Chapter 2 Random variables and distribution functions

The concept of random variables

The concept of random variables

In many trials, outcomes can be expressed by a numerical variable which is defined as taking a sequence of values.

For example,

(1) Let ξ be the nonnegative integers $0, 1, 2, \cdots$, and define it as the number of phone calls some operator receives during a particular interval of time. Then $\xi = 2$ stands for the event $\{$ there are two calls within this interval of time $\}$, while $\xi = 0$ stands for the event $\{$ there are no call within this interval of time $\}$.

(2) All possible measurement values constitute a sample space $\{\omega : \omega \in (a, b)\}$ when we measure length. In turn we can directly use a variable η to express the outcome of measurement: $\eta \in [1.5, 2.5]$ stands for the event $\{$the value of measurement is between 1.5 and 2.5 $\}$.

A random variable ξ is just a function of ω:

$\xi = \xi(\omega), \omega \in \Omega, \xi(\omega) \in \boldsymbol{R}$.

We need to pay attention to probabilities that the variables take on a specific value.

We need to pay attention to probabilities that the variables take on a specific value.

Therefore it is expected that $\{\omega : \xi(\omega) \in (a, b]\}$ is an event.

We need to pay attention to probabilities that the variables take on a specific value.

Therefore it is expected that $\{\omega : \xi(\omega) \in (a, b]\}$ is an event.

Accordingly, it is required that

$\{\omega : \xi(\omega) \in B\}$ is an event for any Borel set $B \in \mathcal{B}$.

Definition Suppose that $\xi(\omega)$ is a real function defined in a probability space $\{\Omega, \mathcal{F}, P\}$, and that for any Borel set B

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F},$$

(equivalently, for any real a, $\{\omega : \xi(\omega) \leq a\} \in \mathcal{F}$). Then we can say that ξ is a random variable, and that $\{P(\xi(\omega) \in B), B \in \mathcal{B}\}$, is a probability distribution associated to ξ.

Theorem

The following statements are equivalent.

1. *For any* x, $\xi^{-1}((-\infty, x]) \in \mathcal{F}$;
2. *For any* x, $\xi^{-1}((-\infty, x)) \in \mathcal{F}$;
3. *For any* $a < b$, $\xi^{-1}((a, b]) \in \mathcal{F}$;
4. *For any* $a < b$, $\xi^{-1}((a, b)) \in \mathcal{F}$;
5. *For any* $a < b$, $\xi^{-1}([a, b)) \in \mathcal{F}$;
6. *For any open or close* B, $\xi^{-1}(B) \in \mathcal{F}$;
7. *For any* $B \in \mathcal{B}$, $\xi^{-1}(B) \in \mathcal{F}$.

2.1 Discrete Random Variables

Discrete random variables

Definition. If a random variable ξ takes at most a set of countably many values (finite or infinite), then we call ξ a discrete random variable.

Distribution sequence of ξ:

For a discrete random variable ξ, let $\{x_j\}$ be the set of all possible values. Write $P(\xi = x_i)$ as $p(x_i)$ or p_i, $i = 1, 2, \cdots$.

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{pmatrix}$$

is said to be distribution sequence (or probability mass function) of ξ .

The properties of a distribution sequence:

$$p(x_i) \geq 0, i = 1, 2, \cdots ,$$

and

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

The properties of a distribution sequence:

$$p(x_i) \geq 0, i = 1, 2, \cdots ,$$

and

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

The probability of an event $\{\xi(\omega) \in B\}$ is

$$P(\xi(\omega) \in B) = \sum_{x_i \in B} p(x_i) \quad B \in \mathcal{B}.$$

Example 1. Suppose that the distribution sequence of random variable ξ is

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \frac{a-1}{4} & \frac{a+1}{4} & 0.1 & 0.2 & 0.2 \end{pmatrix}.$$

(1) Find the constant a;

(2) Find $P(-1 < \xi \leq 2)$.

Solution. (1) From the fact that

$$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.2=1$$

Solution. (1) From the fact that

$$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.2=1$$

it follows that $a=1$, and so the distribution sequence is

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0 & 0.5 & 0.1 & 0.2 & 0.2 \end{pmatrix}.$$

Solution. (1) From the fact that

$$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.2=1$$

it follows that $a=1$, and so the distribution sequence is

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0 & 0.5 & 0.1 & 0.2 & 0.2 \end{pmatrix}.$$

(2)

$$P(-1<\xi\leq 2)=\sum_{-1<x_i\leq 2} p(x_i)$$

Solution. (1) From the fact that

$$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.2=1$$

it follows that $a=1$, and so the distribution sequence is

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0 & 0.5 & 0.1 & 0.2 & 0.2 \end{pmatrix}.$$

(2)

$$P(-1<\xi\leq 2)=\sum_{-1<x_i\leq 2} p(x_i)=0.1+0.2+0.2=0.5.$$

Example 2. Assume that the success probability is p in the Bernoulli probability model, and denote by ξ the number of times that an experiment is conducted until its r-th success. Calculate the distribution sequence of ξ.

