概率论

TimeMachine

2022 年 9 月 8 日

目录

第一章　条件概率与独立性

1.1　两种概型与概率空间的简单介绍

1.1.1　古典概型

古典概型的特征：

1、样本空间中样本点有限，$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$

2、各基本事件等可能，即 $P(\omega) = \frac{1}{n}$

古典概型的计算：

$$P(A) = \frac{m}{n} = \frac{\text{A 包含的样本点数}}{\text{样本空间中样本点总数}}$$

1.1.2　几何概型

几何概型的特征：

1、样本空间中样本点无限

2、样本点落在等测度 (长度、面积、体积 $\cdots$) 区域的概率相等

几何概型的计算：($A_g = \{$任取样本点，位于区域$g \in \Omega$的概率$\}$)

$$P(A_g) = \frac{g\text{的测度}}{\Omega\text{的测度}}$$

1.1.3　概率空间

一个概率空间可以表示为 $(\Omega, \mathcal{F}, P)$，其中

Ω : 样本空间，即样本点 ω 的全体。有 $\Omega = \{\omega_1, \omega_2, \cdots\}$。

$\mathcal{F}$: 事件域，包括所有的事件。

P : 定义的概率。概率的公理化定义如下：

(1)(非负性)$\forall A \in \mathcal{F}, P(A) \geqslant 0$

(2)(规范性)$P(\Omega)=1$

(3)(可列可加性)$A_i\cap A_j=\emptyset, i\neq j$ (即 $A_1,\cdots,A_n,\cdots$ 为两两不相容的事件)

$$P\left(\sum_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n)$$

1.2 定义和链式法则

条件概率 $P(A|B)$：事件 B 发生条件下事件 A 发生的概率，称为事件 A 关于事件 B 的**条件概率** (conditional probability)

有基本公式:

$$P(A|B)=\frac{P(AB)}{P(B)}$$

也可以表示为**链式法则 (乘法公式)** 的形式：

$$P(AB)=P(A|B)P(B)$$

推广到 n 个事件，有链式法则：

$$P\left(\prod_{i=i}^{n}A_i\right)=\prod_{i=1}^{n}P\left(A_i\middle|\prod_{j=1}^{i-1}A_j\right)$$

特别定义 $a>b$ 时，$\prod_{i=a}^{b}A_i$ 为必然事件。或者这样写可能更容易看懂：

$$P\left(A_1A_2\cdots A_n\right)=P\left(A_1\right)\cdot P\left(A_2\mid A_1\right)\cdot P\left(A_3\mid A_1A_2\right)\cdots P\left(A_n\mid A_1A_2\cdots A_{n-1}\right)$$

1.3 全概率公式

在概率空间 $(\Omega,\mathcal{F},P)$ 中，若 $\{A_1,A_2,\cdots,A_n\}(n<\infty$ 或 $n=\infty)$ 构成 Ω 的一个 **分割 (完备事件组)**，满足：

(1)A_i 两两互不相容 (不可能同时发生)，且 $P(A_i)>0$

(2)$\sum\limits_{i=1}^{\infty}A_i=\Omega$

则有**全概率 (total probability) 公式**：$\forall B\in\mathcal{F}$，有

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

证明：

$$
\begin{aligned}
P(B) &= P(B\Omega) \quad (1.1)\\
&= P\left(B\sum_{i=1}^{n} A_i\right) \quad (1.2)\\
&= P\left(\sum_{i=1}^{n} BA_i\right) \quad (1.3)\\
&= \sum_{i=1}^{n} P(BA_i) \quad (1.4)\\
&= \sum_{i=1}^{n} P(A_i)P(B|A_i) \quad (1.5)
\end{aligned}
$$

1.4　贝叶斯公式

贝叶斯 (Bayes) 公式如下：

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum\limits_{k=1}^{\infty} P(A_k)P(B|A_k)}$$

证明：$P(A_i|B) = \dfrac{P(A_iB)}{P(B)}$，分子用链式法则展开，分母用全概率公式展开。

深入了解条件概率的意义：

$P(A_i)$: 不知 B 是否发生，称为**先验 (priori) 概率**

$P(A_i|B)$: 以 B 发生为已知条件，称为**后验 (posteriori) 概率**

1.5　事件独立性

1.5.1　两个事件的独立性

定义：事件 A 与事件 B **相互独立（统计独立，statistical independence）**，如果满足

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

因为此时有

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A)$$

且

$$P(B|A) = \frac{P(AB)}{P(A)} = P(B)$$

如果 A 与 B 不相互独立，也称为**统计相依** (statistical dependence)

1.5.2　多个事件的独立性

对于一组事件 $A_1, A_2, \cdots, A_n$，存在两两独立和整体的相互独立两种概念。不妨先以三个事件 A, B, C 为例进行研究。

两两独立：即 A 与 B 相互独立，B 与 C 相互独立，C 与 A 相互独立。

$$\begin{cases} P(AB) = P(A) \cdot P(B) \\ P(AC) = P(A) \cdot P(C) \\ P(BC) = P(B) \cdot P(C) \end{cases}$$

