Probability & Statistics

Dongjiao Ge

djge@cityu.edu.mo

Session 05

3.5 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.

*Note that some other literatures may use the **cumulative distribution function (c.d.f)** instead of p.m.f. to describe the distribution.*

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

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

The above *c.d.f* can be expressed in terms of the *p.m.f.* of X as follows:

Discrete variables

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

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

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

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:

x	0	1
$P(X = x)$	q	p

- 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 getting a head, then when we toss the coin once getting a head is a *Bernoulli(p)* random variable.

More generally, let A be any event in a probability space S. With A, we associate a random variable I_A (remember that a random variable is just a function on 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.

The expected value and variance of a Bernoulli random variable

Let X ~Bernoulli(p). (Remember that ~ means "has the same *p.m.f*. as".)

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

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

Binomial random variable Bin(n, p)

- A Bernoulli random variable is describing the event $X = 1$ as a 'success'. (we only consider one trail)
- Now a binomial random variable counts the number of successes in n independent trials each associated with a *Bernoulli(p)* random variable.

Suppose that we have a biased coin for which the probability of head appearing is p. We toss the coin n times and count the number of heads obtained. This number is a *Bin(n, p)* random variable.

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

$$P(X=k) = {}^{n}C_{k} q^{n-k} p^{k}$$

where $k = 0,1,2,\ldots,n$, and $q = 1 - p$.

- ${}^{n}C_{k}$--- the number of different ways of obtaining k heads in a sequence of n throws (the number of choices of the k positions in which the heads occur)
- the probability of getting k heads and $n-k$ tails in a particular order is $q^{n-k}\,p^{k}$.

For Bin(4, p) we can list out its probability mass function as:

k	0	1	2	3	4
$P(X=k)$	q^4	$4q^3p$	$6q^2p^2$	$4qp^3$	p^4

- Note: when we add up all the probabilities in the table, we get

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

as it should be: here we used the binomial theorem

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

If $X \sim \mathrm{Bin}(n,p)$, then

$$E(X) = np, \qquad \mathrm{Var}(X) = npq.$$

?

Example Let X be the number of Heads in 50 tosses of a fair coin. Determine EX, $Var(X)$ and $P(X \leq 10)$.

As X is Binomial$(50, \frac{1}{2})$, so $EX = 25$, $\mathrm{Var}(X) = 12.5$, and

$$P(X \leq 10) = \sum_{i=0}^{10} \binom{50}{i} \frac{1}{2^{50}}.$$

We have a coin with probability p of coming down heads, and we toss it n times and count the number X of heads. Then X is our Bin(n, p) random variable. Let X_k be the random variable defined by

$$X_k = \begin{cases} 1 & \text{if we get heads on the } k\text{th toss,} \\ 0 & \text{if we get tails on the } k\text{th toss.} \end{cases}$$

X_i is the indicator variable of the event 'heads on the k-th toss'. Now we have

$$X = X_1 + X_2 + \cdots + X_n$$

$X_1, \ldots, X_n$ are independent Bernoulli(p) random variables (since they are defined by different tosses of a coin). So, as we saw earlier, $E(X_i) = p, Var(X_i) = pq$. Then, by Theorem 3.2, since the variables are independent, we have

$$\begin{aligned} E(X) &= p + p + \cdots + p = np, \\ \operatorname{Var}(X) &= pq + pq + \cdots + pq = npq. \end{aligned}$$

Another method to compute the mean and variance is using the probability generating function

We write p_k for the probability $P(X = k)$. Now the probability generating function of X is the power series

$$G_X(x) = \sum p_k x^k.$$

(The sum is over all values k taken by X.)

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

Proposition 3.4 *Let $G_X(x)$ be the probability generating function of a random variable X. Then*

(a) $[G_X(x)]_{x=1} = 1$;

(b) $E(X) = \left[\frac{\mathrm{d}}{\mathrm{d}x} G_X(x)\right]_{x=1}$;

(c) $\mathrm{Var}(X) = \left[\frac{\mathrm{d}^2}{\mathrm{d}x^2} G_X(x)\right]_{x=1} + E(X) - E(X)^2$.