Solution.

Solution.

$$
\begin{aligned}
& P(\xi = k) \\
= \; & P(\text{ there are } r-1 \text{ successes and } k-r \text{ failures} \\
& \quad \text{in the first } k-1 \text{ trials and} \\
& \quad \text{success in the } k\text{-th trial})
\end{aligned}
$$

Solution.

$$
\begin{aligned}
& P(\xi = k) \\
= \; & P(\text{ there are } r-1 \text{ successes and } k-r \text{ failures} \\
& \quad \text{in the first } k-1 \text{ trials and} \\
& \quad \text{success in the } k\text{-th trial}) \\
= \; & \binom{k-1}{r-1} p^{r-1} q^{k-r} p = \binom{k-1}{r-1} p^r q^{k-r},
\end{aligned}
$$

where $k = r, r+1, r+2, \cdots$. This is called a Pascal distribution.

Typical discrete random variables:

Typical discrete random variables:

1. Degenerate distribution

Assume that a random variable ξ takes only one constant c, that is,

$$P(\xi = c) = 1.$$

2. Two point distribution

If there are two possible values x_1, x_2 in an experiment, then the probability distribution is

$$\begin{pmatrix} x_1 & x_2 \\ p & q \end{pmatrix}, \qquad p, q > 0, p + q = 1.$$

This is called a two point distribution.

Bernoulli distribution:

Bernoulli experiment has only two possible outcomes—-event A occurs or not. If let

$$\xi = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{otherwise,} \end{cases}$$

then its corresponding distribution sequence is

$$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad p, q > 0, p + q = 1.$$

3. The binomial distribution

If a random variable ξ has the following distribution sequence

$$P(\xi = k) = \binom{n}{k} p^k q^{n-k}, \quad p, q > 0, \quad p + q = 1.$$

where $k = 0, 1, 2 \cdots, n$, then we say that ξ obeys a binomial distribution, and simply write it as $\xi \sim B(n, p)$.

- To diagnose whether one gets some kind of disease is a Bernoulli trial. That people get such a disease or not is thought of as being independent, and the probability that a particular person gets such a disease is approximately equal. Thus to check on the disease situations of n people somewhere, on a one by one basis, can be thought of as n repeated Bernoulli trials, and the number of diseased individuals obeys a binomial distribution.

- Consider an insurance company for some kind of disaster (fire disaster). Suppose that a disaster befalls an individual is mutually independent and probabilities associated with such a disaster are equal. Assume that the probability this disaster befalls any one individual is p, then the number of people who encounter this disaster among n people obeys the binomial distribution.

- Consider n machines of the same type. Assume that the probability that each breaks down is p during an interval of time, then the number of machines that break down during this time period obeys the binomial distribution.

Properties of the binomial distribution:

(1)

$$b(k;n,p) = b(n-k;n,1-p),$$

since $\binom{n}{k} = \binom{n}{n-k}$.

$$\binom{n}{k} p^k (1-p)^{n-k} \qquad \binom{n}{n-k} (1-p)^{n-k} p^k.$$

(2) Monotonicity and the best possible number of successes

Fix n, p. Since

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p-k}{kq},$$

when $k < (n+1)p$, $b(k;n,p)$ increases; when $k > (n+1)p$, $b(k;n,p)$ decreases.

When $(n+1)p$ is an integer and $k=(n+1)p$, $b(k;n,p)=b(k-1;n,p)$ attains its maximum. We call $m=(n+1)p$ or $(n+1)p-1$ the best possible number of successes;
When $(n+1)p$ is not an integer the best possible number of successes is

$$m=[(n+1)p].$$

If $\xi \sim B(n, p)$, then

$$P(\xi = k+1) = \frac{p}{q}\frac{n-k}{k+1}P(\xi = k).$$

If $\xi \sim B(n,p)$, then

$$P(\xi = k+1) = \frac{p}{q}\frac{n-k}{k+1}P(\xi = k).$$

Example

Suppose that $\xi \sim B(6, 0.4)$. Compute $P(\xi = k)$, $k = 0, \cdots, 6$.

Solution. $P(\xi = 0) = (0.6)^6 \doteq 0.046656;$

$$
\begin{aligned}
P(\xi = 1) &= \frac{4}{6}\frac{6}{1}P(\xi = 0) \doteq 0.1866 \\
P(\xi = 2) &= \frac{4}{6}\frac{5}{2}P(\xi = 1) \doteq 0.3110 \\
P(\xi = 3) &= \frac{4}{6}\frac{4}{3}P(\xi = 2) \doteq 0.2765 \\
P(\xi = 4) &= \frac{4}{6}\frac{3}{4}P(\xi = 3) \doteq 0.1382 \\
P(\xi = 5) &= \frac{4}{6}\frac{2}{5}P(\xi = 4) \doteq 0.0369 \\
P(\xi = 6) &= \frac{4}{6}\frac{1}{6}P(\xi = 5) \doteq 0.0041
\end{aligned}
$$

(3) Asymptotic behaviors as n goes to ∞

Suppose that p depends on n, which we simply write as p_n.