整体相互独立：即满足

$$P(ABC) = P(A) \cdot P(B) \cdot P(C)$$

推广到 n 个事件，$A_1, A_2, \cdots, A_n$ 相互独立需要满足: $\forall r < n$，$A_1, A_2, \cdots, A_n$ 中任意 r 个事件都相互独立，且

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

或者可以直接这么定义：$A_1, A_2, \cdots, A_n$ 相互独立，如果

$$\forall r \leqslant n (r \in \mathbb{N}_+), P\left(\prod_{i=1}^{r} A_{n_i}\right) = \prod_{i=1}^{r} P(A_{n_i}), 1 \leqslant n_1 < n_2 < \cdots < n_r \leqslant n$$

第二章　随机变量

2.1　随机变量的概念

定义概率空间 $(\Omega, \mathcal{F}, P)$ 上的单值实函数 $\xi(\omega)$ ，即

$$\xi : \Omega \to \mathbb{R}$$

还要求 $\xi(\omega)$ 的任意取值组合对应的样本点集合构成的事件在事件域 $\mathcal{F}$ 中，这样就可以称 $\xi(\omega)$ 为 **随机变量** (random variable)

实在太过抽象，暂且可以认为随机变量就是一个随机值。

2.2　离散型随机变量

离散型随机变量：随机变量 ξ 可取的值至多可列个。

分布列 (distribution sequence):

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

第一行是 ξ 可能取的值，第二行是 ξ 取这些值的概率。

分布列有性质：

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

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

对一些常见离散型随机变量举例如下：

(1) **退化 (degenerate) 分布**

$$\begin{bmatrix} c \\ 1 \end{bmatrix}$$

(2) **两点分布 (伯努利分布，**bernoulli distribution)

$$\begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}, p \in (0,1)$$

(3) **二项 (binomial) 分布**

$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, p \in (0,1), k = 0, 1, \cdots, n$$

记为 $\xi \sim B(n,p)$

(4) **泊松 (Poisson) 分布**

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0, k \in \mathbb{N}$$

记为 $\xi \sim \mathcal{P}(\lambda)$

(5) **几何 (geometry) 分布**

$$P(\xi = k) = p(1-p)^{k-1}, p \in (0,1), k \in \mathbb{N}_+$$

(6) **超几何 (hypergeometry) 分布**

$$P(\xi = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, n \leqslant N, M \leqslant N, k = 0, 1, \cdots, \min\{n, M\}$$

2.3 分布函数与连续型随机向量

2.3.1 分布函数

$$F(x) = P(\xi \leqslant x), -\infty < x < +\infty$$

为随机变量 $\xi(\omega)$ 的**分布函数** (distribution function)

分布函数有公理化定义：

(1) 单调递增 (不要求严格)：$a < b, F(a) \leqslant F(b)$

(2) $\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to +\infty} F(x) = 1$

(3) 处处左极限存在，右连续。即

$$\exists F(x-0)=\lim_{h\to 0^+}F(x-h)$$

$$F(x+0)=\lim_{h\to 0^+}F(x+h)=F(x)$$

注意，如果修改分布函数定义为

$$F(x)=P(\xi<x),-\infty<x<+\infty$$

那么 (3) 应该修改为处处右极限存在，左连续。

2.3.2　连续型随机变量

$\exists$ 一个非负的可积函数 $p(x)$s.t. 分布函数 $F(x)$ 满足

$$F(x)=\int_{-\infty}^{x}p(y)\mathrm{d}y,-\infty<x<+\infty$$

则以 $F(X)$ 为分布函数的 ξ 称为**连续型** (continuous) 随机变量。

$p(x)$ 称为 ξ 的概率密度函数，简称 **密度函数** (density function)

$F(x)$ 是一个变上限积分，可以证明，在 $p(x)$ 的连续点处，有

$$F'(x)=p(x)$$

ξ 落于 $(a,b]$ 的概率为

$$P(a<\xi\leqslant b)=F(b)-F(a) \tag{2.1}$$

$$=\int_{-\infty}^{b}p(y)\mathrm{d}y-\int_{-\infty}^{a}p(y)\mathrm{d}y \tag{2.2}$$

$$=\int_{a}^{b}p(y)\mathrm{d}y \tag{2.3}$$

然而需注意，联系几何概型有类似结论：

$$P(\xi=c)=F(c)-\lim_{h\to 0^+}F(c-h) \tag{2.4}$$

$$=\lim_{h\to 0^+}\int_{c-h}^{c}p(y)\mathrm{d}y=0 \tag{2.5}$$

类似离散型随机变量，连续性随机变量的密度函数有如下性质：

(1) 非负性

$$p(x)\geqslant 0$$

(2) 规范性

$$\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$$

注意，随机变量包括连续型随机变量和离散型随机变量，但随机变量并不总是连续性随机变量或离散型随机变量。如

$$F(x)=\begin{cases}0, & x<0\\ \dfrac{1+x}{2}, & 0\le x<1\\ 1, & x\ge 1\end{cases}$$

