

Some discrete random variables

We now look at five types of discrete random variables, each depending on one or more parameters. We describe for each type the situations in which it arises, and give the p.m.f., the expected value, and the variance. If the variable is tabulated in the *New Cambridge Statistical Tables* [1], we give the table number, and some examples of using the tables. You should have a copy of the tables to follow the examples.

A summary of this information is given on the course information sheet entitled *Discrete random variables*. Make sure that you have a copy of this sheet too.

Before we begin, a comment on the *New Cambridge Statistical Tables* [1]. They don't give the probability mass function (or p.m.f.), but a closely related function called the *cumulative distribution function*. It is defined for a discrete random variable as follows.

Let X be a random variable taking values a_1, a_2, \dots, a_n . We assume that these are arranged in ascending order: $a_1 < a_2 < \dots < a_n$. The *cumulative distribution function*, or c.d.f., of X is given by

$$F_X(a_i) = P(X \leq a_i).$$

We see that it can be expressed in terms of the p.m.f. of X as follows:

$$F_X(a_i) = P(X = a_1) + \dots + P(X = a_i) = \sum_{j=1}^i P(X = a_j).$$

In the other direction, we can recover the p.m.f. from the c.d.f.:

$$P(X = a_i) = F_X(a_i) - F_X(a_{i-1}).$$

Usually the values of a random variable in the tables are integers starting at 0 or sometimes 1. In this case, the equations become

$$\begin{aligned}
F_X(k) &= P(X \leq k) \\
&= P(X = 0) + P(X = 1) + \cdots + P(X = k), \\
P(X = k) &= F_X(k) - F_X(k - 1), \\
P(X \geq k) &= 1 - F_X(k - 1), \\
P(k \leq X \leq l) &= F_X(l) - F_X(k - 1).
\end{aligned}$$

For example, the last equation holds because we obtain the probability that $k \leq X \leq l$ by working out the probability of the values $0, 1, \dots, l$ and subtracting the ones we don't want: $0, 1, \dots, k - 1$.

We won't use the c.d.f. of a discrete random variable except for looking up the tables. It is much more important for continuous random variables!

Bernoulli random variable Bernoulli(p)

A Bernoulli random variable is the simplest type of all. It only takes two values, 0 and 1. So its p.m.f. looks as follows:

$$\begin{array}{c|cc}
x & 0 & 1 \\
\hline
P(X = x) & q & p
\end{array}$$

Here, p is the probability that $X = 1$; it can be any number between 0 and 1. Necessarily q (the probability that $X = 0$) is equal to $1 - p$. So p determines everything.

For a Bernoulli random variable X , we sometimes describe the experiment as a 'trial', the event $X = 1$ as 'success', and the event $X = 0$ as 'failure'.

For example, if a biased coin has probability p of coming down heads, then the number of heads that we get when we toss the coin once is a Bernoulli(p) random variable.

More generally, let A be any event in a probability space \mathcal{S} . With A , we associate a random variable I_A (remember that a random variable is just a function on \mathcal{S}) by the rule

$$I_A(s) = \begin{cases} 1 & \text{if } s \in A; \\ 0 & \text{if } s \notin A. \end{cases}$$

The random variable I_A is called the *indicator variable* of A , because its value indicates whether or not A occurred. It is a Bernoulli(p) random variable, where $p = P(A)$. (The event $I_A = 1$ is just the event A .) Some people write $\mathbf{1}_A$ instead of I_A . This shows that Bernoulli random variables are essentially the same thing as events, so that if we wanted to we could do all probability theory with random variables and never mention events!

Calculation of the expected value and variance of a Bernoulli random variable is easy. Let $X \sim \text{Bernoulli}(p)$. (Remember that \sim means “has the same p.m.f. as” or “is distributed as”.)

$$E(X) = 0 \cdot q + 1 \cdot p = p;$$

$$\text{Var}(X) = 0^2 \cdot q + 1^2 \cdot p - p^2 = p - p^2 = pq.$$

(Remember that $q = 1 - p$.)

Binomial random variable $\text{Bin}(n, p)$

Remember that for a Bernoulli random variable, we describe the event $X = 1$ as a ‘success’. Now a *binomial random variable* counts the number of successes in n independent trials each associated with a Bernoulli(p) random variable.

For example, suppose that we have a biased coin for which the probability of heads is p . We toss the coin n times and count the number of heads obtained. This number is a $\text{Bin}(n, p)$ random variable.

