Chapter 1: Probability Theory

1.1-1.5(omitted)

1.6 Posterior Possibilities

(1) Law of Total Probability & Bayes' Theorem

If we have some portions $A_1, A_2, ..., A_n$ that consist of a sample space S, which is

$$S = A_1 \cup A_2 \cup \cdots \cup A_n \quad (1)$$

Suppose we have known the possibility of these n events, and conditional possibilities $P(B|A_1),\ P(B|A_2), \cdots, P(B|A_n)$, where

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots (A_n \cap B) \quad (2)$$

(**Notes: In fact, formula (2) is an identical equation, which can be expressed in Discrete Math as:**)

$A_1 B + A_2 B + \cdots + A_n B = B(A_1 + A_2 + \cdots + A_n) = B$ (non-restrict proof)

Thus, the possibility of event B happened is

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) P(B|A_i) \quad (3)$$

(Obtained from the formula of conditional possibility)

(**Notes: The reason why $P(B|A_i)$ but not $P(A_i|B)$ appears in the formula is that assume that $P(A_i|B)$ should appear in formula (3).**

Since $P(A_i|B) = \dfrac{P(A_i \cap B)}{P(B)}$, and $P(B)$ is the target we want to find, which conflicts.)

The formula (3) is called **Law of Total Probability**.

Furthermore, for the reason mentioned above about the formula (3), if we would like to find $P(A_i|B)$ for a specific i, how it works? – **apply the formula of conditional possibility**(since we have obtained the formula for $P(B)$)

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)} \quad (4)$$

Formula (4) is called **Bayes' Theorem**

1.7(omitted)

Chapter 2: Random Variables

2.1 Discrete Random Variables

(1) Cumulative Distribution Function

The **cumulative distribution function** of a random variable X is the function

$$F(x) = P(X \leqslant x) = \sum_{y \leqslant x} P(X = y) \quad (5)$$

Notes: From formula (5), you can just take CDF as the summation of a single discrete random variable possibility.

2.2 Continuous & Discrete Random Variables

(1) Probability Density Function

Since the random variables are continuous, it is impossible to find a single discrete random variable possibility. However, the principle of the total sum of all the possibilities is 1 is constant, but just the probability is continuous, which can be expressed with a function $f(x)$.

Thus, **probability density function** $f(x)$ satisfying $f(x) \geqslant 0$ and

$$\int_{\text{state space}} f(x)dx = 1 \quad (6)$$

And the probability $P(a \leqslant X \leqslant b)$ is

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx \quad (7)$$

Note that for a specific value $X = a$, the probability is always 0, since $\int_a^a f(x)dx = 0$. **But to understand it more intuitively, we can tell that since the random variables are continuous, take a single and discrete value is insignificant, and we can consider this action as an event, thus the probability this event happened is 0.**

(2) Cumulative Distribution Function

Analogically, from what we obtained for CDF in discrete random variables, we can also obtain the CDF for continuous random variables:

$$P(X \leqslant x) = \int_{-\infty}^{x} f(y)dy \quad (8)$$

(3) Variance of Random Variables

The definition of variance for discrete random variables is

$$\mathrm{Var}(X) = \sum_{i=1}^{n}(X_i - E(X))^2 p_i = E((X - E(X))^2) \quad (9)$$

(Obtained from the definition of Expectation for discrete random variables)

Thus,

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - 2E(XE(X)) + (E(X))^2 \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X^2) - (E(X))^2\end{aligned}$$

Notice that the final form of the variance in formula (9) also works for continuous random variables, thus

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \int_a^b x^2 f(x)dx - \left(\int_a^b xf(x)dx\right)^2 \quad (10)$$

(4) Quantiles

The pth quantiles of a random variable X with a cumulative distribution function $F(x)$ is defined to be the **value** $\boldsymbol{x}$ for which $F(x) = p$. It is also referred to as $(100p$)th **percentile** of the random variable.