Poisson Theorem. If there exists a positive constant λ such that $np_n \to \lambda$ as $n \to \infty$, then

$$\lim_{n\to\infty} b(k;n,p) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, 2, \cdots .$$

Proof. Set $\lambda_n = np_n$, then $p_n = \lambda_n/n$.

Proof. Set $\lambda_n = np_n$, then $p_n = \lambda_n/n$.Thus we have

$$b(k;n,p) = \binom{n}{k} {p_n}^k (1-p_n)^{n-k}$$

Proof. Set $\lambda_n = np_n$, then $p_n = \lambda_n/n$. Thus we have

$$b(k;n,p) = \binom{n}{k} {p_n}^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

$$\cdot \left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

Proof. Set $\lambda_n = np_n$, then $p_n = \lambda_n/n$.Thus we have

$$b(k;n,p) = \binom{n}{k} {p_n}^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

$$\cdot\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

$$= \frac{\lambda_n^k}{k!}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{(1-\frac{\lambda_n}{n})^n}{(1-\frac{\lambda_n}{n})^k}$$

Proof. Set $\lambda_n = np_n$, then $p_n = \lambda_n/n$.Thus we have

$$b(k;n,p) = \binom{n}{k} {p_n}^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

$$\cdot\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

$$= \frac{\lambda_n^k}{k!}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{(1-\frac{\lambda_n}{n})^n}{(1-\frac{\lambda_n}{n})^k}$$

$$\to \frac{\lambda^k}{k!}e^{-\lambda} \qquad (n\to\infty).$$

Example 3 Somebody shoots a target with the probability 0.001 of hitting it each time. Now he shoots 5000 times, calculate the probability that he hits the target two or more times.

Solution.

Solution. Denote by ξ the number of times he hits, then $\lambda = np = 5$,

$$P(\xi = k) = b(k; 5000, 0.001) \approx \frac{5^k}{k!}e^{-5}.$$

So

Solution. Denote by ξ the number of times he hits, then $\lambda = np = 5$,

$$P(\xi = k) = b(k; 5000, 0.001) \approx \frac{5^k}{k!} e^{-5}.$$

So

$$\begin{aligned} \sum_{k=2}^{5000} P(\xi = k) &= 1 - P(\xi = 0) - P(\xi = 1) \\ &\approx 1 - e^{-5} - 5e^{-5} \approx 0.9596. \end{aligned}$$

Solution. Denote by ξ the number of times he hits, then $\lambda = np = 5$,

$$P(\xi = k) = b(k; 5000, 0.001) \approx \frac{5^k}{k!} e^{-5}.$$

So

$$\begin{aligned} \sum_{k=2}^{5000} P(\xi = k) &= 1 - P(\xi = 0) - P(\xi = 1) \\ &\approx 1 - e^{-5} - 5e^{-5} \approx 0.9596. \end{aligned}$$

$$(1 - b(0; 5000, 0.001) - b(1; 5000, 0.001)$$
$$= 1 - 6.721 \times 10^{-3} - 3.364 \times 10^{-2} = 0.9596)$$

de Moivre-Laplace Theorem.

de Moivre (1732),Laplace (1801): Suppose $\xi_n \sim B(n,p)$, where $p = p_n$ satisfies $np_nq_n \to \infty$. Let

$$j = j_n, x = x(n) = \frac{j - np}{\sqrt{npq}}.$$

Then

$$P_n(x) = P(\xi_n = j) \sim \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2}$$

uniformly in x on every finite interval $[a, b]$ of values.

The relation $a_n \sim b_n$ means that $a_n/b_n \to 1$.The integer $j = j_n$ varies with n, so that $x = x(n)$ remains within a fixed finite interval $[a, b]$ and

$$j = np + x\sqrt{npq} \to \infty, \quad n - j = nq - x\sqrt{npq} \to \infty.$$

The relation $a_n \sim b_n$ means that $a_n/b_n \to 1$.The integer $j = j_n$ varies with n, so that $x = x(n)$ remains within a fixed finite interval $[a, b]$ and

$$j = np + x\sqrt{npq} \to \infty, \quad n - j = nq - x\sqrt{npq} \to \infty.$$

Further

$$P\left(a \leq \frac{\xi_n - np}{\sqrt{npq}} \leq b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

The relation $a_n \sim b_n$ means that $a_n/b_n \to 1$.The integer $j = j_n$ varies with n, so that $x = x(n)$ remains within a fixed finite interval $[a, b]$ and

$$j = np + x\sqrt{npq} \to \infty, \quad n - j = nq - x\sqrt{npq} \to \infty.$$

Further

$$P\left(a \le \frac{\xi_n - np}{\sqrt{npq}} \le b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

这就是Moivre-Laplace中心极限定理. 中心极限定理的一般形式将在第四章学到. 上式右边对应于后面要介绍另一个重要分布-正态分布.

