Section 2.4 Discrete Random Variables

School of Sciences, BUPT

Modeling

It would be very tedious if, every time we had a slightly different problem, we had to determine the probability distributions from scratch. Luckily, there are enough similarities between certain types, or families, of experiments, to make it possible to develop formulas representing their general characteristics.

Contents

Binomial Distribution with Parameters n and p

Suppose that

$$P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad 0 < p < 1.$$

We call X a ***Bernoulli random variable***. We can say X has ***Bernoulli distribution with parameter*** p.

The p.f. of X is written as follows:

$$p(x) = \begin{cases} 1 - p & \text{if } x = 0, \\ p & \text{if } x = 1. \end{cases}$$

It can be easily checked that the mean and variance of X are

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

Binomial Distribution with Parameters n and p

Consider a sequence of Bernoulli trials, i.e.,

(i) the experiment consists of n repeated Bernoulli trials - each trial has only two possible outcomes labeled as success and failure,

(ii) the trials are independent - the outcome of any trial has no effect on the probability of the others,

(iii) the probability of success in each trial is constant which we denote by p.
Similar to establish the Binomial random variable, we can establish

Binomial Distribution with Parameters n and p

Definition

The random variable X that counts the number of successes, k, in the n Bernoulli trials is said to have a ***binomial distribution with parameters*** n ***and*** p*, written as* $X \sim B(n, p)$.

The probability function of the random variable $X \sim B(n, p)$ is given by

$$p(x) = \binom{n}{x} p^x q^{n-x} \quad \text{for } x = 0, 1, \cdots, n, \tag{1}$$

where $q = 1 - p$.

Binomial Distribution with Parameters n and p

Remark

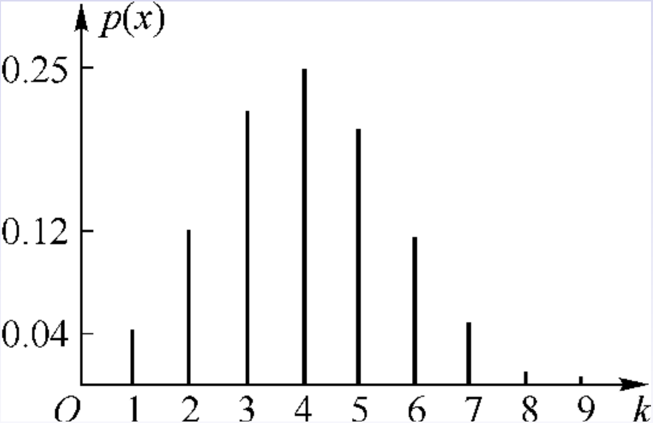
(i) *We have that* $p(x) \geqslant 0$ *for all* x. *By Newton's binomial formula,*

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

as should be.

(ii) *The shape and the position of the probability function* $p(x)$ *depend on the parameters* n *and* p. *When* $p = 0.5$, *the binomial distribution is symmetrical.*

Simeon-Denis Poisson



Binomial Distribution with Parameters n and p

Proposition

Suppose that X is a random variable which has binomial distribution with parameters n and p. Then we have

$$E(X) = np \quad \text{and} \quad Var(X) = np(1-p). \tag{2}$$

Binomial Distribution with Parameters n and p

Proof.

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np[p + (1-p)]^{n-1} = np.
\end{aligned}$$

Binomial Distribution with Parameters n and p

Proof.

$$
\begin{aligned}
& Var(X) \\
&= E(X^2) - [E(X)]^2 \\
&= \sum_{x=0}^{n} x^2 p(x) - (np)^2 = \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} p^x (1-p)^{n-x} - (np)^2 \\
&= \sum_{x=0}^{n} \cdot \frac{x(x-1)n!}{x!(n-x)!} p^x (1-p)^{n-x} + \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} - (np)^2 \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{(n-2)-(x-2)} + np - (np)^2 \\
&= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} (1-p)^{(n-2)-(x-2)} + np - (np)^2 \\
&= n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}
$$

□

Binomial Distribution with Parameters n and p

We can also think about $E(X)$ from another aspect. Since p is the probability of success in each trial, it is reasonable to regard p as the number of success in one trial. So if we repeat the experiment n times independently, the number of success should be np. This will be proved in next chapter.

Binomial Distribution with Parameters n and p

Example

In an airport, five radars are in operation and each radar has a 0.9 probability of detecting an arriving airplane. The radars operate independently of each other.

(a) Calculate the probability that an arriving airplane will be detected by at least four radars.

(b) Knowing that at least three radars detected a given airplane, what is the probability that the five radars detected this airplane?

(c) What is the smallest number n of radars that must be installed if we want an arriving airplane to be detected by at least one radar with probability 0.9999?

