Module 3

Jointly distributed random variables

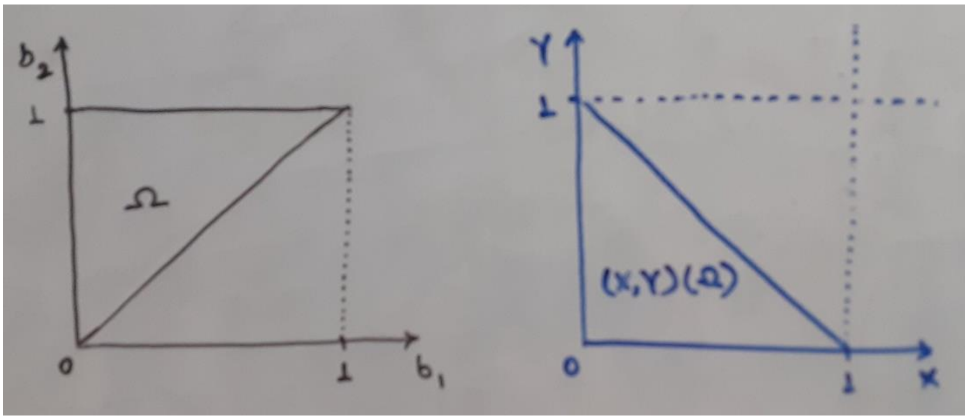
Topics: Jointly distributed random variables, Joint and marginal dist./mass/density functions, Conditional distribution, Conditional mean and variance, Independence of random variables

Jointly distributed random variables

Consider tossing of two fair dice. $\Omega = \{(d_1, d_2): d_i \in \{1,2,\dots,6\} \text{ for } i = 1,2\}$ and classical measure applies. Let X be the sum of the outcomes, i.e., $X = d_1 + d_2$. X is a discrete random variable with $X(\Omega) = \{2,3,\dots,12\}$. Let Y denote the maximum of the outcomes, i.e., $Y = \max(d_1, d_2)$. Y, too, is a discrete random variable with $Y(\Omega) = \{1,2,\dots,6\}$. With our current knowledge of probability, we can study X and Y separately and answer any question about one that does not involve the other. However, we can sense that these two are related, that is, information about one can influence the other. For example, if we know that $Y = 1$, then $X = 2$, or if $Y = 6$, then $X \geq 7$, etc. It makes sense to study X and Y together, i.e., study the vector (X, Y). Observe that (X, Y) is a function from Ω to $\mathbb{R}^2$, as shown in the table below. All the values of (X, Y) listed in the table constitute $(X, Y)(\Omega)$.

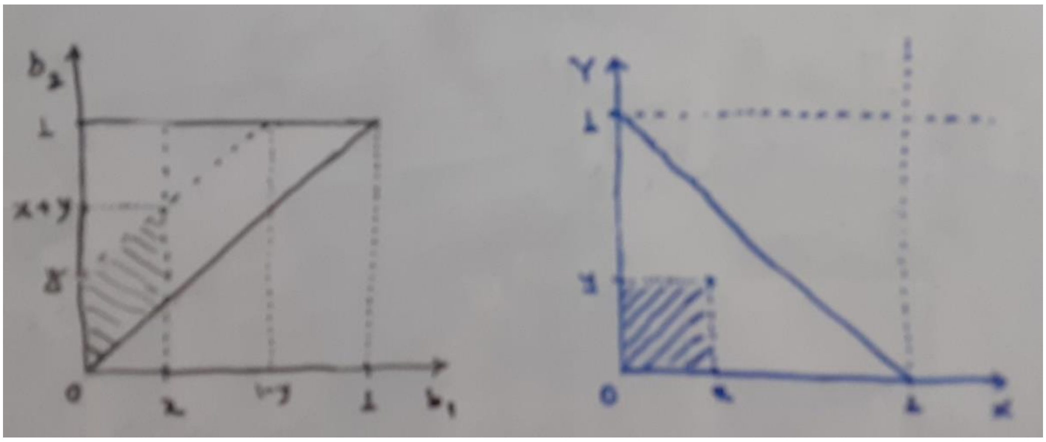
(X,Y)	$d_2 = 1$	$d_2 = 2$	$d_2 = 3$	$d_2 = 4$	$d_2 = 5$	$d_2 = 6$
$d_1 = 1$	(2,1)	(3,2)	(4,3)	(5,4)	(6,5)	(7,6)
$d_1 = 2$	(3,2)	(4,2)	(5,3)	(6,4)	(7,5)	(8,6)
$d_1 = 3$	(4,3)	(5,3)	(6,3)	(7,4)	(8,5)	(9,6)
$d_1 = 4$	(5,4)	(6,4)	(7,4)	(8,4)	(9,5)	(10,6)
$d_1 = 5$	(6,5)	(7,5)	(8,5)	(9,5)	(10,5)	(11,6)
$d_1 = 6$	(7,6)	(8,6)	(9,6)	(10,6)	(11,6)	(12,6)

Functions of the form $(X_1, X_2, \dots, X_n): \Omega \to \mathbb{R}^n$, where $n \geq 2$, are known as random vector or jointly distributed random variables. Our focus in this course will be on the two-variable case. Consider another example – A stick of 1m length is broken into three pieces in a random manner. Let b_1 and b_2 denote the first and second breakpoints with respect to a particular end of the stick. Then $\Omega = \{(b_1, b_2): 0 < b_1 < b_2 < 1\}$ and classical measure applies. Let $X = b_1$ and $Y = b_2 - b_1$ denote lengths of the 1st and 2nd pieces. Then $(X, Y)(\Omega) = \{(x, y): x, y > 0, x + y < 1\}$. Ω and $(X, Y)(\Omega)$ are shown in the diagram below.

