概率论笔记

Contents

1 基本概念

1.1 运算

若 A 代表事件 A 发生, $\overline{A}$ 代表事件没有发生，我们定义如下在随机事件上的关系运算：

包含

$$A \subset B \Leftrightarrow A \to B$$

另外有 $A = B \Leftrightarrow A \subset B \wedge B \subset A$

并集

$$A \cup B \Leftrightarrow A \vee B$$

交集

差集

$$A \cap B \Leftrightarrow A \wedge B$$

$$A - B \Leftrightarrow A \wedge \overline{B}$$

另外有 $A - B = A - AB = A\overline{B}$

对于交集和并集运算，符合以下四种运算律：

交换律

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

结合律

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

分配律

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

对偶律（德摩根定律）

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

1.2 关系

在事件间，存在如下两种关系：

互斥事件

$$A \cap B = \varnothing$$

对立事件

$$A \cap B = \varnothing \wedge A \cup B = \Omega.$$

1.3 频率和概率

定义 1.3.1：

频率

$$f_n(A) = \frac{n_A}{n}$$

概率

$$P(A) = \lim_{n \to \infty} f_n(A) \to P$$

性质：

非负性

$$0 \leq P(A) \leq 1$$

标准性

$$P(\Omega) = 1, P(\varnothing) = 0$$

无限可加性

If $A \cap B = \varnothing$ then

$$P(A \cup B) = P(A) + P(B)$$

互补性

$$P\left(\overline{A}\right) = 1 - P(A)$$

概率之间存在如下运算：

Difference

$P(A - B) = P(A) - P(AB)$. If $B \subset A, P(A - B) = P(A) - P(B)$.

Addition

$$P(A + B) = P(A) + P(B) - P(AB)$$

Multiplication

$$P(AB) = P(A)P(B \mid A)$$

1.4 条件概率

定义 1.4.1: 如果 $P(A) > 0$，则条件概率为 $P(B \mid A) = \frac{P(AB)}{P(A)}$。

依此，我们有两条推广式:
1. $P(B \cup C \mid A) = P(B \mid A) + P(C \mid A) - P(BC \mid A)$
2. $P(B - C \mid A) = P(B \mid A) - P(BC \mid A)$

1.5 全概率与贝叶斯公式

定义 1.5.1 *(完备事件组)*: $S \coloneqq \{A_1, A_2, A_3, ..., A_n\}$ 是一个属于 Ω 的事件组，并且满足 $\forall A_1, A_2 \subset S, A_1 \cap A_2 = \varnothing, \land A_1 \cup A_2 \cup A_3 \cup ... \cup A_n = \Omega$，则 S 为一个完备事件组。

由定义，我们可以设 B 是一个随机事件, $\{A_1, A_2, ..., A_i\}$ 是一个完备事件组，我们有:

公式 1.5.1 *(全概率公式)*:

$$\begin{aligned} P(B) &= \sum_{i=1}^{n} P(A_i)P(B \mid A_i) \\ &= \sum_{i=1}^{n} P(BA_i) \end{aligned}$$

公式 1.5.2 *(贝叶斯公式)*:

$$P(A_i \mid B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

1.6 事件的独立性

若 A，B 是相互独立事件，则有

$$P(A \mid B) = P(A) \Leftrightarrow P(AB) = P(A)P(B) \Leftrightarrow P(B \mid A) = P\left(B \mid \overline{A}\right) \Leftrightarrow P\left(B \mid \overline{A}\right) = P(B)$$

且 $A, \overline{A}, B, \overline{B}$ 也相互独立，此外有

$$\begin{aligned} &\text{若} A, B, C \text{ 互为独立事件} \longrightarrow A, B, C \text{ 两两独立} \\ &\text{若} A, B, C \text{ 互为独立事件} \longrightarrow \text{关于 } A, B \text{ 的加法，乘法，减法} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{以及逆运算 也分别独立与 } C \text{和} \overline{C} \end{aligned}$$

1.7 伯努利概型

定义 1.7.1 *(伯努利实验)*: The experiment have only two probability result $A, \overline{A}$

