

# 第一章 预备知识

随机过程及其在金融中的应用

中国人民大学出版社

# 本章内容

# 样本和样本空间

通常把按照一定想法去做的事件称为试验 (experiment), 把试验的可能结果称为样本点 (sample point), 样本点的集合称为样本空间 (sample space)。通常将样本空间记为  $\Omega$ , 样本点记为  $\omega$ 。

## 举例:

一个均匀的骰子 (dice) 有六个面, 每次抛出这个骰子, 会得到相应的点数 (比如: 3)。这里“抛骰子”就是一个事件, 也就是“试验”; 在抛出骰子后得到的“点数” 3 就是“样本点”  $\omega = 3$ ; 而骰子的可能点数为  $1, 2, 3, \dots, 6$ , 因此对应的样本空间就是  $\Omega = \{1, 2, 3, \dots, 6\}$ 。

# 事件与概率

事件 (event) 是样本空间  $\Omega$  的子集, 且满足三个条件:

- ①  $\Omega$  是事件;
- ② 若  $A$  是事件, 则  $A^c$  是事件;
- ③ 若  $A_i$  是事件, 则  $\bigcup_{i=1}^{\infty} A_i$  是事件。

## 骰子的例子

如果“抛出的点数为 3”是事件  $A$ , 则其对立事件  $A^c$  就是“抛出的点数不为 3”; 另外, “抛一次骰子”是事件, 则“抛无数次骰子”组成的所有事件的并集也是事件。

# 事件

对于事件  $A$ , 如果使用  $\mathbb{P}(A)$  表示事件发生的概率, 则  $\mathbb{P}(A)$  满足以下条件:

- ① 非负性:  $\mathbb{P}(A) \geq 0$ ;
- ② 完备性:  $\mathbb{P}(\Omega) = 1$ ;
- ③ 可列可加性: 对于互不相容的事件  $A_1, A_2, \dots$ , 有:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

# 概率空间

对于样本空间  $\Omega$  和概率  $\mathbb{P}$ , 用  $\mathcal{F}$  表示全体事件时, 称三位一体的  $(\Omega, \mathcal{F}, \mathbb{P})$  为概率空间 (probability space)。  $\mathcal{F}$  称作  $\sigma$ -代数 (sigma-algebra), 相当于样本空间  $\Omega$  的子集的集合。

举例:

假设样本空间  $\Omega = \{1, 2, 3\}$ , 则  $\mathcal{F}$  可表示为:

$$\mathcal{F} = \left\{ \emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega \right\}$$

# $\sigma$ -代数

对于  $\sigma$ -代数  $\mathcal{F}$ , 其中的元素满足以下条件:

- ① 若  $A_i \in \mathcal{F}$ , 则  $A_i^c \in \mathcal{F}$ ;
- ② 若  $A_i, A_j \in \mathcal{F}, \forall i, j$ , 则  $A_i \cap A_j \in \mathcal{F}$ ;
- ③ 若  $A_i, A_j \in \mathcal{F}, \forall i, j$ , 则  $A_i \cup A_j \in \mathcal{F}$ .

说明:

简言之, 对  $\sigma$ -代数  $\mathcal{F}$  中的元素取并集、交集和补集, 结果均在  $\mathcal{F}$  中。(  $\sigma$ -代数对并、交、补运算均封闭。)

# 概率的基本性质

对于事件  $A_i, i = 1, 2, \dots, n$ , 其发生的概率具有如下性质:

- ①  $\mathbb{P}(\emptyset) = 0$ ;
- ② 当且仅当  $A_1, A_2, \dots, A_n$  互不相容时, 下式的等号成立:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i)$$

- ③ 如果  $A_2 \subset A_1$ , 则  $\mathbb{P}(A_1) - \mathbb{P}(A_2) = \mathbb{P}(A_1 - A_2) \geq 0$ ;
- ④  $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 A_2)$ ;
- ⑤ 条件概率公式: 当  $\mathbb{P}(A_1) > 0$  时,

$$\mathbb{P}(A_2|A_1) = \frac{\mathbb{P}(A_1 A_2)}{\mathbb{P}(A_1)}$$



# 乘法公式

乘法公式可看作条件概率公式的直接推论，其表达式如下：

$$\mathbb{P}(B_1 B_2 \cdots B_n) = \mathbb{P}(B_1) \mathbb{P}(B_2 | B_1) \cdots \mathbb{P}(B_n | B_1 B_2 \cdots B_{n-1})$$