A $\text{Bin}(n, p)$ random variable X takes the values $0, 1, 2, \dots, n$, and the p.m.f. of X is given by

$$P(X = m) = {}^n C_m q^{n-m} p^m = b(m; n, p)$$

for $m = 0, 1, 2, \dots, n$, where $q = 1 - p$. This is because there are ${}^n C_m$ different ways of obtaining m heads in a sequence of n throws (the number of choices of the m positions in which the heads occur), and the probability of getting m heads and $n - m$ tails in a particular order is $q^{n-m} p^m$.

Note that we have given a formula rather than a table here. For small values we could tabulate the results; for example, for $\text{Bin}(3, p)$:

m	0	1	2	3
$P(X = m)$	q^3	$3q^2 p$	$3qp^2$	p^3

Adding up the probabilities gives $q^3 + 3q^2 p + 3qp^2 + p^3 = (q + p)^3 = 1$, since $q = 1 - p$. Moreover, we find that

$$\begin{aligned} E(X) &= 0 \cdot q^3 + 1 \cdot 3q^2 p + 2 \cdot 3qp^2 + 3 \cdot p^3 \\ &= 3p(q^2 + 2qp + p^2) \\ &= 3p(q + p)^2 \\ &= 3p, \\ \text{Var}(X) &= 0^2 \cdot q^3 + 1^2 \cdot 3q^2 p + 2^2 \cdot 3qp^2 + 3^2 \cdot p^3 - (3p)^2 \\ &= 3p(q^2 + 4qp + 3p^2) - 9p^2 \end{aligned}$$

$$\begin{aligned}
&= 3p(q+p)(q+3p) - 9p^2 \\
&= 3p(q+3p) - 9p^2 \\
&= 3pq.
\end{aligned}$$

For arbitrary n , when we add up all the probabilities in the table, we get

$$\sum_{m=0}^n {}^n C_m q^{n-m} p^m = (q+p)^n = 1,$$

as it should be: here we used the *binomial theorem*

$$(x+y)^n = \sum_{m=0}^n {}^n C_m x^{n-m} y^m.$$

(This argument explains the name of the binomial random variable!)

The general formula for expected value and variance of $X \sim \text{Bin}(n, p)$ is:

$$E(X) = np, \quad \text{Var}(X) = npq.$$

There are three ways to prove this. The first method is straightforward but hard, taking the explicit calculations that we did above for $n = 3$ and making them work for general n . The second method is more sophisticated, but relatively easy; it also works for very many random variables whose values are integers. However, you can skip it if you wish: I have set it in smaller type for this reason. We will see a third way (yet more sophisticated but even easier) when we have done joint distributions.

The second method uses a gadget called the *probability generating function*. We only use it here for calculating expected values and variances, but if you learn more probability theory you will see other uses for it. Let X be a random variable whose values are non-negative integers. (We don't insist that it takes all possible values; this method is fine for the binomial $\text{Bin}(n, p)$, which takes values between 0 and n .) To save space, we write p_m for the probability $P(X = m)$. Now the *probability generating function* of X is the power series

$$G_X(x) = \sum p_m x^m.$$

(The sum is over all values m taken by X .) It may be abbreviated to $G(x)$ if it is obvious which random variable we are talking about.

We use the notation $[F(x)]_{x=1}$ for the result of substituting $x = 1$ in the series $F(x)$.

Proposition Let $G(x)$ be the probability generating function of a random variable X . Then

- (a) $[G(x)]_{x=1} = 1$;
- (b) $E(X) = \left[\frac{d}{dx} G(x) \right]_{x=1}$;
- (c) $\text{Var}(X) = \left[\frac{d^2}{dx^2} G(x) \right]_{x=1} + E(X) - E(X)^2$.

Part (a) is just the statement that probabilities add up to 1: when we substitute $x = 1$ in the power series for $G(x)$ we just get $\sum p_m$.

For part (b), when we differentiate the series term-by-term (you will learn later in Analysis that this is OK), we get

$$\frac{d}{dx}G(x) = \sum mp_mx^{m-1}.$$

Now putting $x = 1$ in this series we get

$$\sum mp_m = E(X).$$

For part (c), differentiating twice gives

$$\frac{d^2}{dx^2}G(x) = \sum m(m-1)p_mx^{m-2}.$$

Now putting $x = 1$ in this series we get

$$\sum m(m-1)p_m = \sum m^2p_m - \sum mp_m = E(X^2) - E(X).$$

Adding $E(X)$ and subtracting $E(X)^2$ gives $E(X^2) - E(X)^2$, which by definition is $\text{Var}(X)$.

