STAT2001

2018 Term I

Outline

1. Discrete random variables
2. Mathematical expectation
3. Binomial distribution and Hypergeometric distribution
4. Moment generating function
5. Poisson distribution

(Textbook chapters: 2.1 - 2.6)

Definition: For a random experiment with sample space S, a function X that maps each element s in S to a real number x, that is $X(s) = x$, is called a **random variable**.

Definition: If the set $\{x : X(s) = x, s \in S\}$ is a countable subset of real numbers, X is a **Discrete Random Variable**.

1.1 Some Examples

Example 1: Roll a die

$$S = \{1, 2, 3, 4, 5, 6\}$$

If we want to define a discrete random variable that gives "1" for an odd number and "0" for an even number, we define

$$X(s) = \langle \begin{array}{lll} 1 & for & s = 1, 3, 5 \\ 0 & for & s = 2, 4, 6 \end{array}$$

So what is the chance of having $X = 1$? If I tell you that the die is "fair", the answer is clearly $1/2$.

More formally, we define below the **Probability Mass Function (p.m.f.)** of X.

Definition: The Probability Mass Function (p.m.f.) denoted by $f(x)$ of a discrete random variable X is a function satisfying:

Sx for sample space

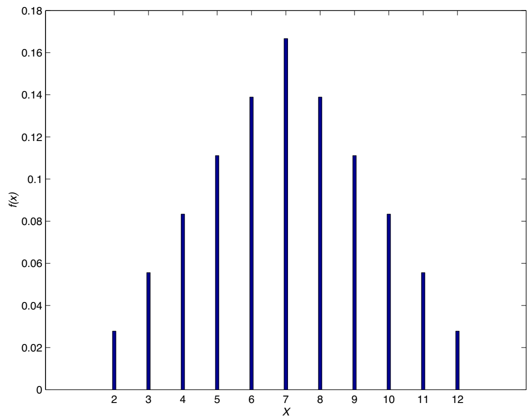
$$\begin{aligned} &(a) \quad f(x) > 0, \qquad x \in S_X; \\ &(b) \quad \sum_{x \in S_X} f(x) = 1; \\ &(c) \quad P(X \in A) = \sum_{x \in A} f(x) \qquad \textit{where} \quad A \subset S_X. \end{aligned}$$

The symbol S_X denotes the *space of* X or also called the *support of* X.

Example 2: Roll a fair die twice. Let X be the sum of the two numbers we get. What is S_X?

$$\begin{aligned} f(2) &= P(X=2) = P[\{(1,1)\}] = 1/36; \\ f(3) &= P(X=3) = P[\{(2,1),(1,2)\}] = 2/36 = 1/18; \quad etc \end{aligned}$$

We can represent the p.m.f. of the above example by a Bar Graph:



Example 3: Randomly tossing a coin 3 times corresponds to the following sample space:

$$S = \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}$$

Let X be the number of heads showing up in 3 tossings. Then, X can only be $0, 1, 2$ or 3. Consider the following events:

$$\{s \in S : X(s) = 0\} = \{(T,T,T)\} \quad q^3$$
$$\{s \in S : X(s) = 1\} = \{(H,T,T),(T,H,T),(T,T,H)\} \quad 3q^2p$$
$$\{s \in S : X(s) = 2\} = \{(H,H,T),(H,T,H),(T,H,H)\} \quad 3p^2q$$
$$\{s \in S : X(s) = 3\} = \{(H,H,H)\} \quad p^3$$

If the coin shows up head with probability p (with $q = 1 - p$) and all tossings are independent to each other, then

$$
\begin{aligned}
\Pr\{X=0\} &= \Pr\{(T,T,T)\} = q^3 \\
\Pr\{X=1\} &= \Pr\{(H,T,T),(T,H,T),(T,T,H)\} \\
&= \Pr\{(H,T,T)\} + \Pr\{(T,H,T)\} + \Pr\{(T,T,H)\} \\
&= 3pq^2 \\
\Pr\{X=2\} &= \Pr\{(H,H,T),(H,T,H),(T,H,H)\} \\
&= \Pr\{(H,H,T)\} + \Pr\{(H,T,H)\} + \Pr\{(T,H,H)\} \\
&= 3p^2q \\
\Pr\{X=3\} &= \Pr\{(H,H,H)\} = p^3
\end{aligned}
$$

1.2 Functions of Discrete Random Variables

Suppose X is a discrete random variable. For any function $g : \Re \to \Re$, define $Y = g(X)$, we have Y also a discrete random varaible. Moreover, the p.m.f. of Y is completely determined by the p.m.f. of X.