公式 1.7.1 *(Binomial probability formula)*:

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

2 随机变量及其分布

2.1 $R.V.$ 和 分布函数

$R.V.$ 是一个从随机试验 E 的样本空间 Ω 到 $\mathbb{R}$ 的一个映射。

2.2 分布函数

定义 2.2.1: 设 X 是一个 $R.V.$, r 是任意实数，则称事件 $\{X \le r\}$ 的概率为 $R.V.\ X$ 的分布函数，计作 $F(r)$。

分布函数有如下性质:

性质:
1. $P\{x_1 < X \le x_2\} = P\{X \le x_2\} - P\{X \le x_1\} = F(x_2) - F(x_1)$.
2. $F(x)$ 是一个不减函数
3. $F(-\infty) = 0, F(+\infty) = 1$.

2.3 离散型随机变量及其分布

例子:

0 – 1 分布

x	0	1
P	p	$1-p$

二项分布

$$P\{X = k\} \coloneqq \tbinom{n}{k} p^k (1-p)^{n-k}, k \in \{x \mid x \in \mathbb{N}^+ \cap [0, n]\}$$

计作 $X \sim B(n, p)$

泊松分布

$$P\{X = k\} \coloneqq \tfrac{\lambda^k e^{-\lambda}}{k!},\ (\lambda > 0) \text{ 计作 } X \sim P(\lambda)$$

几何分布

$$P\{X = k\} \coloneqq (1-p)^{k-1} P$$

超几何分布

$$P\{X = k\} \coloneqq \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\ k \in \{1, 2, 3, ... \min(M, N)\} \text{ 计作 } X \sim H(N, M, n)$$

定理 2.3.1 *(Passion's theorem)*: 当 $X \sim B(n, p)$ 且 n 充分大，p 充分小时

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = np$$

2.4 连续型 $R.V.$

定义 2.4.1 *(distribution function)*:

$$F(X) = \int_{-\infty}^{x} f(t) \,\mathrm{d}t$$

其中 $f(t)$ 是概率密度函数，$F(x)$ 是分布函数

公式 2.4.1 *(Range probability)*: $F\{a < X < b\} = \int_a^b f(x) \,\mathrm{d}x$

定义 2.4.2 *(Uniform distribution)*:

$$f(x)=\begin{cases}\frac{1}{b-a}, & b<x<a\\ 0, & \text{otherwise}\end{cases}$$

定义 2.4.3 *(Index)*:

$$f(x)=\begin{cases}\frac{1}{\lambda}e^{-\frac{x}{\lambda}}, & x>0\\ 0, & \text{otherwize}\end{cases}$$

指数分布具有无记忆性，即 $P\{x>t+T \mid x>t\}=P\{x>T\}$ 。

2.4.1 正态分布

定义 2.4.1.1 *(Normal)*: Marked $X\sim N(\mu,\sigma^2),\sigma>0$

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in\mathbb{R}$$

公式 2.4.1.1:

$$\int_{-\infty}^{+\infty}e^{-\frac{x^2}{A}}\,\mathrm{d}x=\sqrt{A\pi},\ A>0$$

Proof: 设 $X\sim N\left(0,\frac{A}{2}\right)$, 因为概率分布函数具有规范性 $F(+\infty)=1$ 即 $\int_{-\infty}^{+\infty}f(x)=1$. 带入得

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{A\pi}}e^{-\frac{x^2}{A}}\,\mathrm{d}x=1$$

$$\frac{1}{\sqrt{A\pi}}\int_{-\infty}^{+\infty}e^{-\frac{x^2}{A}}\,\mathrm{d}x=1$$

$$\int_{-\infty}^{+\infty}e^{-\frac{x^2}{A}}\,\mathrm{d}x=\sqrt{A\pi}$$

□

定义 2.4.1.2 *(标准正态分布)*: 当 $\mu=0,\sigma^2=1,X\sim N(0,1),x\in\mathbb{R}$时，其为标准正态分布。