Proof. Write $k = n - j$. Notice

$$P_n(x) = \frac{n!}{j!k!} p^j q^k.$$

We apply Stirling's formula

$$m! = \sqrt{2\pi m} \cdot m^m e^{-m} e^{\theta_m}, \quad 0 < \theta_m < \frac{1}{12m}$$

to $P_n(x)$.

Proof. Write $k = n - j$. Notice

$$P_n(x) = \frac{n!}{j!k!} p^j q^k.$$

We apply Stirling's formula

$$m! = \sqrt{2\pi m} \cdot m^m e^{-m} e^{\theta_m}, \quad 0 < \theta_m < \frac{1}{12m}$$

to $P_n(x)$. Thus

$$P_n(x) = \frac{\sqrt{2\pi n} \cdot n^n e^{-n}}{\sqrt{2\pi j} \cdot j^j e^{-j} \sqrt{2\pi k} \cdot k^k e^{-k}} p^j q^k e^{\theta_n - \theta_j - \theta_m},$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{jk}} \left(\frac{np}{j}\right)^j \left(\frac{nq}{k}\right)^k e^{\theta},$$

where, uniformly on $[a, b]$,

$$|\theta| < \frac{1}{12}\left(\frac{1}{n} + \frac{1}{j} + \frac{1}{m}\right),$$

$$\begin{aligned} \frac{jk}{n} =& n\left(p + x\sqrt{\frac{pq}{n}}\right)\left(q - x\sqrt{\frac{pq}{n}}\right) \\ =& npq\left(1 + x(q-p)\sqrt{\frac{1}{npq}} - x^2\frac{1}{n}\right) \sim npq \end{aligned}$$

and $\frac{j}{np} = 1 + x\sqrt{\frac{q}{np}}$, $\frac{k}{nq} = 1 - x\sqrt{\frac{p}{nq}}$,

$$\log\left(\frac{np}{j}\right)^j\left(\frac{nq}{k}\right)^k = -j\log\frac{j}{np} - k\log\frac{k}{nq}$$

$$= -(np + x\sqrt{npq})\left[x\sqrt{\frac{q}{np}} - \frac{1}{2}\frac{qx^2}{np} + O\left(\left(\frac{q}{np}\right)^{3/2}\right)\right]$$

$$-(nq - x\sqrt{npq})\left[-x\sqrt{\frac{p}{nq}} - \frac{1}{2}\frac{px^2}{nq} + O\left(\left(\frac{p}{nq}\right)^{3/2}\right)\right]$$

$$= -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{npq}}\right).$$

Therefore, uniformly on $[a, b]$,

$$
\begin{aligned}
P_n(x) &\sim \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{j}\right)^j \left(\frac{nq}{k}\right)^k \\
&\sim \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2}.
\end{aligned}
$$

The first assertion follows.

Let $x_{nj} = \frac{j-np}{\sqrt{npq}}$, $N_n = \{j : x_{nj} \in [a,b]\}$. Then $\#N_n \sim (b-a)\sqrt{npq}$, $x_{nj} - x_{n,j-1} = 1/\sqrt{npq}$.

Let $x_{nj} = \frac{j-np}{\sqrt{npq}}$, $N_n = \{j : x_{nj} \in [a,b]\}$. Then $\#N_n \sim (b-a)\sqrt{npq}$, $x_{nj} - x_{n,j-1} = 1/\sqrt{npq}$. On account of the first assertion, uniformly in $j \in N_n$,

$$P_n(x_{nj}) \sim \frac{1}{\sqrt{2\pi npq}} e^{-x_{nj}^2/2}$$

Let $x_{nj} = \frac{j-np}{\sqrt{npq}}$, $N_n = \{j : x_{nj} \in [a,b]\}$. Then $\# N_n \sim (b-a)\sqrt{npq}$, $x_{nj} - x_{n,j-1} = 1/\sqrt{npq}$. On account of the first assertion, uniformly in $j \in N_n$,

$$P_n(x_{nj}) \sim \frac{1}{\sqrt{2\pi npq}} e^{-x_{nj}^2/2}$$

and

$$P\left(a \le \frac{\xi_n - np}{\sqrt{npq}} \le b\right) = \sum_j P_n(x_{nj})$$

$$\sim \frac{1}{\sqrt{2\pi}} \cdot \sum_j \frac{1}{\sqrt{npq}} e^{-x_{nj}^2/2} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

4. The Poisson distribution

If a random variable ξ satisfies

$$P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda} \ \ (\lambda > 0), \quad k = 0, 1, 2, \cdots ,$$

we say that ξ obeys a Poisson distribution, or simply write as $\xi \sim P(\lambda)$, where λ is parameter of ξ.

If $A_1, \cdots, A_n$ are independent random events with $P(A_i) = p_n$. Let ξ_n be the number of these events that occur. Then $\xi_n \sim B(n, p_n)$. Poisson theorem tells us that, $\xi_n \dot{\sim} P(np_n)$ if $np_n \to \lambda$, i.e.,

$$P(\xi_n = k) \approx \frac{(np_n)^k}{k!} e^{-np_n}.$$

正因为如此, 泊松分布是用来描述离散型随机现象的一个比较普遍的分布. 人们发现许多随机现象都可以利用泊松分布来描述.

