Dr Isabella Gollini

Probability and Statistics

5 Mean and Variance of Probability Distributions

5.1 The mean of a discrete probability distribution

Let X be a discrete random variable with probability distribution (p_r) $(r = 0, 1, 2, \ldots)$.

- From now on we shall use the abbreviation r.v. for "random variable".

If we repeat n times the experiment associated with the r.v. X, so that a random sample $x_1, x_2, \ldots, x_n$ of size n is obtained from the r.v. X, let $f_n(r)$ denote the number of times that the value r is observed $(r = 0, 1, 2, \ldots)$, i.e., $f_n(r)$ is the frequency of the occurrence of the value r in the n repetitions of the experiment. The sample mean $\bar{x} = \sum_{i=1}^{n} x_i/n$, as defined in Section 1.3, may also be written as

$$\bar{x} = \frac{1}{n}\sum_{r=0}^{\infty} r f_n(r) = \sum_{r=0}^{\infty} r \, \frac{f_n(r)}{n}.$$

If we now consider letting the sample size become arbitrarily large, so that $n \to \infty$, then, according to the relative frequency interpretation of probability,

$$\frac{f_n(r)}{n} \to \Pr(X = r) = p_r \qquad (r = 0, 1, 2, \ldots).$$

Hence, as $n \to \infty$,

$$\bar{x} \to \sum_{r=0}^{\infty} r p_r. \tag{1}$$

The formula in Equation (1) may be thought of as the long-term average value that the r.v. X will take if it is observed an arbitrarily large number of times. This leads to the following definition.

Definition

The *mean* μ or *expected value* or *expectation* $E(X)$ of the r.v. X is given by

$$\mu \equiv E(X) = \sum_{r=0}^{\infty} r p_r. \tag{2}$$

- Equivalently, we may write $\mu = \sum_{r=1}^{\infty} r p_r$.
- We may also refer to μ as the mean of the discrete probability distribution (p_r) $(r = 0, 1, 2, \ldots)$.

- The sample mean $\bar{x}$ and the mean μ of the r.v. X are different, though related, concepts. Combining the results of Equation (1) and Equation (2), we see that

$$\bar{x} \to \mu$$

as $n \to \infty$. As the sample size tends to infinity, the sample mean $\bar{x}$ converges to the mean μ of the r.v. X. Especially when put more rigorously and proved as a result in probability theory, this is known as the "law of large numbers," or in less technical language as the "law of averages."

Example — the binomial distribution

Suppose that $X \sim B(n,p)$. From Equation (2) and the definition of the binomial distribution in Equation (1) of Section 4,

$$\begin{aligned}
\mu \equiv E(X) &= \sum_{r=1}^{n} r \binom{n}{r} p^r q^{n-r} \\
&= \sum_{r=1}^{n} \frac{n!}{(r-1)!(n-r)!} p^r q^{n-r} \\
&= np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} q^{(n-1)-(r-1)} \\
&= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^s q^{n-1-s} \;=\; np,
\end{aligned}$$

where in the last line we have put $s = r - 1$ and then used the fact that the terms in the sum are the probabilities of the $B(n-1,p)$ distribution, so that the sum is equal to 1. Thus the mean of the $B(n,p)$ distribution is given by $\mu = np$. This makes intuitive sense, since in the frequency interpretation of probability p is the long-term proportion of successes in the trials. Hence in n trials we would on average expect to see np successes.

5.2 The expected value of a function of a random variable

Let X again be a discrete r.v. with probability distribution (p_r) $(r = 0, 1, 2, \ldots)$. The expected value of X^2 is given by

$$E(X^2) = \sum_{r=0}^{\infty} r^2 p_r.$$

More generally, let g be any real-valued function. The expected value of $g(X)$ is given by

$$E(g(X)) = \sum_{r=0}^{\infty} g(r) p_r. \tag{3}$$

Let h be another real-valued function and a and b two constants. Then

$$\begin{aligned}
E(ag(X) + bh(X)) &= \sum_{r=0}^{\infty} [ag(r) + bh(r)] p_r \\
&= a \sum_{r=0}^{\infty} g(r) p_r + b \sum_{r=0}^{\infty} h(r) p_r \\
&= aE(g(X)) + bE(h(X)).
\end{aligned}$$

