概率论

Forliage

December 2024

1 古典概型和概率空间

2 随机变量和概率分布

2.1 随机变量及其独立性

2.1.1 随机变量

设 Ω 是试验 S 的样本空间，对于事件 A，示性函数 $I[A]=I_A$ 是 Ω 上的函数，满足

$$\{\omega|I_A(\omega)\leq x\}=\begin{cases}\varnothing,\text{when } x<0\\ \overline{A},\text{when } x\in[0,1)\\ \Omega,\text{when } x\geq 1\end{cases}$$

于是无论 x 取何值，$\{I_A\leq x\}:=\{\omega|I_A(\omega)\leq x\}$ 都是事件。我们将 Ω 上有这种性质的函数称为随机变量。

定义 1.1 设 $(\Omega,\mathcal{F})$ 是可测空间，如果 Ω 上的函数 $X(\omega)$ 满足：对任何实数 x,

$$\{\omega|X(\omega)\leq x\}\in\mathcal{F}\quad(1.1)$$

就称 $X(\omega)$ 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，简称为随机变量。

用 $\mathbb{R}$ 表示全体实数 $(-\infty, +\infty)$，用 $\mathcal{C}$ 表示 $\mathbb{R}$ 的左开右闭的子区间的全体，用 $\mathcal{B}$ 表示 $\mathcal{C}$ 中的子区间们经过交集、余集和可列并的运算及其反复运算得到的集合的全体，则 $\mathcal{B}$ 是 σ 域。通常称 $\mathcal{B}$ 为 **Borel 域**，称 $\mathcal{B}$ 中的元素为 **Borel 集**。

定理 1.1 设 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，则对任何 Borel 集 A，有

$$\{X \in A\} := \{\omega | X(\omega) \in A\} \in \mathcal{F}$$

下面把一维 Borel 集推广的 n 维 Borel 集。用 $\mathbb{R}^n$ 表示 n 维向量空间：

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) | x_i \in \mathbb{R}, 1 \leq i \leq n\}$$

用 $\mathcal{C}^n$ 表示 $\mathbb{R}^n$ 的全体子立方体。用 $\mathcal{B}^n$ 表示 $\mathcal{C}^n$ 中的立方体们经过交集、余集和可列并的运算及其反复运算得到的集合的全体，则 $\mathcal{B}^n$ 是 σ 域，称为 **n 维 Borel 域**；称 $\mathcal{B}^n$ 中的元素是 n 维 Borel 集。这样，$(\mathbb{R}^n, \mathcal{B}^n)$ 是可测空间。

设 $\varphi(x_1, x_2, ..., x_n)$ 是 $\mathbb{R}^n$ 上的 n 元实函数，如果

$$\{(x_1, x_2, ..., x_n) | \varphi(x_1, x_2, ..., x_n) \leq x\} \in \mathcal{B}^n$$

对任何实数 x 成立，则它是 $(\mathbb{R}^n, \mathcal{B}^n)$ 上的随机变量。可测空间 $(\mathbb{R}^n, \mathcal{B}^n)$ 上的随机变量又被称为 Borel 可测函数，简称为可测函数。

定理 1.2 如果 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，$g(x)$ 是可测函数，则 $Y = g(X)$ 是 $(\Omega, \mathcal{F})$ 上的随机变量。

2.1.2 随机变量的独立性

定义 1.2 设 $X_1, X_2, ...$ 是随机变量

(1) 如果对任何实数 $x_1, x_2, ..., x_n$.

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1)...P(X_n \leq x_n)$$

则称随机变量 $X_1, X_2, ..., X_n$ 相互独立。

(2) 如果对任何 $n, X_1, X_2, ..., X_n$ 相互独立，则称随机变量序列 $\{X_j\}$ 相互独立。这时称 $\{X_j\}$ 是独立序列。

定理 1.3 设 $X_1, X_2, ..., X_n$ 相互独立，则对任何 Borel 集 $A_1, A_2, ..., A_n$，事件

$$\{X_1 \in A_1\}, \{X_2 \in A_2\}, ..., \{X_n \in A_n\}$$

相互独立

定理 1.4 设随机变量 $X_1, X_2, ..., X_n$ 相互独立，$g_1(x), g_2(x), ..., g_n(x)$ 是一元实函数，$\varphi(x_1, x_2, ..., x_k)$ 是 k 元实函数，则

(1) 随机变量 $g_1(X_1), g_2(X_2), ..., g_n(X_n)$ 相互独立;

(2) 随机变量 $\varphi(X_1, X_2, ..., X_k), X_{k+1}, X_{k+2}, ..., X_n$ 相互独立。

2.2 离散型随机变量

定义 2.1 如果随机变量 X 只取有限个值 $x_1, x_2, ..., x_n$，或可列个值 $x_1, x_2, ...$，就称 X 是离散型随机变量，简称为离散随机变量。

定义 2.2 设 X 是离散型随机变量，称

$$P(X = x_k) = p_k, k \leq 1 \quad (2.1)$$

为 X 的概率分布，称 $\{p_k\}$ 是概率分布列，简称为分布列。

下面是常用的几种随机变量及其概率分布。

A. 两点分布 $B(1, p)$

如果 X 只取指 0 或 1，概率分布是

$$P(X = 1) = p, P(X = 0) = q, p + q = 1 \quad (2.2)$$

则称 X 服从两点分布（伯努利分布 (Bernoulli) 分布），记作 $X \sim B(1,p)$

任何试验，当只考虑成功与否时，就可以用两点分布的随机变量描述。

B. 二项分布 $B(n,p)$

如果随机变量有如下的概率分布：

$$P(X=k)=\binom{n}{k}p^k q^{n-k}, k=0,1,...,n \quad (2.3)$$

则称 X 服从二项分布，其中 $pq>0, p+q=1$，记作 $X \sim B(n,p)$

二项分布的背景：设试验 S 成功的概率为 p，将试验 S 重复 n 次，用 X 表示成功的次数，则 $X \sim B(n,p)$

C. 泊松 (Poisson) 分布 $\mathcal{P}(\lambda)$

如果随机变量 X 有如下的概率分布：

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,... \quad (2.4)$$

则称 X 服从参数是 λ 的泊松分布，简记为 $X \sim \mathcal{P}(\lambda)$，这里 λ 是正常数。

D. 超几何分布 $H(n,M,N)$ 如果 X 的概率分布是

$$P(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}, m=0,1,...,M \quad (2.5)$$