- In social daily life, the amount of various service requirement, like
    - the number of calls an operator receives during an interval of time,
    - the number of passengers arriving at the bus stop,
    - the number of customers coming to a supermarket or
    - the number of goods sold by a supermarket,

  all obey the Poisson law. Hence the Poisson distribution plays an important role in management science and operational research.

- In biology, with regard to the number of microorganism in some defined region, we can model the number of their offspring based on Poisson law.

- A radioactive substance emits α-particles, and the number of particles reaching a given portion of space during time t is the best-known example of random events obeying the Poisson law.

Example 4 On a certain crossroad the flow of traffic may be assumed to be Possionian. Suppose that the probability that no automobile passes through within one minute is 0.4, find the probability that more than one automobile pass through within 1 minutes.

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.

Note that $P(\xi = 0) = e^{-\lambda} = 0.4$, so we have $\lambda = \ln 5 - \ln 2$.

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.
Note that $P(\xi = 0) = e^{-\lambda} = 0.4$, so we have $\lambda = \ln 5 - \ln 2$. The probability asked for is

$$P(\xi > 1) = \sum_{k=2}^{\infty} P(\xi = k)$$

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.

Note that $P(\xi = 0) = e^{-\lambda} = 0.4$, so we have $\lambda = \ln 5 - \ln 2$. The probability asked for is

$$\begin{aligned} P(\xi > 1) &= \sum_{k=2}^{\infty} P(\xi = k) \\ &= 1 - P(\xi = 0) - P(\xi = 1) \end{aligned}$$

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.

Note that $P(\xi = 0) = e^{-\lambda} = 0.4$, so we have $\lambda = \ln 5 - \ln 2$. The probability asked for is

$$
\begin{aligned}
P(\xi > 1) &= \sum_{k=2}^{\infty} P(\xi = k) \\
&= 1 - P(\xi = 0) - P(\xi = 1) \\
&= 1 - e^{-\lambda} - \lambda e^{-\lambda}
\end{aligned}
$$

Solution. Denote by ξ the number of automobiles passing through the crossroad, and assume $\xi \sim P(\lambda)$.

Note that $P(\xi = 0) = e^{-\lambda} = 0.4$, so we have $\lambda = \ln 5 - \ln 2$. The probability asked for is

$$
\begin{aligned}
P(\xi > 1) &= \sum_{k=2}^{\infty} P(\xi = k) \\
&= 1 - P(\xi = 0) - P(\xi = 1) \\
&= 1 - e^{-\lambda} - \lambda e^{-\lambda} \\
&= \frac{3}{5} - \frac{2}{5} \ln \frac{5}{2} \approx 0.2335.
\end{aligned}
$$

The Poisson theorem tells us that, the Poisson with parameter np is a very good approximation to the distribution of the number of successes in n independent trials when each trial has probability p of being a success, provided that n is large and p small.

In fact, it remains a good approximation even the trials are not independent, provided that their dependence is weak.

Example

In the matching problem (Example 5 in Section 1.3), let A_i be the event that letter i is placed in the corrected envelope. It is easy seen that

$$P(A_i) = \frac{1}{n}, \quad P(A_i|A_j) = \frac{1}{n-1}, \;\; j \neq i.$$

Thus, A_i, $i = 1, 2, \ldots, n$ are not independent, but their dependence, for large n, appears to be weak.

Let ξ_n be a number of the letters that are placed in the corrected envelopes. Notice $np = n \times 1/n = 1$. Then for large n,

Let ξ_n be a number of the letters that are placed in the corrected envelopes. Notice $np = n \times 1/n = 1$. Then for large n,

$$P(\xi_n = k) \approx \frac{1}{k!} e^{-1}.$$

Let ξ_n be a number of the letters that are placed in the corrected envelopes. Notice $np = n \times 1/n = 1$. Then for large n,

$$P(\xi_n = k) \approx \frac{1}{k!}e^{-1}.$$

In fact

$$P(\xi_n = k) = \frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}.$$

Theorem

(Poisson Paradigm) Consider n events, with p_i equal to the probability that event i occurs, $i = 1, 2, \ldots, n$. If all the p_i are "small", and the trials are either independent or at most "weakly dependent", then the number of these events that occur approximately has a Poisson distribution $P(\sum_{i=1}^{n} p_i)$.

例 例如在第一章§2例6提到的生日问题中, 求n个人至少有两人同生日的概率.

例 例如在第一章§2例6提到的生日问题中，求n个人至少有两人同生日的概率.

如果用A_{ij}表示第i和第j个人同生日，那么$\{A_{ij}; 1 \le i < j \le n\}$共有$\binom{n}{2}$个事件，每个事件发生的概率为$P(A_{ij}) = \frac{1}{365}$,

例 例如在第一章§2例6提到的生日问题中，求n个人至少有两人同生日的概率.