(5) Joint Probability Distributions

Analogically, for jointly distributed variables, we also have **probability density function**

$$\sum_i \sum_j p_{ij} = 1(\text{Discrete}) \quad (11)$$

$$\iint_{\text{state space}} f(x,y)\,dxdy = 1(\text{Continuous},\ f(x,y) \geqslant 0) \quad (12)$$

And the probability $P(a \leqslant X \leqslant b,\ c \leqslant Y \leqslant d)$ is

$$P(a \leqslant X \leqslant b,\ c \leqslant Y \leqslant d) = \sum_{a \leqslant i \leqslant b} \sum_{c \leqslant j \leqslant d} p_{ij}\,(\text{Discrete}) \quad (13)$$

$$P(a \leqslant X \leqslant b,\ c \leqslant Y \leqslant d) = \int_a^b \int_c^d f(x,y)\,dydx\,(\text{Continuous})\,(14)$$

For **cumulative distribution function**, we have $F(x,y) = P(X \leqslant x,\ Y \leqslant y)$, where

$$F(x,y) = \sum_{X_i = i \leqslant x} \sum_{Y_j = j \leqslant y} p_{ij}\,(\text{discrete}) \quad (15)$$

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dydx\,(\text{Continuous},\ f(x,y) \geqslant 0) \quad (16)$$

(6) Marginal Probability Distribution

The marginal probability distribution of variable X in terms of jointly distributed variables is defined as

$$P(X = x_i) = \sum_j p_{ij}\,(\text{Discrete}) \quad (17)$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy\ (\text{Continuous}) \quad (18)$$

(7) Conditional Probability Distribution

The conditional probability distribution of variable X in terms of jointly distributed variables is defined as

$$p_{i|Y=y_j} = p(X = x_i | Y = y_j) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{+j}}\,(\text{Discrete}) \quad (19)$$

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)}\,(\text{Continuous}) \quad (20)$$

(8) Independence & Covariance

If two jointly random variables X and Y are **independent**, then

$$p_{ij} = p_{i+} p_{+j} \text{ (Discrete)} \quad (21)$$

$$f(x,y) = f_X(x) f_Y(y) \text{ (Continuous)} \quad (22)$$

The covariance of two random variables X and Y is defined to be

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\ &= E(XY) - 2E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y)\end{aligned} \quad (23)$$

where $\mathrm{Cov}(x,y) \in \mathbb{R}$. $\mathrm{Cov}(x,y) = 0$ if X and Y are independent.

The correlation of two random variables X and Y is defined to be

$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}} \quad (24)$$

where $\mathrm{Cor}(x,y) \in [-1, 1]$. $\mathrm{Cor}(x,y) = 0$ if X and Y are independent.

2.3 Combinations and Functions of Random Variables

(1) Linear Function

A linear function of a random variable X can be defined as $Y = aX + b(a,\ b \in \mathbb{R})$. We can obtain

$$\begin{aligned} E(Y) &= E(aX+b) \\ &= \sum_{i=1}^{n}(ax_i+b)p_i \\ &= aE(X)+b\sum_{i=1}^{n}p_i \\ &= aE(X)+b \end{aligned} \quad (25)$$

Likewise, we can obtain the variance of X

$$\begin{aligned} \mathrm{Var}(Y) &= E(Y^2)-(E(Y))^2 \\ &= E((aX+b)^2)-(aE(X)+b)^2 \\ &= E(a^2X^2+2abX+b^2)-a^2(E(X))^2-2abE(X)-b^2 \\ &= a^2E(X^2)-a^2(E(X))^2 \\ &= a^2\mathrm{Var}(X) \end{aligned} \quad (26)$$

(2) Linear Combinations of Random Variables

Given two random variables X_1 and X_2 we can obtain that

$$E(X_1+X_2) = E(X_1)+E(X_2) \quad (27)$$

Proof: Only prove that when X_1 *and* X_2 *are discrete and independent.*

Assume that $P(X_1 = x_i) = p_i(1 \leqslant i \leqslant m),\ P(X_2 = y_j) = q_j(1 \leqslant j \leqslant n)$

Let $Y = X_1 + X_2$. The prerequisite that Y holds is that X_1 and X_2 can be obtained simultaneously, indicating that $X_1 = x_i$ and $X_2 = y_j$ holds at the same time. Since x_i and y_i are independent each other, the probability of $(x_i + y_j)$ holds is the product of that of x_i and y_i. The number of $(x_i + y_i)$ is the combination of x_i and y_j.