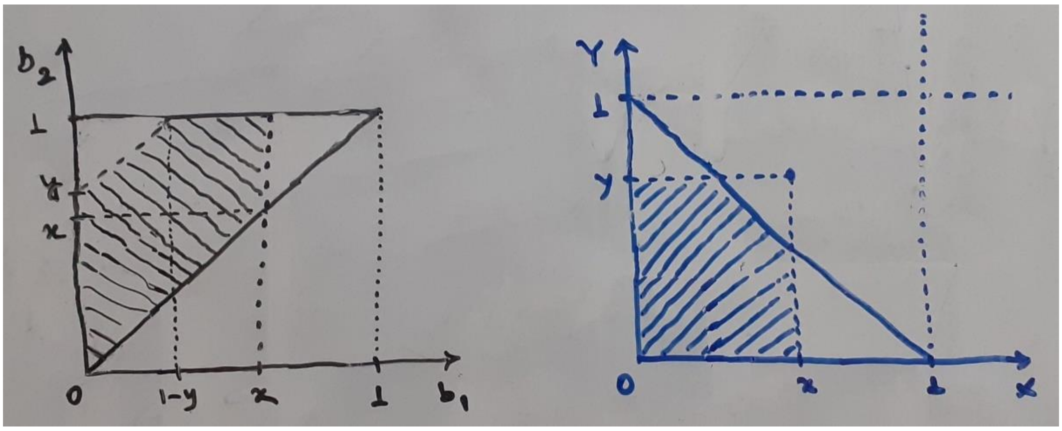


We can calculate probability of any event $A \in (X,Y)(\Omega)$ using the probability measure on Ω as: $P_{(X,Y)(\Omega)}(A) = P_\Omega(\{\omega \in \Omega: (X,Y)(\omega) \in A\})$. The joint distribution function of (X,Y) is defined as: $F_{X,Y}: \mathbb{R}^2 \to [0,1]$ such that $F_{X,Y}(x,y) = P(X \le x, Y \le y) = P_{(X,Y)(\Omega)}\big((-\infty,x] \times (-\infty,y]\big) = P_\Omega(\{\omega \in \Omega: (X,Y)(\omega) \in (-\infty,x] \times (-\infty,y]\})$ for all $(x,y) \in \mathbb{R}^2$. Observe that $(X,Y)(\omega) = \big(X(\omega), Y(\omega)\big)$. Then $F_{X,Y}(x,y) = P_\Omega(\{\omega \in \Omega: X(\omega) \le x, Y(\omega) \le y\})$. It can be used to obtain probability of any $A \in (X,Y)(\Omega)$. Let us obtain the joint distribution function of (X,Y) in the stick breaking example mentioned earlier. Since $(X,Y)(\Omega) = \{(x,y): x,y > 0, x+y < 1\}$, $F_{X,Y}(x,y) = 0$ if $x \le 0$ or $y \le 0$ or both. Also, $F_{X,Y}(x,y) = 1$ for $x,y \ge 1$. Other cases, i.e., $x \in (0,1)$ or $y \in (0,1)$ or both, are considered next.

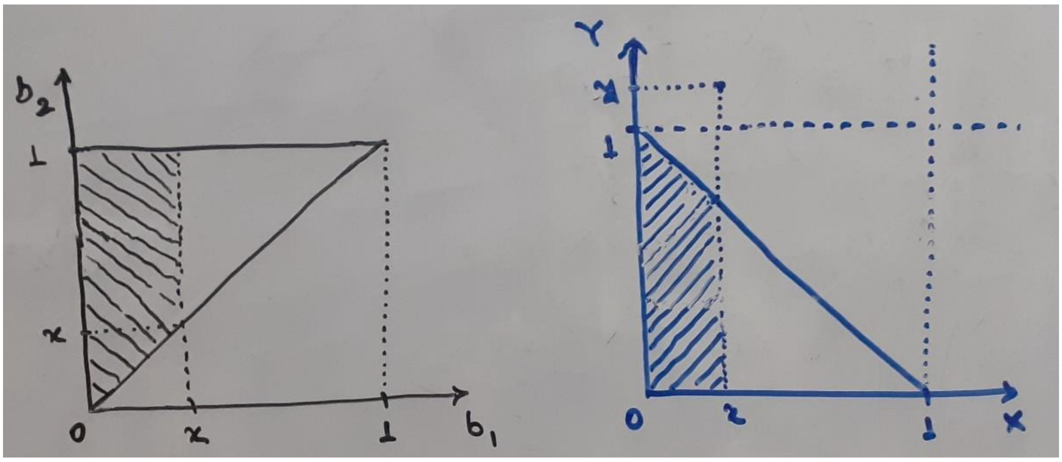
Consider $x \in (0,1)$, $y \in (0,1)$, and $x + y < 1$. The event $\{X \le x, Y \le y\}$ and its equivalent event in Ω are shown in the following diagram. Note that $X \le x \equiv b_1 \le x$ and $Y \le y \equiv b_2 - b_1 \le y$, and $x + y < 1$ ensures that the point of intersection of $b_1 = x$ and $b_2 - b_1 = y$ lies in Ω. Then $F_{X,Y}(x,y) = highlighted\ area/area\ of\ \Omega = xy/0.5 = 2xy$.