就称 X 服从超几何分布，记作 $X \sim H(n,M,N)$

E. 几何分布

如果随机变量 X 有如下分布：

$$P(X=k)=q^{k-1}p, k=1,2,..., pq>0, p+q=1 \quad (2.6)$$

就称 X 服从参数是 p 的几何分布。

定理 2.1 取正整数值的随机变量 X 服从几何分布的充分必要条件是

X 有无记忆性：对每个 $k \geq 1$，

$$P(X = k+1 | X > k) = P(X = 1) \quad (2.7)$$

证明若 (2.6) 式成立，用条件概率公式得到：

$$\begin{aligned} P(X = k+1 | X > k) &= \frac{P(X = k+1, X > k)}{P(X > k)} \\ &= \frac{P(X = k+1)}{P(\bigcup_{j=k+1}^{\infty})\{X = j\}} \\ &= \frac{q^k p}{\sum_{j=k+1}^{\infty} q^{j-1} p} = p = P(X = 1) \end{aligned}$$

若 (2.7) 式成立，设 $r_k = P(X > k)$，利用 $\{X = k+1\} = \{X > k\} - \{X > k+1\}$ 得到

$$p = P(X = k+1 | X > k) = \frac{P(X > k) - P(X > k+1)}{P(X > k)} = 1 - \frac{r_{k+1}}{r_k}$$

于是，用 $r_0 = P(X > 0) = 1$，得到

$$r_{k+1} = (1-p) r_k = (1-p)^2 r_{k-1} = ... = q^{k+1}$$

最后得到

$$P(X = k) = r_{k-1} - r_k = q^{k-1} p$$

2.3 连续型随机变量

定义 3.1 设 X 是随机变量，如果存在非负函数 $f(x)$ 使得对任何 $a < b$，

$$P(a < X \leq b) = \int_a^b f(x)\, dx \quad (3.1)$$

就称 X 是连续型随机变量，称 $f(x)$ 是 X 的概率密度函数，简称为概率密度或密度。

注：可以证明在 $\mathbb{R}$ 上积分等于 1 的非负函数一定是某个随机变量的概率密度函数。

下面介绍几个常见的连续型随机变量。

A. 均匀分布 $\mathcal{U}(a,b)$

对 $a<b$，如果 X 的密度是

$$f(x)=\begin{cases}\frac{1}{b-a}, x\in(a,b)\\ 0, x\notin(a,b)\end{cases} \tag{3.2}$$

就称 X 服从区间 (a,b) 上的均匀分布，记作 $X\sim\mathcal{U}(a,b)$。这里 $\mathcal{U}$ 是 Uniform 的缩写。

对任何 Borel 集，如果 A 的测度 $m(A)=\int_A dx\in(0,\infty)$，可以类似定义 A 上的均匀分布：如果 X 的密度是

$$f(x)=\begin{cases}\frac{1}{m(A)}, x\in A\\ 0, x\notin A\end{cases} \tag{3.3}$$

就称 X 服从 Borel 集 A 上的均匀分布，记作 $X\sim\mathcal{U}(A)$。这时，对任何 Borel 集 B，

$$P(X\in B)=\frac{m(AB)}{m(A)} \tag{3.4}$$

B. 指数分布 $\mathcal{E}(\lambda)$

对正常数 λ，如果 X 的密度是

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, x\geq 0\\ 0, x<0\end{cases} \tag{3.5}$$

就称 X 服从参数为 λ 的指数分布，记作 $X\sim\mathcal{E}(\lambda)$。这里 $\mathcal{E}$ 是 Exponential

的缩写。

当随机变量 X 使得 $P(X<0)=0$，就称 X 是非负随机变量。

定理 3.1 设 X 是连续型非负随机变量，则 X 服从指数分布的充分必要条件是对任何 $s,t\geq 0$，有

$$P(X>s+t|X>s)=P(X>t) \quad (3.6)$$

性质 (3.6) 称为无后效性或无记忆性。无后效性是指数分布的特征。

证明： 设 $X\sim\mathcal{E}(\lambda)$，利用

$$P(X>x)=\int_x^\infty \lambda e^{-\lambda s}\,ds=e^{-\lambda x} \quad (3.7)$$

和条件概率公式得到：

$$\begin{aligned}P(X>s+t|X>s)&=\frac{P(X>s+t)}{P(X>s)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}\\&=e^{-\lambda t}=P(X>t)\end{aligned}$$

设 X 有无后效性 (3.6)，用 $f(x)$ 表示 X 的概率密度，则

$$G(x):=P(X>x)=\int_x^\infty f(t)\,dt \quad (3.8)$$

是连续函数。由无后效性

$$P(X>s+t|X>s)=\frac{P(X>s+t)}{P(X>s)}=P(X>t)$$

得到 $G(s+t)=G(s)G(t)>0$。解函数方程得到 $G(x)=e^{-\lambda x}$，最后，对任

何 $0 \le a < b$，从

$$\begin{aligned} P(a < x \le b) &= P(X > a) - P(X > b) \\ &= e^{-\lambda a} - e^{-\lambda b} = \int_a^b \lambda e^{-\lambda s}\, ds \end{aligned}$$

C. 正态分布 $N(\mu, \sigma^2)$

设 μ 是常数，σ 是正常数。如果 X 的密度是

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R} \quad (3.9)$$

则称 X 服从参数 (μ, σ^2) 的正态分布，记作 $X \sim N(\mu, \sigma^2)$。这里 N 是 Normal 的缩写。特别，当 $X \sim N(0,1)$ 时，称 X 服从标准正态分布。用特定的符号 φ 表示：

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), x \in \mathbb{R} \quad (3.10)$$

正态分布的密度函数呈钟的形状，所以有人将它称作钟型分布。正态密度函数 (3.9) 有如下的简单性质：

(1)$f(x)$ 关于 $x = \mu$ 对称；

(2)$f(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$ 是最大值；

(3)$f(x)$ 在 $x \pm \sigma$ 处有拐点；

对 $X \sim N(\mu, \sigma^2)$，在许多的应用问题中会遇到计算概率 $P(X \le a)$ 的问题。为了使用方便，引入

$$\Phi(a) = \int_{-\infty}^{a} \varphi(x)\, dx \quad (3.11)$$

利用 $\varphi(x)$ 的对称性得到

$$\Phi(x) + \Phi(-x) = 1, x \in \mathbb{R} \quad (3.12)$$

并且，只要 $X \sim N(\mu, \sigma^2)$，就有

$$\begin{aligned} P(X \le a) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \Phi(\frac{a-\mu}{\sigma}) \end{aligned} \tag{3.13}$$

利用公式 (3.12) 和 (3.13) 可以计算，当 $X \sim N(\mu, \sigma^2)$ 时，

$$P(|X-\mu| \le \sigma) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 68.27\%$$

$$P(|X-\mu| \le 2\sigma) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 95.45\%$$

$$P(|X-\mu| \le 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 99.73\%$$

$$P(|X-\mu| \le 6\sigma) = \Phi(6) - \Phi(-6) = 2\Phi(6) - 1 = 99.999999802682\%$$

D.$\Gamma(\alpha, \beta)$ 分布

设 α, β 是正常数，$\Gamma(\alpha)$ 由积分

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx \tag{3.14}$$

定义。如果 X 的密度是

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x \ge 0 \\ 0, x < 0 \end{cases} \tag{3.15}$$