So the expectation operator E is a linear operator in that

$$E(ag(X) + bh(X)) = aE(g(X)) + bE(h(X)). \tag{4}$$

More generally, if X and Y are two r.v.s defined on some sample space S and a and b are two constants then

$$E(aX + bY) = aE(X) + bE(Y).$$

5.3 The variance of a discrete probability distribution

Definition

The *variance* σ^2 or $\text{var}(X)$ of the discrete r.v. X with probability distribution (p_r) $(r = 0, 1, 2, \ldots)$ is given by

$$\sigma^2 \equiv \text{var}(X) = E((X - \mu)^2) = \sum_{r=0}^{\infty} (r - \mu)^2 p_r. \tag{5}$$

- Alternatively, we may refer to σ^2 as the variance of the discrete probability distribution (p_r) $(r = 0, 1, 2, \ldots)$.
- The variance is a measure of the spread of the distribution of X.
- The sample variance s^2 and the variance σ^2 of the r.v. X are different, though related, concepts. Suppose that we repeat n times the experiment associated with the r.v. X, so that a random sample $x_1, x_2, \ldots, x_n$ of size n is obtained from the r.v. X, and then calculate the sample variance s^2. As we let the sample size become arbitrarily large, we find that

  $$s^2 \to \sigma^2$$

  as $n \to \infty$. As the sample size tends to infinity, the sample variance s^2 converges to the variance σ^2 of the r.v. X.

Theorem 1 *An alternative formula for the variance is given by*

$$\sigma^2 \equiv var(X) = E(X^2) - [E(X)]^2 = \sum_{r=0}^{\infty} r^2 p_r - \mu^2. \tag{6}$$

Proof. From Equation (5) and the linearity of the expectation operator,

$$\begin{aligned}
\sigma^2 &= E((X - \mu)^2) \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2.
\end{aligned}$$

5.4 Probability generating functions

Let X again be a discrete random variable with probability distribution (p_r) $(r = 0, 1, 2, \ldots)$. The *probability generating function* (p.g.f.) of this distribution is the function $G(t)$ defined by

$$G(t) = \sum_{r=0}^{\infty} p_r t^r. \tag{7}$$

Note that for any p.g.f. $G(t)$

$$G(1) = \sum_{r=0}^{\infty} p_r = 1.$$

The p.g.f. is a convenient way of summarizing a discrete probability distribution in terms of a single function. It is particularly useful for calculating the mean and variance of a distribution.

Differentiating Equation (7),

$$G'(t) = \sum_{r=1}^{\infty} r p_r t^{r-1}, \tag{8}$$

and, setting $t = 1$, we obtain

$$G'(1) = \sum_{r=1}^{\infty} r p_r = \mu.$$

Hence to find the mean of a discrete probability distribution we may differentiate the p.g.f. and then set $t = 1$.

$$\mu \equiv E(X) = G'(1). \tag{9}$$

Differentiating Equation (8),

$$G''(t) = \sum_{r=2}^{\infty} r(r-1) p_r t^{r-2},$$

and, setting $t = 1$, we obtain

$$G''(1) = \sum_{r=1}^{\infty} r(r-1) p_r = \sum_{r=1}^{\infty} r^2 p_r - \sum_{r=1}^{\infty} r p_r = E(X^2) - E(X). \tag{10}$$

Combining Equations (6) and (10), we obtain

$$\sigma^2 \equiv \mathrm{var}(X) = G''(1) + \mu - \mu^2. \tag{11}$$

Hence to find the variance of a discrete probability distribution we may differentiate the p.g.f. twice, set $t = 1$, and then use the formula of Equation (11).

Example — the binomial distribution (continued)

Suppose that $X \sim B(n, p)$. From Equation (7) and the definition of the binomial distribution in Equation (1) of Section 4,

$$G(t) \;=\; \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} t^r$$

$$
\begin{aligned}
&= \sum_{r=0}^{n} \binom{n}{r} (pt)^r q^{n-r} \\
&= (pt+q)^n,
\end{aligned}
$$

using the binomial theorem. So the p.g.f. of the $B(n,p)$ distribution is given by

$$G(t) = (pt+q)^n. \tag{12}$$

Differentiating Equation (12),

$$G'(t) = np(pt+q)^{n-1}. \tag{13}$$

Setting $t = 1$ in Equation (13) and using the result of Equation (9), for the $B(n,p)$ distribution,

$$\mu = np,$$

which confirms the result obtained in Section 5.1.