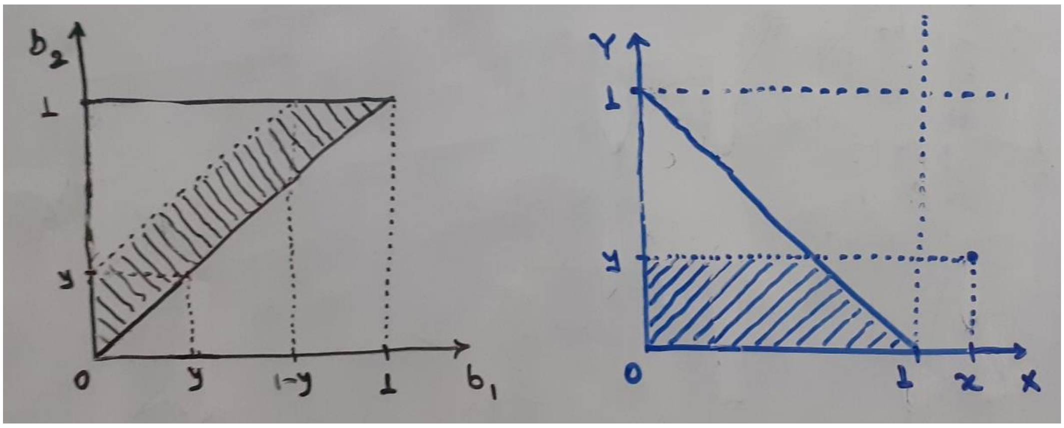
Consider $x \in (0,1)$, $y \in (0,1)$, and $x + y \ge 1$. The event $\{X \le x, Y \le y\}$ and its equivalent event in Ω are shown in the following diagram. It's like the previous case except the point of intersection of $b_1 = x$ and $b_2 - b_1 = y$ lies outside Ω (due to $x + y \ge 1$). So, $F_{X,Y}(x,y) = (0.5 - 0.5(1-y)^2 - 0.5(1-x)^2)/0.5 = 1 - (1-x)^2 - (1-y)^2$.



Consider $x \in (0,1)$ and $y \geq 1$. The event $\{X \leq x, Y \leq y\}$ and its equivalent event in Ω are shown in the following diagram. Note that $X \leq x \equiv b_1 \leq x$ and $Y \leq y \equiv b_2 - b_1 \leq y \equiv b_2 \leq y + b_1$, which is more than 1 (as $y \geq 1$ and $b_1 > 0$) implying no restriction on b_2. Here $F_{X,Y}(x,y) = (0.5 - 0.5(1-x)^2)/0.5 = 1 - (1-x)^2$.



Consider $x \geq 1$ and $y \in (0,1)$. The event $\{X \leq x, Y \leq y\}$ and its equivalent event in Ω are shown in the following diagram. Note that $X \leq x \equiv b_1 \leq x$ imposes no restriction on b_1 (as $x \geq 1$) but $Y \leq y \equiv b_2 - b_1 \leq y \equiv b_2 \leq y + b_1$ restricts b_2 (as shown). Here $F_{X,Y}(x,y) = (0.5 - 0.5(1-y)^2)/0.5 = 1 - (1-y)^2$.



Let us put all expressions in one place. Then we have

$$F_{X,Y}(x,y) = \begin{cases} 0 \ \text{ if } x \leq 0 \text{ or } y \leq 0 \text{ or both} \\ 2xy \ \text{ if } x,y \in (0,1) \text{ and } x+y < 1 \\ 1-(1-x)^2-(1-y)^2 \ \text{ if } x,y \in (0,1) \text{ and } x+y \geq 1 \\ 1-(1-x)^2 \ \text{ if } x \in (0,1) \text{ and } y \geq 1 \\ 1-(1-y)^2 \ \text{ if } x \geq 1 \text{ and } y \in (0,1) \\ 1 \ \text{ if } x,y \geq 1 \end{cases}$$

Obtain the joint distribution function value of (X,Y) at some $(x,y) \in \mathbb{R}^2$ in the dice rolling example mentioned earlier. You will see that $F_{X,Y}(x,y) = 0$ if $x < 2$ or $y < 1$ or both, and 1 if $x \geq 12$ and $y \geq 6$. It's between 0 and 1 at every other $(x,y) \in \mathbb{R}^2$. We can immediately notice some properties of the joint distribution function. First, $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$ for all $x, y \in \mathbb{R}$ and $F_{X,Y}(\infty, \infty) = 1$. Second, $F_{X,Y}(x, y)$ is non-decreasing in both x and y. Third, $F_{X,Y}(x, y)$ is right-continuous in both x and y. *Verify these properties.*

Joint and marginal dist./mass/density functions

From the joint distribution function of (X, Y), we can obtain marginal distribution functions, i.e., distribution functions of X and Y. $F_X(x) = P_{X(\Omega)}\big((-\infty, x]\big) = P_\Omega(\{\omega \in \Omega: X(\omega) \leq x\})$ $= P_\Omega(\{\omega \in \Omega: X(\omega) \leq x, Y(\omega) \in \mathbb{R}\}) = P(X \leq x, Y < \infty) = F_{X,Y}(x, \infty)$ for all $x \in \mathbb{R}$. In a similar manner, $F_Y(y) = F_{X,Y}(\infty, y)$ for all $y \in \mathbb{R}$. In the stick breaking example, $F_{X,Y}(x, \infty)$ $= 0$ if $x \leq 0$, $1 - (1 - x)^2$ if $x \in (0,1)$, and 1 if $x \geq 1$, which must be F_X. Also, $F_{X,Y}(\infty, y)$ $= 0$ if $y \leq 0$, $1 - (1 - y)^2$ if $y \in (0,1)$, and 1 if $y \geq 1$, which must be F_Y. Note that F_X and F_Y are identical, which is intuitive. *Obtain F_X and F_Y directly from Ω using your knowledge of random variables and see if you get the same expressions.*

If X and Y both are discrete random variables, then (X, Y) is referred to as discrete random vector. We define joint mass function for discrete (X, Y) as: $p_{X,Y}: (X, Y)(\Omega) \to [0,1]$ such that $p_{X,Y}(x, y) = P(X = x, Y = y) = P_{(X,Y)(\Omega)}(\{(x, y)\}) = P_\Omega(\{\omega \in \Omega: (X, Y)(\omega) = (x, y)\}) = P_\Omega(\{\omega \in \Omega: X(\omega) = x, Y(\omega) = y\})$ for all $(x, y) \in (X, Y)(\Omega)$. For the dice rolling example, the joint mass function is shown below in a tabular form.