Thus,

$$E(Y) = x_1 p_1 \sum_{j=1}^{n} q_j + x_2 p_2 \sum_{j}^{n} q_j + \ldots + x_m p_m \sum_{j}^{n} q_j + y_1 q_1 \sum_{i=1}^{m} p_i + \ldots + y_n q_n \sum_{i=1}^{m} p_i$$
$$= \sum_{k=1}^{m} x_k p_k \sum_{j=1}^{n} q_j + \sum_{t=1}^{n} y_t q_t \sum_{i=1}^{m} p_i$$
$$= \sum_{k=1}^{m} x_k p_k + \sum_{t=1}^{n} y_t q_t$$
$$= E(X_1) + E(X_2)$$

Also, for the variance in general,

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1,\ X_2) \quad (28)$$

Proof:

Apply the simplified formula for variance, we can obtain

$$\mathrm{Var}(X_1 + X_2) = E[(X_1 + X_2)^2] - [E(X_1 + X_2)]^2$$
$$= E(X^2{}_1 + 2X_1X_2 + X^2{}_2) - [(E(X_1) + E(X_2))^2]$$
$$= E(X^2{}_1) + 2E(X_1X_2) + E(X^2{}_2) - (E(X_1))^2 - 2E(X_1)E(X_2) - (E(X_2))^2$$
$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1,\ X_2)$$

Since $\mathrm{Cov}(X_1,\ X_2) = E(X_1X_2) - E(X_1)E(X_2)$

When X_1 and X_2 are independent each other, indicating that $\mathrm{Cov}(X_1,\ X_2) = 0$, thus,

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

Therefore, if $X_1, X_2, \ldots, X_n$ is a sequence of random variables and $a_1, a_2, \ldots, a_n$ and b are constants, then

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n + b) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) \quad (29)$$

$$\mathrm{Var}(a_1X_1 + a_2X_2 + \ldots + a_nX_n + b) = a^2{}_1\mathrm{Var}(X_1) + a^2{}_2\mathrm{Var}(X_2) + \ldots + a^2{}_n\mathrm{Var}(X_n) \quad (30)$$

Chapter 3: Discrete Probability Distributions

3.1 Binomial Distribution

(1) Bernoulli Trial

A Bernoulli random variable $p(0 \leqslant p \leqslant 1)$ takes the value 1 and 0 with $P(X=1)=p$ and $P(X=0)=1-p$. Thus $E(X)=p$, and $\mathrm{Var}(X)=p(1-p)$.

(2) Binomial Distribution

The Binomial Distribution consists of n **independent** Bernoulli trials that count the number of 'success' in the process, thus you can analogize it as the process of taking a ball n times from a box(put back the ball into the box each time after taking it.). The probability when $X=x$ is

$$P(X=x)=C_n^x p^x(1-p)^{n-x} \quad (31)$$

where p is called 'success probability'.

To obtain the expectation and variance of binomial distribution in a convenient way, from the definition above we can know that

$$E(X)=np,\ \mathrm{Var}(X)=np(1-p) \quad (32)$$

If a random variable X aligns with Binomial Distribution, then the *proportion* of successes $Y=\frac{X}{n}$ has the expectation and variance

$$E(Y)=p,\ \mathrm{Var}(X)=\frac{p(1-p)}{n} \quad (33)$$

(3) Geometric Distribution

The Geometric Distribution consists of *a sequence of* **independent** Bernoulli trials that counts the number of trials performed until the **first** success occurs. The probability when $X=x(x \in \mathbb{N}_+)$ (indicating that first success occurs in the xth trial) is

$$P(X=x)=(1-p)^{x-1}p \quad (34)$$

The expectation and variance of Geometric Distribution are

$$E(X)=\frac{1}{p},\ \mathrm{Var}(X)=\frac{1-p}{p^2} \quad (35)$$

Proof:

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$
$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$
$$= -\,p\sum_{k=1}^{\infty} \frac{d}{dp}[(1-p)^{k}]$$
$$= -\,p\frac{d}{dp}\sum_{k=1}^{\infty}(1-p)^{k} \text{ (additivity of derivative)}$$
$$= -\,p\frac{d}{dp}\left(\frac{1-p}{p}\right)$$
$$= \frac{1}{p}$$

To obtain the variance, we need to compute $E(X^2)$ first.