当  $\mathbb{P}(A) > 0$  时，

$$\mathbb{P}(B_1 B_2 \cdots B_n | A) = \mathbb{P}(B_1 | A) \mathbb{P}(B_2 | B_1 A) \cdots \mathbb{P}(B_n | B_1 B_2 \cdots B_{n-1} A)$$

# 全概率公式

若事件  $A_1, A_2, \dots, A_n$  互不相容, 则当  $\bigcup_{i=1}^n A_i = \Omega$  时, 有:

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(BA_i) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

当  $\mathbb{P}(A) > 0$  时, 有:

$$\mathbb{P}(B|A) = \sum_{i=1}^n \mathbb{P}(B|A_i A)\mathbb{P}(A_i|A)$$

## 全概率公式 (cont.)

全概率公式的意义在于，当直接计算  $\mathbb{P}(B)$  较为困难时，可以通过求小事件的概率，然后相加，从而求得事件  $B$  的概率。而将事件  $B$  进行分割的时候，则是先找到样本空间  $\Omega$  的某个划分 (partition)  $\{A_1, A_2, \dots, A_n\}$ ，进而得到相对应的事件  $B$  的分解，即：

$$B = BA_1 + BA_2 + \cdots + BA_n$$

利用条件概率的计算公式，可得：

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(BA_1) + \mathbb{P}(BA_2) + \cdots + \mathbb{P}(BA_n) \\ &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n) \\ &= \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)\end{aligned}$$

# 贝叶斯公式

若  $A_1, A_2, \dots, A_n$  为一系列互不相容的事件, 并且

$$\bigcup_{i=1}^n A_i = \Omega, \quad \mathbb{P}(A_i) > 0, \quad \forall i$$

则对任一事件  $B$ , 有:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{k=1}^n \mathbb{P}(B|A_k)\mathbb{P}(A_k)}, \quad i = 1, 2, \dots, n$$

## 贝叶斯公式 (cont.)

与全概率公式解决的问题相反，贝叶斯 (Bayesian) 公式是建立在条件概率的基础上，通过确定的结果寻找发生的原因。此处  $\mathbb{P}(A_i)$  ( $i = 1, 2, \dots, n$ ) 表示各种原因发生的可能性大小，故称先验概率 (prior probability)； $\mathbb{P}(A_i|B)$  则反映当结果  $B$  发生之后，再对产生这一结果的各种原因的可能概率进行推断，故称后验概率 (posterior probability)。

# 随机变量

随机变量  $X$  是定义在样本空间  $\Omega$  上的函数。若样本空间  $\Omega$  中的样本点  $\omega$  是离散的, 则相应的  $X(\omega)$  是离散型 (discrete) 随机变量; 若样本点  $\omega$  是连续的, 则相应的  $X(\omega)$  是连续型 (continuous) 随机变量。

## 以金融市场为例

考虑股票价格跳跃次数的样本空间, 其可能的取值为  $\Omega = \{0, 1, 2, \dots\}$ , 不难看出取值是非负的整数, 此时所得到的股价跳跃次数的随机变量就是离散型随机变量; 若考虑股票在某时刻的可能价格的样本空间, 则其取值应当为  $\Omega \in \mathbb{R}^+ \cup \{0\}$ , 此时得到的股价随机变量就是连续型随机变量, 因为对应的样本空间  $\Omega$  取值为非负实数。

## 随机变量 (cont.)

对于离散型随机变量而言，其对应样本点  $\omega$  的概率是非负的，并且样本点所有可能取值下的概率之和等于 1（满足概率的完备性），即：

$$\mathbb{P}(X = \omega) \geq 0, \quad \sum_{\omega \in \Omega} \mathbb{P}(X = \omega) \equiv 1$$

连续型随机变量无法得到类似的性质，由于样本点是连续的，对应的某一个样本点  $\omega$  的概率接近于零，因此需要引入新的概念来刻画连续型随机变量的概率等特征，这就是分布函数。

# 分布函数

对于连续型随机变量  $X$  而言, 其分布函数  $F_X(t)$  (distribution function) 定义为:

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(s) \, ds$$

其中,  $f_X(s)$  是  $X$  的概率密度函数 (probability density function, pdf)。

说明:

分布函数  $F_X(t)$  是单调不减的右连续函数, 并且其取值范围为  $[0, 1]$ 。概率密度函数可以定义在任何连续随机变量上, 比如正态分布、 $F$  分布、 $t$  分布、 $\chi^2$  分布等。



## 分布函数 (cont.)

对于离散型随机变量  $Y$  而言, 其分布函数  $G_Y(t)$  也有类似的定义, 只不过原先的积分符号变成了求和符号罢了, 公式如下:

$$G_Y(t) = \mathbb{P}(Y \leq t) = \sum_{\substack{\omega \leq t \\ \omega \in \Omega}} \mathbb{P}(Y = \omega)$$

此处的  $\mathbb{P}(Y = \omega)$  称作概率质量函数 (probability mass function, pmf)。

概率质量函数可以定义在任何离散型随机变量上, 比如二项分布、负二项分布、泊松分布、几何分布等等。

# 概率质量函数和概率密度函数

- 概率质量函数是对离散型随机变量定义的，本身代表该值的概率；
- 概率密度函数是对连续型随机变量定义的，本身不是概率，只有对连续型随机变量的概率密度函数在某区间内进行积分后才是概率。

要想求得连续型随机变量  $X$  在  $[a, b]$  这一区间上的概率，则计算公式如下：

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(s) \, ds$$

# 随机向量

如果  $X_1, X_2, \dots, X_n$  都是随机变量, 则  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  称作随机向量。相应地, 定义在  $n$  维实数域  $\mathbb{R}^n$  上的  $n$  元函数

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

称作  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  的分布函数。

与前面类似, 对于连续型随机向量  $\mathbf{X}$ , 其在  $\mathbb{R}^n$  上的区域 (domain)  $D$  的概率为:

$$\mathbb{P}(\mathbf{X} \in D) = \underbrace{\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n}}_{n\uparrow} f(\mathbf{x}) \, dx_1 \, dx_2 \cdots dx_n = \int_D f(\mathbf{x}) \, dx_1 \, dx_2 \cdots dx_n$$

其中,  $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$  是  $\mathbf{X}$  的联合密度。

# 随机变量之和的性质：离散型随机变量之和的概率

设随机变量  $X$  和  $Y$  独立，且满足：

$$\mathbb{P}(X = k) = a_k, \quad \mathbb{P}(Y = k) = b_k, \quad k = 0, 1, \dots$$

则  $Z = X + Y = n$ ,  $n \geq 0$  的概率为：

$$\begin{aligned}\mathbb{P}(Z = n) &= \mathbb{P}(X + Y = n) = \sum_{i=0}^n \mathbb{P}(X = i, Y = n - i) \\ &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i) = \sum_{i=0}^n a_i b_{n-i}\end{aligned}$$

令  $\mathbb{P}(Z = n) = c_n$ ，则：
$$c_n = \sum_{i=0}^n a_i b_{n-i}.$$

这里的序列  $\{c_n\}$  称作序列  $\{a_n\}$  和  $\{b_n\}$  的卷积 (convolution)，其中  $n \geq 0$ 。

# 随机变量之和的性质：连续型随机变量之和的分布函数

设随机变量  $X$  和  $Y$  独立，其分布函数分别为  $F(x)$  和  $G(y)$ ，则  $U = X + Y$  有如下分布函数：

$$\begin{aligned} U(t) &= \mathbb{P}(X + Y \leq t) = \sum_{0 \leq s \leq t} \mathbb{P}(X + Y \leq t | Y = s) \mathbb{P}(Y = s) \\ &= \int_0^t \mathbb{P}(X \leq t - s) \, dG(s) = \int_0^t F(t - s) \, dG(s) \end{aligned}$$

由于  $X + Y = Y + X$ ，因此：

$$\begin{aligned} U(t) &= \mathbb{P}(X + Y \leq t) = \sum_{0 \leq s \leq t} \mathbb{P}(X + Y \leq t | X = s) \mathbb{P}(X = s) \\ &= \int_0^t \mathbb{P}(Y \leq t - s) \, dF(s) = \int_0^t G(t - s) \, dF(s) \end{aligned}$$