$p_{X,Y}(x,y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$	$p_X(x)$
$x = 2$	1/36						1/36
$x = 3$		2/36					2/36
$x = 4$		1/36	2/36				3/36
$x = 5$			2/36	2/36			4/36
$x = 6$			1/36	2/36	2/36		5/36
$x = 7$				2/36	2/36	2/36	6/36
$x = 8$				1/36	2/36	2/36	5/36
$x = 9$					2/36	2/36	4/36
$x = 10$					1/36	2/36	3/36
$x = 11$						2/36	2/36
$x = 12$						1/36	1/36
$p_Y(y)$	1/36	3/36	5/36	7/36	9/36	11/36	1

From the joint mass function of (X, Y), we can obtain marginal mass functions, i.e., mass functions of X and Y as follows:

$$p_X(x) = P_{X(\Omega)}(\{x\}) = P_\Omega(\{\omega \in \Omega: X(\omega) = x\}) = P_\Omega(\{\omega \in \Omega: X(\omega) = x, Y(\omega) \in Y(\Omega)\})$$

$$= P_\Omega\left(\bigcup_{y \in Y(\Omega)} \{\omega \in \Omega: X(\omega) = x, Y(\omega) = y\}\right) = \sum_{y \in Y(\Omega)} P_\Omega(\{\omega \in \Omega: X(\omega) = x, Y(\omega) = y\})$$

by the third axiom $\Rightarrow p_X(x) = \sum_{y\in Y(\Omega)} p_{X,Y}(x,y)$ for all $x \in X(\Omega)$

Similarly, $p_Y(y) = \sum_{x\in X(\Omega)} p_{X,Y}(x,y)$ for all $y \in Y(\Omega)$. For the dice rolling example, p_X and p_Y are calculated with the above formulas and shown in the table above. *Obtain p_X and p_Y directly from Ω and see if you get the same mass functions.*

If X and Y both are continuous random variables, then (X,Y) is referred to as continuous random vector. We define joint density function for continuous (X,Y) as: $f_{X,Y}: \mathbb{R}^2 \to [0,\infty)$ such that $\int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dydx = F_{X,Y}(a,b)$ for all $(a,b) \in \mathbb{R}^2$. It can be verified that the $\partial^2 F_{X,Y}(x,y)/\partial x\partial y$, if exists throughout in $\mathbb{R}^2$, qualifies as joint density function. *Verify it yourself*. If the 2nd order partial derivative does not exist in a collection of points in $\mathbb{R}^2$ having zero area (it was zero length for random variable), then joint density function can be obtained as: $f_{X,Y}(x,y) = \partial^2 F_{X,Y}(x,y)/\partial x\partial y$ whenever it exists, else $f_{X,Y}(x,y) =$ any non-negative finite number. In the stick breaking experiment, $\partial^2 F_{X,Y}(x,y)/\partial x\partial y$ exists at all $(x,y) \in \mathbb{R}^2$ except in $\{(x,0): x \geq 0\}$, $\{(0,y): y > 0\}$, and $\{(x,y): x,y > 0, x+y = 1\}$. These are lines, and thus, have zero area. So, the joint density function is: $f_{X,Y}(x,y) = 2$ for $x,y \in (0,1)$ with $x+y < 1$, and 0 everywhere else.

We can obtain marginal density functions, i.e., density functions of X and Y, from the joint density function of (X,Y) in a manner analogous to the discrete case. Let us define $g(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ for all $x \in \mathbb{R}$. Note that g is a non-negative function defined on $\mathbb{R}$. Also, $\int_{-\infty}^{a} g(x)dx = \int_{-\infty}^{a}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\,dx = F_{X,Y}(a,\infty) = F_X(a)$ for all $a \in \mathbb{R}$. Thus, $g(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ qualifies as density function for X. Similarly, $h(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$ is density function for Y. Thus, in the stick breaking experiment,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} \displaystyle\int_0^{1-x} 2dy = 2(1-x) & \text{if } x \in (0,1) \\ \displaystyle\int_{-\infty}^{\infty} 0dy = 0 & \text{otherwise} \end{cases}$$

Similarly, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = 2(1-y)$ for $y \in (0,1)$, and 0 otherwise. *Obtain f_X and f_Y directly from Ω and see if you get the same density functions.*

Conditional distribution

Consider the dice rolling example. Assume that we have the knowledge that $Y = 4$, i.e., the maximum of the two outcomes is 4. Given this information, behaviour of the sum of the outcomes X is different from its usual behaviour, which is described by its mass function p_X. Note that $X|Y = 4$ takes values 5,6,7,8, whereas X takes values 2,3, … ,12. We can capture behaviour of $X|Y = 4$ through conditional mass function of X given that $Y = 4$, denoted by $p_{X|Y=4}$. It is obtained below in a general setting of discrete random vector.

$$p_{X|Y=y}(x) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)} \text{ for } x \in X(\Omega), y \in Y(\Omega)$$

With this, $p_{X|Y=4}(x) = p_{X,Y}(x, 4)/p_Y(4)$, which is $(2/36)/(7/36) = 2/7$ for $x = 5,6,7$ and $(1/36)/(7/36) = 1/7$ for $x = 8$. Values of $p_{X,Y}(x, 4)$ and $p_Y(4)$ are taken from the table mentioned earlier. Observe that the conditional masses are non-negative and add up to 1, just like the mass function. *Show that this observation is always true*. We can obtain conditional distribution by adding up conditional masses; it has all the properties of a distribution function. We can change the role of X and Y; then $p_{Y|X=x}(y) = p_{X,Y}(x, y)/p_X(x)$ for $x \in X(\Omega), y \in Y(\Omega)$. *Obtain $p_{Y|X=6}(y)$ for $y \in Y(\Omega)$ in the dice rolling example*.

