MAS187/AEF258

University of Newcastle upon Tyne

2005-6

Contents

Chapter 8

The Binomial and Poisson Distributions

8.1 The Binomial Distribution

8.1.1 Introduction

In many surveys and experiments we collect data in the form of counts. For example, the number of people in the survey who bought a CD in the past month, the number of people who said they would vote Labour at the next election, the number of defective items in a sample taken from a production line, and so on. All these variables have common features:

1. Each person/item has only two possible (exclusive) responses (Yes/No, Defective/Not defective etc)
   – this is referred to as a *trial* which results in a *success* or *failure*

2. The survey/experiment takes the form of a random sample
   – the responses are independent.

Further suppose that the true probability of a success in the population is p (in which case the probability of a failure is $1 - p$). We are interested in the random variable X, the total number of successes out of n trials. This random variable has a probability distribution in which the probability that $X = r$, that is we get r successes in our n trials, is

$$P(X = r) = {}^n\mathrm{C}_r\, p^r (1 - p)^{n-r}, \quad r = 0, 1, \ldots, n.$$

These probabilities describe how likely we are to get r out of n successes from independent trials, each with success probability p. Note that any number raised to the power zero is one, for example, $0.3^0 = 1$ and $0.654^0 = 1$.

This distribution is known as the *binomial distribution* with index n and probability p. We write this as $X \sim Bin(n, p)$.

8.1.2 Calculating probabilities

For example, we can calculate the probability of getting r threes from 4 rolls of a die as follows. Each roll of the die is a trial which gives a three (success) or “not a three” (failure). The probability of a success is $p = P(\text{three}) = 1/6$. We have $n = 4$ independent trials (rolls of the die). If X is the number of threes obtained then $X \sim Bin(4, 1/6)$ and so

$$
\begin{aligned}
P(X = 0) &= {}^4\mathrm{C}_0 \left(\frac{1}{6}\right)^0 \left(1 - \frac{1}{6}\right)^4 = \left(\frac{5}{6}\right)^4 = 0.4823 \\
P(X = 1) &= {}^4\mathrm{C}_1 \left(\frac{1}{6}\right)^1 \left(1 - \frac{1}{6}\right)^3 = 4 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^3 = 0.3858 \\
P(X = 2) &= {}^4\mathrm{C}_2 \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^2 = 6 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2 = 0.1157 \\
P(X = 3) &= {}^4\mathrm{C}_3 \left(\frac{1}{6}\right)^3 \left(1 - \frac{1}{6}\right)^1 = 4 \times \left(\frac{1}{6}\right)^3 \times \frac{5}{6} = 0.0154 \\
P(X = 4) &= {}^4\mathrm{C}_4 \left(\frac{1}{6}\right)^4 \left(1 - \frac{1}{6}\right)^0 = \left(\frac{1}{6}\right)^4 = 0.0008.
\end{aligned}
$$

This probability distribution shows that most of the time we would get either 0 or 1 successes but, for example, 4 successes would be quite rare.

Consider another example. A salesperson has a 50% chance of making a sale on a customer visit and she arranges 6 visits in a day. What are the probabilities of her making 0,1,2,3,4,5 and 6 sales? Let X denote the number of sales. Assuming the visits result in sales independently, $X \sim Bin(6, 0.5)$ and

No. of sales r	Probability $P(X = r)$	Cumulative Probability $P(X \leq r)$
0	0.015625	0.015625
1	0.093750	0.109375
2	0.234375	0.343750
3	0.312500	0.656250
4	0.234375	0.890625
5	0.093750	0.984375
6	0.015625	1.000000
sum	1.000000	

The formula for binomial probabilities enables us to calculate values for $P(X = r)$. From these, it is straightforward to calculate cumulative probabilities such as the probability of making no more than 2 sales:

$$
\begin{aligned}
P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= 0.015625 + 0.09375 + 0.234375 = 0.34375.
\end{aligned}
$$

These cumulative probabilities are also useful in calculating probabilities such as that of making more than 1 sale:

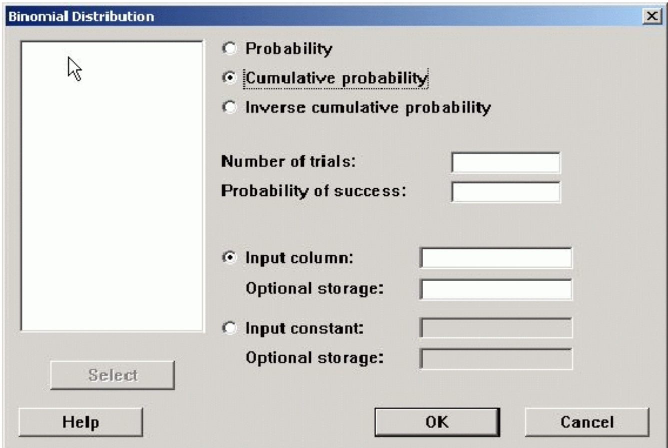
$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.109375 = 0.890625.$$

Fortunately, binomial probabilities can be found in sets of Statistical Tables or calculated using MINITAB.