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1}$$
$$= p\sum_{k=0}^{\infty} k^2 p(1-p)^{k-1}$$
$$= -\,p\frac{d}{dp}\sum_{k=1}^{\infty} k(1-p)^{k}$$
$$= -\,p\frac{d}{dp}\left(\frac{1-p}{p^2}\right)$$
$$= \frac{2-p}{p^2}$$

Thus, we can obtain

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

The Cumulative Distribution Function is

$$F(x)=P(X\leqslant x)=1-(1-p)^{x} \quad (36)$$

Proof:

$$\begin{aligned}P(X\leqslant x)&=\sum_{k=1}^{x}p(1-p)^{k-1}\\&=p\sum_{k=1}^{x}(1-p)^{k-1}\\&=p\frac{1-(1-p)^{x}}{p}\\&=1-(1-p)^{x}\end{aligned}$$

(4) Negative Binomial Distribution

The Negative Binomial Distribution refers to a situation where we succeed r times in x trials(indicating that rth success occurs in the xth trial).($r\in\mathbb{N},\ x\in\mathbb{N}_{+}$) We do not know the exact number of the experiment, which is different from Binomial Distribution that owns a fixed number of the experiment n.

Thus, the probability function is

$$P(X=x)=C_{x-1}^{r-1}p^{r}(1-p)^{x-r}=C_{x-1}^{r-1}p^{r-1}(1-p)^{x-r}\cdot p \quad (37)$$

The expectation and variance of Negative Binomial Distribution is

$$E(X)=\frac{r}{p},\ \mathrm{Var}(X)=\frac{r(1-p)}{p^{2}} \quad (38)$$

Proof:

Before proving the formula, we need to know how to compute C_x^r and C_{x-1}^{r-1}. Thus,

$$C_x^r=\frac{x(x-1)...(x-r+1)}{r(r-1)...\times2\times1},\ C_{x-1}^{r-1}=\frac{(x-1)(x-2)...(x-r+1)}{(r-1)(r-2)...\times2\times1}.$$

Thus, the expectation is

$$
\begin{aligned}
E(X) &= \sum_{x=r}^{\infty} x P(X=x) \\
&= \sum_{x=r}^{\infty} x C_{x-1}^{r-1} p^r (1-p)^{x-r} \\
&= \sum_{x=r}^{\infty} x \cdot \frac{r}{x} C_x^r p^r (1-p)^{x-r} \\
&= r \sum_{x=r}^{\infty} C_x^r p^r (1-p)^{x-r} \\
&= \frac{r}{p} \sum_{x=r}^{\infty} C_x^r p^{r+1} (1-p)^{x-r}
\end{aligned}
$$

Since $\sum_{x=r}^{\infty} C_x^r p^{r+1}(1-p)^{x-r} = 1$, so we can obtain that $E(X) = \frac{r}{p}$.

To obtain the variance in a understandable way, we can consider the negative binomial distribution as r times geometric distributions. Thus, we can obtain that $\operatorname{Var}(X) = \frac{r(1-p)}{p^2}$

(5) Hypergeometric Distribution

This distribution is slightly different from binomial distribution. Assume that we have a case where we have total N products including r products that are disqualified. Now we randomly take n products for checking, the probability of finding x disqualified products is

$$P(X=x)=\frac{C_r^x C_{N-n}^{n-x}}{C_N^n} \quad (39)$$

The expectation of the distribution is $E(X)=\frac{nr}{N}$ (40)

Proof:

Case 1: When $N=r=1$, $E(X)=1=\frac{nr}{N}$

Case 2: When $N\geqslant 2$, $r=1$, $P(X=0)=\frac{C_{N-1}^n}{C_N^n}$, $P(X=1)=\frac{C_1^1 C_{N-1}^{n-1}}{C_N^n}$

