2.4 Bivariate distributions

2.4.1 Definitions

Let X and Y be discrete r.v.s defined on the same probability space $(\mathcal{S}, \mathcal{F}, \mathrm{P})$. Instead of treating them separately, it is often necessary to think of them acting together as a random vector (X, Y) taking values in $\mathcal{R}^2$. The *joint probability function* of (X, Y) is defined as

$$p_{X,Y}(x, y) = \mathrm{P}(\{E \in \mathcal{S} : X(E) = x \text{ and } Y(E) = y\}), \tag{2.28}$$

and is often written as

$$\mathrm{P}(X = x, Y = y) \qquad (x = x_1, x_2, ...x_M; y = y_1, y_2, ..y_N), \tag{2.29}$$

where M, N may be finite or infinite. It satisfies the two conditions

$$\begin{aligned} \mathrm{P}(X = x, Y = y) &\geq 0 \\ \sum_x \sum_y \mathrm{P}(X = x, Y = y) &= 1. \end{aligned} \tag{2.30}$$

Various other functions are related to $\mathrm{P}(X = x, Y = y)$.

The *joint cumulative distribution function* of (X, Y) is given by

$$\begin{aligned} F(u, v) &= \mathrm{P}(X \leq u, Y \leq v), \qquad -\infty < u, v \leq \infty, \\ &= \sum_{x \leq u, y \leq v} \mathrm{P}(X = x, Y = y). \end{aligned} \tag{2.31}$$

The *marginal probability (mass) function of* X is given by

$$\mathrm{P}(X = x_i) = \sum_y \mathrm{P}(X = x_i, Y = y), \qquad i = 1, ...M. \tag{2.32}$$

The *marginal probability (mass) function of* Y is given by

$$\mathrm{P}(Y = y_j) = \sum_x \mathrm{P}(X = x, Y = y_j), \qquad j = 1, ...N. \tag{2.33}$$

The *conditional probability (mass) function of* X *given* $Y = y_j$ is given by

$$\mathrm{P}(X = x_i | Y = y_j) = \frac{\mathrm{P}(X = x_i, Y = y_j)}{\mathrm{P}(Y = y_j)}, \qquad i = 1, ...M. \tag{2.34}$$

The *conditional probability (mass) function of* Y *given* $X = x_i$ is given by

$$\mathrm{P}(Y = y_j | X = x_i) = \frac{\mathrm{P}(X = x_i, Y = y_j)}{\mathrm{P}(X = x_i)}, \qquad j = 1, ...N. \tag{2.35}$$

Expectation

The *expected value* of a function $h(X, Y)$ of the discrete r.v.s (X, Y) can be found directly from the joint probability function of (X, Y) as follows:

$$\mathrm{E}[h(X, Y)] = \sum_x \sum_y h(x, y) \mathrm{P}(X = x, Y = y) \tag{2.36}$$

provided the double series is absolutely convergent. This is the bivariate version of the 'law of the unconscious statistician' discussed earlier.

The *covariance* of X and Y is defined as

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathrm{E}[\{X-\mathrm{E}(X)\}\{Y-\mathrm{E}(Y)\}] \\ &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y). \end{aligned} \tag{2.37}$$

The *correlation coefficient* of X and Y is defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X).\mathrm{Var}(Y)}}. \tag{2.38}$$

2.4.2 Independence

In Chapter 1 the independence of *events* has been defined and discussed: now this concept is extended to random variables. The discrete random variables X and Y are *independent* if and only if the pair of events $\{E \in \mathcal{S} : X(E) = x_i\}$ and $\{E \in \mathcal{S} : Y(E) = y_j\}$ are independent for all x_i, y_j, and we write this condition as

$$\mathrm{P}(X = x_i, Y = y_j) = \mathrm{P}(X = x_i).\mathrm{P}(Y = y_j) \text{ for all } (x_i, y_j). \tag{2.39}$$

It is easily proved that an equivalent statement is: X and Y are independent if and only if there exist functions $f(\cdot)$ and $g(\cdot)$ such that

$$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y) = f(x)g(y) \text{ for all } x, y. \tag{2.40}$$

Example 2.3

A biased coin yields 'heads' in a single toss with probability p. The coin is tossed a random number of times N, where $N \sim \mathrm{Poisson}(\lambda)$. Let X and Y denote the number of heads and tails obtained respectively. Show that X and Y are independent Poisson random variables.