根据分布函数的公理化定义可以判断它确实是一个分布函数。但是对应的随机变量取值在 $[\frac{1}{2},1)$，取值并不可列，因此不是离散型随机变量。它也不是连续型随机变量，因为 $F(x)$ 不连续，比如可以看出应有 $P(\xi=0)=\frac{1}{2}$，与连续性随机变量 $P(\xi=c)=0$ 的性质矛盾。

常见连续性随机变量举例：

(1) **均匀 (uniform) 分布**

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b\\ 0, & \text{其他}.\end{cases}$$

记作 $\xi\sim U(a,b)$

(2) **一元正态 (normal) 分布**

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty,\quad \sigma>0$$

记作 $\xi\sim N(\mu,\sigma^2)$

(3) **指数 (exponential) 分布**

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0,\\ 0, & x<0\end{cases}\quad \lambda>0$$

记作 $\xi\sim \exp(\lambda)$

第三章　随机向量

3.1　随机向量

在同一概率空间 $(\Omega, \mathcal{F}, P)$ 上，有随机变量 $\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega)$ 称

$$\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega))$$

为 **n 维随机向量**。

3.2　离散型随机向量

考虑离散型随机向量 (ξ, η)，其**联合分布**为：

$$P(\xi = x_i, \eta = y_j) = p_{ij}$$

其**边际分布**为：

$$P(\xi = x_i) = p_{i\cdot} = \sum_j p_{ij}$$

$$P(\eta = y_j) = p_{\cdot j} = \sum_i p_{ij}$$

其分布列可以这么画：

$\xi \backslash \eta$	y_1	y_2	$\cdots$	y_n	$\cdots$	$p_{i\cdot}$
x_1	p_{11}	p_{11}	$\cdots$	p_{1n}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$\cdots$	$p_{m\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\vdots$	$p_{\cdot n}$	$\vdots$	

给出一道练习用的例题：

口袋中有 2 个白球 3 个黑球, 连取两次, 每次任取一球. 设 ξ 为第一次得白球数, η 为第二次得白球数. 对 (1) 有放回, (2) 无放回两种情况, 分别求 (ξ, η) 的联合分布.

解：(1)

$$P(\xi=0,\eta=0)=P(\xi=0)P(\eta=0)=\frac{3}{5}\cdot\frac{3}{5}$$

同理

$$P(\xi=0,\eta=1)=\frac{3}{5}\cdot\frac{2}{5}$$
$$P(\xi=1,\eta=0)=\frac{2}{5}\cdot\frac{3}{5}$$
$$P(\xi=1,\eta=1)=\frac{2}{5}\cdot\frac{2}{5}$$

得

$\xi\backslash\eta$	0	1	$p_{i\cdot}$
0	$\frac{3}{5}\cdot\frac{3}{5}$	$\frac{3}{5}\cdot\frac{2}{5}$	$\frac{3}{5}$
1	$\frac{2}{5}\cdot\frac{3}{5}$	$\frac{2}{5}\cdot\frac{2}{5}$	$\frac{2}{5}$
$p_{\cdot j}$	$\frac{3}{5}$	$\frac{2}{5}$	

(2)

$\xi\backslash\eta$	0	1	$p_{i\cdot}$
0	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{3}{5}$
1	$\frac{2}{5}\cdot\frac{3}{4}$	$\frac{2}{5}\cdot\frac{1}{4}$	$\frac{2}{5}$
$p_{\cdot j}$	$\frac{3}{5}$	$\frac{2}{5}$	

3.3 n 元分布函数

$\forall\,(x_1,\cdots,x_n)\in\mathbf{R}^n$, 称 n 元函数

$$F\left(x_1,\cdots,x_n\right)=P\left(\xi_1(\omega)\leq x_1,\xi_2(\omega)\leq x_2,\cdots,\xi_n(\omega)\leq x_n\right)$$

为随机向量 $\xi(\omega)=(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega))$ 的 **(联合) 分布函数**.

以二元联合分布函数为例，其有性质：

(1) 对每个变量单调递增 (不严格)

(2) 对每个变量右连续，左极限存在

(3)$\forall(x,y)$

$$F(x,-\infty)=0,\quad F(-\infty,y)=0,\quad F(\infty,\infty)=1$$

(4)$\forall a_1, a_2, b_1, b_2 \in R, a_1 < b_1, a_2 < b_2$

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geqslant 0$$

考虑边际分布函数 $F_\xi(x)$ 与 $F_\eta(y)$

$$F_\xi(x) = P(\xi < x) \tag{3.1}$$
$$= P(\xi < x, -\infty < y < +\infty) \tag{3.2}$$
$$= F(x, +\infty) \tag{3.3}$$

同理 $F_\eta(y) = F(+\infty, y)$

3.4　连续型随机向量

若存在 n 元可积的非负函数 $p\left(x_1, \cdots, x_n\right)$, 使 n 元分布函数可表示为

$$F\left(x_1, \cdots, x_n\right) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p\left(y_1, \cdots, y_n\right) dy_1 \cdots dy_n$$