Differentiating Equation (13),

$$G''(t) = n(n-1)p^2(pt+q)^{n-2}. \tag{14}$$

Setting $t = 1$ in Equation (14) and using the result of Equation (11),

$$\sigma^2 = n(n-1)p^2 + np - (np)^2 = np - np^2 = npq.$$

Hence, for the $B(n,p)$ distribution,

$$\sigma^2 = npq.$$

5.5 The Poisson distributions

First we note the following mathematical facts.

1. The exponential series: for any number x,

$$\sum_{r=0}^{\infty} \frac{x^r}{r!} = e^x.$$

2. For any number x,

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

Definition

The *Poisson distribution* with parameter μ, where μ is any positive number, is the discrete probability distribution (p_r) with

$$p_r = e^{-\mu} \frac{\mu^r}{r!} \qquad (r = 0, 1, 2, \ldots). \tag{15}$$

Using the exponential series, we may check that the sum of the terms in this distribution really is 1:

$$\sum_{r=0}^{\infty} p_r = \sum_{r=0}^{\infty} e^{-\mu} \frac{\mu^r}{r!} = e^{-\mu} \sum_{r=0}^{\infty} \frac{\mu^r}{r!} = e^{-\mu} e^{\mu} = 1.$$

Again using the exponential series, the corresponding p.g.f. is given by

$$G(t) = \sum_{r=0}^{\infty} e^{-\mu} \frac{\mu^r}{r!} t^r = e^{-\mu} \sum_{r=0}^{\infty} \frac{(\mu t)^r}{r!} = e^{-\mu} e^{\mu t} = e^{\mu(t-1)}.$$

Thus the p.g.f. $G(t)$ of the Poisson distribution with parameter μ is given by

$$G(t) = e^{\mu(t-1)}. \tag{16}$$

Differentiating Equation (16) we obtain

$$G'(t) = \mu e^{\mu(t-1)},$$

and, setting $t = 1$,

$$G'(1) = \mu.$$

Recalling Equation (9), we see that the parameter μ of the Poisson distribution is also its mean, as the notation suggests. Differentiating again, we obtain

$$G''(t) = \mu^2 e^{\mu(t-1)},$$

and, setting $t = 1$,

$$G''(1) = \mu^2.$$

Hence from Equation (11)

$$\sigma^2 = G''(1) + \mu - \mu^2 = \mu^2 + \mu - \mu^2 = \mu.$$

Thus the mean and variance of the Poisson distribution are both equal to the parameter value μ.

One way in which the Poisson distribution arises is as an approximation to the $B(n, p)$ distribution when n is large and np is not large. We may then approximate the $B(n, p)$ distribution by the Poisson distribution with the same mean, i.e., with parameter $\mu = np$.

For fixed μ, consider the binomial distribution with parameters n and μ/n, where $n > \mu$. This binomial distribution, the $B(n, \mu/n)$ distribution, has mean μ and p.g.f.

$$G(t) = \left(\frac{\mu}{n}t + \left(1 - \frac{\mu}{n}\right)\right)^n = \left(1 + \frac{\mu(t-1)}{n}\right)^n \to e^{\mu(t-1)}$$

as $n \to \infty$. The limit on the right hand side is the p.g.f. of the Poisson distribution with parameter μ. As $n \to \infty$, the $B(n, \mu/n)$ distribution converges to the Poisson distribution with parameter μ. This limiting result provides a theoretical basis for the use of the Poisson distribution as an approximation to the binomial distribution.

This use of the Poisson distribution as an approximation to the binomial distribution was of more importance before statistical packages became readily available, when tables were commonly used for evaluation of binomial probabilities. For example Table 1 for the binomial distribution function in the *New Cambridge Statistical Tables* goes up to $n = 20$. For $n > 20$ some form of approximation may be used. An example of the sort of rough guideline that has been suggested is that if $n \geq 50$ and $p \leq 0.1$ then the Poisson distribution may be used to approximate the binomial distribution.

5.6 The use and evaluation of Poisson probabilities

The Poisson distribution is often used where the data are counts of numbers of occurrences of randomly occurring events over some period of time or in some specified area or volume.

- The numbers of occurrences of industrial accidents in some factory might be counted on, say, a weekly basis. The week by week numbers might be thought of as following a Poisson distribution.
- If a Geiger counter is placed in a location where radioactivity remains at a steady level, and the numbers of radioactive particles hitting the counter are recorded on, say, a minute by minute basis, the numbers of hits recorded each minute might be expected to follow a Poisson distribution.
- If misprints in a book occur randomly, the numbers of misprints per page might be taken to follow a Poisson distribution.