Solution Conditioning on the value of $X + Y$, we have

$$\begin{aligned} \mathrm{P}(X = x, Y = y) &= \mathrm{P}(X = x, Y = y | X + Y = x + y)\mathrm{P}(X + Y = x + y) \\ &\quad +\mathrm{P}(X = x, Y = y | X + Y \neq x + y)\mathrm{P}(X + Y \neq x + y). \end{aligned}$$

The second conditional probability is clearly 0, so

$$\begin{aligned} \mathrm{P}(X = x, Y = y) &= \mathrm{P}(X = x, Y = y | X + Y = x + y)\mathrm{P}(X + Y = x + y) \\ &= \tbinom{x+y}{x} p^x q^y . \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} \quad [\text{using (2.201) \& } N = X + Y] \\ &= \frac{(\lambda p)^x (\lambda q)^y}{x! y!} e^{-\lambda}. \end{aligned}$$

But

$$\begin{aligned} \mathrm{P}(X = x) &= \textstyle\sum_{n \geq x} \mathrm{P}(X = x | N = n)\mathrm{P}(N = n) \\ &= \textstyle\sum_{n \geq x} \tbinom{n}{x} p^x q^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} e^{-\lambda} \sum_{n \geq x} \frac{(\lambda q)^{n-x}}{(n-x)!} \\ &= \frac{(\lambda p)^x}{x!} e^{-\lambda} . e^{\lambda q} = \frac{(\lambda p)^x}{x!} e^{-\lambda p}. \end{aligned}$$

Similarly

$$\mathrm{P}(Y = y) = \frac{(\lambda q)^y}{y!} e^{-\lambda q}.$$

Then $\mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x).\mathrm{P}(Y = y)$ for all (x, y), and it follows that X and Y are independent Poisson random variables (with parameters λp and λq respectively). ♢

It can also be shown readily that if X and Y are independent, so too are the random variables $g(X)$ amd $h(Y)$, for any functions g and h: this result is used frequently in problem solving.

If X and Y are independent,

$$\begin{aligned} \mathrm{E}(XY) &= \sum_{x,y} xy\mathrm{P}(X=x, Y=y) \qquad [(2.36)] \\ &= \sum_{x,y} xy\mathrm{P}(X=x)\mathrm{P}(Y=y) \qquad [\text{independence, } (2.39)] \\ &= \sum_{x} x\mathrm{P}(X=x) \sum_{y} y\mathrm{P}(Y=y) \end{aligned}$$

i.e., by (2.10),

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \quad \text{if } X, Y \text{ are independent.} \qquad (2.41)$$

The converse of this result is false i.e. $\mathrm{E}(XY) = \mathrm{E}(X)(\mathrm{E}(Y)$ does *not* imply that X and Y are independent.

It follows immediately that

$$\mathrm{Cov}(X, Y) = \rho(X, Y) = 0 \quad \text{if } X, Y \text{ are independent.} \qquad (2.42)$$

Once again, the converse is false.

A generalisation of the result for $\mathrm{E}(XY)$ is: if X and Y are independent, then, for any functions g and h,

$$\mathrm{E}\{g(X)h(Y)\} = \mathrm{E}\{g(X)\}\mathrm{E}\{h(Y)\}. \qquad (2.43)$$

2.4.3 Conditional expectation

Referring back to the definition of the conditional probability mass function of X, it is natural to define the *conditional expectation* (or *conditional expected value*) of X given $Y = y_j$ as

$$\begin{aligned} \mathrm{E}(X|Y=y_j) &= \sum_i x_i \mathrm{P}(X=x_i|Y=y_j) \\ &= \sum_i x_i \mathrm{P}(X=x_i, Y=y_j)/\mathrm{P}(Y=y_j) \end{aligned} \qquad (2.44)$$

provided the series is absolutely convergent. This definition holds for all values of $y_j (j = 1, 2, ...)$, and is one value taken by the r.v. $\mathrm{E}(X|Y)$. Since $\mathrm{E}(X|Y)$ is a function of Y, we can write down its mean using (2.14): thus