Now, consider the stick breaking example. Assume that we have the knowledge that $Y = 0.4$, i.e., the length of the second piece is 0.4m. Given this information, behaviour of the length of the first piece X is different from its usual behaviour, which is described by its distribution F_X or density f_X. Note that $X|Y = 0.4$ takes values from $(0,0.6)$, whereas X takes values from $(0,1)$. We can capture behaviour of $X|Y = 0.4$ through conditional distribution/density of X given that $Y = 0.4$, denoted by $F_{X|Y=0.4}$ and $f_{X|Y=0.4}$ respectively. Conditional distribution is obtained next in a general setting of continuous random vector.

$$F_{X|Y=y}(x) = P(X \leq x|Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}, \text{which is of } \frac{0}{0} \text{ form } (Why?)$$

To overcome the $0/0$ form, let us replace $\{Y = y\}$ by $\lim_{\delta \to 0^+}\{y - \delta/2 < Y \leq y + \delta/2\}$.

$$\begin{aligned} F_{X|Y=y}(x) &= \lim_{\delta \to 0^+} \frac{P\left(X \leq x, y - \frac{\delta}{2} < Y \leq y + \frac{\delta}{2}\right)}{P\left(y - \frac{\delta}{2} < Y \leq y + \frac{\delta}{2}\right)} = \lim_{\delta \to 0^+} \frac{F_{X,Y}\left(x, y + \frac{\delta}{2}\right) - F_{X,Y}\left(x, y - \frac{\delta}{2}\right)}{F_Y\left(y + \frac{\delta}{2}\right) - F_Y\left(y - \frac{\delta}{2}\right)} \\ &= \lim_{\delta \to 0^+} \frac{\int_{-\infty}^{x} \int_{y-\delta/2}^{y+\delta/2} f_{X,Y}(u, v)dvdu}{\int_{y-\delta/2}^{y+\delta/2} f_Y(w)dw} = \lim_{\delta \to 0^+} \frac{\int_{-\infty}^{x} f_{X,Y}(u, y)\delta du}{f_Y(y)\delta} \\ &= \frac{\int_{-\infty}^{x} f_{X,Y}(u, y)du}{f_Y(y)} = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} du \quad \text{for all } x \in \mathbb{R}, y \in Y(\Omega) \end{aligned}$$

Note that $x \in \mathbb{R}$, but $y \in Y(\Omega)$. This is because Y cannot take a value $y \notin Y(\Omega)$. Conditional distribution exhibits all the properties of a distribution function. *Verify this claim using the above formula*. Conditional distribution does have a simple form like the conditional mass function. However, if we look closely, we see that $f_{X,Y}(u, y)/f_Y(y)$ can be regarded as the conditional density of $X|Y = y$. It meets all the requirements of a density function. Therefore, $f_{X|Y=y}(x) = f_{X,Y}(x, y)/f_Y(y)$ for $x \in \mathbb{R}, y \in Y(\Omega)$ and $f_{Y|X=x}(y) = f_{X,Y}(x, y)/f_X(x)$ for $x \in X(\Omega), y \in \mathbb{R}$. If we integrate the conditional density, we get conditional distribution.

With $Y = 0.4$ given in the stick breaking example, $f_{X|Y=0.4}(x) = f_{X,Y}(x, 0.4)/f_Y(0.4)$. From the previous discussion, $f_{X,Y}(x, 0.4) = 2$ for $x \in (0,0.6)$, and 0 elsewhere. Also, $f_Y(0.4) =$

$2(1-0.4)=1.2$. Then $f_{X|Y=0.4}(x)=2/1.2=1/0.6$ for $x\in(0,0.6)$, and 0 elsewhere. See that $X|Y=0.4$ is uniformly distributed in $(0,0.6)$. This should seem intuitively correct. *Show that $Y|X=x$ for $x\in(0,1)$ is uniformly distributed in $(0,1-x)$.*

We can have conditioning events other than $\{X=x\}$ or $\{Y=y\}$. For example, consider that $Y\leq 4$ is given in the dice rolling example. Using conditional probability,

$$p_{X|Y\leq 4}(x)=\frac{P(X=x,Y\leq 4)}{P(Y\leq 4)}=\frac{\sum_{y=1}^{4}p_{X,Y}(x,y)}{\sum_{y=1}^{4}p_Y(y)}\quad \text{for } x\in X(\Omega)$$

We can plug-in values from the previous table in the above formula and obtain the following:

x	2	3	4	5	6	7	8
$p_{X\|Y\leq 4}(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

Consider $Y\leq 0.4$ is given in the stick breaking example. Using the basic principles, obtain conditional distribution and density functions of X.

Conditional mean and variance

Using conditional mass and density functions, we can obtain conditional mean and variance. Here, we restrict ourselves to conditional events of the form $\{X=x\}$ or $\{Y=y\}$. Conditional mean of X given that $Y=y$ is obtained as follows:

$$E[X|Y=y]=\begin{cases}\displaystyle\sum_{x\in X(\Omega)} x p_{X|Y=y}(x) & \text{if } X \text{ is discrete}\\ \displaystyle\int_{-\infty}^{\infty} x f_{X|Y=y}(x)dx & \text{if } X \text{ is continuous}\end{cases}$$

$E[Y|X=x]$ can be obtained in a similar manner. The following table to the left shows joint and marginal mass functions of the random vector (X,Y), and the table to the right shows conditional mass function and expectation of $X|Y=y$ for all $y\in Y(\Omega)$.