则称它是连续型分布, 并称 $p\left(x_1, \cdots, x_n\right)$ 为 (联合) 密度函数. 显然, 密度函数满足如下条件:

(1) $p\left(x_1, \cdots, x_n\right) \geq 0$

(2)

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p\left(y_1, \cdots, y_n\right) dy_1 \cdots dy_n = 1$$

在这里不多加赘述，只是需要提及一下 n 维正态分布的形式。

设 $\boldsymbol{B} = (b_{ij})$ 为 n 维正定对称矩阵, $|\boldsymbol{B}|$ 为其行列式, $\boldsymbol{B}^{-1}$ 为其逆,

又设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$, $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^T$, 则称

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{B}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$$

为 n 维正态密度函数. 若随机向量 ξ 具有此密度函数, 则称 ξ 服从 n 维正态分布, 记作 $\xi \sim N(\boldsymbol{a}, \boldsymbol{B})$

n=1 时，$\boldsymbol{B} = \sigma^2, \boldsymbol{a} = \mu$，得

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

$n=2$ 时, 记

$$\boldsymbol{B}=\begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

其中 $\sigma_1,\sigma_2>0,|r|<1.\boldsymbol{x}=(x,y)',\boldsymbol{a}=(a,b)'$. 则

$$\boldsymbol{B}^{-1}=\frac{1}{|\boldsymbol{B}|}\begin{pmatrix} \sigma_2^2 & -r\sigma_1\sigma_2 \\ -r\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

故可得

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\left(1-r^2\right)}\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

简记作 $(\xi,\eta)\sim N\left(a,b,\sigma_1^2,\sigma_2^2,r\right)$.

关于正态分布的更多详细内容见附录 C。

第四章　条件分布与独立性

4.1　随机变量的独立性

设 $F(x,y), F_\xi(x)$ 和 $F_\eta(y)$ 分别为 (ξ,η) 的联合分布函数及其边际分布函数, 如果对一切 x,y 都有

$$F(x,y)=F_\xi(x)F_\eta(y)$$

成立, 则称 ξ 与 η 相互独立.

4.2　离散型随机变量的独立性

如果离散型随机向量 (ξ,η) 的联合分布列满足

$$P\left(\xi=x_i,\eta=y_j\right)=P\left(\xi=x_i\right)P\left(\eta=y_j\right),\quad i,j=1,2,\cdots,$$

即

$$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$$

则称 ξ 与 η **相互独立** (independent). 否则, 称 ξ 与 η **相依** (dependent).
可以这么推导：

$$F(x,y)=P(\xi\leq x,\eta\leq y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}P\left(\xi=x_i,\eta=y_j\right)$$

$$F_\xi(x)F_\eta(y)=\sum_{x_i\leqslant x}P\left(\xi=x_i\right)\sum_{y_j\leqslant y}P\left(\eta=y_j\right)$$

根据

$$F(x,y)=F_\xi(x)F_\eta(y)$$

令 $x_1<x_2<\cdots,y_1<y_2<\cdots$, 有

$$F(x_1,y_1)=F_\xi(x_1)F_\eta(y_1)$$

即

$$p_{11} = p_{1\cdot} \cdot p_{\cdot 1}$$

考虑数学归纳，如果 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ 对 $\forall i \leqslant m, j \leqslant n (i, j \in \mathbb{N}_+)$ 成立

那么根据

$$F(x_{m+1}, y_n) = F_\xi(x_{m+1}) F_\eta(y_n)$$

有

$$\sum_{i \leqslant m+1} \sum_{j \leqslant n} p_{ij} = \sum_{i \leqslant m+1} p_{i\cdot} \sum_{j \leqslant n} p_{\cdot j}$$

根据已有条件，除去相等项，可以得到

$$p_{m+1,n} = p_{m+1,\cdot} \cdot p_{\cdot,n}$$

同理，根据

$$F(x_m, y_{n+1}) = F_\xi(x_m) F_\eta(y_{n+1})$$

可以得到

$$p_{m,n+1} = p_{m,\cdot} \cdot p_{\cdot,n+1}$$

因此对于离散型随机变量，用分布函数定义的随机变量的独立性条件

$$F(x, y) = F_\xi(x) F_\eta(y)$$

可以通过数学归纳推出其特有独立性条件

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$$

同理，很方便地可以用后者反推出前者。因此在离散型随机变量中，这两种定义是等价的。

4.3　连续型随机变量的独立性

设 $p(x, y)$ 与 $p_\xi(x), p_\eta(y)$ 分别为连续型随机向量 (ξ, η) 的联合密度和边际密度, 则 ξ, η 相互独立的充要条件是

$$p(x, y) = p_\xi(x) p_\eta(y)$$

证明:$\forall x, y$,

$$
\begin{aligned}
&F(x,y) = F_\xi(x)F_\eta(y)\\
&\Longleftrightarrow \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)dudv = \int_{-\infty}^{x} p_\xi(u)du \int_{-\infty}^{y} p_\eta(v)dv\\
&\Longleftrightarrow \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)dudv = \int_{-\infty}^{x}\int_{-\infty}^{y} p_\xi(u)p_\eta(v)dudv\\
&\Longleftrightarrow p(x,y) = p_\xi(x)p_\eta(y)
\end{aligned}
$$