Thus, $E(X)=\frac{C_1^1 C_{N-1}^{n-1}}{C_N^n}=\frac{C_{N-1}^{n-1}}{C_N^n}=\frac{n}{N}=\frac{nr}{N}$

Case 3: When $N\geqslant 2$, $r\geqslant 2$

When $n<N-r$, we can obtain

$$E(X)=\sum_{k=0}^{\min\{n,r\}}\frac{kC_r^k C_{N-r}^{n-k}}{C_N^n}=\frac{r}{C_N^n}\sum_{k=1}^{\min\{n,r\}}C_{r-1}^{k-1}C_{N-r}^{n-k}=\frac{nr}{N}$$

When $n>N-r$, we can obtain

$$E(X)=\sum_{k=n-N+r}^{\min\{n,r\}}\frac{kC_r^k C_{N-r}^{n-k}}{C_N^n}=\frac{r}{C_N^n}\sum_{k=n-N+r}^{\min\{n,r\}}C_{r-1}^{k-1}C_{N-r}^{n-k}=\frac{nr}{N}$$

The variance of distribution is $\operatorname{Var}(X)=\frac{nr}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$ (41)

Proof:

$$
\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - [E(X)]^2 \\
&= \sum_{k=0}^{\min\{n,r\}} \frac{k^2 C_r^k C_{N-r}^{n-k}}{C_N^n} - \frac{n^2 r^2}{N^2} \\
&= \frac{1}{C_N^n} \sum_{k=0}^{\min\{n,r\}} kr C_{r-1}^{k-1} C_{N-r}^{n-k} - \frac{n^2 r^2}{N^2} \\
&= \frac{r}{C_N^n} \sum_{k=1}^{\min\{n,r\}} [(k-1) C_{r-1}^{k-1} C_{N-r}^{n-k} + C_{r-1}^{k-1} C_{N-r}^{n-k}] - \frac{n^2 r^2}{N^2} \\
&= \frac{r}{C_N^n} \sum_{k=2}^{\min\{n,r\}} (r-1) C_{r-2}^{k-2} C_{N-r}^{n-k} + \frac{r}{C_N^n} \cdot C_{N-1}^{n-1} - \frac{n^2 r^2}{N^2} \\
&= \frac{r(r-1)}{C_N^n} \cdot C_{N-2}^{n-2} + \frac{r}{C_N^n} \cdot C_{N-1}^{n-1} - \frac{n^2 r^2}{N^2} \\
&= \frac{r(r-1) \cdot n(n-1)}{N(N-1)} + \frac{nr}{N} - \frac{n^2 r^2}{N^2} \\
&= \frac{nr}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}
\end{aligned}
$$

In the two proofs above, we apply two corollaries to help us finish them.

$$
\sum_{k=1}^{\min\{n,r\}} C_{r-1}^{k-1} C_{N-r}^{n-k} = C_{N-1}^{n-1} (n \leqslant N-r)
$$

$$
\sum_{k=n-N+r}^{\min\{n,r\}} C_{r-1}^{k-1} C_{N-r}^{n-k} = C_{N-1}^{n-1} (n \geqslant N-r)
$$

(6) Poisson Distribution

Poisson Distribution is usually used to define a random variable that *counts* the number of "events" that occur within certain specified boundaries. Its mass function is

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\ (42)$$

We can prove this function by its definition

Proof:

Assume that we have a period of time $[0,1]$ and take a great number $n\,(n\in\mathbb{N}_{+})$. Thus, we divide this period of time into n pieces, and we can obtain

$$l_1=\left[0,\frac{1}{n}\right],\ l_2=\left[\frac{1}{n},\frac{2}{n}\right],\ldots,\ l_n=\left[\frac{n-1}{n},1\right]$$

Our goal is to derive the possibility that an event occurred in every piece l_i, and we have two assumptions

① The probability of an event occurred is proportional to the range of time divided in each piece, denoted by $\frac{\lambda}{n}$.

② The occurrence of the event in each piece is mutually independent.

