Elements of Discrete Probability
3. Important Discrete Distributions

1 Introduction

In this chapter we analyze in some detail some significant distributions that model a large number of concrete situations. As usual, we are basically concerned with discrete random variables. For brevity, we shall refer to "density functions", although the denotation "mass function" is appropriate for discrete random variables.

2 Uniform density

The uniform distribution is, in some sense, the simplest known distribution. Consider a box with n balls in it labeled from 1 to n. Assume that there is equal likelihood of choosing any ball from the box. Let the discrete random variable X denote the number on the ball drawn from the box. The probability that X is equal to any one value is uniformly distributed over $1, \ldots, n$, so

$$\Pr(X = x) = \begin{cases} 1/n, & 1 \leq x \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

More generally, if a random variable X can take on all integer values on the interval $[a, b]$, the corresponding density function is

$$P(X = x) = \begin{cases} \frac{1}{b-a+1}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

3 Bernoulli density

Consider an experiment that can only have one of two outcomes, such as flipping a coin. Such an experiment is known as a *Bernoulli trial*. For example, consider a die rolling experiment. We define random variable X to be 0 if the die comes up even and 1 if the die comes up odd. Such a random variable is sometimes called an *indicator random variable*, and is often quite useful for modelling more complex experiments as we will see later. Often, when referring to Bernoulli trials, we designate one outcome event as "success" and one as "failure", and we let p denote the probability of success. In a Bernoulli trial, clearly one outcome has probability p and the other has probability $1 - p = q$. Such a variable is said to be Bernoulli-distributed.

Notice that the mean value of this distribution is trivially p and that the variance is $(0 - p)^2(1 - p) + (1 - p)^2 p = p(1 - p)$.

4 Binomial density

It is of great interest to analyze experiments consisting of repeated trials, that is, of a sequence of n (independent) executions of the same trial.

Example Consider an experiment consisting of n independent flips of a weighted coin such that flipping heads (call this event H) has probability p and flipping tails (T) has probability $1 - p$. Let S_n be the random variable expressing the number of times we have the outcome H in a sequence of n trials.

What is the probability of S_n taking on any specific value?

Say $S_n = i$. That means that in the sequence of n coin-flips (trials), i of them are H and $n - i$ of them are T. In how many ways can we arrange n trials so that exactly i of them are heads? This is simply $\binom{n}{i}$.

So we know that there are $\binom{n}{i}$ events for which $S_n = i$. Now we need to determine the probability of these events. We know that given some set of i collectively independent events, $\omega_1, \omega_2, \ldots, \omega_i$, $\Pr(\omega_1 \cap \omega_2 \cap \ldots \cap \omega_i) = \Pr(\omega_1) \Pr(\omega_2) \cdots \Pr(\omega_i)$. Using this theorem, the probability of a specific sequence of$i$ events H and $(n - i)$ events T is $p^i(1 - p)^{n-i}$.

This gives us the sought density:

$$B(n, p; i) = \Pr(S_n = i) = \binom{n}{i} p^i (1 - p)^{n-i},$$

which is known as the *binomial density.*

The expectation of the binomial random variable is given by

$$
\begin{aligned}
\mu &= \sum_{j=1}^{n} jB(n,p;i) \\
&= \sum_{j=1}^{n} j\frac{n!}{j!(n-j)!}p^j q^{n-j} \\
&= np\sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!(n-j)!}p^{j-1}q^{n-j} \\
&= np\sum_{i=0}^{n-1} \binom{n-1}{i} p^i q^{n-1-} \\
&= np\sum_{j=0}^{n-1} B(n-1,p;i) = np
\end{aligned}
$$

since $\sum_{j=0}^{n-1} B(n-1,p;i) = 1$. Its variance is given by

$$
\begin{aligned}
\sigma^2 &= \sum_{j=1}^{n} j^2 B(n,p;j) - (np)^2 \\
&= \sum_{j=1}^{n} (j^2 - j)\frac{n!}{j!(n-j)!}p^j q^{n-j} + np - (np)^2 \\
&= n(n-1)p^2 \sum_{j=2}^{n} \frac{(n-2)!}{(j-2)!(n-j)!}p^{j-2}q^{n-j} + np - (np)^2 \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} B(n-2,p;i) + np - (np)^2 = -np^2 + np = npq
\end{aligned}
$$

Typically we are interested in evaluating the probability that a binomially distributed random variable lies within a given range of values, as illustrated by the following example:

Example 1 A machine produces screws, 1% of which are defective, on the average. Find the probability that a box of 200 screws contains at most 4 defective screws.

In this case we model "success" as the nondefective screw (so that $p = .99$) and the production of a box as a sequence of 200 Bernoulli trials. The sought probability is

$$
P = \sum_{j=196}^{200} B(200, 0.99; j)
$$

This is a laborious calculation, because it involves 5 binomial coefficients with very large integers. This tedious calculation yields a probability value 0.9466.

In this case, the range of interest [196,200] is at one extreme of the range [0,200] and is referred to as a *tail*.

Example 2 On the average 10% of air passengers don't show up for their flight. An airline is forced to overbook (of course, with moderation) in order to be profitable. On a 200-seat flight, how many reservations should be accepted in order that the probability of at most 8 vacant seats is at least 0.85%?