$$\begin{aligned} \mathrm{E}[\mathrm{E}(X|Y)] &= \sum_j \mathrm{E}(X|Y=y_j).\mathrm{P}(Y=y_j) \\ &= \sum_j \sum_i x_i \mathrm{P}(X=x_i|Y=y_j)\mathrm{P}(Y=y_j) \qquad [(2.44)] \\ &= \sum_j \sum_i x_i \frac{\mathrm{P}(X=x_i, Y=y_j)}{\mathrm{P}(Y=y_j)} \mathrm{P}(Y=y_j) \qquad [(2.34)] \\ &= \sum_j \sum_i x_i \mathrm{P}(X=x_i, Y=y_j) \\ &= \sum_i x_i \sum_j \mathrm{P}(X=x_i, Y=y_j) \\ &= \sum_i x_i \mathrm{P}(X=x_i) \qquad [(2.30b)] \end{aligned}$$

i.e.

$$\text{E}[\text{E}(X|Y) = \text{E}(X). \qquad (2.45)$$

This result is very useful in practice: it often enables us to compute expectations easily by first conditioning on some random variable Y and using

$$\text{E}(X) = \sum_j \text{E}(X|Y = y_j).\text{P}(Y = y_j). \qquad (2.46)$$

(There are similar definitions and results for $\text{E}(Y|X = x_i)$ and the r.v. $\text{E}(Y|X)$.)

Example 2.4 (Ross)

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel which takes him to safety after 2 hours of travel. The second door leads to a tunnel which returns him to the mine after 3 hours of travel. The third door leads to a tunnel which returns him to the mine after 5 hours. Assuming he is at all times equally likely to choose any of the doors, what is the expected length of time until the miner reaches safety?

Solution Let

X: time to reach safety (hours)

Y: door intially chosen (1,2 or 3)

Then

$$\begin{aligned} \text{E}(X) &= \text{E}(X|Y = 1)\text{P}(Y = 1) + \text{E}(X|Y = 2)\text{P}(Y = 2) + \text{E}(X|Y = 3)\text{P}(Y = 3) \\ &= \tfrac{1}{3}\{\text{E}(X|Y = 1) + \text{E}(X|Y = 2) + \text{E}(X|Y = 3)\}. \end{aligned}$$

Now

$$\begin{aligned} \text{E}(X|Y = 1) &= 2 \\ \text{E}(X|Y = 2) &= 3 + \text{E}(X) \\ \text{E}(X|Y = 3) &= 5 + \text{E}(X) \qquad \text{(why?)} \end{aligned}$$

So

$$\text{E}(X) = \frac{1}{3}\{2 + 3 + \text{E}(X) + 5 + \text{E}(X)\} \qquad \text{or } \text{E}(X) = 10.$$ □

It follows from the definitions that, if X and Y are *independent* r.v.s, then

$$\begin{aligned} \text{E}(X|Y) &= \text{E}(X) \\ \text{and } \text{E}(Y|X) &= \text{E}(Y) \qquad \text{(both constants).} \end{aligned} \qquad (2.47)$$

2.5 Transformations and Relations

In many situations we are interested in the probability distribution of some function of X and Y. The usual procedure is to attempt to express the relevant probabilities in terms of the joint probability function of (X, Y). Two examples illustrate this.

Example 2.5 (*Discrete Convolution*)

Suppose X and Y are independent count random variables. Find the probability distribution of the r.v. $Z = X + Y$. Hence show that the sum of two independent Poisson r.v.s is also Poisson distributed.

Solution The event '$Z = z$' can be decomposed into the union of mutually exclusive events:

$$(Z = z) = (X = 0, Y = z) \cup (X = 1, Y = z - 1) \cup \cdots \cup (X = z, Y = 0) \qquad \text{for } z = 0, 1, 2, ...$$

Then we have $\text{P}(Z = z) = \sum_{x=0}^{z} \text{P}(X = x, Y = z - x)$ or, invoking independence [(2.39)],

$$\mathrm{P}(Z=z)=\sum_{x=0}^{z}\mathrm{P}(X=x).\mathrm{P}(Y=z-x). \tag{2.48}$$

This summation is known as the (discrete) *convolution* of the distributions $p_X(x)$ and $p_Y(y)$.