就称 X 服从参数为 (α, β) 的 Gamma 分布，记作 $X \sim \Gamma(\alpha, \beta)$

2.4 概率分布函数

定义 4.1 对随机变量 X，称 x 的函数

$$F(x) = P(X \leq x)$$

为 X 的概率分布函数，简称为分布函数。

下面是分布函数 $F(x)$ 的常用性质:

(1) F 单调不减右连续;

(2) $F(\infty) = 1, F(-\infty) = 0$

证明: (1) 单调不减由

$$F(y) - F(x) = P(x < X \leq y) \geq 0$$

得到 $(x < y)$。由 F 的单调性和概率 P 的连续型得到

$$\begin{aligned}\lim_{n\to\infty} F(x + \frac{1}{n} &= \lim_{n\to\infty} P(X \leq x + \frac{1}{n}) \\ &= P(\bigcap_{n=1}^{\infty}\{X \leq x + \frac{1}{n}\}) \\ &= P(X \leq x) = F(x)\end{aligned}$$

(2) 由 F 的单调性和概率 P 的连续性得到

$$\begin{aligned}F(\infty) = \lim_{n\to\infty} F(n) &= \lim_{n\to\infty} P(X \leq n) \\ &= P(\bigcup_{n=1}^{\infty}\{X \leq n\}) = P(X < \infty) = 1\end{aligned}$$

同理可证 $F(-\infty) = 0$。

定理 4.1 设 X 的分布函数 F 连续，数集 A 中任何两点之间的距离大

于某个正数 δ，如果在 A 外导数 $F'(x)$ 存在且连续，则

$$f(x)=\begin{cases}F'(x), x\notin A\\ 0, x\in A\end{cases}\qquad(4.1)$$

是 X 的密度函数。

证明：对任何 $a<b$，不妨设 (a,b) 中只有集合 A 中的 $a_1,a_2,...,a_k$ 并且

$$a_0=a<a_1<a_2<...<a_k<b=a_{k+1}$$

这时

$$\begin{aligned}P(a<X\le b)&=\sum_{j=0}^{k}P(a_j<X\le a_{j+1})=\sum_{j=0}^{k}[F(a_{j+1})-F(a_j)]\\&=\sum_{j=0}^{k}\int_{a_j}^{a_{j+1}}f(t)\,dt=\int_a^b f(t)\,dt\end{aligned}$$

由定义知 $f(x)$ 是 X 的概率密度。

2.5 随机变量函数的分布

定理 5.1 设 X 有密度函数 $f(x), D\subset\mathbb{R}=(-\infty,\infty), Y=g(X), P(Y\in D)=1$. 如果

(1) 对 $y\in D, \{Y=y\}=\bigcup_{i=1}^{n}\{X=h_i(y)\}$;

(2) 每个 $h_i(y)$ 是 D 到其值域 D_i 的可逆映射，在 D 内有连续的导数;

(3)$D_1,D_2,..,D_n$ 互不相交，

则 Y 有密度函数

$$f_Y(y)=\begin{cases}\sum_{i=1}^{n}f(h_i(y))|h_i'(y)|, y\in D\\ 0, y\notin D\end{cases}\qquad(5.1)$$

*** 证明**设 $f_Y(y)$ 由 (5.1) 式定义。对 $a < b$ 和 $A = (a, b]$，当 $AD = \varnothing$ 时，有 $A \subset \overline{D}$，于是

$$P(a < Y \leq b) = 0 = \int_A f_Y(y)\, dy$$

下面设 $AD \neq \varnothing$。设 $h_i(y)$ 把 AD 映射到 $A_i \subset D_i$，则该映射是可逆的，且 $A_1, A_2, ..., A_n$ 互不相交。

根据已知条件得到：

$$\{Y \in AD\} = \bigcup_{i=1}^{n} \{X \in A_i\}$$

事件 $\{X \in A_i\}(i = 1, 2, ..., n)$ 互不相容。得到：

$$\begin{aligned} P(a < Y \leq b) = P(Y \in AD) &= \sum_{i=1}^{n} P(X \in A_i) = \sum_{i=1}^{n} \int_{A_i} f(x)\, dx \\ &= \sum_{i=1}^{n} \int_{AD} f(h_i(y))|h_i'(y)|\, dy = \int_{AD} f_Y(y)\, dy \\ &= \int_A f_Y(y)\, dy \end{aligned}$$

再由概率密度定义知 $f_Y(y)$ 是 $Y = g(X)$ 的密度。

3 随机向量及其分布

3.1 随机向量及其联合分布

定义 1.1 如果 $X_1, X_2, ..., X_n$ 都是概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，就称 $\mathbf{X} = (X_1, X_2, ..., X_n)$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机向量，简称为随机向量。

定义 1.2 设 $\mathbf{X}=(X_1,X_2,...,X_n)$ 是随机向量，称 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1,x_2,...,x_n)=P(X_1\le x_1,X_2\le x_2,...,X_n\le x_n)$$

为 $\mathbf{X}=(X_1,X_2,...,X_n)$ 的联合概率分布函数，简称为分布函数或联合分布。

容易证明，联合分布函数 $F(x_1,x_2,...,x_n)$ 是 $\mathbb{R}^n$ 上的右连续函数，关于每个自变元 x_j 单调不减。对于 $1\le k<n$，从 $(X_1,X_2,...,X_n)$ 的联合分布容易得到 $(X_1,X_2,...,X_n)$ 的联合分布如下：

$$\begin{aligned}F_k(x_1,x_2,...,x_k)&=P(X_1\le x_1,X_2\le x_2,...,X_k\le x_k)\\&=P(X_1\le x_1,X_2\le x_2,...,X_k\le x_k,X_{k+1}\le\infty,...,X_n\le\infty)\\&=F(x_1,x_2,...,x_k,\infty,...,\infty)\end{aligned}$$

$(X_1,X_2,...,X_k)$ 的分布函数 $F_k(x_1,x_2,...,x_k)$ 又称为边缘分布。不难看出，$F(x_1,x_2,...,x_n)$ 有 2^n-2 个边缘分布。边缘分布由联合分布 $F(x_1,x_2,...,x_n)$ 唯一决定，但是反之不成立。

3.2 离散型随机向量及其分布

3.2.1 离散型随机变量

定义 2.1 如果 $X_1,X_2,...,X_n$ 都是离散型的随机变量，则称 $\mathbf{X}=(X_1,X_2,...,X_n)$ 都是离散型随机变量。如果 $\mathbf{X}$ 所有的不同取指是

$$\mathbf{x}(j_1,j_2,...,j_n)=(x_1(j_1),x_2(j_2),...,x_n(j_n)),j_1,j_2,...,j_n\ge 1$$

则称

$$p_{j_1,j_2,...,j_n}=P(\mathbf{X}=\mathbf{x}(j_1,j_2,...,j_n)),j_1,j_2,...,j_n\ge 1$$