Where the Poisson distribution is to be used, the appropriate value of the parameter μ, the expected number of occurrences or events, has to be specified first, but that is the only entity that needs to be specified.

Given μ, Poisson probabilities may be calculated in a number of ways. Because the formula for the Poisson distribution is quite simple, we may readily use it for direct calculations:

$$\begin{aligned} p_0 &= e^{-\mu}, \\ p_1 &= e^{-\mu}\,\mu, \\ p_2 &= e^{-\mu}\,\mu^2/2, \end{aligned}$$

etc. More generally we may use the recurrence relation,

$$p_r = \frac{\mu}{r}\,p_{r-1} \qquad (r = 1, 2, 3, \ldots)$$

Alternatively, Table 2 in the *New Cambridge Statistical Tables* gives values of the Poisson distribution function,

$$F_r = \sum_{i=0}^{r} e^{-\mu}\,\frac{\mu^i}{i!},$$

for values of μ in the range $0 \leq \mu \leq 20$. In Excel the `POISSON` function may be used.

To calculate Poisson probabilities using `R`, you can use the function `ppois` to get the cumulative distribution function or `dpois` to calculated terms from the probability density function. The values you have to specify in the function are (in this order) r, and μ. By default the function `ppois` gives $F_r = \Pr(X \leq r)$, but adding the argument `lower.tail = FALSE` it returns $\Pr(X > r)$. Type `?ppois` to learn more about this function.

Example (cf *Clarke and Cooke*)

The mean number of misprints per page in a book is 1.2. What is the probability that on a particular page there are three or more misprints?

Let the r.v. X denote the number of misprints on the page. The required probability is $\Pr(X \geq 3)$. We assume that X has the Poisson distribution with mean $\mu = 1.2$.

$$\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - F_2 = 1 - 0.8795 = 0.1205,$$

where the value $F_2 = 0.8795$ for $\mu = 1.2$ has been read from Table 2 in the *New Cambridge Statistical Tables*. In Excel, the formula `= 1 - POISSON(2,1.2,TRUE)` gives the value 0.120512901.

Alternatively, the calculations may be carried out in `R`, as shown in the following output. You can notice that `1 - ppois(r, mu)` and `ppois(r, mu, lower.tail = FALSE)` return the same result: 0.1205129.

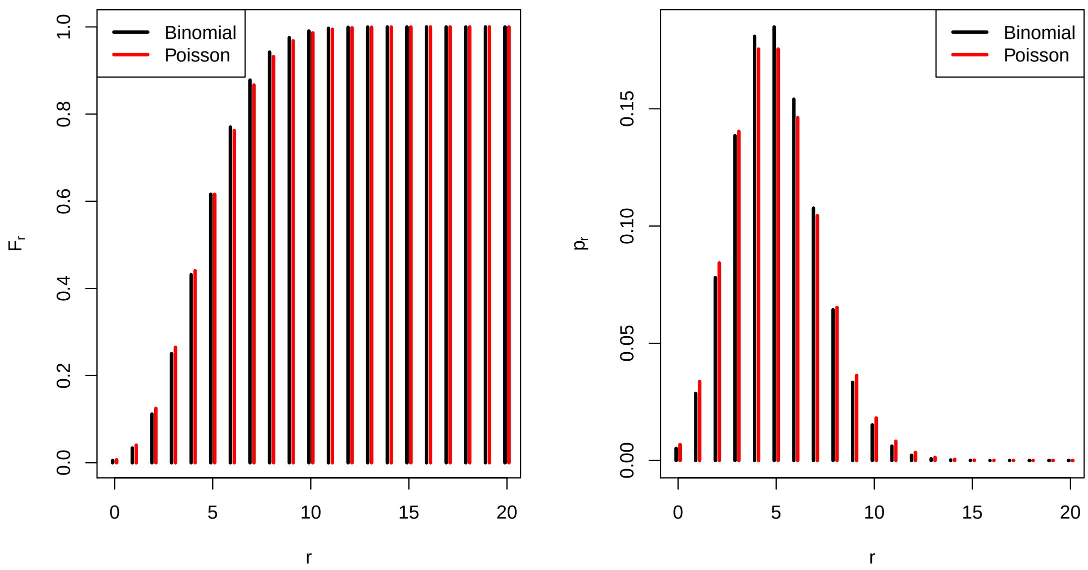
```
r <- 2
mu <- 1.2

1 - ppois(r, mu)

## [1] 0.1205129

ppois(r, mu, lower.tail = FALSE)

## [1] 0.1205129
```

If we were asked to consider the distribution of the total number of misprints in the first 5 pages, say, then we would still use a Poisson distribution but with $\mu = 5 \times 1.2 = 6$, the expected number of misprints in the 5 pages.