Now let the independent r.v.s X and Y be such that $X \sim \text{Poisson}(\lambda_1)$; $Y \sim \text{Poisson}(\lambda_2)$ i.e.

$$\mathrm{P}(X=x)=\frac{{\lambda_1}^x e^{-\lambda_1}}{x!}, \quad x>0; \qquad \mathrm{P}(Y=y)=\frac{{\lambda_2}^x e^{-\lambda_2}}{y!}, \quad y>0.$$

Then, for $Z = X + Y$,

$$\begin{aligned}
\mathrm{P}(Z=z) &= \sum_{x=0}^{z}\frac{{\lambda_1}^x e^{-\lambda_1}}{x!}.\frac{{\lambda_2}^{z-x} e^{-\lambda_2}}{(z-x)!} \\
&= \sum_{x=0}^{z}\frac{z!}{x!(z-x)!}{\lambda_1}^x{\lambda_2}^{z-x}.\frac{e^{-(\lambda_1+\lambda_2)}}{z!} \\
&= \frac{(\lambda_1+\lambda_2)^z}{z!}e^{-(\lambda_1+\lambda_2)}, \qquad z=0,1,2,...
\end{aligned}$$

i.e.

$$Z \sim \text{Poisson}(\lambda_1+\lambda_2). \qquad \square$$

Example 2.6

Given count r.v.s (X,Y), obtain an expression for $\mathrm{P}(X<Y)$.

Solution Again, we decompose the event of interest into the union of mutually exclusive events:

$$\begin{aligned}
(X<Y) &= (X=0,Y=1)\cup(X=0,Y=2)\cup\cdots \\
&\quad \cup(X=1,Y=2)\cup(X=1,Y=3)\cup\cdots \\
&\quad \cup(X=2,Y=3)\cup(X=2,Y=4)\cup\cdots \\
&\quad \cdots\cdots \\
&= \bigcup_{x=0,\dots,\infty;y=x+1,\dots,\infty}(X=x,Y=y)
\end{aligned}$$

So

$$\mathrm{P}(X<Y)=\sum_{x=0}^{\infty}\sum_{y=x+1}^{\infty}\mathrm{P}(X=x,Y=y). \qquad \square$$

2.6 Multivariate distributions

2.6.1 Definitions

The basic definitions for the *multivariate* situation – where we consider a p-vector of r.v.s $(X_1, X_2, ..., X_p)$ – are obvious generalisations of those for the bivariate case. Thus the *joint probability function* is

$$\begin{aligned} &\mathrm{P}(X_1 = x_1, X_2 = x_2, ..., X_p = x_p) = \\ &\qquad \mathrm{P}(\{E \in \mathcal{S} : X_1(E) = x_1 \text{ and } X_2(E) = x_2 \text{ and ... and } X_p(E) = x_p\}) \end{aligned} \tag{2.49}$$

and has the properties

$$\begin{aligned} \mathrm{P}(X_1 = x_1, ..., X_p = x_p) &\geq 0 \qquad \text{for all } (x_1, ..., x_p) \\ \text{and } \sum_{x_1} \cdots \sum_{x_p} \mathrm{P}(X_1 = x_1, ..., X_p = x_p) &= 1. \end{aligned} \tag{2.50}$$

The *marginal probability function* of X_i is given by

$$\mathrm{P}(X_i = x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \sum_{x_p} \mathrm{P}(X_1 = x_1, ..., X_p = x_p) \qquad \text{for all } x_i. \tag{2.51}$$

The probability function of any subset of $(X_1, ..., X_p)$ is found in a similar way.

Conditional probability functions can be defined by analogy with the bivariate case, and *expected values* of functions of $(X_1, ..., X_p)$ are found as for bivariate functions.

2.6.2 Multinomial distribution

This is the most important discrete multivariate distribution, and is deduced by arguments familiar from the case of the binomial distribution. Consider n repeated independent trials, where each trial results in one of the outcomes $E_1, ..., E_k$ with

$$\mathrm{P}(E_i \text{ occurs in a trial}) = p_i, \qquad \sum_{i=1}^{k} p_i = 1.$$

Let X_i = number of times the outcome E_i occurs in the n trials.