如果用A_{ij}表示第i和第j个人同生日，那么$\{A_{ij}; 1 \le i < j \le n\}$共有$\binom{n}{2}$个事件，每个事件发生的概率为$P(A_{ij}) = \frac{1}{365}$，用泊松分布$P(\lambda)$，$\lambda = \binom{n}{2}\frac{1}{365}$，来近似这些事件发生次数$\xi$的分布得，$n$个人生日互不相同的概率为

$$P(\xi = 0) \approx e^{-\lambda} = \exp\left\{-\frac{n(n-1)}{2 \times 365}\right\}.$$

这与我们在第一章§2例6中得到的结论相同.

Poisson process. Consider a situation where "events" (E) occur at certain points in time. We will consider the number of these events occurring in a certain time interval.

Let us assume that for some positive constant λ the following assumptions hold true:

1. The probability that exactly 1 event occurs in each interval of length h is equal and equal to $\lambda h + o(h)$.
2. The probability that 2 or more events occur in an interval of length h is equal to $o(h)$.
3. For any integers, n, $j_1, \ldots, j_n$, and any set of n nonoverlapping intervals, if we denote E_i to be the event that exactly j_i of the events under consideration occur in the ith of these intervals, then events E_1,E_2,. . ., E_n are independent:

$$E_i = \{\text{在第}i\text{个时间段内}E\text{发生}j_i\text{次}\}.$$

Let $N(t)$ be number of events occurring in time interval $[0, t]$. Then

1. For any $t_1 < t_2 < \ldots < t_n$ and nonnegative integers $j_1, \ldots, j_n$, $\{N(t_1) = j_1\}$, $\{N(t_2) - N(t_1) = j_2\}, \ldots,$ $\{N(t_n) - N(t_{n-1}) = j_n\}$ are independent;
2. $N(s+t) - N(s)$ 与$N(t)$同分布;
3. $N(t) \sim P(\lambda t)$.

Proof. We want to compute $P(N(t) = k)$, we start by breaking the interval $[0, t]$ into n nonoverlapping subintervals each of length t/n.

Proof. We want to compute $P(N(t)=k)$, we start by breaking the interval $[0,t]$ into n nonoverlapping subintervals each of length t/n. Then the probability of a subinterval containing exactly 1 event is $p_n=\lambda\frac{t}{n}+o(\frac{t}{n})$, and the probability of a subinterval containing 2 or more events is $r_n=o(\frac{t}{n})$.

Proof. We want to compute $P(N(t) = k)$, we start by breaking the interval $[0, t]$ into n nonoverlapping subintervals each of length t/n. Then the probability of a subinterval containing exactly 1 event is $p_n = \lambda \frac{t}{n} + o(\frac{t}{n})$, and the probability of a subinterval containing 2 or more events is $r_n = o(\frac{t}{n})$.

Let A be the event that *k of the n subintervals contain exactly 1 event and the other $n-k$ contian 0 events*, and B be the event that *at least 1 subinterval contains 2 or more events*.

Then

$$P(N(t) = k) = P(A) + P(\{N(t) = k\} \cap B).$$

Then

$$P(N(t) = k) = P(A) + P(\{N(t) = k\} \cap B).$$

It is obvious that

$$\begin{aligned} P(\{N(t) = k\} \cap B) &\le P(B) \le nr_n \\ &= no(t/n) = o(1) \ \ as \ \ n \to \infty. \end{aligned}$$

and

Then

$$P(N(t)=k)=P(A)+P(\{N(t)=k\}\cap B).$$

It is obvious that

$$\begin{aligned} P(\{N(t)=k\}\cap B) \le P(B) \le nr_n & \\ &= no(t/n)=o(1) \ \ as \ \ n\to\infty. \end{aligned}$$

and

$$P(A)=b(k;n,p_n)\to \frac{(\lambda t)^k}{k!}e^{-\lambda t},$$

because

$$np_n=\lambda t+no(t/n)\to \lambda t.$$

The proof of (2) is completed.

5. The geometric distribution

If a random variable ξ satisfied

$$P(\xi = k) = pq^{k-1}, \quad p + q = 1, \quad p, q > 0,$$

where $k = 1, 2, \cdots$, then we say that ξ obeys a geometric distribution.

5. The geometric distribution

If a random variable ξ satisfied

$$P(\xi = k) = pq^{k-1}, \quad p + q = 1, \quad p, q > 0,$$

where $k = 1, 2, \cdots$, then we say that ξ obeys a geometric distribution.

Bernoulli probability model with p. Then the number ξ of experiments required in order to attain the first success obeys the geometric distribution.