是 $\mathbf{X}$ 的概率分布。

设 $\{p_{j_1,j_2,...,j_n}\}$ 是 $\mathbf{X}$ 的概率分布，则有

(1) $p_{j_1,j_2,...,j_n}\ge 0$;

(2) $\sum\limits_{j_1,j_2,...,j_n} p_{j_1,j_2,...,j_n} = 1$;

(3) 对任何整数 $k \in (0,n), (X_1, X_2, ..., X_k)$ 有概率分布

$$\begin{aligned} &P(X_1 = x_1(j_1), X_2 = x_2(j_2), ..., X_k = x_k(j_k)) \\ &= \sum_{j_{k+1},j_{k+2},...,j_n} p_{j,j_2,...,j_n}, j_1, j_2, ..., j_k \geq 1 \end{aligned}$$

称上式为 $\mathbf{X}$ 的边缘分布。

3.2.2 离散型随机变量的独立性

定理 2.1 设离散型的随机向量 $\mathbf{X} = (X_1, X_2, ..., X_n)$ 有概率分布，则 $X_1, X_2, ..., X_n$ 相互独立的充分必要条件是:对任何 $(x_1(j_1), x_2(j_2), ..., x_n(j_n))$，有

$$\begin{aligned} &P(X_1 = x_1(j_1), X_2 = x_2(j_2), ..., X_n = x_n(j_n)) \\ &= P(X_1 = x_1(j_1))P(X_2 = x_2(j_2))...P(X_n = x_n(j_n)) \end{aligned}$$

推论 2.2 设离散型随机向量 (X, Y) 的所有不同取值是

$$(x_i, y_j), i, j \geq 1$$

则 X, Y 相互独立的充分必要条件是对任何 (x_i, y_j),

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

3.3 连续型随机向量及其联合密度

3.3.1 联合概率密度

定义 3.1 设 $\mathbf{X} = (X_1, X_2, ..., X_n)$ 是随机向量，如果有 $\mathbb{R}^n$ 上的非负可积函数 $f(x_1, x_2, ..., x_n)$，使得对 $\mathbb{R}^n$ 的任何子立方体:

$$D = \{(x_1, x_2, ..., x_n) | a_i < x_i \leq b_i, 1 \leq i \leq n\}$$

有

$$P(\mathbf{X} \in D) = \int_D f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n$$

就称 $\mathbf{X}$ 是连续型随机向量，并称 $f(x_1, x_2, ..., x_n)$ 是 $\mathbf{X}$ 的联合概率密度函数，简称为联合密度或概率密度。

定理 3.1 (Fubini 定理) 设 D 是 $\mathbb{R}^n$ 的子区域，$\varphi(x_1, x_2, ..., x_n)$ 是 D 上的非负函数或绝对可积函数（指取绝对值后积分有限），则对区域 D 上的 n 重积分

$$\int_D \varphi(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n$$

可以进行累次积分计算，且积分的次序可以交换。

3.3.2 边缘密度

对 $\mathbb{R}^k$ 的子立方体

$$D_k = \{(x_1, x_2, ..., x_k) | a_i < x_i \leq b_i, 1 \leq i \leq k\}$$

定义

$$D_k \times \mathbb{R}^{n-k} = \{(x_1, ..., x_n) | (x_, ..., x_k) \in D_k, (x_{k+1}, ..., x_n) \in \mathbb{R}^{n-k}\}$$

则有

$$\begin{aligned} &P((X_1, X_2, ..., X_k) \in D_k) = P(\mathbf{X} \in D^k \times \mathbb{R}^{n-k}) \\ &= \int_{D_k \times \mathbb{R}^{n-k}} f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n \\ &= \int_{D_k} \left(\int_{\mathbb{R}^{n-k}} f(x_1, x_2, ..., x_n)\, dx_{k+1}\, dx_{k+2} ... dx_n \right) dx_1\, dx_2 ... dx_k \end{aligned}$$

式中的被积函数

$$f_k(x_1, x_2, ..., x_k) = \int_{\mathbb{R}^{n-k}} f(x_1, x_2, ..., x_n)\, dx_{k+1}\, dx_{k+2} ... dx_n$$

是 $(X_1, X_2, ..., X_k)$ 的联合密度，称上式是随机向量 $(X_1, X_2, ..., X_n)$ 的边缘概率密度函数，简称为边缘密度。

3.3.3 联合分布与联合密度

定理 3.2 设 (X, Y) 有连续的分布函数 $F(x, y)$，定义

$$f(x, y) = \begin{cases} \frac{\partial^2 F(x,y)}{\partial x \partial y} & \text{当该混合偏导数存在} \\ 0 \quad \text{其它} \end{cases}$$

如果

$$\int_{\mathbb{R}^2} f(x, y)\, dx\, dy = 1$$

则 $f(x, y)$ 是 (X, Y) 的联合密度。

3.3.4 连续型随机变量的独立性

定理 3.3 设对每个 $i \quad (1 \leq i \leq n)$，随机变量 X_i 有概率密度 $f_i(x_i)$，则 $X_1, X_2, ..., X_n$ 相互独立的充分必要条件是随机向量 $\mathbf{X} = (X_1, X_2, ..., X_n)$ 有联合密度

$$f_1(x_1) f_2(x_2) ... f_n(x_n), \quad (x_1, x_2, ..., x_n) \in \mathbb{R}^n$$

3.4 随机向量函数的分布

3.4.1 随机向量的分布

设随机向量 $(X_1, X_2, ..., X_n)$ 有联合密度 $f(x_1, x_2, ..., x_n)$, $g_i(x_1, x_2, ..., x_n)$ 是 n 元函数，则

$$Y_i = g_i(X_1, X_2, ..., X_n), \quad 1 \le i \le m$$

是随机变量，于是可以计算 $(Y_1, Y_2, ..., Y_m)$ 的联合分布。

引入 Borel 集

$$C = \{(x_1, x_2, ..., x_n) | g_i(x_1, x_2, ..., x_n) \le y_i, 1 \le i \le m\}$$

得到

$$\begin{aligned} &P(Y_1 \le y_1, Y_2 \le y_2, ..., Y_m \le y_m) \\ &\quad = P(g_1(X_1, X_2, ..., X_n) \le y_1, ..., g_m(X_1, X_2, ..., X_n) \le y_m) \\ &\quad = P((X_1, X_2, ..., X_n) \in C) \\ &\quad = \int_C f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n \end{aligned}$$

即有

$$\begin{aligned} F_Y(y_1, y_2, .., y_m) &= \int_C f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n \\ &= \int_{\mathbb{R}^n} I_C f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n \end{aligned}$$

其中 I_C 是集合 C 的示性函数。

如果式中的 $F_Y(y_1, y_2, ..., y_m)$ 是 $\mathbb{R}^m$ 上的连续函数，定义

$$f_Y(y_1, y_2, ..., y_m) = \begin{cases} \frac{\partial^m F_Y(y_1, y_2, ..., y_m)}{\partial y_1 \partial y_2 ... \partial y_m} & \text{当该高阶偏导数存在} \\ 0 \quad \text{其它} \end{cases}$$