$p_{X,Y}(x,y)$	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=0$	1/9	0	0	1/9
$x=1$	1/9	1/9	0	2/9
$x=2$	1/9	1/9	1/9	3/9
$x=3$	0	1/9	1/9	2/9
$x=4$	0	0	1/9	1/9
$p_Y(y)$	1/3	1/3	1/3	1

$p_{X\|Y=y}(x)$	$y=1$	$y=2$	$y=3$
$x=0$	1/3	0	0
$x=1$	1/3	1/3	0
$x=2$	1/3	1/3	1/3
$x=3$	0	1/3	1/3
$x=4$	0	0	1/3
$E[X\|Y=y]$	1	2	3

Observe that $E[X|Y=y]$ changes with y, which takes values 1,2,3 with equal probabilities. Therefore, $E[X|Y=y]$ can be viewed as a real-valued function of random variable Y. So, $E[X|Y=y]$ is a random variable, unlike $E[X]=\sum_{x=0}^{4}xp_X(x)=2$, which is a constant. The law of unconscious statistician is applicable in the context of conditional distribution as well. Then we can obtain expectation of $E[X|Y=y]$ as follows:

$$E\big[E[X|Y=y]\big] = \sum_{y=1}^{3} E[X|Y=y]p_Y(y) = 1\times\frac{1}{3} + 2\times\frac{1}{3} + 3\times\frac{1}{3} = 2, \text{same as } E[X].$$

It is no coincidence that $E\big[E[X|Y=y]\big] = E[X]$, as shown below. We can use subscripts for the E-operator to differentiate one expectation from another.

$$\begin{aligned}
E_Y\big[E[X|Y=y]\big] &= \sum_{y\in Y(\Omega)} E[X|Y=y]p_Y(y) = \sum_{y\in Y(\Omega)}\left(\sum_{x\in X(\Omega)} x p_{X|Y=y}(x)\right)p_Y(y) \\
&= \sum_{y\in Y(\Omega)}\sum_{x\in X(\Omega)} x\cdot\frac{p_{X,Y}(x,y)}{p_Y(y)}\cdot p_Y(y) = \sum_{x\in X(\Omega)} x\left(\sum_{y\in Y(\Omega)} p_{X,Y}(x,y)\right) \\
&= \sum_{x\in X(\Omega)} x p_X(x) = E[X], \text{as claimed.}
\end{aligned}$$

Following the above steps, show that $E_Y\big[E[X|Y=y]\big] = E[X]$ *for the continuous case*. We can change the role of X and Y to obtain $E_X\big[E[Y|X=x]\big] = E[Y]$. We can take a step further and show that $E_Y\Big[E[X^k|Y=y]\Big] = E[X^k]$ and $E_X\Big[E[Y^k|X=x]\Big] = E[Y^k]$ for all $k \geq 1$.

Conditional variance of X given that $Y=y$ is defined as the expectation of squared distance of X from $E[X|Y=y]$ given that $Y=y$, i.e., $Var(X|Y=y) = E[(X-E[X|Y=y])^2|Y=y]$. Using the law of unconscious statistician,

$$\begin{aligned}
Var(X|Y=y) &= E[(X-E[X|Y=y])^2|Y=y] \\
&= E[X^2 - 2XE[X|Y=y] + (E[X|Y=y])^2|Y=y] \\
&= E[X^2|Y=y] - 2E[X|Y=y]E[X|Y=y] + E^2[X|Y=y] \\
&= E[X^2|Y=y] - E^2[X|Y=y]
\end{aligned}$$

The following table to the left shows $E[X^2|Y=y]$ for the example mentioned earlier. The table to the right shows $Var(X|Y=y)$ for all $y\in Y(\Omega)$.

$p_{X\|Y=y}(x)$	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=0$	1/3	0	0	1/9
$x=1$	1/3	1/3	0	2/9
$x=2$	1/3	1/3	1/3	3/9
$x=3$	0	1/3	1/3	2/9
$x=4$	0	0	1/3	1/9
$E[X^2\|Y=y]$	5/3	14/3	29/3	

	$y=1$	$y=2$	$y=3$
From the previous table,			
$p_Y(y)$	1/3	1/3	1/3
$E[X\|Y=y]$	1	2	3
$Var(X\|Y=y)$	2/3	2/3	2/3
$(E[X\|Y=y] - E[X])^2$	$(1-2)^2$	$(2-2)^2$	$(3-2)^2$

$Var(X|Y=y)$ changes with y, and thus, can be viewed as a real-valued function of Y. Its expectation $E_Y[Var(X|Y=y)] = (2/3)\cdot(1/3) + (2/3)\cdot(1/3) + (2/3)\cdot(1/3) = 2/3$, however, is not same as $Var(X) = E[X^2] - E^2[X] = \sum_{x=0}^{4} x^2 p_X(x) - 2^2 = 48/9 - 4 = 4/3$. This happens because $Var(X)$ measures distance of X from $E[X]$, whereas $Var(X|Y = y)$ measures distance of X from $E[X|Y = y]$. For each $y \in Y(\Omega)$, the unaccounted squared distance is $(E[X|Y = y] - E[X])^2$, which is shown in the last row of the above table to the right. The expected unaccounted squared distance $(-1)^2 \cdot (1/3) + (0)^2 \cdot (1/3) + (1)^2 \cdot (1/3) = 2/3$, when added to $E_Y[Var(X|Y = y)] = 2/3$, gives $Var(X) = 4/3$.