Then the joint probability function of $(X_1, ..., X_k)$ is given by

$$\mathrm{P}(X_1 = x_1, ..., X_k = x_k) = \frac{n!}{x_1! x_2! ... x_k!} {p_1}^{x_1} ... {p_k}^{x_k}, \tag{2.52a}$$

where the $x_1, ..., x_k$ are counts between 0 and n such that

$$\sum_{i=1}^{k} x_i = n. \tag{2.52b}$$

For consider the event

$$\underset{x_1}{E_1....E_1} \qquad \underset{x_2}{E_2....E_2} \qquad .... \qquad \underset{x_k}{E_k....E_k} \qquad \text{times}$$

It has probability

$${p_1}^{x_1} {p_2}^{x_2} .... {p_k}^{x_k}, \qquad \sum_i x_i = n, \quad \sum_i p_i = 1.$$

Any event with x_1 outcomes E_1, x_2 outcomes E_2,..... and x_k outcomes E_k in a given order also has this probability. There are $\frac{n!}{x_1!...x_k!}$ different arrangements of such a set of outcomes and these are mutually exclusive: the event $(X_1 = x_1, ..., X_k = x_k)$ is the union of these mutually exclusive arrangements. Hence the above result.

The marginal probability distribution of X_i is Binomial with parameters n and p_i, and hence

$$\begin{aligned} \mathrm{E}(X_i) &= np_i \\ \mathrm{Var}(X_i) &= np_i(1-p_i). \end{aligned} \tag{2.53}$$

Also, we shall prove later that

$$\mathrm{Cov}(X_i, X_j) = -np_ip_j, \qquad i \neq j. \tag{2.54}$$

2.6.3 Independence

For convenience, write $I = \{1, ..., p\}$ so that we are considering the r.v.s $\{X_i : i \in I\}$. These r.v.s are called *independent* if the events $\{X_i = x_i\}, i \in I$ are independent for all possible choices of the set $\{x_i : i \in I\}$ of values of the r.v.s. In other words, the r.v.s are independent if and only if

$$\mathrm{P}(X_i = x_i \text{ for all } i \in J) = \prod_{i \in J} \mathrm{P}(X_i = x_i) \tag{2.55}$$

for all subsets J of I and all sets $\{x_i : i \in I\}$.

Note that a set of r.v.s which are pairwise independent are not necessarily independent.

2.6.4 Linear combinations

Linear combinations of random variables occur frequently in probability analysis. The principal results are as follows:

$$\mathrm{E}\left[\sum_{i=1}^n a_iX_i\right] = \sum_{i=1}^n a_i\mathrm{E}(X_i) \tag{2.56}$$

(whether or nor the r.v.s are independent);

$$\begin{aligned} \mathrm{Var}\left[\sum_{i=1}^n a_iX_i\right] &= \sum_{i=1}^n a_i^2\mathrm{Var}(X_i) + 2\sum_{i<j} a_ia_j\mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n a_i^2\mathrm{Var}(X_i) \qquad \text{if the r.v.s are independent;} \end{aligned} \tag{2.57}$$

$$\mathrm{Cov}\left[\sum_{i=1}^n a_iX_i, \sum_{j=1}^m b_jX_j\right] = \sum_{i=1}^n\sum_{j=1}^m a_ib_j\mathrm{Cov}(X_i, X_j) \tag{2.58}$$

where $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$ by definition.

2.7 Indicator random variables

Some probability problems can be solved more easily by using *indicator* random variables, along with the above results concerning linear combinations.

An *indicator random variable* X for an event A takes the value 1 if A occurs and the value 0 if A does not occur. Thus we have:

$$\begin{array}{l} \mathrm{P}(X=1)=\mathrm{P}(A) \\ \mathrm{P}(X=0)=\mathrm{P}(\overline{A})=1-\mathrm{P}(A) \\ \mathrm{E}(X)=1.\mathrm{P}(X=1)+0.\mathrm{P}(X=0)=\mathrm{P}(A) \\ \mathrm{E}(X^2)=1^2.\mathrm{P}(X=1)+0^2.\mathrm{P}(X=0)=\mathrm{P}(A) \\ \mathrm{Var}(X)=\mathrm{P}(A)-[\mathrm{P}(A)]^2. \end{array} \quad (2.59)$$

Clearly $1-X$ is the indicator r.v. for the event $\overline{A}$.