Example 4:
Suppose $P(X = -1) = 0.2; P(X = 0) = 0.3; P(X = 1) = 0.5$

Define $Y = 3X + 1$, we have $P(Y = -2) = 0.2; P(Y = 1) = 0.3; P(Y = 4) = 0.5;$

Define $Y = X^2$, we have $P(Y = 1) = 0.7; P(Y = 0) = 0.3$.

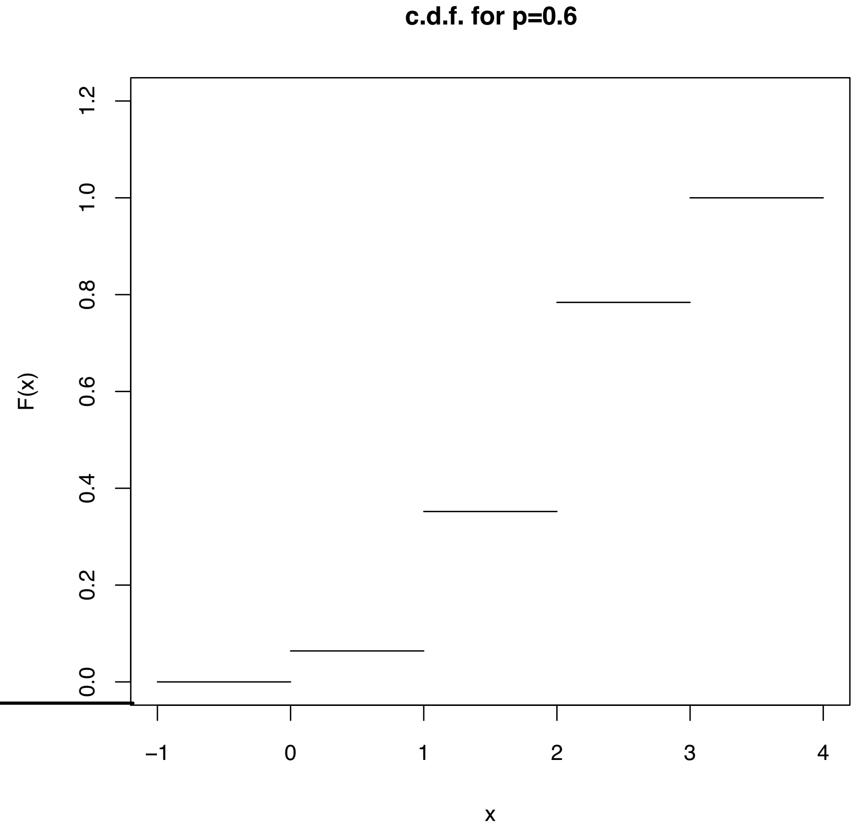
1.3 Cumulative Distribution Function (c.d.f.)

increasing

This is another function commonly used to summarise the distribution of a random variable: $F(x) = P(\underbrace{X \leq x})$

Example 5: Consider Example 3 again:

$$F(x) = 0 \text{ for } x < 0$$
$$F(x) = q^3 \text{ for } x \in [0, 1)$$
$$F(x) = q^3 + \ 3pq^2 \text{ for } x \in [1, 2)$$
$$F(x) = q^3 + 3pq^2 + 3p^2q \text{ for } x \in [2, 3)$$
$$F(x) = 1 \text{ for } x \geq 3$$

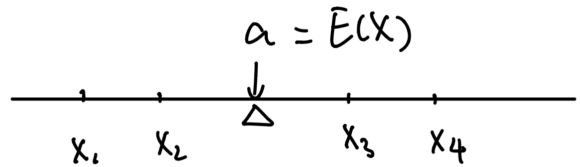

c.d.f. for p=0.6
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−1
0
1
2
3
4
x

Consider throwing a fair die a large number of times, what will you expect to be the average of the numbers you get?

It should be approximately $\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \ldots + \frac{1}{6} \times 6 = 3.5$. This is what we called the *Mean* or *Expected value.*

Definition: For a discrete random variable X with p.m.f. $f(x)$, the *expected value* of X is defined as:

$$E(X) = \sum_{x \in S_X} x f(x)$$



$(a-x_1)f(x_1) + (a-x_2)f(x_2) + \cdots + (a-x_4)f(x_4)$

$= 0$

provided that the sum exists.

Remark: If S_X is a finite set, clearly $E(X)$ always exists. However it is possible that $E(X)$ does not exist when S_X is countably infinite. Consider the case that X takes the value 2^k with probability 2^{-k}, for k=1,2,...

We often denote $E(X)$ by μ for convenience. More generally, for any function $u(x)$, the expected value of $u(X)$ is defined as:

$$E(u(X)) = \sum_{x \in S_X} u(x) f(x)$$

provided that the sum exists.