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}-e^{\frac{x^2}{2}}$$

$$\Phi(x)=F(x)=\int_{-\infty}^{x}\varphi(t)\,\mathrm{d}t$$

定义 2.4.1.3 *(标准化)*: 若 $X\sim N(\mu,\sigma^2)$ 不满足标准正态分布，则 $\frac{X-\mu}{\sigma}\sim N(0,1)$

根据标准化，如果我们想要计算一个满足非标准化的正态分布的随机变量在范围 $(a,b]$上的概率，我们可以 X 先将其标准化为 $\frac{X-\mu}{\sigma}$ 并计算 $\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$ 即可。

定义 2.4.1.4 *(Quantile 分位点)*: μ_α 表示 $P\{x > \mu_\alpha\} = \alpha$.
并且有 $\mu_{1-\alpha} = \mu(-\alpha)$.

2.5 随机变量的分布

对于离散型随机变量，我们可以先求出取值，在分别对对应的取值求出概率。而对于连续性随机变量，重点是求其密度函数：即已知 $X\text{\textasciitilde}f_X(x), Y = g(X)$，求 $R.V.Y$ 的分布函数 $f_Y(y)$。

首先介绍根据分布函数求 $R.V.Y$ 的 密度函数的方法：

公式 2.5.1 *(根据分布函数法)*:

1. 首先，找到密度函数 $f_Y(y)$ 的分段点，一般有如下两种情况
   1. $f_X(x)$ 的分段点，带入 $g(x)$ 后得到的 y 的值，和
   2. $y = g(x)$ 的最值
2. 其次，根据以上分段点，求出区间 $(l, r]$ 的 $F_Y(y)$：

$$F_Y(y) = P\{Y \le y\} = P\{g(x) \le y\} = \int_l^r f_X(x)\,\mathrm{d}x$$

3. 最后，对求出的分布函数求导即可得到随机变量 y 的密度函数 $f_Y(y) = {F_Y}'(y)$。

除此之外，还可以用下方的定理中的公式来进行求解：

定理 2.5.1: 设 $f_X(x)$ 随机变量 X 的密度函数，对于随机变量 Y 有 $Y = g(X)$，且 $g(X)$ 为单调函数，令 $x = h(y)$ 是 $y = g(x)$ 的反函数，α, β 分别是 $g(x)$ 的最小值和最大值。则 $Y = g(X)$ 的密度函数 $f_Y(y)$ 为：

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & \alpha < y < \beta \\ 0, & \text{otherwize} \end{cases}$$

3 多维 $R.V.$

3.1 二维随机变量和联合分布函数

定义 3.1.1 *(double R.V.)*: 设随机试验 E 的样本空间 $\Omega = \{e\}$，$X = X(e), Y = Y(e)$ 的定义在 Ω 上的随机变量，则 (X, Y) 为定义在 Ω 上的二维随机变量。

定义 3.1.2 *(Union Distribution Function)*: 设 $x, y \in \mathbb{R}$，则 x, y 的联合分布函数为事件 $\{X \le x\}$ 与事件 $\{Y \le y\}$ 同时发生的概率为二维随机变量的联合分布函数

$$F(x, y) = P\{(X \le x) \cup (Y \le y)\} \overset{\text{计作}}{=} P\{X \le x, Y \le y\}$$

随机变量的分布函数的几何意义为在二维坐标轴中，$F(x, y)$ 为 $X \le x$ 与 $Y \le y$ 所围成的矩形区域的面积。易得，点 (X, Y) 落在 $\{(x, y) \mid x_1 < x \le x_2, y_1 < y \le y_2\}$ 区域的概率为 $F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$ 。

3.2 连续型随机变量

定义 3.2.1 *(Density function)*: 联合概率密度指的是对于二维随机变量 (X, Y)，其概率分布函数为

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \,\mathrm{d}u \,\mathrm{d}v$$