4.4　离散型随机变量的条件分布

设 (ξ,η) 的联合分布列为 $P\left(\xi=x_i,\eta=y_j\right)=p_{ij}, i,j=1,2,\cdots$. 若已知 $\xi=x_i(P(\xi=x_i>0)$, 则

$$P\left(\eta=y_j \mid \xi=x_i\right)=\frac{P\left(\xi=x_i,\eta=y_j\right)}{P\left(\xi=x_i\right)}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

即

$$p_{\eta|\xi}(y_j|x_i)=\frac{p_{ij}}{p_{i\cdot}}$$

如果令事件 $A=\{\eta=y_j\}$，$B=\{\xi=x_i\}$，由乘法公式可知这是很自然的。

可见，ξ 和 η 相互独立的等价条件是

$$P(\eta=y_j \mid \xi=x_i)=P(\eta=y_j)$$

定义 $\xi=x_i$ 的条件下 η 的**条件分布函数**:

$$P(\eta\leqslant y|\xi=x_i)=\sum_{j:y_j\leqslant y} p_{\eta|\xi}(y_j|x_i)$$

4.5　连续性随机变量的条件分布

首先获得条件分布函数，通过条件分布函数求导获得概率密度函数。

$$\begin{aligned}
P(Y \leqslant y|X = x) &= \lim_{\epsilon \to 0} P(Y \leqslant y|x - \epsilon < X \leqslant x + \epsilon) && (4.1)\\
&= \lim_{\epsilon \to 0} \frac{P(Y \leqslant y, x - \epsilon < X \leqslant x + \epsilon)}{P(x - \epsilon < X \leqslant x + \epsilon)} && (4.2)\\
&= \lim_{\epsilon \to 0} \frac{\dfrac{1}{2\epsilon}\displaystyle\int_{x-\epsilon}^{x+\epsilon}\int_{-\infty}^{y} p(u, v)\mathrm{d}u\mathrm{d}v}{\dfrac{1}{2\epsilon}\displaystyle\int_{x-\epsilon}^{x+\epsilon} p_X(u)\mathrm{d}u} && (4.3)\\
&= \frac{\int_{-\infty}^{y} p(x, v)\mathrm{d}v}{p_X(x)} && (4.4)
\end{aligned}$$

求导得到

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}$$

同理可以得到

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}$$

第五章　数学期望

5.1　离散型随机变量的数学期望

设离散型随机变量 $\xi \sim$

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

如果满足前提条件 $\sum_k x_k p_k$ 绝对收敛 ($\sum_k x_k p_k < \infty$)，则定义 **数学期望** (mathematical expectation) 或 **均值** (mean) 为

$$E\xi = \sum_k x_k p_k$$

(前提条件为了保证 $E\xi$ 的和不受求和次序的影响)

习题：计算泊松分布 ($\xi \sim \mathcal{P}(\lambda)$) 的数学期望

答案：$E\xi = \lambda$

5.2　连续型随机变量的数学期望

设连续型随机变量 ξ 有密度函数 $p(x)$，且满足前提条件

$$\int_{-\infty}^{+\infty} |x| p(x) \mathrm{d}x < \infty$$

（即 $\int_{-\infty}^{+\infty} x p(x) \mathrm{d}x$ 绝对收敛）则称

$$E\xi = \int_{-\infty}^{+\infty} x p(x) \mathrm{d}x$$

为 ξ 的数学期望。

无论是连续型还是离散型随机变量，如果前提条件 (绝对收敛) 不满足，都称数学期望不存在。

习题：计算指数分布 ($\xi \sim \exp(\lambda)$) 的数学期望

答案：$E\xi = \frac{1}{\lambda}$

习题 *：计算正态分布 ($\xi \sim N(a, \sigma^2)$) 的数学期望

答案：$E\xi = a$，过程可见附录 C

5.3　一般随机变量的数学期望

设连续型随机变量 ξ 有分布函数 $F(x)$，且满足前提条件

$$\int_{-\infty}^{+\infty} |x| \mathrm{d}Fx < \infty$$

则称

$$E\xi = \int_{-\infty}^{+\infty} x \mathrm{d}F(x)$$

为 ξ 的数学期望。前提条件若不满足，则数学期望不存在。

需要注意的是，这里的积分不是黎曼 (Riemann) 积分，而是新定义的一种积分，名为斯梯尔吉斯 (Stieltjes) 积分，在此不加赘述。所以在此处可以稍稍看一看它的形式，而不必直接计算，因为很容易把黎曼积分的思想套在这个积分上，而这是有可能出错的。（大佬可以忽视）

5.4　数学期望的性质

性质 1$a \leqslant \xi \leqslant b \Rightarrow \exists E\xi, a \leqslant E\xi \leqslant b$