从而：

$$\int_0^t F(t - s) \, dG(s) = \int_0^t G(t - s) \, dF(s)$$

# 随机变量之和的性质：连续型随机变量之和的密度函数

设随机变量  $X$  和  $Y$  独立，其概率密度函数分别为  $f(x)$  和  $g(y)$ ，则  $U = X + Y$  有如下概率密度函数  $u(t)$ ：

$$\begin{aligned} u(t) &= \int_0^t f(t-s)g(s) \, ds = (f * g)(s) \\ &= \int_0^t g(t-s)f(s) \, ds = (g * f)(s) \end{aligned}$$

$(f * g)(s)$  称为函数  $f$  和  $g$  的卷积。

需要说明的是，卷积运算满足交换律，即  $f * g = g * f$ 。

## 举例:

设随机变量  $X$  和  $Y$  独立, 其分布函数分别为:

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0; \quad G_Y(y) = y/2, \quad 0 \leq y \leq 2$$

求  $U = X + Y$  的分布函数。

### 思路

分布函数求解的公式如下:

$$U(t) = \mathbb{P}(X + Y \leq t) = \int_0^t F_X(t - s) \, dG_Y(s)$$

这里要特别注意, 由于  $y \in [0, 2]$ , 因此需要根据  $t$  的取值分情况来讨论。

当  $t < 2$  时:

$$\begin{aligned} U(t) &= \int_0^t F_X(t-s) \, dG_Y(s) = \int_0^t \left[1 - e^{-\lambda(t-s)}\right] \, d\left(\frac{s}{2}\right) \\ &= \frac{1}{2} \left( \int_0^t ds - e^{-\lambda t} \int_0^t e^{\lambda s} ds \right) = \frac{1}{2} \left( t - \frac{1}{\lambda} + \frac{e^{-\lambda t}}{\lambda} \right) \end{aligned}$$

当  $t \geq 2$  时:

$$\begin{aligned} U(t) &= \int_0^2 F_X(t-s) \, dG_Y(s) = \int_0^2 \left[1 - e^{-\lambda(t-s)}\right] \, d\left(\frac{s}{2}\right) \\ &= \frac{1}{2} \left( \int_0^2 ds - e^{-\lambda t} \int_0^2 e^{\lambda s} ds \right) = 1 - \frac{e^{-\lambda t}}{2\lambda} (e^{2\lambda} - 1) \end{aligned}$$

因此:

$$U(t) = \begin{cases} \frac{1}{2} \left( t - \frac{1}{\lambda} + \frac{e^{-\lambda t}}{\lambda} \right), & t < 2 \\ 1 - \frac{e^{-\lambda t}}{2\lambda} (e^{2\lambda} - 1), & t \geq 2 \end{cases}$$



## 举例:

设随机变量  $X$  和  $Y$  独立, 其概率密度函数分别为:

$$f_X(x) = \lambda e^{-\lambda x}; \quad f_Y(y) = \lambda e^{-\lambda y}$$

求  $U = X + Y$  的概率密度函数。

解答:

运用卷积求解如下:

$$\begin{aligned} f_U(t) &= \int_0^t f_X(t-s)f_Y(s) \, ds = \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s} \, ds \\ &= \lambda^2 \int_0^t e^{-\lambda t} \, ds = \lambda^2 e^{-\lambda t} s \Big|_{s=0}^{s=t} = \lambda^2 t \cdot e^{-\lambda t} \end{aligned}$$

# 随机变量的数字特征

对于随机变量而言，我们通常关心它的重要数字特征，比如期望、方差、偏度、峰度等等。本节主要介绍期望和方差，以及更一般的矩母函数和特征函数。

期望 (expectation) 用于反映随机变量平均取值的大小。根据随机变量的不同，可以得到离散型随机变量和连续型随机变量的期望。

# 离散型随机变量的期望

设随机变量  $X$  有离散的概率分布如下：

$$p_i = \mathbb{P}(X = x_i), \quad i = 1, \dots, n$$

则：

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) = \sum_{i=1}^n x_i p_i$$

## 举例：泊松分布的期望

假设随机变量  $X$  服从参数为  $\lambda$  的泊松分布，相应的概率质量函数如下：

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \lambda > 0, n = 0, 1, 2, \dots$$