The memoryless property: If $\xi \sim \text{gemo}(p)$, then

$$P\left(\xi = m + k \middle| \xi > m\right) = P(\xi = k) = pq^{k-1}.$$

Equivalently,

$$P\left(\xi > m + k \middle| \xi > m\right) = P(\xi > k) = q^{k}.$$

Indeed,

$$P(\xi > m) = \sum_{i=1}^{\infty} P(\xi = m+i)$$

Indeed,

$$P(\xi > m) = \sum_{i=1}^{\infty} P(\xi = m+i) = \sum_{i=1}^{\infty} pq^{m+i-1} = q^m.$$

Indeed,

$$P(\xi > m) = \sum_{i=1}^{\infty} P(\xi = m+i) = \sum_{i=1}^{\infty} pq^{m+i-1} = q^m.$$

So,

$$P(\xi = m + k | \xi > m)$$

Indeed,

$$P(\xi > m) = \sum_{i=1}^{\infty} P(\xi = m+i) = \sum_{i=1}^{\infty} pq^{m+i-1} = q^m.$$

So,

$$\begin{aligned} & P(\xi = m + k | \xi > m) \\ = \ & \frac{P(\xi = m + k, \xi > m)}{P(\xi > m)} = \frac{P(\xi = m + k)}{P(\xi > m)} \end{aligned}$$

Indeed,

$$P(\xi > m) = \sum_{i=1}^{\infty} P(\xi = m+i) = \sum_{i=1}^{\infty} pq^{m+i-1} = q^m.$$

So,

$$\begin{aligned} & P(\xi = m + k | \xi > m) \\ = & \frac{P(\xi = m + k, \xi > m)}{P(\xi > m)} = \frac{P(\xi = m + k)}{P(\xi > m)} \\ = & \frac{pq^{m+k-1}}{q^m} = pq^{k-1}. \end{aligned}$$

6. The hypergeometric distribution

Let n, N and M be positive integers with $n \leq N$ and $M \leq N$. The hypergeometric distribution is defined as follows

$$P(\xi = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

$$k = 0, 1, 2, \cdots, \min(n, M).$$

Consider a sampling inspection of product quality without replacement. If there are M defects in N products, then the number of defects found in n sampling products obeys a hypergeometric distribution.

There is a close relation between the binomial distribution and the hypergeometric distribution. If n, k are fixed, then as $N \to \infty, M/N \to p$ we have

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k q^{n-k}, \quad N \to \infty.$$

There is a close relation between the binomial distribution and the hypergeometric distribution. If n, k are fixed, then as $N \to \infty, M/N \to p$ we have

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k q^{n-k}, \quad N \to \infty.$$

Hence when N is sufficiently large, a hypergeometric distribution can be approximately calculated by using a binomial distribution as a proxy.

In fact,

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{M!}{(M-k)!k!}\frac{(N-M)!}{(N-M-n+k)!(n-k)!}\frac{(N-n)!n!}{N!}$$

In fact,

$$
\begin{aligned}
\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} &= \frac{M!}{(M-k)!k!}\,\frac{(N-M)!}{(N-M-n+k)!(n-k)!}\,\frac{(N-n)!n!}{N!} \\
&= \binom{n}{k}\frac{M!}{(M-k)!}\,\frac{(N-M)!}{(N-M-n+k)!}\,\frac{(N-n)!}{N!}
\end{aligned}
$$

In fact,

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{M!}{(M-k)!k!}\frac{(N-M)!}{(N-M-n+k)!(n-k)!}\frac{(N-n)!n!}{N!}$$

$$= \binom{n}{k}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-n+k)!}\frac{(N-n)!}{N!}$$

$$= \binom{n}{k}\frac{M\cdots(M-k+1)\times(N-M)\cdots(N-M-n+k+1)}{N(N-1)\cdots(N-n+1)}$$

In fact,

$$
\begin{aligned}
\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} &= \frac{M!}{(M-k)!k!}\frac{(N-M)!}{(N-M-n+k)!(n-k)!}\frac{(N-n)!n!}{N!} \\
&= \binom{n}{k}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-n+k)!}\frac{(N-n)!}{N!} \\
&= \binom{n}{k}\frac{M\cdots(M-k+1)\times(N-M)\cdots(N-M-n+k+1)}{N(N-1)\cdots(N-n+1)} \\
&= \binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k}\times \\
&\quad \frac{(1-\frac{1}{M})\cdots(1-\frac{k-1}{M})\cdot(1-\frac{1}{N-M})\cdots(1-\frac{n-k-1}{N-M})}{(1-\frac{1}{N})\cdots(1-\frac{n-1}{N})}
\end{aligned}
$$

In fact,

$$
\begin{aligned}
&\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{M!}{(M-k)!k!}\frac{(N-M)!}{(N-M-n+k)!(n-k)!}\frac{(N-n)!n!}{N!} \\
= \quad & \binom{n}{k}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-n+k)!}\frac{(N-n)!}{N!} \\
= \quad & \binom{n}{k}\frac{M\cdots(M-k+1)\times(N-M)\cdots(N-M-n+k+1)}{N(N-1)\cdots(N-n+1)} \\
= \quad & \binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k} \times \\
& \frac{(1-\frac{1}{M})\cdots(1-\frac{k-1}{M})\cdot(1-\frac{1}{N-M})\cdots(1-\frac{n-k-1}{N-M})}{(1-\frac{1}{N})\cdots(1-\frac{n-1}{N})} \\
\to \quad & \binom{n}{k}p^k q^{n-k}, \quad N\to\infty.
\end{aligned}
$$