Probabilities of binomial events can be calculated in MINITAB as follows. If $X \sim Bin(n, p)$ then probabilities $P(X = r)$ and cumulative probabilities $P(X \leq r)$ can be obtained using the following commands:

```
Calc > Probability Distributions > Binomial
```

This opens the following dialogue box



1. Select Probability for $P(X = r)$ or Cumulative Probability for $P(X \leq r)$.
2. Enter the Number of trials (n).
3. Enter the Probability of success (p).
4. Check the `Input constant:` button
5. Enter the Input constant (r)
6. Click OK.

8.1.3 Mean and variance

If X is a random variable with a binomial $Bin(n, p)$ distribution then its mean and variance are

$$E(X) = np, \qquad Var(X) = np(1 - p).$$

For example, if $X \sim Bin(4, 1/6)$ then

$$E(X) = np = 4 \times \frac{1}{6} = \frac{2}{3} = 0.6667$$

and

$$Var(X) = np(1-p) = 4 \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{9} \simeq 0.5556.$$

Also

$$SD(X) = \sqrt{Var(X)} = \sqrt{\frac{5}{9}} = 0.7454.$$

8.2 The Poisson Distribution

8.2.1 Introduction

The *Poisson distribution* is a very important discrete probability distribution which arises in many different contexts. We can think of a Poisson distribution as what becomes of a binomial distribution if we keep the mean fixed but let n become very large and p become very small, i.e. a large number of trials with a small probability of success in each. In general, it is used to model data which are counts of (random) events in a certain area or time interval, without a known fixed upper limit.

For example, consider the number of calls made in a 1 minute interval to an Internet service provider (ISP). The ISP has thousands of subscribers, but each one will call with a very small probability. If the ISP knows that on average 5 calls will be made in the interval, the actual number of calls will be a Poisson random variable, with mean 5.

If X is a random variable with a Poisson distribution with parameter λ (Greek lower case *lambda*) then the probability that $X = r$ is

$$P(X = r) = \frac{\lambda^r\, e^{-\lambda}}{r!}, \quad r = 0, 1, 2, \ldots\ .$$

We write $X \sim Po(\lambda)$. The parameter λ has a very simple interpretation as the rate at which events occur. The distribution has mean and variance

$$E(X) = \lambda, \qquad Var(X) = \lambda.$$

8.2.2 Calculating probabilities

Returning to the ISP example, suppose we want to know the probabilities of different numbers of calls made to the ISP. Let X be the number of calls made in a minute. Then $X \sim P(5)$ and, for example, the probability of receiving 4 calls is

$$P(X = 4) = \frac{5^4 e^{-5}}{4!} = 0.1755.$$

We can use the formula for Poisson probabilities to calculate the probability of all possible outcomes:

r	Probability $P(X=r)$	Cumulative Probability $P(X \leq r)$
0	0.0067	0.0067
1	0.0337	0.0404
2	0.0843	0.1247
3	0.1403	0.2650
4	0.1755	0.4405
5	0.1755	0.6160
6	0.1462	0.7622
7	0.1044	0.8666
8	0.0653	0.9319
9	0.0363	0.9682
10	0.0181	0.9863
$\vdots$	$\vdots$	$\vdots$
sum	1.000000	

Therefore the probability of receiving between 2 and 8 calls is

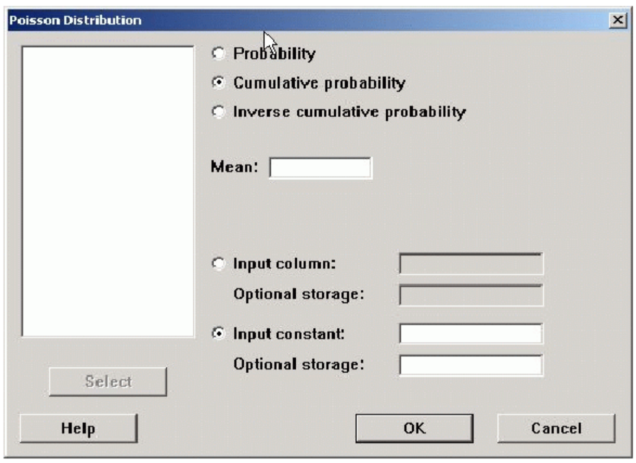
$$P(2 \leq X \leq 8) = P(X \leq 8) - P(X \leq 1) = 0.9319 - 0.0404 = 0.8915$$

and so this event is very likely. Probability calculations such as this enable ISPs to assess the likely demand for their service and hence the resources they need to provide the service.