$E(X) = -0.2 + 0.5 = 0.3$

Example 6:

Suppose $P(X = -1) = 0.2$; $P(X = 0) = 0.3$; $P(X = 1) = 0.5$

Define $Y = X^2$, we have,

$$E(Y) = E(X^2) = (-1)^2 \times 0.2 + (0)^2 \times 0.3 + (1)^2 \times 0.5 = 0.7$$

The Expected value or mean of a random variable is a measure of the ”middle part” of the random variable’s distribution (or so-called a measure of *central tendency*).

To measure the variation or dispersion of the distribution of the random variable X (about its own mean), the following concept of the *variance* of the random variable is needed:

$$Var(X) = \sigma^2 = E[X - \mu]^2$$

It should be noted that μ is in the same unit of X while σ^2 would be in unit^2. A more interpretable measure of variation (which is in the same unit as X) is the *standard deviation* of X:

$$SD(X) = \sigma = \sqrt{Var(X)}$$

2.1 Properties of Expectation:

(a) If c is a constant, $E(c) = c$;

(b) If c is a constant and u is a function, $E[cu(X)] = cE[u(X)]$

(c) If c_1 and c_2 are constants and u_1 and u_2 are functions, then

$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$$

A convenient formula for calculating Variance of X :

$$\begin{aligned}\sigma^2 &= Var(X) = E\left[(X-\mu)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right] = E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

Example 7:

Consider the following probability mass function of a discrete random variable X:

x	-1	0	1	2	3
$f(x)$	0.1	0.5	0.2	0.1	0.1

$$Var(X) = E[X^2] - \{E(X)\}^2 = 1.6 - (0.6)^2 = 1.24$$

and

$$\sigma = \sqrt{1.24} = 1.114.$$

We take this chance to introduce the sample counterparts of mean, variance and standard deviation. In the above discussion, we start with a known probability mass function and then calculate mean, variance and standard deviation. It is of course possible in a lot of random experiments such as tossing a coin with known probability of success.

However, in many other situations, we can only collect sample data and may not know the underlying distribution of a random variable. For example, if we want to know the average (that is the mean) number of years of schooling of all citizens in HK, we can only afford to ask maybe 1,000 people and hope that the sample is a good representative of the whole population. We then calculate sample counterparts for mean, variance and standard deviation to describe the middle part and dispersion of the sample which we hope, can also be a good description of the population.

Consider a sample of n observations $\{x_1, ..., x_n\}$,

1. Sample mean $= \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;
2. Sample variance $= s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$;
3. Sample standard deviation $= s$.

3.1 Bernoulli experiment

A random experiment that the outcome can be classified in one of two mutually exclusive and exhaustive ways. For convenience, we can call one class "success" and the other class "failure" depending on the question we want to address. The probability of "success" is denoted by p.

Example 8: Toss a coin and defining a success be showing a "head".

Example 9: Throw a die and defining a success be showing an odd number.

3.2 Binomial distribution

If we repeat the same Bernoulli experiment independently for n times, we are performing a Binomial experiment. The subject of interest is a random variable X equals the number of "successes" in these n trials.

Example 10: Toss a coin 100 times and define X to be the number of "tails" observed.

Example 11: Throw a die 100 times and define X to be the number of even numbers observed.

We call X in the Binomial Experiment a Binomial random variable (or We say that X follows a Binomial distribution). Symbolically, we write $X \sim b(n, p)$.

If $n = 1$, we call it a Bernoulli random variable.

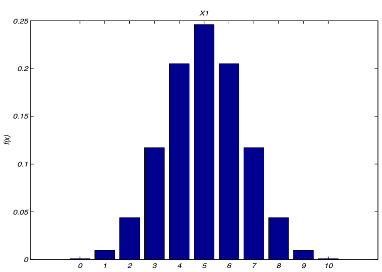
3.3 Probability Mass Function of Binomial distribution:

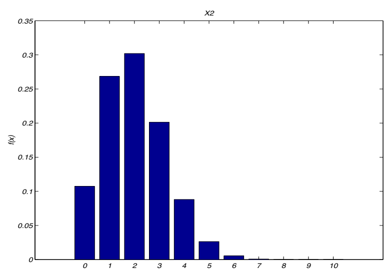
$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad where \;\; x = 0, 1, ..., n$$

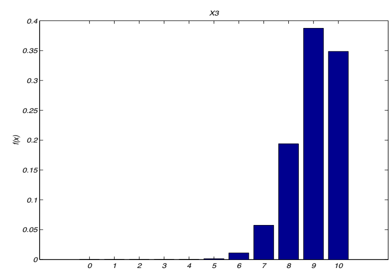
→ 二项式定理展开后的某项

Example 12: $X_1 \sim b(10, 0.5), X_2 \sim b(10, 0.2), X_3 \sim b(10, 0.9)$. Different shapes of p.m.f. can be observed

$$\sum f(x) = (1+1-p)^n .$$







Example 13:

It is known that all items produced by a certain machine will be defective with probability 0.1, independently of each other.