In general, the expected unaccounted squared distance can be viewed as:

$$\begin{aligned}
E_Y[(E[X|Y = y] - E[X])^2] &= E_Y\left[E^2[X|Y = y] - 2E[X|Y = y]E[X] + E^2[X]\right] \\
&= E_Y\left[E^2[X|Y = y]\right] - 2E[X]E_Y\left[E[X|Y = y]\right] + E^2[X] \\
&= E_Y\left[E^2[X|Y = y]\right] - E^2[X] = E_Y\left[E^2[X|Y = y]\right] - E_Y^2\left[E[X|Y = y]\right] \\
&= Var_Y(E[X|Y = y]),
\end{aligned}$$

with $E[X|Y = y]$ treated as the random variable. Note that its randomness is governed by Y. So, $Var_Y(E[X|Y = y]) = E_Y\left[E^2[X|Y = y]\right] - E^2[X]$. Also,

$$\begin{aligned}
E_Y[Var(X|Y = y)] &= E_Y\left[E[X^2|Y = y] - E^2[X|Y = y]\right] \\
&= E_Y\left[E[X^2|Y = y]\right] - E_Y\left[E^2[X|Y = y]\right] = E[X^2] - E_Y\left[E^2[X|Y = y]\right]
\end{aligned}$$

Adding these two quantities, we obtain $E_Y[Var(X|Y = y)] + Var_Y(E[X|Y = y]) = E[X^2] - E^2[X] = Var(X)$, as was observed in the above example.

Independence of random variables

Earlier, we defined independence of events using conditional probability. Now, we extend the concept of independence to the random variables. Two random variables X and Y are said to be independent if events about X are independent of the events about Y. Except for simple cases, it's impossible to check independence of all such events. Fortunately, this requirement translates into a simple condition involving distribution/mass/density function.

Consider events of the form $\{X \le x\}$ and $\{Y \le y\}$ for arbitrary $x, y \in \mathbb{R}$. Then these events must be independent for X, Y to be independent. So, a necessary condition for independence of X, Y is: $P(X \le x, Y \le y) = P(X \le x)P(Y \le y) \equiv F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$, i.e., the joint distribution function must be product of the marginal distribution functions. This condition is sufficient as well, i.e., it implies independence of events of other forms. For example, consider $\{a_1 < X \le a_2\}$ and $\{b_1 < Y \le b_2\}$ for arbitrary $a_1 < a_2$ and $b_1 < b_2$.

$$\begin{aligned}
&P(a_1 < X \le a_2, b_1 < Y \le b_2) \\
&\quad = P(X \le a_2, Y \le b_2) - P(X \le a_1, Y \le b_2) - P(X \le a_2, Y \le b_1) \\
&\quad + P(X \le a_1, Y \le b_1); \text{draw a diagram to understand this better} \\
&\quad = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1) \\
&\quad = F_X(a_2)F_Y(b_2) - F_X(a_1)F_Y(b_2) - F_X(a_2)F_Y(b_1) + F_X(a_1)F_Y(b_1) \\
&\quad = \left(F_X(a_2) - F_X(a_1)\right)\left(F_Y(b_2) - F_Y(b_1)\right) = P(a_1 < X \le a_2)P(b_1 < Y \le b_2),
\end{aligned}$$

as desired. One can consider other forms of events and show that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$ implies independence of all such events. *Consider $X \sim U(0,10)$ and $Y \sim U(-5,5)$ to be independent. Determine $F_{X,Y}(x,y)$ for all $x, y \in \mathbb{R}$.*

For discrete random variables, the condition $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$ is equivalent to $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $(x,y) \in (X,Y)(\Omega)$. Let us assume $(X,Y)(\Omega) = \{a_1, a_2, \dots\} \times \{b_1, b_2, \dots\}$. Let $a_1 < a_2 < \cdots$ and $b_1 < b_2 < \cdots$. Then

$$F_{X,Y}(a_1,b_1) = F_X(a_1)F_Y(b_1) \Rightarrow \sum_{x \le a_1} \sum_{y \le b_1} p_{X,Y}(x,y) = \left(\sum_{x \le a_1} p_X(x)\right)\left(\sum_{y \le b_1} p_Y(y)\right)$$
$$\Rightarrow p_{X,Y}(a_1,b_1) = p_X(a_1)p_Y(b_1)$$

With $p_{X,Y}(a_1,b_1) = p_X(a_1)p_Y(b_1)$ established, we consider a_2, b_1 next.

$$F_{X,Y}(a_2,b_1) = F_X(a_2)F_Y(b_1) \Rightarrow \sum_{x \le a_2} \sum_{y \le b_1} p_{X,Y}(x,y) = \left(\sum_{x \le a_2} p_X(x)\right)\left(\sum_{y \le b_1} p_Y(y)\right)$$
$$\Rightarrow p_{X,Y}(a_1,b_1) + p_{X,Y}(a_2,b_1) = \big(p_X(a_1) + p_X(a_2)\big)p_Y(b_1)$$
$$\Rightarrow p_{X,Y}(a_2,b_1) = p_X(a_2)p_Y(b_1)$$

Proceeding in the same manner, we can show that $p_{X,Y}(a_i,b_1) = p_X(a_i)p_Y(b_1)$ for all $i = 1,2, \dots$. We can go in the other direction and show that $p_{X,Y}(a_1,b_j) = p_X(a_1)p_Y(b_j)$ for all $j = 1,2, \dots$. Let us consider a_2, b_2 next.

$$F_{X,Y}(a_2,b_2) = F_X(a_2)F_Y(b_2) \Rightarrow \sum_{x \le a_2} \sum_{y \le b_2} p_{X,Y}(x,y) = \left(\sum_{x \le a_2} p_X(x)\right)\left(\sum_{y \le b_2} p_Y(y)\right)$$
$$\Rightarrow p_{X,Y}(a_1,b_1) + p_{X,Y}(a_2,b_1) + p_{X,Y}(a_1,b_2) + p_{X,Y}(a_2,b_2)$$
$$= \big(p_X(a_1) + p_X(a_2)\big)\big(p_Y(b_1) + p_Y(b_2)\big) \Rightarrow p_{X,Y}(a_2,b_2) = p_X(a_2)p_Y(b_2)$$

We can proceed as before and show that $p_{X,Y}(a_i,b_2) = p_X(a_i)p_Y(b_2)$ for all $i = 1,2, \dots$ and $p_{X,Y}(a_2,b_j) = p_X(a_2)p_Y(b_j)$ for all $j = 1,2, \dots$. Continuing this way, we get the result that $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $(x,y) \in (X,Y)(\Omega)$.