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

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

$$\frac{\mathrm{d}}{\mathrm{d}x} G_X(x) = \sum k p_k x^{k-1}.$$

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

$$\sum k p_k = E(X).$$

For part (c), differentiating twice gives

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} G_X(x) = \sum k(k-1) p_k x^{k-2}.$$

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

$$\sum k(k-1) p_k = \sum k^2 p_k - \sum k p_k = E(X^2) - E(X).$$

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

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

$$p_k = P(X=k) = {}^nC_k q^{n-k} p^k,$$

so the probability generating function is

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

by the Binomial Theorem. Putting $x=1$ gives $(q+p)^n = 1$, in agreement with Proposition 3.4(a).

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

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

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, $X \sim Bin(n, p)$, 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 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 Hg(n,M,N).

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

$$P(X=k)=\frac{{}^{M}C_k \cdot {}^{N-M}C_{n-k}}{{}^{N}Cn}.$$

- the number of samples of n balls from N --- ${}^{N}C_n$
- the number of ways of choosing k of the M red balls and $n-k$ of the $N-M$ others --- ${}^{M}C_k \cdot {}^{N-M}C_{n-k}$
- 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):

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

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 hat, then $E(X) = np$, and $Var(X)$ 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 hat 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 $Hg(n, M, N)$ is approximately $Bin(n, M/N)$ if n is small compared to M and $N - M$.

Geometric random variable 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 $1,,2,3,\ldots$
(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 Geom(p) random variable is given by

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

where $q = 1 - p$.

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

We add these probabilities up:

$$\sum_{k=1}^{\infty} q^{k-1}p = p + qp + q^2p + \cdots = \frac{p}{1-q} = 1,$$

since the series is a geometric progression with first term p *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 $X \sim Geom(p)$, the result will be that

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

We have

$$G_X(x) = \sum_{k=1}^{\infty} q^{k-1} p x^k = \frac{px}{1-qx},$$

again by summing a geometric progression. Differentiating, we get

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

Putting x = 1, we obtain

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

Differentiating again gives 2pq/(1−qx)3, so

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

Poisson random variable Poisson(λ)

The Poisson random variable is very closely connected with continuous things.

Suppose that ‘incidents’ occur at random times, but at a steady rate overall.

- The radioactive decay example: 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.

The p.m.f. for a $Poisson(\lambda)$ variable X:

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

Note that we have

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

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

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

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

$$G_X(x) = \sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!} \mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda(x-1)},$$

again using the series for the exponential function.

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

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

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 $Bin(1000, 1/1000)$. But it turns out to be $Poisson(\lambda)$, 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 $\mathrm{Bin}(n, p)$ can be approximated by $\mathrm{Poisson}(\lambda)$.

Theorem*. Poisson approximation to Binomial. When n is large, p is small, and $\lambda = np$ is of moderate size, $Binomial(n, p)$ is approximately $Poisson(\lambda)$:

> *If X is Binomial(n,p), with $p = \frac{\lambda}{n}$, then, as $n \to \infty$,*
>
> $$P(X = i) \to e^{-\lambda}\frac{\lambda^i}{i!},$$
>
> *for each fixed $i = 0, 1, 2, \ldots$.*

Proof.

$$\begin{aligned}
P(X = i) &= \binom{n}{i}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\ldots(n-i+1)}{i!}\cdot\frac{\lambda^i}{n^i}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-i} \\
&= \frac{\lambda^i}{i!}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\frac{n(n-1)\ldots(n-i+1)}{n^i}\cdot\frac{1}{\left(1-\frac{\lambda}{n}\right)^i} \\
&\to \frac{\lambda^i}{i!}\cdot e^{-\lambda}\cdot 1\cdot 1,
\end{aligned}$$

as $n \to \infty$. $\square$

EX1 Denote by d the dominant gene and by r the recessive gene at a single locus. Then dd is called the pure dominant genotype, dr is called the hybrid, and rr the pure recessive genotype. The two genotypes with at least one dominant gene, dd and dr, result in the phenotype of the dominant gene, while rr results in a recessive phenotype. Assuming that both parents are hybrid and have n children, what is the probability that at least two will have the recessive phenotype? Each child, independently, gets one of the genes at random from each parent.

EX2 You roll a fair die, your opponent tosses a fair coin independently. If you roll 6 you win; if you do not roll 6 and your opponent tosses Heads you lose; otherwise, this round ends and the game repeats. On the average, how many rounds does the game last?

EX3 Suppose that the probability that a person is killed by lightning in a year is, independently, 1/(500 million). Assume that the US population is 300 million.
(1) Compute P(3 or more people will be killed by lightning next year) exactly.
(2) Approximate the above probability.