What is the probability that in a sample of three items, at most one will be defective?

Ans:

Let X be the number of defective items in the sample. $X \sim b(3, 0.1)$. Hence $P(X = 0) + P(X = 1) = \begin{pmatrix} 3 \\ 0 \end{pmatrix} 0.1^0(1 - 0.1)^{3-0} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} 0.1^1(1 - 0.1)^{3-1} = 0.972$

3.4 Expectation and Variance of Binomial distribution:

For $X \sim b(n, p)$,

$$\mu = np \quad and \quad \sigma^2 = np(1-p)$$

Proof:

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \end{aligned}$$

Let $k = x - 1$,

$$\begin{aligned} E(X) &= \sum_{k=0}^{n-1} \frac{n!}{k!(n-1-k)!} p^{k+1}(1-p)^{n-1-k} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^{k}(1-p)^{n-1-k} \\ &= np(p+1-p)^{n-1} \\ &= np \end{aligned}$$

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

$$= \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k}$$

To find $Var(X)$, we first find the value of $E(X(X-1))$,

$$\begin{aligned}
E(X(X-1)) &= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x} \\
&= \sum_{k=0}^{n-2} \frac{n!}{(k)!(n-2-k)!}p^{k+2}(1-p)^{n-2-k} \\
&= n(n-1)p^2\sum_{k=0}^{n-2} \frac{(n-2)!}{(k)!(n-2-k)!}p^{k}(1-p)^{n-2-k} \\
&= n(n-1)p^2(p+1-p)^{n-2} \\
&= n(n-1)p^2
\end{aligned}$$

Thus,

$$Var(X) = E(X^2)-\mu^2 = E(X(X-1))+E(X)-\mu^2 = n(n-1)p^2+np-n^2p^2 = np(1-p)$$

Example 14:
For the machine example above, what is the expected number of defectives and the standard deviation of number of defectives?
Ans: $\mu = 3 \times 0.1 = 0.3$
$\sigma = \sqrt{3 \times 0.1 \times 0.9} = \sqrt{0.27} = 0.52$

3.5 Hypergeometric Distribution

Suppose we have a box containing N_1 white balls and N_2 black balls. Suppose we randomly select n balls ($n < N_1 + N_2 = N$) *without replacement.* Let X be the number of balls being white. X follows a Hypergeometric distribution with p.m.f.:

$$f(x) = P(X = x) = \frac{\begin{pmatrix} N_1 \\ x \end{pmatrix}\begin{pmatrix} N_2 \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

超几何分布

the support of X is $\{x \textit{ being nonnegative integer}: \ x \le n, x \le N_1, n-x \le N_2\}$.

Question: If $N_1 = 10, N_2 = 20, n = 5$, find $P(X = 2)$.

Ans: $P(X = 2) = \dfrac{\begin{pmatrix} 10 \\ 2 \end{pmatrix}\begin{pmatrix} 20 \\ 3 \end{pmatrix}}{\begin{pmatrix} 30 \\ 5 \end{pmatrix}} = \frac{45\times 1140}{142{,}506} = 0.36$

Remark:
It is sometimes confusing whether we should use Binomial distribution or Hypergeometric distribution to model the number of "successes". A simple example can illustrate how we can distinguish them:

Suppose I have a box of 10 white balls and 20 black balls. If I repeatedly draw a ball from the box 5 times *with replacement* and I define X to be the number of white balls I get. Then $X \sim b(5, 1/3)$.

Suppose now I repeatedly draw a ball five times from the box *without replacement* and define X to be the number of white balls I get. Then X follows a hypergeometric distribution with $N = 30$, $N_1 = 10$, $N_2 = 20$, $n = 5$.

The difference lies in *with/without replacement.* With replacement, the draws are clearly independent and $p = 1/3$ does not change throughout. Without replacement, the draws are not independent.

The *mean* and *variance* of a Hypergeometric random variable X are:

$$E(X) = n\left(\frac{N_1}{N}\right)$$

$$Var(X) = n\left(\frac{N_1}{N}\right)\left(\frac{N_2}{N}\right)\boxed{\left(\frac{N-n}{N-1}\right)}$$

$E(X)=n\frac{N_1}{N}$

$V(X)=n\frac{N_1}{n}\frac{N_2}{n}\left(\frac{N-n}{N-1}\right)$

Proof for $E(X)$ can be found in example 2.2-5 of the textbook and that for $Var(X)$ is left as an exercise for you.