其中的非负函数 $f(x, y)$ 即为联合概率密度。

性质: 有如下性质:
1. 非负性: $f(x, y) \geq 0$
2. 规范性: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \,\mathrm{d}u \,\mathrm{d}v = F(\infty, \infty) = 1$
3. 设 G 是平面 xOy 上的闭区域，则点 (X, Y) 落在 G 区域上的概率为

$$F\{(X, Y) \in G\} = \iint\limits_{G} f(x, y) \,\mathrm{d}x \,\mathrm{d}y$$

一些分布的常见形式
1. 均匀分布

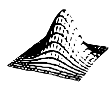
$$f(u, v) = \begin{cases} \frac{1}{S_G}, (u, v) \in G \\ 0, \text{其他} \end{cases}$$

1. 二维正态分布

(2) 二维正态分布

若 (X, Y) 的 $f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma^2} \right] \right\}$

其中 $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho \leq 1 \Rightarrow$ 则称 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2; \rho)$



3.3 边缘分布

定义 3.3.1 *(Distribution function)*: 设二维随机变量 (X, Y) 的联合分布函数为 $F(X, Y)$，Ω 为完备事件组，则 $F_X(x) = P\{X \leq x, \Omega\}, F_Y(y) = P\{\Omega, Y \leq y\}$ 分别为 二维随机变量关于 X 或 Y 的边缘分布函数。

定义 3.3.2 *(Distribution Law)*: 已知 二维随机变量 (X, Y) 的分布律为 $P\{X, Y\} = P\{X = x_i \cap Y = y_i\}$ 则关于 X 的边缘分布律为

$$P_X(x) = \sum^{i} P(x, y_i)$$

定义 3.3.3 *(Density function)*: 设 (X, Y) 的密度函数为 $f(x, y)$ 则关于 X 的边缘密度函数为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y$$

同理

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x$$

为关于 Y 的边缘密度函数。

3.4 条件分布

定义 3.4.1 *(law)*: 有二维随机变量 (X,Y)

$$P\{X = x_i \mid Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P_Y\{Y = y_j\}}$$

即为随机变量 X 在 $Y = y_j$ 下的条件分布律

定义 3.4.2 *(Density function)*: 有二维随机变量 (X,Y) 及其联合概率密度$f(x,y)$，固定 $Y = y$，则随机变量 X 在 $Y = y$ 条件下的概率密度函数为 $f_{X \mid Y} = \frac{f(x,y)}{f_Y(y)}$

3.5 独立性

定义 3.5.1: 有二维随机变量 (X,Y)，$F(x,y) = F_X(y)F_Y(x)$ 满足独立。对于离散型随机变量，独立性在于是否满足 $P(x,y) = P_X(x)P_Y(y)$。对于连续性随机变量，在于其密度函数是否满足 $f(x,y) = f_X(x)f_Y(y)$ 。

3.6 二维连续型随机变量的分布

对于离散型随机变量，依旧是先求取值，再求概率。而对于两个连续型随机变量，我们有如下方法:

公式 3.6.1 *(分布函数法)*: 有二维随机变量 (X,Y) 及其联合概率分布 $f(x,y)$，已知 $Z = g(X,Y)$，则 Z 的概率分布为

$$F_Z(z) = P\{Z \le z\} = P\{g(X,Y) \le z\} = P\{(X,Y) \mid g(X,Y) \le z\} = \iint\limits_G f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

或使用卷积公式:

公式 3.6.2 *(卷积公式)*: 若随机变量 X,Z,Y 存在 $Z = X + Y$ 关系，则

X,Y 不独立	X,Y 独立
$f_Z(z) = \int_{-\infty}^{+\infty} f(x,z-x)\,\mathrm{d}x$	$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\,\mathrm{d}x$

Proof: 对于 $Z = X + Y$

$$F_Z(z)=\int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z-x}f(x,y)\,\mathrm{d}y$$

令 $y+x=t$ 将二重积分中对 y 的积分换为对 t 的积分得

$$\int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z}f(x,t-x)\,\mathrm{d}t$$

改变积分次序后得

$$\int_{-\infty}^{z}\left[\int_{-\infty}^{+\infty}f(x,t-x)\,\mathrm{d}x\right]\mathrm{d}t$$