则当

$$\int_{\mathbb{R}^m} f_Y(y_1, y_2, ..., y_m)\, dy_1\, dy_2 ... dy_m = 1$$

时，$f_Y(y_1, y_2, ..., y_m)\, dy_1\, dy_2 ... d_m$ 是 $(Y_1, Y_2, ..., Y_m)$ 的联合密度。

3.4.2 随机向量函数的联合密度

定理 4.1 设 (X, Y) 有联合密度 $f(x, y), U = u(X, Y), V = v(X, Y)$ 是 (X, Y) 的函数，D 是平面上的区域使得 $P((X, Y) \in D) = 1$。如果存在 D 上的函数 $x_i = x_i(u, v), y_i = y_i(u, v)(i = 1, 2, ..., n)$ 使得

(1) 对 $(u, v) \in D$，有 $\{U = u, V = v\} = \bigcup_{i=1}^{n} \{X = x_i, Y = y_i\}$;

(2) Δ_i:

$$\begin{cases} x_i = x_i(u, v) \\ y_i = y_i(u, v) \end{cases}$$

是 D 到其值域 D_i 的可逆映射，有连续的偏导数，并且雅可比行列式的绝对值:

$$\left| \frac{\partial(x_i, y_i)}{\partial(u, v)} \right| \neq 0 \quad (u, v) \in D_i, i = 1, 2, ..., n$$

(3) 集合 $D_1, D_2, ..., D_n$ 互不相交，则 (U, V) 有联合密度

$$g(u, v) = \begin{cases} \sum_{i=1}^{n} f(x_i(u, v), y_i(u, v)) \left| \frac{\partial(x_i, y_i)}{\partial(u, v)} \right|, & (u, v) \in D \\ 0. \quad (u, v) \in \overline{D} \end{cases}$$

定理 4.3 设 $\mathbf{X} = (X_1, X_2, ..., X_n)$ 有联合密度 $f(\mathbf{x}), \mathbf{U} = (u_1(\mathbf{X}), u_2(\mathbf{X}), ..., u_n(\mathbf{X}))$ 是 $\mathbf{X}$ 的函数，D 是 $\mathbb{R}^n$ 的区域使得 $P(\mathbf{U} \in D) = 1$。如果有 D 上的 n 元向量值函数 $\mathbf{x}_i(\mathbf{u})$ 使得

(1) 对 $\mathbf{u} \in D$，有 $\{\mathbf{U}{=}\mathbf{u}\} = \bigcup_{i=1}^{m} \{\mathbf{X}{=}\mathbf{x}_i(\mathbf{u})\}$;

(2) 每个 $\mathbf{x}_i(\mathbf{u})$ 是 D 到其值域 D_i 的可逆映射，有连续的偏导数，并且

雅可比行列式的绝对值

$$\left|\frac{\partial \mathbf{x}_i}{\partial \mathbf{u}}\right| \neq 0, \mathbf{u} \in D_i, i = 1, 2, ..., m$$

(3) 集合 $D_1, D_2, ..., D_m$ 互不相交，则 $\mathbf{U}$ 有联合密度

$$g(\mathbf{u}) = \begin{cases} \sum_{i=1}^{n} f(\mathbf{x}_i(\mathbf{u})) \left|\frac{\partial \mathbf{x}_i}{\partial \mathbf{u}}\right|, \mathbf{u} \in D \\ 0, \quad \mathbf{u} \in \overline{D} \end{cases}$$

3.5 条件分布和条件密度

3.5.1 离散型的情况

设 (X, Y) 是离散型随机向量，有概率分布

$$p_{ij} = P(X = x_i, Y = y_j) > 0 \quad i, j = 1, 2, ...$$

则 X, Y 分别有边缘分布

$$p_i = P(X = x_i), q_j = P(Y = y_j), \quad i, j = 1, 2, ...$$

对每个固定的 j，由条件概率公式得到条件概率

$$P(X = x_1 | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{q_j}, \quad i = 1, 2, ...$$

定义 5.1 上式称为条件 $Y = y_j$ 下，X 的条件概率分布，简称为条件分布。

定理 5.1 X, Y 独立的充分必要条件是对任何 $i, j \geq 1$,

$$P(X = x_i | Y = y_j) = p_i$$

如果随机变量 $X_1, X_2, ..., X_n$ 相互独立并且有相同的分布函数，则称 $X_1, X_2, ..., X_n$ 独立同分布。

3.5.2 连续型的情况

定义 5.2 设随机向量 (X, Y) 有联合密度 $f(x, y), Y$ 有边缘密度 $f_Y(y)$。若在 y（确定的 y）处 $f_Y(y) > 0$，则称

$$P(X \leq x | Y = y) = \frac{\int_{-\infty}^{x} f(s, y)\, ds}{f_Y(y)}, x \in \mathbb{R}$$

为条件 $Y = y$ 下，X 的条件分布函数，简称为条件分布，记作 $F_{X|Y}(x|y)$。称

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, x \in \mathbb{R}$$

为条件 $Y = y$ 下，X 的条件概率密度，简称为条件密度。

对使得 $f_Y(y) > 0$ 的 y:

(1) $F_{X|Y}(x|y) = P(X \leq x | Y = y) = \int_{-\infty}^{x} f_{X|Y}(s|y)\, ds, x \in \mathbb{R}$

(2) 如果 $F_{X|Y}(x|y)$ 关于 x 连续，且除去至多可列个点外有连续的导数，则

$$f_{X|Y}(x|y) = \begin{cases} \frac{\partial F_{X|Y}(x|y)}{\partial x} & \text{当偏导数存在} \\ 0 \quad \text{其它} \end{cases}$$

是条件 $Y = y$ 下，X 的条件密度。

定理 5.2 X, Y 独立的充分必要条件是对 $y \in \{y | f_Y(y) > 0\}$,

$$f_{X|Y}(x|y) = f_X(x) \quad x \in \mathbb{R}$$

定义 5.3 设 $\mathbf{X} = (X_1, X_2, .., X_n)$ 和 $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)$ 是随机向量，$f(\mathbf{x,y})$ 是随机向量 $(\mathbf{X,Y})$ 的联合概率密度，这时 $\mathbf{Y}$ 有边缘密度

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{x,y})\, d\mathbf{x}$$

若在 $\mathbf{y}$（确定的 $\mathbf{y}$）处 $f_{\mathbf{Y}}(\mathbf{y})>0$，则称

$$P(\mathbf{X}\leq\mathbf{x}|\mathbf{Y}{=}\mathbf{y})=\frac{\int_{\mathbf{s}\leq\mathbf{x}}f(\mathbf{s},\mathbf{y})\,d\mathbf{s}}{f_{\mathbf{Y}}(\mathbf{y})},\mathbf{x}\in\mathbb{R}^n$$