Support formulas: ① $iC_n^i = nC_{n-1}^{i-1}$ (用组合数公式证明

② $C_n^0 C_m^k + C_n^1 C_m^{k-1} + \cdots + C_n^k C_m^0 = C_{n+m}^k$ (二项式展开对应项系数项等)

eg: $(1+x)^{n+m} = (1+x)^n(1+x)^m$

In section 2, we learned that *Mean* μ and *Variance* σ^2 are important quantities to describe a distribution. *Mean* and *Variance* are also called the *first moment* and the *second moment about the mean* respectively. Why *Mean* is called the *first moment*?

Consider a discrete distribution with finite support $\{x_1, x_2, ..., x_k\}$. It is like we distribute k masses on a real number line (imagine a lever with pivot at the origin). The directed distance from origin and the weight of the ith mass are x_i and $f(x_i)$ respectively.

In mechanics, the product of a distance and its weight is called a *moment* that measures a turning effect. Thus the collective moment of the distribution about the origin is simply $\sum_{i=1}^{k} x_i f(x_i)$, that is the definition of *Mean* (μ). Hence μ is called the *first moment.*

Imagine that if we now move the pivot of the lever to μ, what happens? There will be no turning effect anymore from the masses. In fact we can easily calculate the *first moment about* μ to be zero.

$$\sum_{i=1}^{k}(x_i - \mu)f(x_i) = \sum_{i=1}^{k} x_i f(x_i) - \mu = \mu - \mu = 0$$

On the other hand, *Variance* (σ^2) is called the *second moment about* μ. The reason is that in the definition of σ^2 ($\sigma^2 = \sum_{x \in S}(x - \mu)^2 f(x)$), the distances about μ are raised to the second power.

However in section 2, we show that $\sigma^2 = E(X^2) - \mu^2$. Since it is easier to calculate $E(X^2)$ and we can get the variance from it and the first moment, statisticians call $E(X^2)$ to be the *second moment*.

We can generalize this idea to define the *rth moment* of a distribution to be $E(X^r)$.

Certainly we can calculate each moment of a distribution based on its definition. An alternative approach is to first derive the *Moment Generating Function* of a distribution and then generate all moments from it.

Definition: Let X be a discrete random variable with p.m.f. $f(x)$. If there exists $h > 0$ such that

$E(e^{tX}) = \sum_{x \in S_X} e^{tx} f(x)$

$$E(\exp(tX)) = \sum_{x \in S_X} \exp(tx) f(x)$$

exists and is finite for $-h < t < h$, then the function

$$M(t) = E(\exp(tX))$$

is called the **Moment Generating Function (m.g.f.)** of X.

矩生成函数

How can $M(t)$ be used to produce the moments of X?
Suppose we can interchange the sum and differentiation (technical details out of scope here),

$$M'(t) = \frac{d}{dt}E(\exp(tX)) = \frac{d}{dt}\left[\sum_{x\in S_X} \overset{e^{tx}}{\exp(tx)} f(x)\right] = \sum_{x\in S_X} x\exp(tx)f(x),$$

thus,

$$M'(0) = \sum_{x\in S_X} xf(x) = E(X)$$

Similarly,

$$\boxed{M^{(r)}(0) = E(X^r)}$$

Therefore, if the m.g.f of X is particularly simple, we have a efficient way of calculating its moments through successive differentiations.

Example 15

Find the first and second moments of Binomial distribution:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\begin{aligned} M(t) &= \sum_{x=0}^{n} \exp(tx) \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (p\exp(t))^x (1-p)^{n-x} \\ &= (p\exp(t) + (1-p))^n \end{aligned}$$

Thus

$$M'(t) = n\left[(1-p) + p\exp(t)\right]^{n-1} (p\exp(t)) \Longrightarrow E(X) = M'(0) = np$$

$$\begin{aligned} M''(t) &= n(n-1)\left[(1-p) + p\exp(t)\right]^{n-2} (p\exp(t))^2 \\ &+ n\left[(1-p) + p\exp(t)\right]^{n-1} (p\exp(t)) \end{aligned}$$

$$\Longrightarrow E\left(X^2\right) = M''(0) = n(n-1)p^2 + np = n^2p^2 + np(1-p) \Rightarrow Var(X) = np(1-p)$$

Let's introduce two more commonly used distributions here.