The probability of "success" is $p = 0.9$ (the passenger shows up). The unknown number of reservations is the length n of the sequence of Bernoulli trials in which we require 192 successes with probability at least 0.85, i.e., we seek the smallest n for which

$$\sum_{j=192}^{n} B(n, 0.9; j) \geq 0.85$$

This case is more complicated because the evaluation of the tail must be done for increasing values of $n > 200$, stopping at the smallest integer that satisfies the inequality.

The above examples motivate for the convenience of seeking effective ways to approximate the binomial distribution with more tractable functions.

5 Poisson density

Consider the situation of a process that generates, on average, some number λ of occurrences of a specific event in a fixed time span T (for example, 3000 airplane landings in a day at a busy airport). Each event happens randomly (i.e., the spacing between two consecutive happenings is anything but constant), consistently with the given average. Suppose that we wish to observe the process more frequently, say, with a time span T/n. Clearly we expect that the average number of occurrence in this shorter span is λ/n. Assume that we increase the frequency of our observations (i.e., the number n). In each of these short intervals of time we have a probability λ/n of observing events, and $1 - \lambda/n$ of not observing any, and as n increases λ/n becomes the probability of observing *just* one event in the very small interval, the probability of more than 1 being utterly negligible. If we neglect this small probability altogether, then the occurrence of the event is essentially a Bernoulli random variable and our observations are a sequence of repeated trials. What is the probability p_k of observing exactly k events in the n observations? From our knowledge of the binomial distribution this probability is

$$p_k = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

which can be rewritten as

$$\begin{aligned} p_k &= \frac{1}{k!} n(n-1)\dots(n-k+1)\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{n}{n}\frac{n-1}{n}\dots\frac{n-k+1}{n}\left(1-\frac{\lambda}{n}\right)^{-k} \end{aligned}$$

If we now let the integer n grow unbounded, i.e., take the limit for n going to infinity, we obtain:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lim_{n\to\infty}\frac{n}{n}\frac{n-1}{n}\dots\frac{n-k+1}{n} = 1$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$$

so that we may conclude

$$\lim_{n\to\infty}\binom{n}{k}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$

The value of the limit is defined as the *Poisson density* with parameter λ, i.e.,

$$P(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$$

Therefore, we see that the the Poisson is the limiting distribution of the binomial, i.e.,

$$\lim_{n\to\infty} B(n,p;k) = P(k;np)$$

We shall find this conclusion particularly useful, because the Poisson is numerically much more tractable than the binomial. The following example illustrates the use of this technique.

Example 3 A machine produces screws, 1% of which are defective. Find the probability that in a box of 200 screws there are no defectives.

Here we have $n = 200$ Bernoulli trials, with success probability $p = .01$. The probability that there are no defective screws is

$$(1-0.01)^{200} = (.99)^{200} = 0.1340$$

The Poisson approximation to this is given by

$$e^{-200(.01)} = e^{-2} = 0.1353$$

Example 4 Consider a transistor production run. The quality of the manufacturing process is such that on the average 1 out of 10,000 transistors is defective. Suppose that 40,000 transistors are shipped to a customer. What is the probability that the batch contains more than 3 faulty transistors?

Again we model the assembly of the batch as a sequence of 40,000 trials with success probability $1 - 10^{-4}$, so that

$$\begin{aligned}
\Pr(faulty > 3) &= \sum_{j=4}^{40{,}000} \binom{40,000}{j} (1-10^{-4})^j (10^{-4})^{40{,}000-j} \\
&= 1 - \sum_{j=0}^{3} \binom{40,000}{j} (10^{-4})^j (1-10^{-4})^{40{,}000-j}
\end{aligned}$$

The very tedious calculation of the binomial coefficients gives the result 0.5667.... However, noting that $np = 40,000 \times 10^{-4} = 4 = \lambda$, we replace the above calculation with the much simpler one

$$\Pr(faulty > 3) = 1 - e^{-4} \sum_{j=0}^{3} \frac{4^j}{j!}$$

which gives the result 0.5666.... in excellent agreement with the exact evaluation.

It must be stressed, however, that the Poisson distribution models an extremely wide number of phenomena in its own right, such as the number of calls into a telephone hotline in a given time interval, the number of misprints in a book, or the number of field mice per square acre of crop land.

We now evaluate the two first moments of the Poisson distribution.

$$\mu = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum -k = 1^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \lambda e^{-\lambda} sum_{j=0}^{\infty} \lambda^j j! = \lambda$$

which is exactly the value we expected.

$$\begin{aligned}
\sigma^2 &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} - \lambda^2 = e^{-\lambda} \sum_{j=0}^{\infty} (j+1) \frac{\lambda^{j+1}}{j!} - \lambda^2 \\
&= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} + \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}$$

Finally we notice that, while for the Poisson distribution $\mu = \sigma^2 = \lambda$, for the binomial $\mu = np$ and $\sigma^2 = npq$; thus, we expect that the approximation of the binomial by the Poisson will be more accurate when $q \approx 1$, i.e., for small p.