Binomial Distribution with Parameters n and p

Solution. Let X be the number of radars that detect the airplane.
(a) We have that $X \sim B(5,\ 0.9)$. We find Table of Binomial Probabilities in Appendix, that

$$P(X \geqslant 4) = \binom{5}{4}(0.9)^4(0.1) + (0.9)^5 = 0.5905 + 0.3280 = 0.9185.$$

(b) We want

$$\begin{aligned} P(X = 5 | X \geqslant 3) =& \frac{P(\{X = 5\} \cap \{X \geqslant 3\})}{P(\{X \geqslant 3\})} = \frac{P(X = 5)}{P(X \geqslant 3)} \\ =& \frac{0.5905}{0.5905 + 0.3280 + 0.0729} \approx 0.596. \end{aligned}$$

Binomial Distribution with Parameters n and p

Solution. (c)We now have that $X \sim B(n,\ 0.9)$ and we seek (the smallest) n such that $P(X \geqslant 1) = 0.9999$. We have:

$$
\begin{aligned}
P(X \geqslant 1) =& 1 - P(X = 0) = 1 - \binom{n}{0}(0.9)^0(0.1)^n = 1 - (0.1)^n \\
=& 0.9999 \iff (0.1)^n = 0.0001 = (0.1)^4.
\end{aligned}
$$

Thus, we may write that $n_{min} = 4$. $\square$

Binomial Distribution with Parameters n and p

Example

Suppose that a baseball player's batting average is 0.33 (1 for 3 on average). Consider the case where the player either gets a hit or makes an out (forget about walks here!). On average how many hits does the player get in 100 at bats? What's the standard deviation for the number of hits in 100 at bats?

Solution. It is obvious that the probability for a hit is $p = 0.33$ and the probability for no hit is $q = 1 - p = 0.67$. Let X be the hit number. Then $X \sim B(100, 0.33)$.

$$\mu = np = 100 \cdot 0.33 = 33 \text{ hits} \qquad \text{and}$$

$$\sigma = \sqrt{npq} = \sqrt{100 \cdot 0.33 \cdot 0.67} \approx 4.7 \text{ hits}.$$

$\square$

Binomial Distribution with Parameters n and p

Example

Given that 5% of a population are left-handed, calculate the probability that a random sample of 100 people contains 2 or more left-handed people.

Solution. Let X be the number of left handed people in a sample of 100. Then $X \sim B(100, 0.05)$. So

$$P(X = x) = \binom{100}{x} 0.05^x 0.95^{100-x}, \quad x = 0, 1, \cdots, 100$$

What we want is $P(X \geqslant 2)$.

$$\begin{aligned} P(X \geqslant 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{100}{0} 0.95^{100} - \binom{100}{1} 0.05 \cdot 0.95^{99}. \end{aligned}$$

$\square$

Binomial Distribution with Parameters n and p

The calculation will make us crazy if we have no computer.

Theorem

(***Poisson limit theorem***) *Suppose that* λ *is a constant and* n *is a positive integer. If* $\lim_{n\to\infty} np_n = \lambda$, *then we have*

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

for any fixed nonnegative integer x.

Binomial Distribution with Parameters n and p

Proof. We rewrite $p_n = \lambda/n$ ($p \to +\infty$). Now we have

$$\begin{aligned}
&\lim_{n\to+\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
=&\lim_{n\to+\infty} \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
=&\lim_{n\to+\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
=&\lim_{n\to+\infty} \frac{\lambda^x}{x!} \frac{n(n-1)\cdots(n-x+1)}{n\cdot n\cdots n} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
=&\lim_{n\to+\infty} \frac{\lambda^x}{x!} \left[1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right] \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
=&\frac{\lambda^x}{x!}\cdot 1\cdot e^{-\lambda}\cdot 1 = \frac{e^{-\lambda}\lambda^x}{x!}.
\end{aligned}$$

$\square$

Simeon-Denis Poisson

Binomial Distribution with Parameters n and p

In general, if n is large (say > 50, (> 20)) and p is small (say < 0.1 (< 0.05)), then we can use Poisson limit theorem.

Now by using Poisson approximation, we can get
$\lambda = 100 \cdot 0.05 = 5$, and

$$\begin{aligned} P(X \geqslant 2) &= 1 - \binom{100}{0} 0.95^{100} - \binom{100}{1} 0.05 \cdot 0.95^{99} \\ &\approx 1 - e^{-5}\frac{5^0}{0!} - e^{-5}\frac{5^1}{1!} \\ &\approx 1 - 0.0067 - 0.0337 \approx \ 0.9596. \end{aligned}$$

Contents

Poisson Distribution with Parameter λ

Definition

We say that the discrete random variable X whose probability function is given by equation

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, \cdots . \tag{3}$$

has a ***Poisson distribution with parameter*** $\lambda > 0$. *We write that* $X \sim P(\lambda)$.