为条件下 $\mathbf{Y}{=}\mathbf{y}$ 下，$\mathbf{X}$ 的条件分布函数，简称为条件分布，记作 $F_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$。称

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})=\frac{f(\mathbf{x},\mathbf{y})}{f_Y(\mathbf{y})},\mathbf{x}\in\mathbb{R}^n$$

为条件 $\mathbf{Y}{=}\mathbf{y}$ 下，$\mathbf{X}$ 的条件概率密度，简称为条件密度。

4 数学期望和方差

4.1 数学期望

定义 1.1 设 X 有概率分布

$$p_j=P(X=x_j),\quad j=0,1,...$$

只要级数 $\sum\limits_{j=0}^{\infty}|x_j|p_j$ 收敛，就称

$$E(X)=\sum_{j=0}^{\infty}x_jp_j$$

为 X 的数学期望或均值。

对于有概率密度 $f(x)$ 的连续型随机变量 X，我们也用 $f(x)$ 和横轴所夹面积的（横坐标的几何）质心定义 X 的数学期望。设 μ 是所述的质心，如果

$$\int_{-\infty}^{\infty}|x|f(x)\,dx<\infty$$

就有

$$\int_{-\infty}^{\infty}(x-\mu)f(x)\,dx=\int_{-\infty}^{\infty}xf(x)\,dx-\mu=0$$

于是 $\mu=\int_{-\infty}^{\infty}xf(x)\,dx$ 是所述的质心。

定义 1.2 设 X 是有概率密度 $f(x)$ 的随机变量，称

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

为 X 的数学期望或均值。

下面计算几个常见分布的数学期望:

(1) 两点分布 $B(1,p)$

由 $P(X=1)=p, P(X=0)=1-p$，则

$$E(X)=1\cdot p+0\cdot(1-p)=p$$

(2) 二项分布 $B(n,p)$

设 $q=1-p$，由

$$p_j=P(X=j)=\binom{n}{j}p^jq^{n-j},\quad 0\le j\le n$$

得到:

$$\begin{aligned}E(X)&=\sum_{j=0}^{n}j\binom{n}{j}p^jq^{n-j}=np\sum_{j=1}^{n}\binom{n-1}{j-1}p^{j-1}q^{n-j}\\&=np\sum_{k=0}^{n-1}\binom{n-1}{k}p^kq^{n-1-k}=np(p+q)^{n-1}\\&=np\end{aligned}$$

说明在 n 次独立重复试验中，p 越大，平均成功的次数越多。

(3) 泊松分布 $\mathcal{P}(\lambda)$

由

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,...$$

得到

$$E(X)=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda$$

说明参数 λ 是泊松分布 $\mathcal{P}(\lambda)$ 的数学期望。

(4) 几何分布

服从几何分布的随机变量 X 有离散分布

$$P(X=j)=pq^{j-1},\quad j=1,2,...$$

X 的数学期望

$$\begin{aligned}E(X)&=\sum_{j=1}^{\infty}jpq^{j-1}=p\left(\sum_{j=0}^{\infty}q^j\right)'\\&=p\left(\frac{1}{1-q}\right)'=\frac{1}{p}\end{aligned}$$

结论说明单次试验中的成功概率 p 越小，首次成功所需要的平均试验次数就越多。

(5) 均匀分布 $\mathcal{U}(a,b)$

由 X 的概率密度

$$f(x)=\frac{1}{b-a}I_{(a,b)}$$

得到

$$E(X)=\int_{-\infty}^{\infty}xf(x)\,dx=\int_a^b\frac{x}{b-a}\,dx=\frac{a+b}{2}$$

(6) 指数分布 $\mathcal{E}(\lambda)$

由 X 的概率密度

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

得到

$$E(X)=\int_{-\infty}^{\infty}xf(x)\,dx=\int_{0}^{\infty}x\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda}$$

(7) 正态分布 $N(\mu,\sigma^2)$

由 X 的概率密度

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

得到

$$\begin{aligned}E(X)&=\int_{-\infty}^{\infty}xf(x)\,dx=\int_{-\infty}^{\infty}\mu f(x)\,dx+\int_{-\infty}^{\infty}(x-\mu)f(x)\,dx\\&=\mu+\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx\\&=\mu+\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}te^{-\frac{t^2}{2\sigma^2}}\,dt\\&=\mu\end{aligned}$$

(8) $\Gamma(\alpha,\beta)$ **分布**

由 X 的概率密度

$$f(x)=\frac{\beta^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta x},\quad x>0$$

得到

$$\begin{aligned}E(X)&=\int_{-\infty}^{\infty}xf(x)\,dx=\int_{0}^{\infty}\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha}e^{-\beta x}\,dx\\&=\frac{1}{\Gamma(\alpha)\beta}\int_{0}^{\infty}t^{\alpha}e^{-t}\,dt=\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta}\\&=\frac{\alpha}{\beta}\end{aligned}$$

4.2 数学期望的性质

4.2.1 随机向量函数的数学期望

定理 2.1 设 $X=(X_1,X_2,...,X_n)$ 是随机向量，$x=(x_1,x_2,...,x_n)\in\mathbb{R}^n$.
(1) 如果 X 有联合密度 $f(x)=f(x_1,x_2,...,x_n)$，实函数 $g(x)$ 使得

$$\int_{\mathbb{R}^n}|g(x)|f(x)\,dx_1\,dx_2...\,dx_n<\infty$$

则 $Y=g(X)$ 有数学期望

$$E(Y)=\int_{\mathbb{R}^n}g(x)f(x)\,dx_1\,dx_2...\,dx_n$$

(2) 如果 X 是离散型随机向量，有概率分布

$$p_{j_1,j_2,...,j_n}=P(X=(x_1(j_1),x_2(j_2),...,x_n(j_n))),\quad j_1,j_2,...,j_n\geq 1$$

实函数 $h(x)$ 使得

$$\sum_{j_1,j_2,...,j_n}|h(x_1(j_1),x_2(j_2),...,x_n(j_n))|p_{j_1,j_2,...,j_n}<\infty$$

则 $Y=h(X)$ 有数学期望

$$E(Y)=\sum_{j_1,j_2,...,j_n}h(x_1(j_1),x_2(j_2),...,x_n(j_n))p_{j_1,j_2,...,j_n}$$

4.2.2 数学期望的性质

定理 2.2 设 $E(|X_j|)<\infty(1\leq j\leq n),c_0,c_1,...,c_n$ 是常数，则有以下结果:

