RANDOM VARIABLES (Discrete Case)

When each outcome (simple event) of a random experiment is assigned a *number* (in this chapter, an integer), this defines a RANDOM VARIABLE (RV). In the same experiment, we can define several random variables, and call them X, Y, Z, etc. (capital letters from the end of the alphabet).

EXAMPLE: Using the experiment of rolling two dice, we can define X as the *total* number of dots, and Y as the *larger* of the two numbers. This means assigning numbers to individual simple events in the following fashion:

X:

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Y:

1	2	3	4	5	6
2	2	3	4	5	6
3	3	3	4	5	6
4	4	4	4	5	6
5	5	5	5	5	6
6	6	6	6	6	6

■

Note the difference between *events* and *random variables*: an event is effectively assigning, to each outcome, of either 'yes' or 'no' (a Boolean value).

(Probability) Distribution of a RV

is a *table* (or formula) giving us the information about (i) all possible values of the RV (ii) and their individual probabilities. Thus, for example, our X and Y have the following (probability) distributions:

$X =$	2	3	4	5	6	7	8	9	10	11	12
Pr:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

and

$Y =$	1	2	3	4	5	6
Pr:	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

The probabilities must add up to 1.

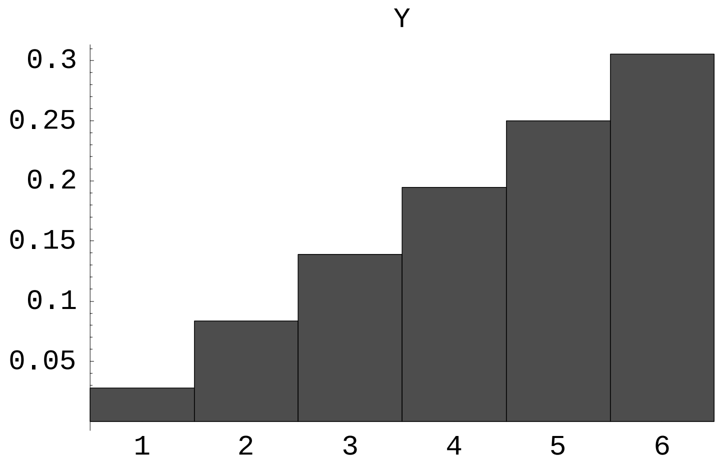
Eventually, we will find it more convenient to express the same information using a *formula* (called probability function) instead of a table. Thus, for example the distribution of X can be specified by: $\Pr(X = i) \equiv f_x(i) = \frac{6-|i-7|}{36}$, where $i = 2,\ 3, ...12$. Similarly: $\Pr(Y = i) \equiv f_y(i) = \frac{2i-1}{36}$, where $i = 1,\ 2, ...6$. Formulas become more convenient when dealing with RVs having too many (sometimes infinitely many) values.

Thus, for example, if we go back to the experiment of flipping a coin till a head appears, and define X as the total number of tosses, we have a choice of either

$X =$	1	2	3	4	.....	i	.....
Pr:	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{1}{2^3}$	$\frac{1}{2^4}$	.....	$\frac{1}{2^i}$	.....

or $f_x(i) = \frac{1}{2^i}$, where $i = 1,\ 2,\ 3, ...$

Sometimes it's useful to have a graph (`histogram`) of a distribution. Probabilities are usually displayed as vertical bars, each *centered* on the actual value.



Distribution Function of a RV is defined by: $F_x(k) = \Pr(X \leq k)$ i.e. a table or a formula providing *cumulative* probabilities. Thus, using our previous example:

$Y =$	1	2	3	4	5	6
F:	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	1

Obviously, $F(k)$ cannot decrease with increasing k, and the last value must be equal to 1 (when there is no last value, it is the limit which must equal to 1).

EXAMPLE: For the total number of tosses to get the first head, we get $F(k) = \sum_{i=1}^{k} (\frac{1}{2})^i = 1 - (\frac{1}{2})^k$, where $k = 1, 2, 3,...$ Note that $\lim_{k \to \infty} F(k) = 1$.

Multivariate (or **joint**) **distribution**

of *several* RV's.. When there are only *two* of them (the `bi-variate` case), all we need is a two-dimensional *table* which spells out, for every possible *pair* of values, the corresponding probability. Alternately, these probabilities can be computed from a `joint probability function` $f(i, j) \equiv \Pr(X = i \cap Y = j)$, and the *range* of possible i and j values (this can get tricky).

EXAMPLE: A coin is flipped three times, X is the total number of tails, Y is the number of heads up to the first tail.