Now let us apply this to the binomial random variable $X \sim \text{Bin}(n, p)$. We have

$$p_m = P(X = m) = {}^nC_m q^{n-m} p^m,$$

so the probability generating function is

$$\sum_{m=0}^n {}^nC_m q^{n-m} p^m x^m = (q + px)^n,$$

by the Binomial Theorem. Putting $x = 1$ gives $(q + p)^n = 1$, in agreement with the Proposition.

Differentiating once, using the Chain Rule, we get $np(q + px)^{n-1}$. Putting $x = 1$ we find that

$$E(X) = np.$$

Differentiating again, we get $n(n-1)p^2(q + px)^{n-2}$. Putting $x = 1$ gives $n(n-1)p^2$. Now adding $E(X) - E(X)^2$, we get

$$\text{Var}(X) = n(n-1)p^2 + np - n^2p^2 = np - np^2 = npq.$$

The binomial random variable is tabulated in Table 1 of the *Cambridge Statistical Tables* [1]. As explained earlier, the tables give the cumulative distribution function.

For example, suppose that the probability that a certain coin comes down heads is 0.45. If the coin is tossed 15 times, what is the probability of five or fewer heads? Turning to the page $n = 15$ in Table 1 and looking at the row 0.45, you read off the answer 0.2608. What is the probability of exactly five heads? This is $P(5 \text{ or fewer}) - P(4 \text{ or fewer})$, and from tables the answer is $0.2608 - 0.1204 = 0.1404$.

The tables only go up to $p = 0.5$. For larger values of p , use the fact that the number of failures in $\text{Bin}(n, p)$ is equal to the number of successes in $\text{Bin}(n, 1 - p)$. So the probability of five heads in 15 tosses of a coin with $p = 0.55$ is $0.9745 - 0.9231 = 0.0514$. More formally, if $X \sim \text{Bin}(15, 0.55)$, and $Y = 15 - X$, then $Y \sim \text{Bin}(15, 0.45)$.

Another interpretation of the binomial random variable concerns sampling. Suppose that we have N balls in a box, of which M are red. We sample n balls from the box with replacement; let the random variable X be the number of red balls in the sample. What is the distribution of X ? Since each ball has probability M/N of being red, and different choices are independent, $X \sim \text{Bin}(n, p)$, where $p = M/N$ is the proportion of red balls in the sample.

What about sampling without replacement? This leads us to our next random variable:

Hypergeometric random variable $\text{Hg}(n, M, N)$

Suppose that we have N balls in a box, of which M are red. We sample n balls from the box *without replacement*. Let the random variable X be the number of red balls in the sample. Such an X is called a *hypergeometric* random variable $\text{Hg}(n, M, N)$.

The random variable X can take any of the values $0, 1, 2, \dots, n$. Its p.m.f. is given by the formula

$$P(X = m) = \frac{{}^M C_m \times {}^{N-M} C_{n-m}}{{}^N C_n}.$$

For the number of samples of n balls from N is ${}^N C_n$; the number of ways of choosing m of the M red balls and $n - m$ of the $N - M$ others is ${}^M C_m \times {}^{N-M} C_{n-m}$; and all choices are equally likely.

The expected value and variance of a hypergeometric random variable are as follows (we won't go into the proofs):

$$E(X) = n \left(\frac{M}{N} \right), \quad \text{Var}(X) = n \left(\frac{M}{N} \right) \left(\frac{N-M}{N} \right) \left(\frac{N-n}{N-1} \right).$$

You should compare these to the values for a binomial random variable. If we let $p = M/N$ be the proportion of red balls in the box, then $E(X) = np$, and $\text{Var}(X)$ is equal to npq multiplied by a 'correction factor' $(N - n)/(N - 1)$.

In particular, if the numbers M and $N - M$ of red and non-red balls in the box are both very large compared to the size n of the sample, then the difference between sampling with and without replacement is very small, and indeed the 'correction factor' is close to 1. So we can say that $\text{Hg}(n, M, N)$ is approximately $\text{Bin}(n, M/N)$ if n is small compared to M and $N - M$.