求泊松分布的期望  $\mathbb{E}(X)$ 。

# 泊松分布的期望 (cont.)

根据期望的定义, 可得:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X=n) = \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n}{(n-1)!} e^{-\lambda} = \lambda \sum_{n=0}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \lambda\end{aligned}$$

注意, 这里最后一步化简用到了  $e^x$  的泰勒展开式, 即:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

# 定理

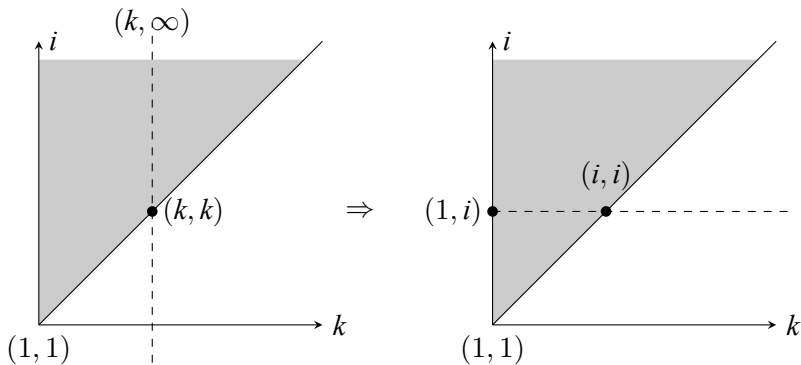
设  $X$  为取值是非负整数的随机变量，其期望值的计算公式如下：

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

证明思路：

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) &= \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} \mathbb{P}(X = i) = \sum_{i=1}^{\infty} \underbrace{\sum_{k=1}^i \mathbb{P}(X = i)}_{i \uparrow \mathbb{P}(X=i)} \\ &= \sum_{i=1}^{\infty} i \cdot \mathbb{P}(X = i) = \mathbb{E}(X) \end{aligned}$$

# 求和符号交换的图示



# 连续型随机变量的期望

设  $X$  是有密度函数  $f_X(x)$  的随机变量, 即密度函数满足:

$$F_X(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^a f_X(x) \, dx$$

则:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx$$



## 举例：指数分布的期望

假设随机变量  $X$  服从参数为  $\lambda$  的指数分布，相应的概率密度函数如下：

$$f_X(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, x \geq 0$$

求指数分布的期望  $\mathbb{E}(X)$ 。

## 指数分布的期望 (cont.)

根据期望的定义, 可得:

$$\mathbb{E}(X) = \int_0^{\infty} x f_X(x) \, dx = \int_0^{\infty} \lambda x e^{-\lambda x} \, dx$$

记  $u = -\lambda x$ , 则:

$$\text{上式} = \frac{1}{\lambda} \int_0^{-\infty} u e^u \, du = \frac{1}{\lambda} \left[ e^u (u - 1) \right]_0^{-\infty} = \frac{1}{\lambda}$$

# 定理

设  $X$  是非负的随机变量, 则:

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X > x) \, dx = \int_0^{\infty} [1 - F_X(x)] \, dx$$

证明思路:

$$\begin{aligned} \int_0^{\infty} \mathbb{P}(X > x) \, dx &= \int_0^{\infty} dx \int_x^{\infty} f_X(y) \, dy = \int_0^{\infty} dy \underbrace{\int_0^y f_X(y) \, dx}_{f_X(y) \text{ 可从积分中提出}} \\ &= \int_0^{\infty} y f_X(y) \, dy = \mathbb{E}(X) \end{aligned}$$

# 方差

方差 (variance) 是反映随机变量离散程度的指标。在概率统计的教科书中, 方差计算的恒等式如下:

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

接下来分别利用这个恒等式来计算离散型随机变量和连续型随机变量的方差。

## 举例：泊松分布的方差

假设随机变量  $X$  服从参数为  $\lambda$  的泊松分布，相应的概率质量函数如下：

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \lambda > 0, n = 0, 1, 2, \dots$$

求泊松分布的方差  $\text{Var}(X)$ 。

思路：

前面已经计算出泊松分布的期望  $\mathbb{E}(X) = \lambda$ ，要求出方差，还需要计算  $\mathbb{E}(X^2)$ 。

# 泊松分布的方差 (cont.)

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{n=0}^{\infty} n^2 \mathbb{P}(X=n) = \sum_{n=0}^{\infty} n^2 \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\&= \sum_{n=0}^{\infty} \frac{n\lambda^n}{(n-1)!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{n\lambda^{n-1}}{(n-1)!} \\&= \lambda e^{-\lambda} \cdot (\lambda + 1) e^{\lambda} \\&= \lambda(\lambda + 1)\end{aligned}$$