Now, we need to establish that $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $(x,y) \in (X,Y)(\Omega)$ implies $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$. For arbitrary $a, b \in \mathbb{R}$,

$$F_{X,Y}(a,b) = \sum_{x \le a} \sum_{y \le b} p_{X,Y}(x,y) = \sum_{x \le a} \sum_{y \le b} p_X(x)p_Y(y) = \left(\sum_{x \le a} p_X(x)\right)\left(\sum_{y \le b} p_Y(y)\right)$$
$$= F_X(a)F_Y(b), \text{as desired.}$$

The analogous condition for continuous random variables is: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$. *Check for independence of X, Y in the dice rolling and stick breaking examples.*

It shall be noted that $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $(x,y) \in (X,Y)(\Omega)$ is equivalent to $p_{X|Y=y}(x) = p_X(x)$ as well as to $p_{Y|X=x}(y) = p_Y(y)$ for all $(x,y) \in (X,Y)(\Omega)$. Likewise, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x,y \in \mathbb{R}$ is equivalent to $f_{X|Y=y}(x) = f_X(x)$ for all $x \in \mathbb{R}$ and $y \in Y(\Omega)$ as well as to $f_{Y|X=x}(y) = f_Y(y)$ for all $x \in X(\Omega)$ and $y \in \mathbb{R}$. Note the similarity between independence of events and independence of random variables.

We can extend the idea of independence to multiple random variables. Random variables X, Y, Z are said to be independent if events about one or more of these random variables are independent of the events about one or more the remaining random variables. This translates into $F_{X,Y,Z}(x,y,z) = F_X(x)F_Y(y)F_Z(z)$ for all $x,y,z \in \mathbb{R}$. We need not check for the similar conditions involving two random variables, because those are implied. For example,

$$\lim_{z\to\infty} F_{X,Y,Z}(x,y,z) = \lim_{z\to\infty} F_X(x)F_Y(y)F_Z(z) \Rightarrow F_{X,Y,Z}(x,y,\infty) = F_X(x)F_Y(y)F_Z(\infty)$$
$$\Rightarrow F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ for all } x,y \in \mathbb{R}.$$

In the discrete case, $F_{X,Y,Z}(x,y,z) = F_X(x)F_Y(y)F_Z(z)$ for all $x,y,z \in \mathbb{R}$ is equivalent to $p_{X,Y,Z}(x,y,z) = p_X(x)p_Y(y)p_Z(z)$ for all $(x,y,z) \in (X,Y,Z)(\Omega)$. The analogous condition for the continuous case is $f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z)$ for all $x,y,z \in \mathbb{R}$.

The idea of independence extends to real-valued functions of independent random variables. Let $g_1, g_2, \dots, g_n$ be real-valued functions and $X_1, X_2, \dots, X_n$ denote independent random variables. Then $g_1(X_1), g_2(X_2), \dots, g_n(X_n)$ are independent too.

Practice problems

Book-1: A Modern Introduction to Probability and Statistics by Dekking et al.
Book-3: A First Course in Probability by Ross

Joint and marginal distributions
Book-1, Chapter-9, Exercise No. 4, 6, 9, 12, 15

Conditional distribution and expectation
Book-3, Chapter-6, Problem No. 40, 41, 42

Conditional mean and variance
Book-3, Chapter-7, Problem No. 48, 50

Independence of random variables
Book-1, Chapter-9, Exercise No. 5, 7, 17

Problems from Book-3 are provided in the next page.

6.40. The joint probability mass function of X and Y is given by

$$p(1,1) = \frac{1}{8} \qquad p(1,2) = \frac{1}{4}$$
$$p(2,1) = \frac{1}{8} \qquad p(2,2) = \frac{1}{2}$$

(a) Compute the conditional mass function of X given $Y = i, i = 1, 2$.
(b) Are X and Y independent?
(c) Compute $P\{XY \leq 3\}, P\{X + Y > 2\}$, $P\{X/Y > 1\}$.

6.41. The joint density function of X and Y is given by

$$f(x, y) = xe^{-x(y+1)} \quad x > 0,\ y > 0$$

(a) Find the conditional density of X, given $Y = y$, and that of Y, given $X = x$.
(b) Find the density function of $Z = XY$.

6.42. The joint density of X and Y is

$$f(x, y) = c(x^2 - y^2)e^{-x} \quad 0 \leq x < \infty,\ -x \leq y \leq x$$

Find the conditional distribution of Y, given $X = x$.

7.48. A fair die is successively rolled. Let X and Y denote, respectively, the number of rolls necessary to obtain a 6 and a 5. Find
(a) $E[X]$;
(b) $E[X|Y = 1]$;
(c) $E[X|Y = 5]$.

7.50. The joint density of X and Y is given by

$$f(x, y) = \frac{e^{-x/y}e^{-y}}{y}, \quad 0 < x < \infty, \quad 0 < y < \infty$$

Compute $E[X^2|Y = y]$.