Geometric Distribution

Suppose you toss a coin repeatedly until a "head" comes up. Further assume the probability of getting a "head" is p. What is the probability mass function of X defined as the number of tossing required?

$$f(x) = P(X = x) = p(1-p)^{x-1}, \qquad where \quad x = 1, 2, \ldots$$

$E(X) = \frac{1}{p}$ $\quad Var(X) = \frac{1-p}{p^2}$

$\frac{a_1(1-p^n)}{1-p}$.

$P(Y=x) = p(1-p)^x$ (Y 表示成功前的次数).

$\frac{1}{p}$

$E(Y) = \frac{1-p}{p}$ $\quad Var(Y) = \frac{1-p}{p^2}$

$\frac{1(1-(1-p)^x)}{1-(1-p)}$.

$M(x) = \sum e^{tx} p(1-p)^{x-1}$

$\frac{1-(1-p)^n}{p}$.

$m'(t) = \sum x e^{tx} p(1-p)^{x-1}$

$m'(0) = \sum x p(1-p)^{x-1}$

Negative Binomial Distribution

It includes the Geometric distribution as a special case. Suppose now you define X to be the number of tossings you need to get r "heads". X follows a negative binomial distribution with the following p.m.f.:

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad where \;\; x = r, r+1, \ldots$$

We can find the mean and variance of X using moment generating function technique in a coming example. Here we simply write them down:

$$EX = \frac{r}{p}; \;\; Var(X) = \frac{r(1-p)}{p^2}$$

Negative Binomial distribution reduces to Geometric distribution when $r = 1$.

Example 16:
Suppose that during practice, a basketball player can make a successful free throw 80% of the time. What is the probability that he can make 3 successful shots by no more than 5 trials?

Let X be the number of trials he need. X follows Negative Binomial with $p = 0.8, r = 3$

$P(X = 3) + P(X = 4) + P(X = 5)$

$= \binom{3-1}{3-1} 0.8^3(1-0.8)^{3-3} + \binom{4-1}{3-1} 0.8^3(1-0.8)^{4-3} + \binom{5-1}{3-1} 0.8^3(1-0.8)^{5-3}$

$= 0.512 + 0.3072 + 0.1229 = 0.942$

Y = num of successful trials in 5 throw

Example 17
Find the mean and variance of Negative Binomial distribution:

$$
\begin{aligned}
M(t) &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\
&= (pe^t)^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} [(1-p)e^t]^{x-r} \\
&= \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \ \ \textit{where} \ \ (1-p)e^t < 1,
\end{aligned}
$$

The last line can be obtained using 2 different methods:
Method 1: We can follow section 2.5 of the textbook to use the following mathematical result:

$$(1-w)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} w^k, \qquad -1 < w < 1.$$

or $$(1-w)^{-r} = \sum^{\infty} \binom{k-1}{r-1} w^{k-r}, \quad -1 < w < 1$$

$x=r$

$p^* = 1-(1-p)e^t \Rightarrow 1-p^* = (1-p)e^t.$

Method 2: We observe that the second line can be written as:

$$M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r} \sum_{x=r}^{\infty} \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} [1-(1-p)e^t]^r [(1-p)e^t]^{x-r}$$

We identify that the above summation is the sum of all probabilities of a Negative Binomial distribution with parameters $p^* = 1-(1-p)e^t$ and $r^* = r$ when $0 < p^* < 1$. And therefore the sum must be 1. That gives the same result as method 1.

Therefore, after differentiation and evaluation at $t = 0$,

$$M'(0) = \frac{r}{p} \; and \; M''(0) = rp^{-2}(r+1-p)$$

Thus,

$$EX = \frac{r}{p}, \quad Var(X) = M''(0) + (M'(0))^2 = \frac{r(1-p)}{p^2}$$

Cumulate generating function $R(t) = \ln(M(t))$

$R'(0) = E(X)$

$R''(0) = Var(X)$

Besides calculating moments, another reason that Moment Generating Function is important is that it can uniquely characterize a distribution. In other words, if we know that the moment generating functions of 2 random variables are the same, we know that their distributions are the same.

Example 18: Suppose that the m.g.f. of X is $M(t) = (0.3 + 0.7e^t)^{10}$. We know that X follows $b(10, 0.7)$.

Example 19: If X has m.g.f. $M(t) = \frac{1}{2}e^t + \frac{1}{3}e^{2t} + \frac{1}{6}e^{3t}$, we know that X follows the discrete distribution:

$$P(X = 1) = 1/2; P(X = 2) = 1/3; P(X = 3) = 1/6.$$

Remark: This characteristic of m.g.f. is particularly useful in theoretical proofs in probability theories. In the last chapter of this course, we shall make use of this property to give a proof of Central Limit Theorem.