特别地，$\xi = c \Rightarrow E\xi = Ec = c$

性质 1’$|\xi| < \eta, \exists E\eta \Rightarrow \exists E\xi, |E\xi| \leqslant E|\xi| \leqslant E\eta$

性质 2 $\exists E\xi_1, \cdots, E\xi_n \Rightarrow \forall$ 常数 $c_1, \cdots, c_n, b,$

$$\exists E\left(\sum_{i=1}^{n} c_i \xi_i + b\right)$$

且

$$E\left(\sum_{i=1}^{n} c_i \xi_i + b\right) = \sum_{i=1}^{n} c_i E\xi_i + b$$

特别地,

$$E\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}E\xi_i,\quad E(c\xi)=cE\xi$$

性质 3$\xi_1,\cdots,\xi_n$ 相互独立, $\exists E\xi_i, i=1,\cdots,n$, 则

$$E\left(\xi_1\cdots\xi_n\right)=E\xi_1\cdots E\xi_n$$

性质 4(有界收敛定理) 假设对任意 $\omega\in\Omega$ 有 $\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)$, 并且对一切 $n\geq 1$, $|\xi_n|\leq M$, 其中 M 为常数, 则

$$\lim_{n\to\infty}E\xi_n=E\xi$$

感兴趣的话可以自己证明一下。在这里象征性地证明一下最常用的性质 2:

$$E\left(\sum_{i=1}^{n}c_i\xi_i+b\right)=E(\xi+b)\cdots\cdots\xi=\sum_{i=1}^{n}c_i\xi_i \tag{5.1}$$

$$=\int_{-\infty}^{+\infty}(x+b)p_\xi(x)\mathrm{d}x \tag{5.2}$$

$$=\int_{-\infty}^{+\infty}xp_\xi(x)\mathrm{d}x+b\cdots\cdots\int_{-\infty}^{+\infty}p_\xi(x)\mathrm{d}x=1 \tag{5.3}$$

$$=\int\limits_{R^n}\left(\sum_{i=1}^{n}c_ix_i\right)p_{\xi_1,\xi_2,\cdots,\xi_n}(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n+b \tag{5.4}$$

$$=\sum_{i=1}^{n}\left(c_i\int\limits_{R^n}x_ip_{\xi_1,\xi_2,\cdots,\xi_n}(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n\right)+b \tag{5.5}$$

$$=\sum_{i=1}^{n}\left(c_i\int_{-\infty}^{+\infty}x_ip_{\xi_i}(x_i)\mathrm{d}x_i\right)+b \tag{5.6}$$

$$=\sum_{i=1}^{n}c_iE\xi_i+b \tag{5.7}$$

以上证明是针对连续型随机变量。下面对离散型随机变量进行证明:

$$E\left(\sum_{i=1}^{n} c_i\xi_i + b\right) = E(\xi + b)\cdots\cdots\xi = \sum_{i=1}^{n} c_i\xi_i \tag{5.8}$$
$$= \sum_k (x_k + b)p_k \tag{5.9}$$
$$= \sum_k x_k p_k + b\cdots\cdots\sum_k p_k = 1 \tag{5.10}$$
$$= \sum_k \left(\sum_{i=1}^{n} c_i x_{ik}\right) p_k + b \tag{5.11}$$
$$= \sum_{i=1}^{n} c_i \sum_k x_{ik}p_k + b \tag{5.12}$$
$$= \sum_{i=1}^{n} c_i E\xi_i + b \tag{5.13}$$

注意到，性质 2 其实就是线性性质。那么我们浮想联翩，线性代数的 DNA 动了。建立一个线性空间 V，V 包括所有存在数学期望的一元随机变量。那么 $E: V \to R$ 就是一个线性变换。

进一步地，我们定义一个内积：$\forall\alpha,\beta\in V$, $(\alpha,\beta)=E\alpha\beta$

首先根据内积的公理化定义验证它是内积。$\forall\alpha,\beta,\gamma\in V,\forall\lambda\in\mathbb{R}$

(1) 正定性:$E\alpha^2\geqslant 0$, 当且仅当 α 服从 $P(\alpha=0)=1$ 的退化分布时（定义这种随机变量为这个线性空间的零元），有 $E\alpha^2=0$

(2) 对称性: 显然有 $E\alpha\beta=E\beta\alpha$

(3) 加性:$E(\alpha+\beta)\gamma=E\alpha\gamma+E\beta\gamma$

(4) 齐性:$E(\lambda\alpha)\beta=\lambda E\alpha\beta$

加性和齐性由性质 2 证得。因此 $E\alpha\beta$ 可以成为线性空间 V 上的一个内积，那么就有 Cauchy-Schwarz 不等式:

$$(E\alpha\beta)^2\leqslant E\alpha^2\cdot E\beta^2$$

特别地，$\forall X,Y\in V$，有 $\exists EX,EY$，则 $\exists E(X-EX),E(Y-EY)$，则 $X-EX,Y-EY\in V$，有