因此：

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

## 举例：指数分布的方差

假设随机变量  $X$  服从参数为  $\lambda$  的指数分布，相应的概率密度函数如下：

$$f_X(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, x \geq 0$$

求指数分布的方差  $\text{Var}(X)$ 。

思路：

前面已经计算出指数分布的期望  $\mathbb{E}(X) = \frac{1}{\lambda}$ ，要求出方差，还需要计算  $\mathbb{E}(X^2)$ 。

# 指数分布的方差 (cont.)

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^{\infty} x^2 f_X(x) \, dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} \, dx \\ &= \frac{1}{\lambda} \int_0^{\infty} (\lambda x)^2 e^{-\lambda x} \, dx\end{aligned}$$

记  $u = -\lambda x$ , 则:

$$\begin{aligned}\text{上式} &= -\frac{1}{\lambda^2} \int_0^{-\infty} u^2 e^u \, du \\ &= -\frac{1}{\lambda^2} \left[ e^u (u^2 - 2u + 2) \right]_0^{-\infty} = \frac{2}{\lambda^2}\end{aligned}$$

因此:

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$



# 矩母函数

矩母函数 (moment generating function, mgf) 是一种构造函数。对于任何满足概率密度函数为  $f_X(x)$  的随机变量  $X$ , 其矩母函数的定义如下:

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) \, dx$$

根据泰勒展开式:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

由此可得: 
$$e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$$

# 矩母函数 (cont.)

矩母函数可以相应进行如下展开：

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x) \, dx = \int_{-\infty}^{\infty} f_X(x) \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} \, dx \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \int_{-\infty}^{\infty} x^k f_X(x) \, dx = \sum_{k=0}^{\infty} \frac{t^k}{k!} \cdot \mathbb{E}(X^k) \\ &= 1 + t \cdot \mathbb{E}(X) + \frac{t^2}{2!} \cdot \mathbb{E}(X^2) + \cdots + \frac{t^n}{n!} \cdot \mathbb{E}(X^n) + \cdots \end{aligned}$$

由此可见，矩母函数包含了随机变量  $X$  的各阶矩  $\mathbb{E}(X^n)$ ,

$n = 1, 2, 3, \dots$

## 矩母函数 (cont.)

若对  $M_X(t)$  关于  $t$  求导, 可得:

$$\frac{dM_X(t)}{dt} = \int_{-\infty}^{\infty} e^{tx} \cdot x f_X(x) dx$$

类似地:

$$\frac{d^n M_X(t)}{dt^n} = \int_{-\infty}^{\infty} e^{tx} \cdot x^n f_X(x) dx$$

当  $t = 0$  时:

$$\left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0} = \int_{-\infty}^{\infty} x^n f_X(x) dx = \mathbb{E}(X^n)$$

因此, 可以通过对矩母函数关于  $t$  求  $n$  阶导的方式, 并令  $t = 0$ , 进而求出随机变量的  $n$  阶矩。

## 举例：泊松分布的各阶矩

假设随机变量  $X$  服从速率为  $\lambda$  的指数分布，其概率密度函数为：

$$f_X(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, x \geq 0$$

求其各阶矩。

思路：

通过矩母函数加以求解。

# 泊松分布的各阶矩 (cont.)

由矩母函数的定义可得:

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tx}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda}{t-\lambda} e^{-(\lambda-t)x} \Big|_{x=0}^{x=\infty} = \frac{\lambda}{\lambda-t}, \quad \lambda > t \end{aligned}$$

# 泊松分布的各阶矩 (cont.)

相应地:

$$\begin{aligned}\frac{dM_X(t)}{dt} &= \frac{\lambda}{(\lambda - t)^2} \Rightarrow \mathbb{E}(X) = \left. \frac{dM_X(t)}{dt} \right|_{t=0} = \frac{1}{\lambda} \\ \frac{d^2 M_X(t)}{dt^2} &= \frac{2\lambda}{(\lambda - t)^3} \Rightarrow \mathbb{E}(X^2) = \left. \frac{d^2 M_X(t)}{dt^2} \right|_{t=0} = \frac{2}{\lambda^2} \\ \frac{d^3 M_X(t)}{dt^3} &= \frac{2 \cdot 3\lambda}{(\lambda - t)^4} \Rightarrow \mathbb{E}(X^3) = \left. \frac{d^3 M_X(t)}{dt^3} \right|_{t=0} = \frac{6}{\lambda^3}\end{aligned}$$