Outcome:	Prob:	X:	Y:
HHH	$\frac{1}{8}$	0	3
HHT	$\frac{1}{8}$	1	2
HTH	$\frac{1}{8}$	1	1
HTT	$\frac{1}{8}$	2	1
THH	$\frac{1}{8}$	1	0
THT	$\frac{1}{8}$	2	0
TTH	$\frac{1}{8}$	2	0
TTT	$\frac{1}{8}$	3	0

The joint distribution of X and Y follows:

$Y =$ / $X =$	0	1	2	3
0	0	0	0	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0
2	$\frac{2}{8}$	$\frac{1}{8}$	0	0
3	$\frac{1}{8}$	0	0	0

Note that once we know the joint distribution of X and Y, we don't need to know what experiment it came from.

The joint distribution function is defined as $F(i, j) = \Pr(X \le i \cap Y \le j)$. We won't use it much.

Marginal distribution of X (and, similarly, of Y) is, effectively the *ordinary* (`univariate`) distribution of X (as if Y never existed). It can be obtained from the bivariate (joint) distribution by adding the probabilities in each row (total probability formula), e.g.:

$Y =$ / $X =$	0	1	2	3	
0	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
2	$\frac{2}{8}$	$\frac{1}{8}$	0	0	$\frac{3}{8}$
3	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$

Similarly, one can find the marginal distribution of Y by adding the probabilities in each column.

A bivariate distribution is often given to us via the corresponding *joint probability* `function`. At this level, we try to convert that information into an explicit table whenever possible (tables are easier to deal with).

EXAMPLE: Consider the following bivariate probability function of two random variables X and Y:

$$f_{xy}(i,j) = c \cdot (2i + j^2) \text{ where } \begin{array}{ll} 0 \le i \le 2 & \text{(Marginal range)} \\ i \le j \le 4 - i & \text{(Conditional range)} \end{array}$$

Find the value of c, the marginal distribution of Y and (based on this) $\Pr(Y \le 2)$.

Solution:

$Y =$	0	1	2	3	4
$X = 0$	$0c$	$1c$	$4c$	$9c$	$16c$
1	$\times$	$3c$	$6c$	$11c$	$\times$
2	$\times$	$\times$	$8c$	$\times$	$\times$

which clearly implies that $c = \frac{1}{58}$, the marginal distribution of Y is

$Y =$	1	2	3	4
Pr	$\frac{4}{58}$	$\frac{18}{58}$	$\frac{20}{58}$	$\frac{16}{58}$

and $\Pr(Y \le 2) = \frac{22}{58} = 37.93\%$. Note that the ranges could be 'reversed', thus: $1 \le j \le 4$ (marginal) and $0 \le i \le 2 - |j-2|$ (conditional).

Another EXAMPLE of 'reversing' two ranges of a bivariate distribution: $0 \le i \le 2$ (marginal) and $i \le j \le i+2$ (conditional) can be also expressed as: $0 \le j \le 4$ (marginal) and $\max(0, j-2) \le i \le \min(2, j)$ (conditional). Verify that these two (seemingly different) sets of inequalities describe the *same* set of i, j values.

Independence of X and Y is almost always a consequence of 'natural' independence. It means that $\Pr(X = i \cap Y = j) = \Pr(X = i) \times \Pr(Y = j)$ for *every* possible combination of i and j. This implies that, when X and Y are independent, their individual, *marginal* distributions is all we need.

All of these concepts can be extended to *three or more* RVs. This time, explicit tables are no longer practical, and we have to rely on `joint probability functions`

$$f_{xyz}(i, j, k) \equiv \Pr(X = i \cap Y = j \cap Z = k)$$

Stipulating the permissible *ranges* of i, j and k now gets even more tricky:.there are *6* distinct ways of doing that, can you see why?

Based on $f_{xy}(i, j)$, we should be able to tell instantly whether the corresponding RVs are *independent* or not. They are, if and only if *both* of the following conditions are met: (i) $f_{xy}(i, j)$ can be written as a *product* of a function of i times a function of j (let's call such function 'separable'), and (ii) the i and j *ranges* are (algebraically) independent of each other (i.e. in each range, both limits are *fixed* numbers). This extends easily to the case of 3 or more RVs.

EXAMPLES:

1. $f(i,j) = \frac{i+j}{24}$, where $1 \le i \le 3$ and $1 \le j \le i$, clearly implies that X and Y are not independent (both conditions violated).

2. $f(i,j,k) = \frac{i \cdot j \cdot k}{108}$, where $1 \le i \le 3$, $1 \le j \le 3$ and $1 \le k \le 2$. Yes, X, Y and Z *are* independent (it is easy to establish the corresponding marginals).

