random variable X be the portion of a flood insurance claim for flooding damage to the house and Y the portion of the claim for flooding damage to the rest of the property. The joint density function of X and Y is given by $f(x, y) = 3 - 2x - y$ for $0 < x, y < 1$ and $x + y < 1$. What are the marginal densities of X and Y ?*

Problem 2. *A business trip is equally likely to take $2, 3$, or 4 days. After a d-day trip, the change in the traveler's weight, measured as an integer number of pounds, is uniformly distributed between $-d$ and d pounds. For one such trip, denote the number of days by D and the change in weight by W. Find the joint PMF $P_{D,W}(d, w)$.*

Problem 3. *If the random variables X and Y have joint density function:*

$$f(x,y) = \begin{cases} \frac{xy}{96}, & 0 < x < 4, 1 < y < 5 \\ 0, & \textit{otherwise} \end{cases}$$

find the density function of $U = X + 2Y$.

Problem 4. *Let X and Y be two independent random variables each having the standard normal distribution. What is the probability distribution function of $X^2 + Y^2$?*

Problem 5. *An ice cream company orders supplies by fax. Depending on the size of the order, a fax can be either*

1 page for a short order,
2 pages for a long order.

The company has three different suppliers:

The vanilla supplier is 20 miles away.
The chocolate supplier is 100 miles away.
The strawberry supplier is 300 miles away.

An experiment consists of monitoring an order and observing N, the number of pages, and D, the distance the order is transmitted. The following probability model describes the experiment:

	van.	choc.	straw.
short	0.2	0.2	0.2
long	0.1	0.2	0.1

(a) What is the joint PMF $P_{N,D}(n, d)$ of the number of pages and the distance?
(b) What is $E[D]$, the expected distance of an order?

(c) Find $P_{D|N}(d|2)$*, the conditional PMF of the distance when the order requires 2 pages.*

(d) Write $E[D|N=2]$*, the expected distance given that the order requires 2 pages.*

(e) Are the random variables D *and* N *independent?*

(f) The price per page of sending a fax is one cent per mile transmitted. C *cents is the price of one fax. What is* $E[C]$*, the expected price of one fax?*

Bibliography

[1] R. Yates and D. Goodman. *Probability and Stochastic processes*, Wiley&Sons, 2005.

[2] D. Montgomery and G. Runger. *Applied Statistics and Probability for Engineers*, Wiley, 2014.

[3] C. Ariesanu. *Lecture Notes on Special Mathematics*, 2020.