要求 Z 的密度函数，对变上限积分求导可得

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)\,\mathrm{d}x$$

若 (X,Y) 独立，又有

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\,\mathrm{d}x$$

□

4 随机变量的数字特征

4.1 数学期望

定义 4.1.1 *(Discrete type)*: 对于离散型随机变量 X，设 x_i 为其分布律的第 i 个取值，相应概率为 p_i，则其数学期望（均值）为:

$$E(X)=\sum^{i}x_ip_i$$

定义 4.1.2 *(continuous type)*: 若连续型随机变量 X 的密度函数为 $f_X(x)$ 则它的均值为

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)\,\mathrm{d}x$$

若要求 $Y=g(X)$ 的均值，则

$$E(g(X))=\int_{-\infty}^{+\infty}g(x)f_X(x)\,\mathrm{d}x$$

若 $Z=g(X,Y)$ 且二维随机变量 (X,Y) 有联合概率密度 $f(x,y)$ 则

$$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

性质: 有常数 C，随机变量 X 与 Y:

1. $E(C)=C$

2. $E(C+X)=C+E(X)$
3. $E(CX)=CE(X)$
4. $E(X+Y)=E(X)+E(Y)$
5. 若 X,Y 独立，则 $E(XY=E(X)E(Y)$

4.2 方差

定义 4.2.1: 方差为随机变量与其均值的距离的平方的均值即 $D(X)=E\left[(X-E(X))^2\right]$，其用来表示 X 偏离其均值 $E(X)$ 的程度大小。且方差 $D(X)\geq 0$。

公式 4.2.1: $D(X)=E(X^2)-E^2(X)$

性质: C 为常数
1. $D(C)=0$
2. $D(CX)=C^2D(X)$
3. $D(X+Y)=D(X)+D(Y)-2[E(XY)-E(X)E(Y)]$

4.3 常见形式

定理 4.3.1 *(0 - 1 分布)*: 若随机变量 X 服从 $0-1$ 分布，则

$$E(X)=p,D(X)=p(1-p)$$

定理 4.3.2 *(二项分布)*: 若随机变量 X 服从二项分布即 $X{\sim}B(n,p)$ 则

$$E(X)=np,D(X)=np(1-p)$$

Proof: 若随机变量 $X{\sim}B(n,p)$ 则其含义为 n 重伯努利实验中成功的次数即
$X=X_1+X_2+...+X_n$，其中 X_i 表示第 i 次伯努利实验，每次伯努利实验独立且都有相同的 p，即 $E(X_i)=p$，则

$$\begin{aligned} E(X)&=E(X_1+X_2+...+X_n)=np \\ D(X)&=D(X_1+X_2+...+X_n)=np(1-p) \end{aligned}$$

□

定理 4.3.3 *(Passion 分布)*: 若随机变量 X 服从参数为 λ 的泊松分布即 $X{\sim}P(\lambda)$ 则

$$E(X)=D(X)=\lambda$$

定理 4.3.4 *(Uniform 分布)*: 若 $X{\sim}U(a,b)$ 则

$$E(X)=\frac{a+b}{2},D(X)=\frac{(b-a)^2}{12}$$

定理 4.3.5 *(Index 分布)*: 若 $X{\sim}e(\lambda)$ 则

$$E(X)=\frac{1}{\lambda},D(X)=\frac{1}{\lambda^2}$$

定理 4.3.6 *(Normal 分布)*: 若 $X \sim N(\mu, \sigma^2)$ 则

$$E(X) = \mu, D(X) = \sigma^2$$

4.4 协方差

定义 4.4.1 *(协方差)*: 有二维随机变量 (X, Y)，称 $E[(X - E(X))(Y - E(Y))]$ 为随机变量 (X, Y) 的协方差，通常计作 $\mathrm{cov}(X, Y)$ 即

$$\begin{aligned}\mathrm{cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

特别的，相同变量的协方差为其方差 $\mathrm{cov}(X, X) = D(X)$。

已知方差的性质 3：$D(X + Y) = D(X) + D(Y) - 2[E(XY) - E(X)E(Y)]$，我们将协方差的计算公式带入可得到 $D(X + Y) = D(X) + D(Y) - 2\mathrm{cov}(X, Y)$。