From these assumptions above, we can know that $X\sim B\left(n,\frac{\lambda}{n}\right)$. Thus, we can obtain

$$P(X=i)=C_n^i\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$

When $n\to\infty$, we have

$$\begin{aligned}\lim_{n\to\infty}P(X=i)&=\lim_{n\to\infty}C_n^i\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}\\&=\lambda^i\lim_{n\to\infty}C_n^i\left(\frac{1}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}\\&=\frac{\lambda^i}{i!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-i}\\&=\frac{e^{-\lambda}\lambda^i}{i!}\end{aligned}$$

Thus, we prove the function.

The expectation and variance of the distribution is $E(X)=\mathrm{Var}(X)=\lambda$ (43)

(7) Multinomial Distribution

Multinomial Distribution meets the situation where a sequence of n independent trials where each trial can have k outcomes that occur with constant probability $p_1, p_2, \ldots, p_k$ where $\sum_{i=1}^{k} p_i = 1$. We have random variables $X_1, X_2, \ldots, X_k$ that count the number of the occurrences of each kind of outcome. Its mass function can be obtained by definition, which is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} {p_1}^{x_1} {p_2}^{x_2} \ldots \cdot {p_k}^{x_k} \quad (44)$$

Since the multinomial distribution is a special case of binomial distribution that owns multiple outcomes in each trial, for random variables $X_1, X_2, \ldots, X_k$, the expectations and variances are

$$E(X_i) = np_i,\ \mathrm{Var}(X_i) = np_i(1 - p_i) \quad (45)$$

Chapter 4: Continuous Probability Distributions

4.1 The Uniform Distribution

The Uniform Distribution refers to a probability distribution with a **constant** probability distribution function between two points a and b (where a and b are constants), denoted by $X \sim U(a,\ b)$.

According to the definition above and the total sum of probability is 1, we can obtain that $f(x)=\frac{1}{b-a}$, so that the cumulative distribution function is

$$F(x)=\int_{y=a}^{x} f(y)dy=\int_{y=a}^{x}\frac{dy}{b-a}=\frac{x-a}{b-a} \quad (46)$$

where $a \leqslant x \leqslant b$.

Thus, its expectation and variance are $E(X)=\frac{a+b}{2}$, $\operatorname{Var}(X)=\frac{(b-a)^2}{12}$

Proof:

From the definition of expectation, we can obtain

$$\begin{aligned} E(X) &= \int_a^b x \cdot \frac{dx}{b-a} \\ &= \frac{1}{2(b-a)}(b^2-a^2) \\ &= \frac{a+b}{2} \end{aligned}$$

Thus, we can obtain

$$\begin{aligned} E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{1}{3(b-a)}(b^3-a^3) \\ &= \frac{a^2+ab+b^2}{3} \end{aligned}$$

Where $b^3-a^3=(b-a)(a^2+ab+b^2)$

Thus, we can obtain the variance is

$$\operatorname{Var}(X)=E(X^2)-[E(X)]^2=\frac{a^2+ab+b^2}{3}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$$

4.2 The Exponential Distribution

The Exponential Distribution is often used to model failure or waiting times and interarrival times. Thus, its state space is $x \geqslant 0$. Its probability density function is

$$f(x) = \lambda e^{-\lambda x} \quad (47)$$

Thus, we can obtain the cumulative distribution function is

$$F(x) = \int_0^x f(y)\,dy = \int_0^x \lambda e^{-\lambda y}\,dy = 1 - e^{-\lambda x} \quad (48)$$

The expectation and variance are $E(X) = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$ (49)

Proof:

We can obtain the expectation is

$$\begin{aligned} E(X) &= \int_0^\infty x\lambda e^{-\lambda x}\,dx \\ &= \frac{1}{\lambda}\int_0^\infty u e^{-u}\,du \text{ (let } u = \lambda x) \\ &= \frac{1}{\lambda} \end{aligned}$$

Thus, we can obtain

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx \\ &= \lim_{t\to\infty}\int_0^t x^2 \lambda e^{-\lambda x}\,dx \\ &= -\lim_{t\to\infty}\left[x^2 e^{-\lambda x} + \frac{2(\lambda x + 1)e^{-\lambda x}}{\lambda^2}\right]_0^t \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Thus, we can obtain the variance is

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$