Let Y be the indicator r.v. for the event B. Then the various combinations involving A and B have indicator r.v.s as follows:

Event	Indicator r.v.
$A \cap B$	XY
$\overline{A} \cap \overline{B}$	$(1-X)(1-Y)$
$A \cup B$	$1-(1-X)(1-Y)$
$A \cup B$ (A, B mutually exclusive)	$X+Y$

EXAMPLES

Example 2.6

Derive the generalised addition law (1.16) for events $A_1, A_2, ...A_n$ using indicator r.v.s.

Solution Let X_i be the indicator r.v. for A_i. Then we deduce the following indicator r.v.s:

$$\begin{array}{rll} 1-X_i & \text{for} & \overline{A}_i; \\ (1-X_1)....(1-X_n) & \text{for} & \overline{A}_1 \cap .... \cap \overline{A}_n = \overline{A_1 \cup .... \cup A_n} \\ 1-(1-X_1)....(1-X_n) & \text{for} & A_1 \cup .... \cup A_n. \end{array}$$

Hence

$$\begin{aligned} \mathrm{P}(A_1 \cup ....A_n) &= \mathrm{E}[1-\{1-(1-X_1)....(1-X_n)\}] \\ &= \mathrm{E}[\sum_i X_i - \sum_{i<j} X_iX_j + \sum_{i<j<k} X_iX_jX_k - .... + (-1)^{n+1}X_1....X_n] \\ &= \sum_i \mathrm{E}(X_i) - \sum_{i<j} \mathrm{E}(X_iX_j) + ... + (-1)^{n+1}\mathrm{E}(X_1....X_n) \\ &= \sum_i \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i \cap A_j) + ... + (-1)^{n+1}\mathrm{P}(A_1 \cap .... \cap A_n). \end{aligned}$$

The last line follows because, for example,

$$\mathrm{E}(X_iX_j) = 1.1.\mathrm{P}(X_i=1, X_j=1) + 0 = \mathrm{P}(A_i \cap A_j) \qquad \square$$

Example 2.7 *(Lift problem)*

Use indicator r.v.s to solve the lift problem with 3 people and floors (Ex. 1.8).

Solution Let

$$X = \begin{cases} 1, & \text{if exactly one person gets off at each floor} \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y_i = \begin{cases} 1, & \text{if no-one gets off at floor } i \\ 0, & \text{otherwise.} \end{cases}$$

Then $X = (1 - Y_1)(1 - Y_2)(1 - Y_3)$ and

$$\begin{aligned} \text{P(one person gets off at each floor)} &= \text{P}(X = 1) \\ &= \text{E}(X) \\ &= \text{E}[1 - \{Y_1 + Y_2 + Y_3\} + \{Y_1Y_2 + Y_1Y_3 + Y_2Y_3\} - Y_1Y_2Y_3] \\ &= 1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23} - p_{123} \end{aligned}$$

where

$$\begin{aligned} p_i &= \text{P}(Y_i = 1) &&= \left(\tfrac{2}{3}\right)^3, && i = 1, 2, 3 \\ p_{ij} &= \text{P}(Y_i = 1, Y_j = 1) &&= \left(\tfrac{1}{3}\right)^3, && i \neq j \\ p_{123} &= \text{P}(Y_1 = 1, Y_2 = 1, Y_3 = 1) &&= 0 \end{aligned}$$

So the required probability is

$$1 - 3\left(\frac{2}{3}\right)^3 + 3\left(\frac{1}{3}\right)^3 = \frac{2}{9}.$$

□

Example 2.8

Consider the generalisation of the tokens-in-cereal collecting problem (Ex. 1.2) to N different card types.
(a) Find the expected number of different types of cards that are contained in a collection of n cards.
(b) Find the expected number of cards a family needs to collect before obtaining a complete set of at least one of each type.
(c)Find the expected number of cards of a particular type which a family will have by the time a complete set has been collected.