协方差有以下性质

性质:

1. $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
2. $\mathrm{cov}(aX, bY) = ab\ \mathrm{cov}(X, Y)$
   a, b 为常数
3. $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$
4. 若 X, Y 相互独立，则
   $\mathrm{cov}(X, Y) = 0$

定义 4.4.2 *(相关系数)*: 有随机变量 X, Y，则其相关系数为:

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

相关系数有以下性质

性质:

1. $-1 \le \rho_{XY} \le 1$
2. 相关性
   1. 若相关系数 $\rho_{XY} = 0$ 则称 X, Y 不相关
   2. 若 $\rho_{XY} = 1$，则称 X, Y 为正相关，$y = ax + b, a > 0$
   3. 若 $\rho_{XY} = -1$，则称 X, Y 为负相关，$y = ax + b, a < 0$

5 大数定律和中心极限定理

定义 5.1 *(切比雪夫不等式)*: 有随机变量 X 及其均值 $E(X)$ 方差 $D(X)$，存在任意正数 ε 有

$$P\{\mid X - E(X) \mid \ge \varepsilon\} \le \frac{D(X)}{\varepsilon^2}$$

另有

$$P\{\mid X-E(X)\mid<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$

定义 5.2 *(* 切比雪夫大数定律 *)*: 设 $X_1,X_2,...,X_n$ 是相互独立的随机变量序列，且 $E(X_i),D(X_i)$ 均存在，且 $D(X_i)\leq C$，记 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$ 则对于任意正数 ε，有

$$\lim_{n\to+\infty}P\left\{\mid \overline{X}-E\left(\overline{X}\right)\mid<\varepsilon\right\}=1$$

$$\Leftrightarrow \overline{X} \text{ 依概率收敛到 } E\left(\overline{X}\right) \text{ 即: } \overline{X}\xrightarrow{P}E\left(\overline{X}\right)$$

定义 5.3 *(* 伯努利大数定律 *)*: 设 n_A 为 n 次独立重复试验中事件 A 发生的次数，且 $P(A)=p$，则对于任意正数 ε，有

$$\lim_{n\to+\infty}P\left\{\mid \frac{n_A}{n}-p\mid<\varepsilon\right\}=1$$

$$\Leftrightarrow \frac{n_A}{n}\xrightarrow{P}p$$

定义 5.4 *(* 辛钦大数定律 *)*: 有随机变量序列 $X_1,X_2,...,X_n$，随机变量间相互独立且服从同一分布，$E(X_i)$ 存在，则对于任意正数 ε 有

$$\lim_{n\to+\infty}P\left\{\mid \overline{X}-E\left(\overline{X}\right)\mid<\varepsilon\right\}=1$$

$$\Leftrightarrow \overline{X}\xrightarrow{P}E\left(\overline{X}\right)$$

6 附录 1：常见的分布类型的期望与方差

3. 泊松分布

若 $X\sim P(\lambda)$，即 $P\{X=k\}=\frac{\lambda^k\cdot e^{-\lambda}}{k!}$，$k=0,1,\cdots\cdots$，则 $E(X)=\lambda$，$D(X)=\lambda$

解释、① $E(X)=\sum\limits_{k=0}^{\infty}kP\{X=k\}=\sum\limits_{k=0}^{\infty}k\cdot\frac{\lambda^k e^{-\lambda}}{k!}=\sum\limits_{k=1}^{\infty}k\cdot\frac{\lambda^k e^{-\lambda}}{k!}$

$$=\sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\overset{令\, n=k-1}{=\!=}\lambda e^{-\lambda}\overbrace{\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}}^{=e^{\lambda}}=\lambda$$