$$E(X-EX)(Y-EY)\leq\left(E(X-EX)^2E(Y-EY)^2\right)^{1/2}$$

这将是下一节 Pearson 相关系数一个重要性质的依据。

5.5 条件期望

进入二元关联的考虑，给定不同的 $\eta=y$，$\xi=x$ 的后验概率有所不同，因而会影响其期望。这种情况下的期望就成为条件期望。

当然，需要注意这里是用离散型随机变量举例进行的一个理解，并不严格。如连续型随机变量的期望还需定义。

一般地，若 $\eta=y$ 时，ξ 有条件分布函数 $F_{\xi|\eta}(x|y)$，那么定义随机变量 ξ 的此时的条件期望为

$$E(\xi|\eta=y)=\int_{-\infty}^{+\infty}x\mathrm{d}F_{\xi|\eta}(x|y)$$

对于离散型随机变量，可以导出其条件期望：

$$E(\xi|\eta=y)=\sum_i x_i p_{\xi|\eta}(x|y)=\sum_i x_i P(\xi=x|\eta=y)$$

对于连续型随机变量，可以导出其条件期望：

$$E(\xi|\eta=y)=\int_{-\infty}^{+\infty}xp_{\xi|\eta}(x|y)\mathrm{d}x$$

5.6 全期望公式

全期望公式是一个很有趣的公式，它可以写成

$$E(\xi)=E[E(\xi|\eta)]$$

在连续型随机变量的情况下，我们可以做一个简单的推导

$$\begin{aligned}
E[E(\xi|\eta)] &= \int_{-\infty}^{+\infty}E(\xi|\eta)p_Y(y)\mathrm{d}y && (5.14)\\
&= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}xp_{\xi|\eta}(x|y)\mathrm{d}x\right)p_Y(y)\mathrm{d}y && (5.15)\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\frac{p(x,y)}{p_Y(y)}\cdot p_Y(y)\mathrm{d}x\mathrm{d}y && (5.16)\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)\mathrm{d}x\mathrm{d}y && (5.17)\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)\mathrm{d}y\mathrm{d}x && (5.18)\\
&= \int_{-\infty}^{+\infty}xp_X(x)\mathrm{d}x=E(\xi) && (5.19)
\end{aligned}$$

在离散型随机变量的情况下，

$$E[E(\xi|\eta)] = \sum_j E(\xi|\eta)P(\eta = y_j) \tag{5.20}$$

$$= \sum_j (\sum_i x_i P(\xi = x_i|\eta = y_j))P(\eta = y_j) \tag{5.21}$$

$$= \sum_j \sum_i x_i P(\xi = x_i|\eta = y_j)P(\eta = y_j) \tag{5.22}$$

$$= \sum_i x_i \left(\sum_j P(\xi = x_i|\eta = y_j)P(\eta = y_j) \right) \tag{5.23}$$

$$= \sum_i x_i P(\xi = x_i) = E(\xi) \tag{5.24}$$

条件期望也有一系列性质，在此不再列举，只举其比较有趣的性质：Cauchy-Schwarz 不等式：

$$E(XY|Z) \leqslant \sqrt{E(X^2|Z)}\sqrt{E(Y^2|Z)}$$

条件期望是具有线性的（显然，可以自己写写），那么内积的齐性和加性就满足了。对称性和正定性当然也满足。那么定义二元运算

$$(X, Y) = E(XY|Z)$$

就成为一种 V 上的内积。(V 就是 5.5.4 末尾定义的线性空间) 所以根据内积的 Cauchy-Schwarz 不等式就可以得证。

第六章　方差、协方差与相关系数

6.1　方差及其性质

6.1.1　方差

一切为了描述数据的离散程度！

离差 (deviation)：$\xi - E\xi$

离差取期望时，只要 ξ 期望存在，那么将会正负相消。因此考虑

方差 (variance)：$E(\xi - E\xi)^2$ 来描述期望的离散程度，即

$$Var\xi(D\xi) = E(\xi - E\xi)^2, \text{当它存在且为有限值时}$$

为统一量纲，有时使用**标准差 (standard deviation)**:$\sqrt{Var\xi}$

计算方差可以直接使用定义，也可以使用重要的**方差公式**：

$$Var\xi = E\xi^2 - (E\xi)^2$$

证明：

$$E(\xi - E\xi)^2 = E(\xi^2 - 2\xi E\xi + (E\xi)^2) \tag{6.1}$$

$$= E\xi^2 - 2E\xi \cdot E\xi + (E\xi)^2 \tag{6.2}$$

$$= E\xi^2 - (E\xi)^2 \tag{6.3}$$

6.1.2　切比雪夫 (Chebyshev) 不等式

若随机变量 ξ 的方差存在, 则 $\forall \varepsilon > 0$ 有

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{Var\xi}{\varepsilon^2}$$

证明设 ξ 的分布函数为 $F(x)$, 则

$$
\begin{aligned}
P(|\xi - E\xi| \geq \varepsilon) &= \int_{|x-E\xi|\geq\varepsilon} dF(x) \\
&\leq \int_{|x-E\xi|\geq\varepsilon} \frac{(x-E\xi)^2}{\varepsilon^2} dF(x) \\
&\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty} (x-E\xi)^2 dF(x) \\
&= \frac{Var\xi}{\varepsilon^2}
\end{aligned}
$$

根据切比雪夫不等式，可以利用方差粗糙估计随机变量落在偏离均值一定范围内的概率。

6.1.3　方差的性质

性质 1 $Var\xi = 0$ 的充要条件是 $P(\xi = c) = 1$, 其中 c 是常数.