因此:

$$\mathbb{E}(X^n) = \frac{n!}{\lambda^n}$$

## 矩母函数的不足

通过矩母函数可以相对容易地求出服从某个概率分布的随机变量之各阶矩。但是该方法仍有不足之处，它依赖于矩母函数关于  $t$  进行  $n$  阶求导后，在  $t = 0$  处有定义，否则会出现无法计算各阶矩的问题。

为解决这个问题，引入特征函数 (characteristic function, cf)，其定义如下：

$$\phi_X(t) = \mathbb{E}(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f_X(x) \, dx$$

特征函数的重要用途在于，若随机变量  $X$  的概率分布  $f_X(x)$  无法直接算出，可以首先计算出它的特征函数  $\phi_X(t)$ ，再通过傅立叶变换 (Fourier transform) 最终算出概率分布  $f_X(x)$ 。

# 依概率收敛

设  $\{X_n\}$ ,  $n \in \mathbb{N}$  是一个随机变量序列。若存在一个随机变量  $X$ , 使得  $\forall \varepsilon > 0$ , 均有:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

则称序列  $\{X_n\}$  依概率收敛 (convergence in probability) 于  $X$ , 记作:

$$X_n \xrightarrow{P} X \quad \text{或者} \quad \text{plim}_{n \rightarrow \infty} X_n = X$$

依概率收敛可以理解为: 任意指定一个正数  $\varepsilon$ , 无论  $n$  取多大,  $X_n$  与  $X$  差距的绝对值大于  $\varepsilon$  的可能次数仍然是存在的, 但只要  $n$  足够大, 这种例外情形的次数占比将逐渐趋于 0。



# 弱大数定律

设  $\{X_n\}$  是一个随机变量序列,  $\mathbb{E}(X_n)$  存在。令  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , 若

$$\bar{X}_n - \mathbb{E}(\bar{X}_n) \xrightarrow{P} 0$$

则称随机变量序列  $\{X_n\}$  服从弱大数定律 (weak law of large numbers)。

# 以概率 1 收敛

设  $\{X_n\}$ ,  $n \in \mathbb{N}$  是一个随机变量序列, 若存在一个随机变量  $X$ , 使得

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \bigcup_{i=n}^{\infty} |X_i - X| > \varepsilon \right) = 0$$

则称序列  $\{X_n\}$  以概率 1 收敛 (convergence with probability 1) 于  $X$ , 也称作 “几乎处处收敛” (almost surely convergence), 记作:

$$X_n \xrightarrow{\text{a.s.}} X, \quad \text{或} \quad X_n \rightarrow X, \quad \text{a.s.}$$

该定义的等价形式是:

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} X_n = X \right) = 1$$

## 以概率 1 收敛 (cont.)

以概率 1 收敛可以理解为：只要  $n$  足够大，任意指定一个正数  $\varepsilon$ ，总能找到一个  $N$ ，使得  $n > N$  时， $X_n$  与  $X$  差距的绝对值不再大于  $\varepsilon$ 。正因如此，以概率 1 收敛比依概率收敛的条件强，也就是说以概率 1 收敛意味着一定依概率收敛，反之则不然。

# 强大数定律

设  $\{X_n\}$  是一个随机变量序列,  $\mathbb{E}(X_n)$  存在。令  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , 若

$$\bar{X}_n - \mathbb{E}(X_n) \xrightarrow{a.s.} 0 \quad \text{或} \quad \mathbb{P} \left[ \lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X_n) \right] = 1$$

则称随机变量序列  $\{X_n\}$  服从强大数定律 (strong law of large numbers)。

所谓大数定律, 实际上是想说明当对一个随机变量进行无限次采样时, 得到的平均值会无限接近真实的期望值。只不过强大数定律是想说明, 采样的次数越多, 平均值几乎一定越接近真实期望值 (这意味着不可能出现反方向的偏离); 而弱大数定律则是想说明, 采样的次数越多, 平均值接近真实期望值的可能性越大 (这意味着也有极小的可能出现反方向的偏离)。