(1) 线性组合 $Y = c_0 + c_1X_1 + c_2X_2 + ... + c_nX_n$ 的数学期望存在，并且

$$E(Y) = c_0 + c_1E(X_1) + c_2E(X_2) + ... + c_nE(X_n)$$

(2) 如果 $\{X_j\}$ 相互独立，则 $Z = X_1X_2...X_n$ 的数学期望存在，并且

$$E(X_1X_2...X_n) = E(X_1)E(X_2)...E(X_n)$$

(3) 如果 $X_1 \leq X_2$ a.s.，则 $E(X_1) \leq E(X_2)$

定义 2.1 设 X 是随机变量, m 是正整数。如果 $E|X|^m < \infty$, 就称 EX^m 为 X 的 m 阶原点矩，称 $E(X - EX)^m$ 为 X 的 m 阶中心矩。当 $m > 2$ 时，我们将原点矩和中心矩统称为高阶矩。

设 $\mu = EX$ 存在, 易证 $E|X|^m < \infty$ 的充分必要条件是 $E|X - \mu|^m < \infty$。

4.3 随机变量的方差

4.3.1 方差的定义

定义 3.1 如果随机变量 X 的数学期望 $\mu = EX$ 有限，则称

$$E(X - \mu)^2$$

为 X 的方差 (variance)，记作 $var(X)$ 或 σ_{XX}. 当 $var(X) < \infty$ 时，称 X 的方差有限。称 $\sigma_{XX} = \sqrt{var(X)}$ 为 X 的标准差。

当 X 有离散分布 $P(X = x_j)(j = 1, 2, ...)$ 时，利用公式得到

$$var(X) = \sum_{j=1}^{\infty}(x_j - \mu)^2 P(X = x_j)$$

当 X 有概率密度 $f(x)$ 时，利用公式得到

$$var(X) = \int_{-\infty}^{\infty}(x - \mu)^2 f(x)\, dx$$

容易化简得到

$$var(X) = E(X^2 - 2X\mu + \mu^2) = EX^2 - (EX)^2$$

下面计算几个常见分布的方差:

(1) 两点分布 $B(1,p)$

由 $P(X=1)=p, P(X=0)=1-p=q$ 以及 $X^2=X$ 和 $EX=p$ 得到

$$var(X) = EX^2 - (EX)^2 = p - p^2 = pq$$

(2) 二项分布 $B(n,p)$

设 $q=1-p$, 由 $EX=np$

$$P(X=j) = \binom{n}{j} p^j q^{n-j}, \quad 0 \le j \le n$$

得到

$$\begin{aligned} E(X^2) &= E[X(X-1)] + EX = \sum_{j=0}^{n} \binom{n}{j} j(j-1) p^j q^{n-j} + np \\ &= p^2 \frac{d^2}{dx^2} \sum_{j=0}^{n} \binom{n}{j} x^j q^{n-j}|_{x=p} + np \\ &= p^2 \frac{d^2}{dx^2} (x+q)^n|_{x=p} + np \\ &= n(n-1)p^2 + np \end{aligned}$$

最后得到

$$var(X) = n(n-1)p^2 + np - (np)^2 = npq$$

(3) 泊松分布 $\mathcal{P}(\lambda)$

由 $EX=\lambda$

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, ...$$

得到

$$E(X^2) = E[X(X-1)] + EX = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \lambda$$
$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \lambda$$
$$= \lambda^2 + \lambda$$

得到

$$var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

(4) 几何分布

服从几何分布的随机变量 X 有离散分布

$$P(X = j) = pq^{j-1}, \quad j = 1, 2, ...$$

用 $EX = 1/p$ 得到

$$E(X^2) = E[X(X-1)] + EX = \sum_{j=1}^{\infty} j(j-1)pq^{j-1} + \frac{1}{p}$$
$$= pq\left(\sum_{j=0}^{\infty} q^j\right)'' + \frac{1}{p} = pq\left(\frac{1}{1-q}\right)'' + \frac{1}{p}$$
$$= \frac{2pq}{(1-q)^3} + \frac{1}{p}$$
$$= \frac{2q}{p^2} + \frac{1}{p}$$

得到

$$var(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$$

(5) 均匀分布 $\mathcal{U}(a, b)$

由 X 的概率密度 $f(x)=\frac{1}{b-a}I_{(a,b)}$，数学期望 $EX=\frac{a+b}{2}$，

$$E(X^2)=\int_a^b \frac{x^2}{b-a}\,dx=\frac{b^3-a^3}{3(b-a)}$$

得到

$$var(X)=\frac{b^3-a^3}{3(b-a)}-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}$$

(6) 指数分布 $\mathcal{E}(\lambda)$

$$\begin{aligned}E(X^2)&=\int_0^\infty x^2\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda^2}\int_0^\infty t^2e^{-t}\,dt\\&=\frac{1}{\lambda^2}\Gamma(3)\\&=\frac{2}{\lambda^2}\end{aligned}$$

得到

$$var(X)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$

(7) 正态分布 $N(\mu,\sigma^2)$

$$\begin{aligned}var(X)&=\int_{-\infty}^{\infty}(x-\mu)^2f(x)\,dx\\&=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx\\&=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2\exp\left(-\frac{t^2}{2}\right)\,dt\\&=\sigma^2\end{aligned}$$

(8) $\Gamma(\alpha,\beta)$ **分布**

由

$$\begin{aligned} E(X^2) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\beta x}\, dx \\ &= \frac{1}{\Gamma(\alpha)\beta^2} \int_0^\infty t^{\alpha+1} e^{-t}\, dt \\ &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\beta^2} \\ &= \frac{\alpha(\alpha+1)}{\beta^2} \end{aligned}$$

得到

$$var(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}$$

4.3.2 方差的性质

定理 3.1 设 $EX = \mu, var(X) < \infty, \mu_j = EX_j, var(X_j) < \infty (1 \le j \le n)$，则

(1) $var(a + bX) = b^2 var(X)$;

(2) $var(X) = E(X-\mu)^2 < E(X-c)^2$, 只要$c \neq \mu$;

(3) $var(X) = 0$ 的充分必要条件是 $X = \mu$ a.s.;

(4) $var\left(\sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n [E(X_i X_j) - \mu_i \mu_j]$;

(5) 当 $X_1, X_2, ..., X_n$ 相互独立，$var\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n var(X_j)$

4.3.3 两个不等式

定理 3.2 (马尔可夫 (Markov) 不等式) 对随机变量 X 和 $\epsilon > 0$，有

$$P(|X| \ge \epsilon) \le \frac{1}{\epsilon^\alpha} E|X|^\alpha, \quad \alpha > 0$$

取 $\alpha = 2$，用 $X - EX$ 代替 X 就得到

$$P(|X - EX| \geq \epsilon) \leq \frac{1}{\epsilon^2} var(X), \quad \epsilon > 0$$

称 $\alpha = 2$ 时的马尔可夫不等式为切比雪夫 (Chebyshev) 不等式。

定理 3.3 (内积不等式) 设 $EX^2 < \infty, EY^2 < \infty$，则有

$$|E(XY)| \leq \sqrt{EX^2 EY^2}$$

式中等号成立的充分必要条件是有不全为零的常数 a, b，使得 $aX + bY = 0$ a.s..