② $E(X^2)=\sum\limits_{k=0}^{\infty}k^2P\{X=k\}=\sum\limits_{k=0}^{\infty}k^2\cdot\frac{\lambda^k e^{-\lambda}}{k!}=\sum\limits_{k=1}^{\infty}k^2\frac{\lambda^k e^{-\lambda}}{k!}$

$$=\sum_{k=1}^{\infty}[(k^2-k)+k]\frac{\lambda^k e^{-\lambda}}{k!}=\sum_{k=1}^{\infty}k(k-1)\frac{\lambda^k e^{-\lambda}}{k!}+\overbrace{\sum_{k=1}^{\infty}k\frac{\lambda^k e^{-\lambda}}{k!}}^{=E(X)=\lambda}$$

$$=\sum_{k=1}^{\infty}(k-1)\frac{\lambda^k e^{-\lambda}}{(k-1)!}+\lambda=\sum_{k=2}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-2)!}+\lambda=\lambda^2\cdot e^{-\lambda}e^{\lambda}+\lambda=\lambda^2+\lambda$$

4. 均匀分布

设 $X \sim U(a, b)$，则 $E(X) = \frac{a+b}{2}$ $=$ 中点，$D(X) = \frac{(b-a)^2}{12}$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & \text{其它} \end{cases}$$

解释：① $E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{a+b}{2}$

② $D(X) = E(X^2) - E^2(X) = \int_a^b x^2 \frac{1}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2$ $\frac{1}{3}x^3\Big|_a^b$

$= \frac{b^3-a^3}{3} \cdot \frac{1}{b-a} - \frac{(a+b)^2}{4}$

$= \frac{(b-a)^2}{12}$ ✓

5 指数分布

设 $X \sim e(\lambda)$，则 $E(X) = \frac{1}{\lambda}$，$D(X) = \frac{1}{\lambda^2}$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x>0 \\ 0, & x \le 0 \end{cases}$$

解释：① $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x}\,dx \xlongequal{\text{分部法}} \frac{1}{\lambda}$

② $E(X^2) = \int_0^{+\infty} x^2 \cdot \lambda e^{-\lambda x}\,dx = -\int_0^{+\infty} \underbrace{x^2}_{u}\, d\,\underbrace{e^{-\lambda x}}_{v}$

$= -\left[x^2 \cdot e^{-\lambda x}\Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda x}\,dx^2\right]$

$= 2\int_0^{+\infty} x \cdot e^{-\lambda x}\,dx$

$= 2\frac{1}{\lambda}\underbrace{\int_0^{+\infty} x \cdot \lambda e^{-\lambda x}\,dx}_{\to \frac{1}{\lambda}} = \frac{2}{\lambda^2}$

$\therefore D(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

6. 正态分布

$X\sim N(\mu,\sigma^2)$，则 $E(X)=\mu$，$D(X)=\sigma^2$ $\qquad f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty<x<+\infty$

解释：① $E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \xlongequal[x=\sigma t+\mu]{\text{令}\frac{x-\mu}{\sigma}=t}\int_{-\infty}^{+\infty}(\mu+\sigma t)\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{t^2}{2}}\sigma\cdot dt$

$$=\mu\underbrace{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt}_{1}+\underbrace{\int_{-\infty}^{+\infty}\sigma t\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt}_{0}=\mu+0=\mu$$

② $E(X^2)=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \xlongequal{\text{令}\frac{x-\mu}{\sigma}=t}\int_{-\infty}^{+\infty}(\mu+\sigma t)^2\cdot\frac{1}{\sqrt{2\pi}\cancel{\sigma}}e^{-\frac{t^2}{2}}\cdot\cancel{\sigma}dt$

$$=\int_{-\infty}^{+\infty}(\mu^2+\overset{0}{\underset{\text{奇}}{2\mu\sigma t}}+\sigma^2t^2)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

$$=\mu^2\underbrace{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt}_{1}+\sigma^2\underbrace{\int_{-\infty}^{+\infty}t^2\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt}_{\text{分部法}=1}=\mu^2+\sigma^2$$

故 $D(X)=E(X^2)-E^2(X)=\mu^2+\sigma^2-\mu^2=\sigma^2$