Consider the example from the last notes of choosing five sheep from 24, of which 6 are shorn. The number X of shorn sheep in the sample is a $\text{Hg}(5, 6, 24)$ random

variable. We calculated in the last notes that $E(X) = 1.2501$ and $\text{Var}(X) = 0.7743$, but noted that these figures were affected by rounding errors. The formulae above show that the exact values should be

$$E(X) = \frac{5}{4} \quad \text{and} \quad \text{Var}(X) = \frac{5}{4} \times \frac{3}{4} \times \frac{19}{23} = \frac{285}{368}.$$

Geometric random variable $\text{Geom}(p)$

The geometric random variable is like the binomial but with a different stopping rule. We have again a coin whose probability of heads is p . Now, instead of tossing it a fixed number of times and counting the heads, we toss it until it comes down heads for the first time, and count the number of times we have tossed the coin. Thus, the values of the variable are the positive integers $1, 2, 3, \dots$. (In theory we might never get a head and toss the coin infinitely often, but if $p > 0$ this possibility is ‘infinitely unlikely’, i.e. has probability zero, as we will see.) We always assume that $0 < p < 1$.

More generally, the number of independent Bernoulli trials required until the first success is obtained is a geometric random variable.

The p.m.f. of a $\text{Geom}(p)$ random variable is given by

$$P(X = m) = q^{m-1}p,$$

where $q = 1 - p$. For the event $X = m$ means that we get tails on the first $m - 1$ tosses and heads on the m th, and this event has probability $q^{m-1}p$, since ‘tails’ has probability q and different tosses are independent.

Let’s add up these probabilities:

$$\sum_{m=1}^{\infty} q^{m-1}p = p + qp + q^2p + \dots = p(1 + q + q^2 + \dots) = \frac{p}{1 - q} = 1,$$

since the series in parentheses is a geometric progression with first term 1 and common ratio q , where $q < 1$. (Just as the binomial theorem shows that probabilities sum to 1 for a binomial random variable, and gives its name to the random variable, so the geometric progression does for the geometric random variable.)

We calculate the expected value and the variance using the probability generating function. If $X \sim \text{Geom}(p)$, the result will be that

$$E(X) = 1/p, \quad \text{Var}(X) = q/p^2.$$

We have

$$G(x) = \sum_{m=1}^{\infty} q^{m-1}px^m = \frac{px}{1 - qx},$$

again by summing a geometric progression. Differentiating, we get

$$\frac{d}{dx}G(x) = \frac{(1-qx)p + pxq}{(1-qx)^2} = \frac{p}{(1-qx)^2}.$$

Putting $x = 1$, we obtain

$$E(X) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Differentiating again gives $2pq/(1-qx)^3$, so

$$\text{Var}(X) = \frac{2pq}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

For example, if we toss a fair coin until heads is obtained, the expected number of tosses until the first head is 2 (so the expected number of tails is 1); and the variance of this number is also 2.

Poisson random variable Poisson(λ)

The Poisson random variable, unlike the ones we have seen before, is very closely connected with continuous things.

Suppose that ‘incidents’ occur at random times, but at a steady rate overall. The best example is radioactive decay: atomic nuclei decay randomly, but the average number λ which will decay in a given interval is constant. The Poisson random variable X counts the number of ‘incidents’ which occur in a given interval. So if, on average, there are 2.4 nuclear decays per second, then the number of decays in one second starting now is a Poisson(2.4) random variable.

Another example might be the number of telephone calls a minute to a busy telephone number, or the number of people joining the queue at the bus-stop in the next minute.

The p.m.f. for a Poisson(λ) variable X is given by the formula

$$P(X = m) = e^{-\lambda} \frac{\lambda^m}{m!}$$

for $m = 0, 1, \dots$. It is derived from the binomial distribution. I do not expect you to reproduce this derivation, so I am putting it in small type.

Suppose that the incidents happen at the rate of λ per minute. Choose n large enough that it is very unlikely for two or more incidents to happen in one $(1/n)$ -th of a minute. Then the number that happen in such a small interval of time is approximately Bernoulli(p) for some suitable p . If the incidents in the n different parts of the minute are mutually independent and X is the number of incidents in one minute then $X \sim \text{Bin}(n, p)$. Then $E(X) = np$. But we know that $E(X) = \lambda$ so $p = \lambda/n$.