Probabilities of Poisson events can be calculated in MINITAB as follows. If $X \sim Po(\lambda)$ then probabilities $P(X = r)$ and cumulative probabilities $P(X \leq r)$ can be obtained using the following commands:

```
Calc > Probability Distributions > Poisson
```

This opens the following dialogue box



1. Select Probability for $P(X = r)$ or Cumulative Probability for $P(X \leq r)$.

2. Enter the Mean (λ).

3. Check the `Input constant:` button
4. Enter the Input constant (r)
5. Click OK.

8.2.3 The Poisson distribution as an approximation to the binomial distribution

When we want to calculate probabilities in a binomial distribution with large n and small p it is often convenient to approximate the binomial probabilities by Poisson probabilities. We match the means of the distributions: $\lambda = np$.

For example, an insurance company has 1,000 customers. In a particular month, each customer has a probability of 0.003 of making a claim and all customers are independent. The distribution of the number of claims (assuming no customer will make more than one claim in a month) is then Bin$(1000,\ 0.003)$. This distribution has mean $1000 \times 0.003 = 3$. We can calculate approximate probabilities using the Poisson(3) distribution. For example, the probability that there are no claims in a month is approximately

$$P(X = 0) = \frac{3^0 e^{-3}}{0!} = e^{-3} = 0.050.$$

8.3 Exercises 8

1. An operator at a call centre has 20 calls to make in an hour. History suggests that they will be answered 85% of the time. Let X be the number of answered calls in an hour.

   (a) What probability distribution does X have?

   (b) What are the mean and standard deviation of X?

   (c) Calculate the probability of getting a response exactly 9 times.

   (d) Calculate the probability of getting fewer than 2 responses.

2. Calls are received at a telephone exchange at random times at an average rate of 10 per minute. Let X be the number of calls received in one minute.

   (a) What probability distribution does X have?

   (b) What are the mean and standard deviation of X?

   (c) Calculate the probability that there are 12 calls in one minute.

   (d) Calculate the probability there are no more than 2 calls in a minute.

3. If X_1 and X_2 are independent Poisson random variables with means λ_1 and λ_2 respectively, then $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

   The number of sales made by a small business in a day is a Poisson random variable with mean 2. The number of sales made on one day is independent of the number of sales made on any other day.

   (a) What is the distribution of the total number of sales in a 5-day period?

   (b) What is the probability that the business makes more than 12 sales in a 5-day period?

4. A machine is used to produce components. Each time it produces a component there is a chance that the component will be defective. When the machine is working correctly the probability that a component is defective is 0.05. Sometimes, though, the machine requires adjustment and, when this is the case, the probability that a component is defective is 0.2. Given the state of the machine, components are independent of each other. At the time in question there is a probability of 0.1 that the machine requires adjustment. Components produced by the machine are tested and either accepted or rejected. A component which is not defective is accepted with probability 0.97 and (falsely) rejected with probability 0.03. A defective component is (falsely) accepted with probability 0.15 and rejected with probability 0.85. Given the state of the machine, the acceptance of one component is independent of the acceptance of another component.

   (a) Find the conditional probability that a component is accepted given that the machine is working correctly.

   (b) Find the conditional probability that a component is rejected given that the machine is working correctly.

   (c) Find the conditional probability that a component is accepted given that the machine requires adjustment.

(d) Find the conditional probability that a component is rejected given that the machine requires adjustment.

(e) Find the conditional probability that, out of a sample of 5 components, 2 are accepted and 3 are rejected, given that the machine is working correctly.

(f) Find the conditional probability that, out of a sample of 5 components, 2 are accepted and 3 are rejected, given that the machine requires adjustment.

(g) Find the probability that the machine is working correctly and, out of a sample of 5 components, 2 are accepted and 3 are rejected.

(h) Find the probability that the machine requires adjustment and, out of a sample of 5 components, 2 are accepted and 3 are rejected.

(i) Find the probability that, out of a sample of 5 components, 2 are accepted and 3 are rejected.

(j) Find the conditional probability that the machine is working correctly given that, out of a sample of 5 components, 2 are accepted and 3 are rejected.