Some calculations to compare the $B(50, 0.1)$ distribution with the corresponding Poisson distribution with mean $\mu = n \times p = 5$ are shown in the output below. The range of integer values 0:20 is put into the variable `r`. The cumulative distribution function (F_r) for the $B(50, 0.1)$ distribution up to $r = 20$ is computed and put into first column `Binom_Fr` of the table `Tab`. The cumulative distribution function (F_r) for the Poisson distribution with mean $\mu = 5$ is put into column `Pois_Fr`. The probability density function (p_r) for the $B(50, 0.1)$ distribution is put into column `Binom_pr`. The probability density function (p_r) for the Poisson distribution with mean 5 is put into column `Pois_pr`.

```r
r <- 0:20
n <- 50
p <- 0.1

Binom_Fr <- pbinom(r, n, p)
Binom_pr <- dbinom(r, n, p)

mu <- n * p

Pois_Fr <- ppois(r, mu)
Pois_pr <- dpois(r, mu)
Tab <- cbind(Binom_Fr, Pois_Fr, Binom_pr, Pois_pr)
rownames(Tab) <- r
Tab

##       Binom_Fr     Pois_Fr     Binom_pr      Pois_pr
## 0  0.005153775 0.006737947 5.153775e-03 6.737947e-03
## 1  0.033785860 0.040427682 2.863208e-02 3.368973e-02
## 2  0.111728756 0.124652019 7.794290e-02 8.422434e-02
## 3  0.250293906 0.265025915 1.385651e-01 1.403739e-01
## 4  0.431198407 0.440493285 1.809045e-01 1.754674e-01
## 5  0.616123008 0.615960655 1.849246e-01 1.754674e-01
## 6  0.770226842 0.762183463 1.541038e-01 1.462228e-01
## 7  0.877854916 0.866628326 1.076281e-01 1.044449e-01
## 8  0.942132794 0.931906365 6.427788e-02 6.527804e-02
## 9  0.975462064 0.968171943 3.332927e-02 3.626558e-02
## 10 0.990645398 0.986304731 1.518333e-02 1.813279e-02
## 11 0.996780079 0.994546908 6.134680e-03 8.242177e-03
## 12 0.998995380 0.997981148 2.215301e-03 3.434240e-03
## 13 0.999714880 0.999302010 7.194996e-04 1.320862e-03
## 14 0.999926161 0.999773746 2.112816e-04 4.717363e-04
## 15 0.999982503 0.999930992 5.634176e-05 1.572454e-04
## 16 0.999996197 0.999980131 1.369418e-05 4.913920e-05
## 17 0.999999240 0.999994584 3.043151e-06 1.445271e-05
## 18 0.999999860 0.999998598 6.199011e-07 4.014640e-06
## 19 0.999999976 0.999999655 1.160049e-07 1.056484e-06
## 20 0.999999996 0.999999919 1.997862e-08 2.641211e-07
```

```r
par(mfrow = c(1, 2))

plot(r - 0.1, Tab[,1],
     type = "h", lwd = 3,
     xlab = "r", ylab = expression(F[r]))
lines(r + 0.1, Tab[,2],
      type = "h", lwd = 3, col = 2)
legend("topleft", c("Binomial", "Poisson"),
       col = 1:2, lwd = 3)

plot(r - 0.1, Tab[,3],
     type = "h", lwd = 3,
     xlab = "r", ylab = expression(p[r]))
lines(r + 0.1, Tab[,4],
      type = "h", lwd = 3, col = 2)
legend("topright", c("Binomial", "Poisson"),
       col = 1:2, lwd = 3)
```



Extra Exercise

Telephone calls arrive in an office at a rate of 5 calls per hour. Find the probability there are:

- exactly 2 calls in one hour
- exactly 1 call in 15 minutes
- 10 or fewer calls in 2 hours