# 柯尔莫哥洛夫 (Kolmogorov) 大数定律

设  $\{X_n\}$  是相互独立的随机变量序列, 并且  $\sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} < \infty$ , 则

$\{X_n\}$  服从强大数定律, 即:

$$\mathbb{P} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{\infty} [X_k - \mathbb{E}(X_k)] = 0 \right\} = 1$$

# 柯尔莫哥洛夫大数定律 (cont.)

更进一步，在柯尔莫哥洛夫大数定律的基础上，假设  $\{X_n\}$  同分布，并且  $\mathbb{E}(X_n) = \mu$ ，令  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ，则：

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \bar{X}_n = \mu \right) = 1$$

换句话说，如果随机变量序列  $\{X_n\}$  独立同分布 (independent and identically distributed, iid)，则随着样本的增大，样本的均值将以概率 1 收敛于总体的均值。这是后面研究随机过程性质的重要基础。

# 均方收敛

设  $\{X_n\}$ ,  $n \in \mathbb{N}$  是一个随机变量序列,  $X$  是一个随机变量。若

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0$$

则称序列  $\{X_n\}$  均方收敛 (convergence in mean square) 于  $X$ , 记作:

$$X_n \xrightarrow{\text{m.s.}} X$$

注意:

均方收敛是从随机变量的二阶矩的角度考虑收敛问题, 其约束很严格, 且不能与以概率 1 收敛互相推导。

# 依分布收敛

设  $\{F_n(x)\}$  是随机变量序列  $\{X_n\}$ ,  $n \in \mathbb{N}$  的分布函数列, 如果存在一个单调不减函数  $F(x)$ , 使得在  $F(x)$  所有的连续点  $x$  上, 有:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

则称随机变量序列  $X_n$  依分布收敛 (convergence with distribution) 于  $X$ , 记作:

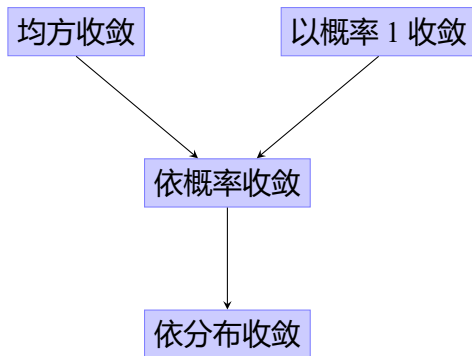
$$X_n \xrightarrow{L} X$$

称函数列  $\{F_n(x)\}$  弱收敛于  $F(x)$ 。

依分布收敛的约束非常小, 此种形式的收敛并未考虑随机变量的值, 只考察两者的分布函数是否相近。



# 几种收敛的关系



# 随机过程的概念

设  $(\Omega, \mathcal{F}, \mathbb{P})$  是一个概率空间,  $T$  为一个参数集。若对每一个  $t \in T$ , 均有定义在概率空间上的一个随机变量  $X_t(\omega)$ ,  $\omega \in \Omega$  与之对应, 则称  $\{X_t : t \in T\}$  为  $(\Omega, \mathcal{F}, \mathbb{P})$  上的一个随机过程 (stochastic process)。这里的  $t$  通常理解成时间, 相应的参数集  $T$  就是时间参数;  $X_t$  可看作过程在时刻  $t$  的状态。  $X_t$  的取值范围称作状态空间  $\Omega$ 。

# 随机过程的概念 (cont.)

对于  $X_t(\omega)$ ,

- 若固定  $\omega$ , 则  $X_t$  称为样本函数或轨道, 是随机过程的一次实现 (realization);
- 若固定  $t$ , 则  $X(\omega)$  称为一个随机变量;
- 所有可能出现的结果的总体  $\{X_1(\omega), X_2(\omega), \dots, X_n(\omega), \dots\}, \omega \in \Omega$  构成一个随机过程。

随机过程这个词包含了两重含义: “随机” 意味着某时刻出现结果的不确定性, 但是可能的结果一定在状态空间内; “过程” 意味着还要考虑随机现象随时间的演化规律。正因如此, 随机过程可看作所有样本函数的集合 (assemble)。

# 以股价为例说明随机过程的含义