A random variable X with $S_X = \{0, 1, 2, 3, ...\}$ is said to be a Poisson random variable with parameter λ, if for some $\lambda > 0$, average number of occurrences in the period.

$$f(x) = P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!} \qquad where \quad x = 0, 1, 2, 3, ...$$

With this p.m.f., we can easily calculate the m.g.f., mean and variance

$$M(t) = \sum_{x=0}^{\infty} \exp(tx) \frac{\lambda^x \exp(-\lambda)}{x!} = \sum_{x=0}^{\infty} \frac{(\lambda \exp(t))^x \exp(-\lambda)}{x!}$$

$M'(t) = e^{\lambda(e^t-1)} \lambda e^t$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda \exp(t))^x}{x!} = e^{-\lambda} \exp(\lambda e^t) = \exp(\lambda(\exp(t) - 1))$$

$\Rightarrow M'(0) = \lambda.$

$E(X) = \lambda$ $\quad V(X) = EX^2 - (EX)^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$

$M''(0) = \lambda^2 + \lambda$

$$\therefore E(X) = \lambda, Var(X) = \lambda$$

Poisson random variables are commonly used to model count data.

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \underline{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}$$

Taylor expansion

Sum of Poisson (λe^t) $= 1$

$$M(t) = \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x e^{-\lambda}}{x!} = e^{-\lambda} e^{\lambda e^t} \boxed{\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x e^{-\lambda e^t}}{x!}}$$

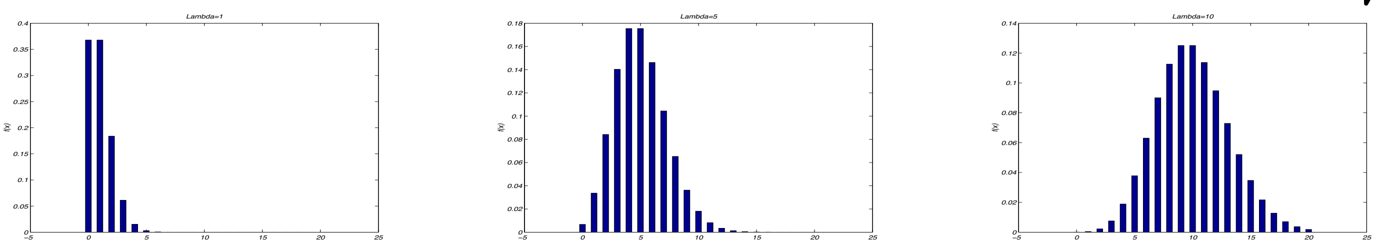
$= e^{\lambda(e^t-1)}$

Example 20: If the number of accidents occuring in a tunnel every week can be modelled by a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur in this week?

Ans: $P(X = 0) = \frac{3^0 \exp(-3)}{0!} = \exp(-3) = 0.05$

Remark: The parameter λ equals to both the expected value and variance of X. The following are shapes of Poisson distributions with $\lambda = 1, 5,$ and 10:

$\lambda \uparrow$: more and more symmetric => related to CLT



5.1 Poisson Process

If a poisson random variable is used to model counts in a given time period, it can be viewed as a limiting result of an *approximate* Poisson Process. (The difference of Poisson Process and *approximate* Poisson Process is due to some Mathematical rigors and we do not bother that in this course. Hereafter, we use Poisson Process to stand for *approximate* Poisson Process for brievity).

Definition: Let the number of changes that occur in a given continuous interval be counted. We have a Poisson process with parameter $\lambda > 0$ if the followings are satisfied:

(a) The numbers of changes occuring in nonoverlapping intervals are independent;

(b) The probability of exactly one change in a sufficiently short interval, h is approximately λh;

(c) The probability of two or more changes in a very short interval is essentially zero.

For instance, we can use Poisson Process to model the arrival of customers. We first divide a given time period (say with length 1) into n (suppose a large number) subintervals. By (b) and (c), for each subinterval, we have probability $\lambda\left(\frac{1}{n}\right)$ of have one customer arrival and probability $1-\lambda\left(\frac{1}{n}\right)$ of no custumer arrival. Clearly it is a Bernoulli trial with $p=\frac{\lambda}{n}$.