Conditional Distribution of X, given that Y has been observed to have a specific value, which we denote $\mathbf{j}$ (no longer a 'variable') is defined via its `conditional probability function` as follows

$$f_x(i \mid Y = \mathbf{j}) \equiv \Pr(X = i | Y = \mathbf{j}) = \frac{\Pr(X = i \cap Y = \mathbf{j})}{\Pr(Y = \mathbf{j})}$$

where i varies over its *conditional* range, given the value of $\mathbf{j}$. In practice, this means pulling out the corresponding row (column) of the bivariate table, and 're-normalizing' its probabilities (each is divided by their total).

Note that conditional distributions have all the properties of ordinary distributions.

EXAMPLE: Using our original example, we can easily construct

$Y \mid X = 2$	0	1
Prob:	$\frac{2}{3}$	$\frac{1}{3}$

by taking the probabilities in the $X = 2$ row, and dividing each of them by the corresponding *marginal* probability of $X = 2$. The values with zero probability should be discarded. Similarly:

$X \mid Y = 0$	1	2	3
Prob:	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Things get more tricky when dealing with *three* (or more) RVs. One can define a `conditional` distribution of *one*, given values of the other two, say:

$$\Pr(X = i \mid Y = \mathbf{j} \cap Z = \mathbf{k}) = \frac{f_{xyz}(i, \mathbf{j}, \mathbf{k})}{f_{yz}(\mathbf{j}, \mathbf{k})} \equiv f_x(i \mid Y = \mathbf{j}, Z = \mathbf{k})$$

or a conditional (joint) distribution of *two* of them, given a value of the third:

$$\Pr(X = i \cap Y = j \mid Z = \mathbf{k}) = \frac{f_{xyz}(i, j, \mathbf{k})}{f_z(\mathbf{k})} \equiv f_{xy}(i, j \mid Z = \mathbf{k})$$

Mutual **independence** implies that all *conditional* distributions are the same as the corresponding marginal distribution (i.e. ignore/remove the condition). For example, when X, Y and Z are mutually independent, $f_x(i\,|Y=\mathbf{j}) \equiv f_x(i)$, $f_x(i\,|Y=\mathbf{j}, Z=\mathbf{k}) \equiv f_x(i)$, etc.

Transforming RVs

When X is a random variable, any 'transformation' of X (i.e. an *expression* involving X, such as $\frac{X}{2}+1$) defines a new random variable (say V), with its own distribution.

EXAMPLE: If X has a distribution given by

$X =$	0	1	2	3
Prob:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

then, to build a distribution of $V = \frac{X}{2}+1$, one simply replaces the first-row values of the previous table, thus:

$V =$	1	$\frac{3}{2}$	2	$\frac{5}{2}$
Prob:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Similarly, if the new RV is $U = (X-2)^2$ (one can define any number of new RVs based on the same X), using the same approach the new table looks:

$U =$	4	1	0	1
Prob:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Here we don't like the duplication of values and their general 'disorder', so the same table should be presented as:

$U =$	0	1	2	3	4
Prob:	$\frac{3}{8}$	$\frac{4}{8}$	0	0	$\frac{1}{8}$

■

The most important is the so called **linear transformation** of X, i.e.

$$Y = aX + b$$

where a and b are two constants. Note that the *shape* (in terms of a histogram) of the Y-distribution is the *same* as that of X, only the horizontal scale has different tick marks (the new random variable is effectively the old random variable on a new scale - when X is temperature in Celsius, Y 'transforms' it to Fahrenheit).

Transforming *Two* RVs (into a *single* one)

EXAMPLE: If X and Y have the distribution of our old bivariate example, and $W = |X - Y|$, we can easily construct the (univariate) distribution of W by first building a table which shows the *value* of W for each X, Y combination:

$Y =$ / $X =$	0	1	2	3
0	$0 \leftthreetimes 0$	$1 \leftthreetimes 0$	$2 \leftthreetimes 0$	$3 \leftthreetimes \frac{1}{8}$
1	$1 \leftthreetimes \frac{1}{8}$	$0 \leftthreetimes \frac{1}{8}$	$1 \leftthreetimes \frac{1}{8}$	$2 \leftthreetimes 0$
2	$2 \leftthreetimes \frac{2}{8}$	$1 \leftthreetimes \frac{1}{8}$	$0 \leftthreetimes 0$	$1 \leftthreetimes 0$
3	$3 \leftthreetimes \frac{1}{8}$	$2 \leftthreetimes 0$	$1 \leftthreetimes 0$	$0 \leftthreetimes 0$

and then collect the probabilities of each unique value of W, from the smallest to the largest, to yield

$W =$	0	1	2	3
Prob:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{2}{8}$

Transforming RVs is a lot more fun in the continuous case.