Solution

(a) Let

$$X = \text{number of different types in a collection of } n \text{ cards}$$

and let

$$I_i = \begin{cases} 1, & \text{if at least one type } i \text{ card in collection} \\ 0, & \text{otherwise.} \end{cases} \qquad i = 1, ..., n.$$

Then

$$X = I_1 + ... + I_N.$$

Now

$$\begin{aligned} \text{E}(I_i) &= \text{P}(I_i = 1) = 1 - \text{P(no type } i \text{ cards in collection of } n) \\ &= 1 - \left(\frac{N-1}{N}\right)^n, \quad i = 1, ..., N. \end{aligned}$$

So

$$\text{E}(X) = \sum_{i=1}^{N} \text{E}(I_i) = N\left[1 - \left(\frac{N-1}{N}\right)^n\right].$$

(b) Let

$$X = \text{number of cards collected before a complete set is obtained,}$$

and

$$Y_i = \text{number of } \textit{additional} \text{ cards that need to be obtained after } i \text{ distinct cards have been collected, in order to obtain another distinct type } (i = 0, ..., N-1).$$

When i distinct cards have already been collected, a new card obtained will be of a distinct type with probability $(N-i)/N$. So Y_i is a geometric r.v. with parameter $\frac{(N-i)}{N}$, i.e.

$$\mathrm{P}(Y_i = k) = \left(\frac{N-i}{N}\right)\left(\frac{i}{N}\right)^{k-1}, \quad k \geq 1,$$

Hence from (2.21b)

$$\mathrm{E}(Y_i) = \frac{N}{N-i}.$$

Now

$$X = Y_0 + Y_1 + \cdots + Y_{N-1}.$$

So

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=0}^{N-1} \mathrm{E}(Y_i) = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1} \\ &= N\left(1 + \cdots + \frac{1}{N-1} + \frac{1}{N}\right). \end{aligned}$$

(c) Let

$$X_i = \text{number of cards of type } i \text{ acquired.}$$

Then

$$\mathrm{E}(X) = \mathrm{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{E}(X_i).$$

By symmetry, $\mathrm{E}(X_i)$ will be the same for all i, so

$$\mathrm{E}(X_i) = \frac{\mathrm{E}(X)}{N} = \left(1 + \cdots + \frac{1}{N-1} + \frac{1}{N}\right)$$

from part (b). ◊

Example 2.9

Suppose that $(X_1, ..., X_p)$ has the multinomial distribution

$$P(X_1 = x_1, ..., X_k = x_k) = \frac{n!}{x_1!x_2!...x_k!} {p_1}^{x_1}...{p_k}^{x_k},$$

where $\sum_{i=1}^{k} x_i = n, \quad \sum_{i=1}^{k} p_i = 1$. Show that

$$\text{Cov}(X_i, X_j) = -np_ip_j, \qquad i \neq j.$$

Solution Consider the rth trial: let

$$I_{ri} = \begin{cases} 1, & \text{if } r\text{th trial has outcome } E_i \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\text{Cov}(I_{ri}, I_{rj}) = \text{E}(I_{ri}.I_{rj}) - \text{E}(I_{ri}).\text{E}(I_{rj}).$$

Now

$$\begin{aligned} \text{E}(I_{ri}.I_{rj}) &= 0.0\text{P}(I_{ri} = 0, I_{rj} = 0) \\ &\quad +0.1\text{P}(I_{ri} = 0, I_{rj} = 1) \\ &\quad +1.0\text{P}(I_{ri} = 1, I_{rj} = 0) \\ &\quad +1.1\text{P}(I_{ri} = 1, I_{rj} = 1) \\ &= 0, \quad i \neq j \qquad (\text{since P}(I_{ri} = 1, I_{rj} = 1) = 0 \text{ when } i \neq j). \end{aligned}$$

So

$$\text{Cov}(I_{ri}, I_{rj}) = -\text{E}(I_{ri}).\text{E}(I_{rj}) = -p_ip_j, \quad i \neq j.$$

Also, from the independence of the trials,

$$\text{Cov}(I_{ri}, I_{sj}) = 0 \quad \text{when } r \neq s.$$

Now the number of times that E_i occurs in the n trials is

$$X_i = I_{1i} + I_{2i} + \cdots + I_{ni}.$$

So

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \text{Cov}(\sum_{r=1}^{n} I_{ri}, \sum_{s=1}^{n} I_{sj}) \\ &= \sum_{r=1}^{n}\sum_{s=1}^{n} \text{Cov}(I_{ri}, I_{sj}) \\ &= \sum_{r=1}^{n} \text{Cov}(I_{ri}, I_{rj}) \\ &= \sum_{r=1}^{n} (-p_ip_j) \\ &= -np_ip_j, \qquad i \neq j. \end{aligned}$$

(This negative correlation is not unexpected, for we anticipate that, when X_i is large, X_j will tend to be small). ◊