Now,

$$\begin{aligned}
 P(X = m) &= {}^n C_m \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\
 &= \frac{n!}{m!(n-m)!} \frac{\lambda^m}{n^m} \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^m} \\
 &= \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-m+1)}{n \cdot n \cdot n \cdots n} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^m}
 \end{aligned}$$

Now let n tend to ∞ . Each of the m ratios in the first term tends to 1. The second term has no n in it, so it stays as $\lambda^m/m!$. For the third term, we need to use the fact that

$$\left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}$$

as $n \rightarrow \infty$, which you will learn in Calculus. In the fourth term, $\lambda/n \rightarrow 0$, so $(1 - \lambda/n) \rightarrow 1$ so the whole term tends to $1/1^m$, which is 1. Thus the limiting value of $P(X = m)$ is indeed $e^{-\lambda}\lambda^m/m!$.

In the lectures I gave the example of bristles falling out of my hairbrush when I brush my (thick and tangly) hair. Initially there are N bristles, so it is reasonable to suppose that the random variable X , which counts how many bristles fall out when I brush my hair, is $\text{Bin}(N, p)$ for some probability p . But what happens when there are only 10 bristles left? The X should be a random variable with first parameter 10, but what should the probability be? It seems reasonable to suppose that I apply constant force K when I brush my hair. With N bristles, that force is spread over all of them, so p should be (proportional to) K/N . When there are only 10 bristles, the same force is spread over just those 10 bristles, so the probability should be (proportional to) $K/10$. Thus, in general, when there are n bristles left then

$$X \sim \text{Bin}\left(n, \frac{K}{n}\right).$$

Replacing K by λ and taking the limit as $n \rightarrow \infty$ gives the same result as above.

Let's check that these probabilities add up to one. We get

$$\left(\sum_{m=0}^{\infty} \frac{\lambda^m}{m!}\right) e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1,$$

since the expression in brackets is the sum of the exponential series.

By analogy with what happened for the binomial and geometric random variables, you might have expected that this random variable would be called 'exponential'. Unfortunately, this name has been given to a closely-related continuous random variable which we will meet later. However, if you speak a little French, you might use as a mnemonic the fact that if I go fishing, and the fish are biting at the rate of λ per hour on average, then the number of fish I will catch in the next hour is a $\text{Poisson}(\lambda)$ random variable.

The expected value and variance of a Poisson(λ) random variable X are given by

$$E(X) = \text{Var}(X) = \lambda.$$

Again we use the probability generating function. If $X \sim \text{Poisson}(\lambda)$, then

$$G(x) = \sum_{m=0}^{\infty} e^{-\lambda} \frac{(\lambda x)^m}{m!} = e^{\lambda(x-1)},$$

again using the series for the exponential function.

Differentiation gives $\lambda e^{\lambda(x-1)}$, so $E(X) = \lambda$. Differentiating again gives $\lambda^2 e^{\lambda(x-1)}$, so

$$\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

The line graphs on the next page illustrate how the binomial distribution tends towards the Poisson when $\lambda = 5$.

The cumulative distribution function of a Poisson random variable is tabulated in Table 2 of the *New Cambridge Statistical Tables*. So, for example, we find from the tables that, if 2.4 fish bite per hour on average, then the probability that I will catch no fish in the next hour is 0.0907, while the probability that I catch at five or fewer is 0.9643 (so that the probability that I catch six or more is 0.0357).

There is another situation in which the Poisson distribution arises. Suppose I am looking for some very rare event which only occurs once in 1000 trials on average. So I conduct 1000 independent trials. How many occurrences of the event do I see? This number is really a binomial random variable $\text{Bin}(1000, 1/1000)$. But we have seen above that this is $\text{Poisson}(1)$, to a very good approximation. So, for example, the probability that the event doesn't occur is about $1/e$.

The general rule is:

If n is large, p is small, and $np = \lambda$, then $\text{Bin}(n, p)$ can be approximated by $\text{Poisson}(\lambda)$.

For example, if the lake contains 2400 fish and if each fish bites independently with probability $1/1000$ in an hour, then the number of fish I catch in an hour is $\text{Bin}(2400, 1/1000)$, which is approximately $\text{Poisson}(2.4)$.

[1] D. V. Lindley and W. F. Scott, *New Cambridge Statistical Tables*, Cambridge University Press.