With (a), the total number of customers (X) arrived in the given period follows $b(n,\frac{\lambda}{n})$ approximately. That is

$$P(X=x)\approx\frac{n!}{(n-x)!x!}\left(\frac{\lambda}{n}\right)^x(1-\frac{\lambda}{n})^{n-x}$$

By taking $n\to\infty$, we have

$$\begin{aligned}P(X=x) &= \lim_{n\to\infty}\frac{n!}{(n-x)!x!}\left(\frac{\lambda}{n}\right)^x(1-\frac{\lambda}{n})^{n-x}=\frac{\lambda^x}{x!}\lim_{n\to\infty}\frac{n!}{(n-x)!}\left(\frac{1}{n}\right)^x(1-\frac{\lambda}{n})^{n-x}\\ &= \frac{\lambda^x}{x!}\lim_{n\to\infty}\frac{n\,(n-1)\ldots(n-x+1)}{n^x}(1-\frac{\lambda}{n})^n\left(1-\frac{\lambda}{n}\right)^{-x}=\frac{\lambda^x}{x!}\exp\left(-\lambda\right)\end{aligned}$$

Thus we get back the Poisson distribution.

In the derivation above, we set the interval that we have interest to be of length 1. Clearly, the same argument holds for arbitrary interval of length t. Thus for a Poisson process with parameter λ, the number of occurrences in an interval of length t, we call that X, has p.m.f.:

$$P(X = x) = \frac{(\lambda t)^x \exp(-\lambda t)}{x!}$$

Example 21: Telephone calls enter a college switchboard on the average of two every 3 minutes. If one assumes a Poisson process, what is the probability of five or more calls arriving in a 9-minute period?

$$1 - \sum_{0}^{4} \frac{6^x e^{-6}}{x!}.$$

Let X denote the number of calls in a 9-minute period

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \frac{(2\times3)^x \exp(-2\times3)}{x!} = 1 - 0.285 = 0.715$$

5.2 Relationship with Binomial Distribution

From the above derivation of Poisson distribution from Poisson process, we see that there is a close relationship between a Poisson random variable with parameter λ and a Binomial random variable follows $b(n, \frac{\lambda}{n})$.

We actually see that the two p.m.f. are approximately the same given very large n. Therefore, a Binomial distribution $b(n,p)$ can be approximated by a Poisson distribution with parameter np provided n is large and p is small.

Example 20: Suppose we want to calculate $P(X \leq 20)$ for $X \sim b(100, 0.02)$, we need to calculate $\sum_{x=0}^{20} \frac{100!}{(100-x)!x!} (0.02)^x (1-0.02)^{100-x}$ which is complicated. Otherwise, we can approximate by $P(X \leq 20) \approx P(Y \leq 20)$ for Y Poisson with $\lambda = 100 \times 0.02 = 2$ by calculating $\sum_{x=0}^{20} \frac{(2)^x \exp(-2)}{x!}$ which is relatively easier. It turns out that $P(X \leq 20) = 0.859$ and $P(Y \leq 20) = 0.857$

$$\lim_{n\to\infty}\left(1 + \frac{b}{n}\right)^n = \lim_{n\to\infty} e^{n\ln(1+b/n)},$$

where b is a constant.

Since the exponential function is continuous, the limit can be taken to the exponent. That is,

$$\lim_{n\to\infty} \exp[n\ln(1 + b/n)] = \exp[\lim_{n\to\infty} n\ln(1 + b/n)].$$

By L'Hôpital's rule, the limit in the exponent is equal to

$$\lim_{n\to\infty} \frac{\ln(1 + b/n)}{1/n} = \lim_{n\to\infty} \frac{\dfrac{-b/n^2}{1 + b/n}}{-1/n^2} = \lim_{n\to\infty} \frac{b}{1 + b/n} = b.$$

Since this limit is equal to b, the original limit is

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

Applications of this limit in probability occur with $b = -1$, yielding

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

A function $f(x)$ possessing derivatives of all orders at $x = b$ can be expanded in the following **Taylor series**:

$$f(x) = f(b) + \frac{f'(b)}{1!}(x - b) + \frac{f''(b)}{2!}(x - b)^2 + \frac{f'''(b)}{3!}(x - b)^3 + \cdots .$$

If $b = 0$, we obtain the special case that is often called the **Maclaurin series**;

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots .$$

For example, if $f(x) = e^x$, so that all derivatives of $f(x) = e^x$ are $f^{(r)}(x) = e^x$, then $f^{(r)}(0) = 1$, for $r = 1, 2, 3, \ldots$. Thus, the Maclaurin series expansion of $f(x) = e^x$ is

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots .$$

THEOREM (L'Hospital's Rule): Suppose f and g are differentiable and $g'(x) \neq 0$ near a (except possibly at a). Suppose that

$$\lim_{x\to a} f(x) = 0 \quad \text{and} \quad \lim_{x\to a} g(x) = 0$$

or that

$$\lim_{x\to a} f(x) = \pm\infty \quad \text{and} \quad \lim_{x\to a} g(x) = \pm\infty$$

(In other words, we have an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$.) Then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

if the limit on the right side exists (or is ∞ or $-\infty$).