4.4 协方差和相关系数

4.4.1 协方差和相关系数

定义 4.1 设 $\mu_X = EX, \mu_Y = EY$ 存在，

(1) 当 $E|(X - \mu_X)(Y - \mu_Y)| < \infty$ 时，称

$$E[(X - \mu_X)(Y - \mu_Y)]$$

为随机变量 X, Y 的协方差，记作 $\mathrm{cov}(X, Y)$ 或 σ_{XY}. 当 $\mathrm{cov}(X, Y) = 0$ 时，称 X, Y 不相关。

(2) 当 $0 < \sigma_{XY} < \infty$ 时，称

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

为 X, Y 的相关系数。有时也用 $\rho(X, Y)$ 表示相关系数 ρ_{XY}.

在以上定义中，引入 X, Y 的标准化：

$$\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}$$

则有

$$\rho_{XY} = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$

下面是计算协方差的常用公式:

$$\sigma_{XY} = E(XY) - (EX)(EY)$$

定理 4.1 设 ρ_{XY} 是 X, Y 的相关系数，则有

(1) $|\rho_{XY}| \leq 1$;

(2) $|\rho_{XY}| = 1$ 的充分必要条件是有常数 a, b 使得 $P(Y = a + bX) = 1$;

(3) 如果 X, Y 独立，则 X, Y 不相关。

当 $|\rho_{XY}| = 1$ 成立时，X, Y 有线性关系，也称 X, Y 线性相关。

4.4.2 协方差矩阵

设 $X = (X_1, X_2, ..., X_n)$ 是随机变量，如果对每个 $i, \mu_i = EX_i$ 存在，就称 X 的数学期望存在，并且定义

$$EX = (EX_1, EX_2, ..., EX_n)$$

引入 $\mu = EX$ 时，还有

$$\mu = (\mu_1, \mu_2, ..., \mu_n)$$

如果每个随机变量 X_{ij} 的数学期望 EX_{ij} 存在 $(1 \leq i \leq m, 1 \leq j \leq n)$，就称以 X_{ij} 为第 (i, j) 元素的矩阵

$$Y = \begin{pmatrix} X_{11} & X_{12} & ... & X_{1n} \\ X_{21} & X_{22} & ... & X_{2n} \\ ... & ... & & ... \\ X_{m1} & X_{m2} & ... & X_{mn} \end{pmatrix}$$

的数学期望存在，并且定义

$$EY = \begin{pmatrix} EX_{11} & EX_{12} & ... & EX_{1n} \\ EX_{21} & EX_{22} & ... & EX_{2n} \\ ... & ... & & ... \\ EX_{m1} & EX_{m2} & ... & EX_{mn} \end{pmatrix}$$

设 X, Y 如上定义，且数学期望都存在。容易证明，对任何常数向量 $a = (a_1, a_2, ..., a_n), k \times m$ 常数矩阵 A 和 $n \times j$ 常数矩阵 B，有以下的结果：

$$\begin{cases} E(aX^T) = aEX^T \\ (EY)^T = E(Y)^T \\ E(AY) = AEY \\ E(YB) = E(Y)B \\ E(AYB) = AE(Y)B \end{cases}$$

定义 4.2 如果随机向量 $X = (X_1, X_2, ..., X_n)$ 的数学期望 $\mu = EX$ 存在，每个 X_i 的方差 $\mathrm{var}(X_i) < \infty$，则称

$$\Sigma = E[(X - \mu)^T (X - \mu)] = (\sigma_{ij})_{n \times n}$$

为 X 的协方差矩阵，其中

$$\sigma_{ij} = \mathrm{cov}(X_i, X_j)$$

是 X_i, X_j 的协方差。

协方差矩阵 Σ 是对称矩阵。如果行列式 $\det(\Sigma) = 0$，就称 Σ 是退化的。

定理 4.2 设 Σ 是 $X = (X_1, X_2, ..., X_n)$ 的协方差矩阵，则

(1) Σ 是非负定矩阵;

(2) Σ 退化的充分必要条件是有不全为零的常数 $a_1, a_2, ..., a_n$ 使得

$$\sum_{i=1}^{n} a_i(X_i - \mu_i) = 0 \quad a.s.$$

其中 $\mu_i = EX_i$

5 大数定律及中心极限定理

5.1 大数定律

定义设 $\{X_i, i \geq 1\}$ 为一随机变量序列，若存在常数系列 $\{c_n, n \geq 1\}$，使得对任意的 $\epsilon > 0$，有

$$\lim_{n \to +\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - c)n\right| \geq \epsilon\right\} = 0$$

或

$$\lim_{n \to +\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - c_n\right| < \epsilon\right\} = 1$$

成立，即当 $n \to +\infty$ 时，有 $\frac{1}{n}\sum_{i=1}^{n} X_i - c_n \to 0$，则称随机变量序列 $\{X_i, i \geq 1\}$ 服从弱大数定律 (weak law of large numbers)，简称服从大数定律。

特别地，当 $c_n = c(n = 1, 2, ...)$ 时，可写为

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to c, \quad n \to +\infty$$

定理 (伯努利大数定律) 设 n_A 为 n 重伯努利试验中事件 A 发生的次数，$p(0 < p < 1)$ 为事件 A 在每次试验中发生的概率，即 $P(A) = p$，则对任意的 $\epsilon > 0$，有

$$\lim_{n \to +\infty} P\left\{\left|\frac{n_A}{n} - p\right| \geq \epsilon\right\} = 0$$

定理 (Khinchin 大数定律) 设 $\{X_i, i \geq 1\}$ 为独立同分布的随机变量序列，且数学期望存在，记为 μ，则对任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right\} = 0$$

即 $\frac{1}{n}\sum\limits_{i=1}^{n} X_i \to \mu(n \to +\infty)$，并认为此时随机变量序列 $\{X_i, i \geq 1\}$ 服从大数定律。

5.2 中心极限定理

Lindeberg-Levy 中心极限定理设 $\{X_i, i \geq 1\}$ 为独立同分布的随机变量序列，且数学期望 $E(X_i) = \mu$ 和方差 $\mathrm{var}(X_i) = \sigma^2$ 均存在 $(\sigma > 0)$，则对任意的 $x \in \mathbb{R}$，有

$$\begin{aligned}\lim_{n\to+\infty} P\left\{\frac{\sum\limits_{i=1}^{n} X_i - E\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{\mathrm{var}\left(\sum\limits_{i=1}^{n} X_i\right)}} \leq x\right\} &= \lim_{n\to+\infty} P\left\{\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \Phi(x)\end{aligned}$$

De Moivre-Laplace 中心极限定理设 n_A 为在 n 重伯努利试验中事件 A 发生的次数，p 为事件 A 在每次试验中发生的概率，即 $P(A) = p(0 < p < 1)$，则对任意的 $x \in \mathbb{R}$，有

$$\lim_{n\to+\infty} P\left\{\frac{n_A - np}{\sqrt{np(1-p)}} \leq x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \Phi(x)$$