7. The Zeta (or Zipf) distribution

A random variable is said to have a zeta (sometimes called the Zipf) distribution if its probability mass function is given by

$$P(\xi = k) = \frac{C}{k^{\alpha+1}}, \quad k = 1, 2, \cdots$$

for some $\alpha > 0$, where

$$C = \left[\sum_{k=1}^{\infty} \frac{1}{k^{\alpha+1}}\right]^{-1}.$$

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

is known as the Riemann zeta function (G.F.B. Riemann is a German mathematician).

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

is known as the Riemann zeta function (G.F.B. Riemann is a German mathematician).

The zeta distribution was used by the Italian economist Pareto to describe the distribution of family incomes in a given country. However, it was G. K. Zipf who applied there distributions in a wide variety of different areas and, in doing so, popularized their use.

More about discrete random variables and random variables

Theorem

Suppose that X and Y are random variables. Then cX, $X \pm Y$, XY, $X \vee Y$ and $X \wedge Y$ are all random variables.

Proof. We only give the proofs for $X+Y$ and XY here.

$$\{X+Y<x\} = \bigcup_{r: r \text{ is an irrational number}} \{Y \le r\} \cap \{X < x-r\}.$$

$$\begin{aligned}\{XY < x\} = & \{0 < x\} \cap \{Y = 0\} \\ & + \{XY < x\} \cap \{Y > 0\} \\ & + \{XY < x\} \cap \{Y < 0\}.\end{aligned}$$

$$\{XY < x\} \cap \{Y > 0\} = \bigcup_{r:r>0 \text{is an irrational number}} \{0 < Y \le r\} \cap \{X < x/r\}.$$

$$\{XY < x\} \cap \{Y < 0\} = \bigcup_{r:r>0 \text{is an irrational number}} \{-r \le Y < 0\} \cap \{X > -x/r\}.$$

Theorem

Suppose that $\{X_n\}$ *is a sequence of random variables. Suppose that for every* ω, $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ *exists and is finite. Then* X *is a random variable.*

Proof. We have

$$X = \limsup_{n\to\infty} X_n = \inf_{n\geq 1} \sup_{m\geq n} X_m.$$

So

$$\{X \leq x\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{X_m \leq x\}.$$

Theorem

The following statements are equivalent.

- X *is a discrete random variables;*
- $X = \sum_{m=1}^{\infty} x_m I_{A_m}$ *for disjoint sets* $A_m \in \mathcal{F}$, $\sum_{m=1}^{\infty} A_m = \Omega$.

Theorem

The following statements are equivalent.

- X *is a discrete random variables;*
- $X = \sum_{m=1}^{\infty} x_m I_{A_m}$ *for disjoint sets* $A_m \in \mathcal{F}$, $\sum_{m=1}^{\infty} A_m = \Omega$.

Proof. $A_m = \{\omega : X(\omega) = x_m\}$.

A discrete random variable with finite many values is called a simple random variables.

A discrete random variable with finite many values is called a simple random variables.

Theorem

Suppose X, Y are discrete (or simple) random variables. Then cX, $X \pm Y$, XY, $X \vee Y$ and $X \wedge Y$ are all discrete (or simple) random variables.

Proof. Suppose $X = \sum_i a_i I_{A_i}$ for disjoint sets A_i and $Y = \sum_j b_j I_{B_j}$ for disjoint sets B_j. Then

$$X \pm Y = \sum_{i,j} (a_i \pm b_j) I_{A_i B_j}.$$

So $X \pm Y$ is a discrete (or simple) random variables. The other proofs are similar.

Theorem

1. *For a nonnegative random variable X, there is a non-decreasing sequence of simple random variables $\{X_n\}$ for which $0 \leq X_n(\omega) \nearrow X(\omega)$ for every ω;*
2. *For any random variable X, there is a sequence of simple random variables $\{X_n\}$ for which $X_n(\omega) \to X(\omega)$ and $|X_n(\omega)| \leq |X(\omega)|$ for every ω.*

Proof. (1). Suppose $X \geq 0$. For $n \geq 2$, define

$$X_n(\omega) = \begin{cases} n, & \text{if } X(\omega) > n; \\ 0, & \text{if } X(\omega) = 0; \\ \frac{m}{2^n}, & \text{if } \frac{m}{2^n} < X(\omega) \leq \frac{m+1}{2^n}, \\ & \qquad m = 0, 1, \ldots, n2^n - 1. \end{cases}$$

Then $0 \leq X_n(\omega) \nearrow X(\omega)$.

(2). For general X, we define Y_n for $X^+ = \max(X,0)$ and Z_n for $X^- = \max(-X,0)$ as in (1). Then $X_n = Y_n - Z_n$ is desired.