Poisson Distribution with Parameter λ

Remark

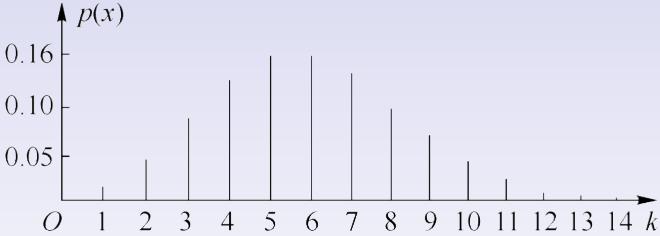
(i) *It can be easily shown that*

$$\sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = 1$$

by using the formula $e^x = 1 + x + \frac{x^2}{2!} + \cdots$.

(ii) *The shape of the probability function* $p(x)$ *depends on the value of the parameter* λ*. We observe that the distributions are centered roughly on* λ*, exhibit positive skew (that decreases a* λ *increases) and the variance (spread) increases as* λ *increases.*

Simeon-Denis Poisson



Poisson Distribution with Parameter λ

Remark

(iii) *Table of Poisson Probabilities in Appendix gives many values of this function.*

(iv)*In general, the Poisson approximation should be good if* $n > 20$ *and* $p < 0.05$.
If the value of p *is greater than 1/2, then we must consider the number of failures (with probability* $1 - p < 1/2$*) rather than the number of successes before performing the approximation by the Poisson distribution.*

Poisson Distribution with Parameter λ

Proposition

Suppose that X is a random variable which has Poisson distribution with parameter λ. Then we have

$$E(X) = Var(X) = \lambda. \tag{4}$$

Proof.

$$\begin{aligned} E(X) = &\sum_{x=0}^{+\infty} xp(x) = \sum_{x=0}^{+\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} \\ =&\lambda \sum_{x=1}^{+\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \sum_{x'=0}^{+\infty} \frac{e^{-\lambda}\lambda^{x'}}{x'!} = \lambda. \end{aligned}$$

Poisson Distribution with Parameter λ

Proof.

$$
\begin{aligned}
Var(X) = E(X^2) - [E(X)]^2 &= \sum_{x=0}^{+\infty} x^2 p(x) - \lambda^2 \\
&= \sum_{x=0}^{+\infty} x(x-1) \cdot p(x) + \sum_{x=1}^{+\infty} x p(x) - \lambda^2 = \lambda.
\end{aligned}
$$

$\square$

Poisson Distribution with Parameter λ

Consider the following examples:

(i) Let X be the number of typos on a printed page.
(ii) Let X be the number of cars passing through an intersection in one minute.
(iii) Let X be the number of Alaskan salmon caught in a squid driftnet.
(iv) Let X be the number of customers at an ATM in 10-minute intervals.
(v) Let X be the number of students arriving during office hours.

Common characteristic: they all denote the number of times an event occurs in an interval of time (or space).

They are all Poisson random variables.

Poisson Distribution with Parameter λ

Example

Births in a hospital occur randomly at an average rate of 1.5 births per hour. What is the probability of observing 4 births in a given hour at the hospital?

Solution. Let X be the number of births in a given hour. Since the events occur randomly and mean rate is $\lambda = 1.5$, we assume $X \sim P(1.5)$. We can now use the equation (3) to calculate the probability of observing exactly 4 births in a given hour.

$$P(X = 4) = e^{-1.5} \frac{1.5^4}{4!} = 0.0471.$$

$\square$

Poisson Distribution with Parameter λ

Example

Suppose that in a large population the proportion of people that have a certain disease is 0.01. Determine the probability that in a random group of 200 people at least four people will have the disease.

Solution. Let X be the number of people having the disease among the 200 people in the random group. It is obvious that $X \sim B(200,\ 0.01)$. This distribution can be approximated by a Poisson distribution for which the mean is $\lambda = np = 2$. Let Y denote a random variable having this Poisson distribution, i.e., $Y \sim P(2)$. Then it can be found from the Table of the Poisson probabilities in Appendix at the end of this book that

Poisson Distribution with Parameter λ

Solution.

$$\begin{aligned} & P(X \geqslant 4) \approx P(Y \geqslant 4) \\ &= 1 - P(Y = 0) - P(Y = 1) - P(Y = 2) - P(Y = 3) \\ &= 1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} - \frac{e^{-2}2^2}{2!} - \frac{e^{-2}2^3}{3!} \\ &= 1 - 0.1353 - 0.2707 - 0.2707 - 0.1805 = 0.1428. \end{aligned}$$

Hence, the probability that at least four people will have the disease is approximately 0.1428. The actual value is 0.1420. $\square$

The idea of using one distribution to approximate another is widespread throughout statistics.

Thank you for your patience !