证明：显然条件充分。反之，如果 $Var\xi = 0$, 记 $E\xi = c$, 由切贝雪夫不等式

$$P(|\xi - E\xi| \geq \varepsilon) = 0$$

对 $\forall\varepsilon > 0$ 成立。从而

$$
\begin{aligned}
P(\xi = c) &= 1 - P(|\xi - c| > 0) && (6.4) \\
&= 1 - \lim_{n\to\infty} P\left(|\xi - c| > \frac{1}{n}\right) = 1 && (6.5)
\end{aligned}
$$

性质 2 设 c, b 都是常数, 则

$$Var(c\xi + b) = c^2 Var\xi$$

证明：

$$
\begin{aligned}
Var(c\xi + b) &= E(c\xi + b - E(c\xi + b))^2 && (6.6) \\
&= E(c\xi + b - cE\xi - b)^2 && (6.7) \\
&= c^2 E(\xi - E\xi)^2 = c^2 Var\xi && (6.8)
\end{aligned}
$$

性质 3 若 $c \neq E\xi$, 则 $Var\xi < E(\xi - c)^2$.

证明：注意到

$$Var\xi = E\xi^2 - (E\xi)^2$$

和

$$E(\xi-c)^2=E\xi^2-2cE\xi+c^2$$

两边相减得

$$Var\xi-E(\xi-c)^2=-(E\xi-c)^2<0$$

这说明随机变量 ξ 对数学期望 $E\xi$ 的离散度最小.

性质 4

$$Var\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}Var\xi_i+2\sum_{1\le i<j\le n}E\left(\xi_i-E\xi_i\right)\left(\xi_j-E\xi_j\right)$$

特别地, 若 $\xi_1,\xi_2,\cdots,\xi_n$ 两两独立, 则

$$Var\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}Var\xi_i$$

证明：

$$Var\left(\sum_{i=1}^{n}\xi_i\right)=E\left(\sum_{i=1}^{n}\xi_i-E\sum_{i=1}^{n}\xi_i\right)^2=E\left(\sum_{i=1}^{n}(\xi_i-E\xi)\right)^2 \tag{6.9}$$

$$=E\left[\sum_{i=1}^{n}\left(\xi_i-E\xi_i\right)^2+2\sum_{1\le i<j\le n}\left(\xi_i-E\xi_i\right)\left(\xi_j-E\xi_j\right)\right] \tag{6.10}$$

$$=\sum_{i=1}^{n}Var\xi_i+2\sum_{1\le i<j\le n}E\left(\xi_i-E\xi_i\right)\left(\xi_j-E\xi_j\right) \tag{6.11}$$

当 $\xi_1,\xi_2,\cdots,\xi_n$ 两两独立时, 易证 $\xi_1-E\xi_1,\cdots,\xi_n-E\xi_n$ 也两两独立, 故

$$E\left(\xi_i-E\xi_i\right)\left(\xi_j-E\xi_j\right)=E\left(\xi_i-E\xi_i\right)\cdot E\left(\xi_j-E\xi_j\right)=0$$

交叉项为 0，则成立

$$Var\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}Var\xi_i$$

习题：求二项分布 $(\xi\sim B(n,p))$ 的方差

答案：$Var\xi=npq$（提示：$\xi=\sum_{i=1}^{n}\xi_i,\xi_i$ 服从两点分布且相互独立）

习题 *：求一元正态分布 $(\xi\sim N(a,\sigma^2))$ 的方差

答案：$Var\xi=\sigma^2$，详细过程见附录 C

6.2 协方差及其性质

6.2.1 协方差

对于随机向量，需要研究各分量之间的关系。

设 ξ_i 和 ξ_j 的联合分布函数为 $F_{ij}(x,y)$. 若 $E\left|(\xi_i - E\xi_i)(\xi_j - E\xi_j)\right| < \infty$, 称

$$E(\xi_i - E\xi_i)(\xi_j - E\xi_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E\xi_i)(y - E\xi_j)\,dF_{ij}(x,y)$$

为 ξ_i 和 ξ_j 的**协方差** (covariance), 记作 $Cov(\xi_i, \xi_j)$.

因此协方差就是方差性质 4 当中的交叉项，因此方差性质 4 可以改写为

$$Var\left(\sum_{i=1}^{n}\xi_i\right) = \sum_{i=1}^{n} Var\xi_i + 2\sum_{1\le i<j\le n} Cov(\xi_i, \xi_j)$$

6.2.2 协方差的性质

性质 1 $Cov(\xi,\eta) = Cov(\eta,\xi) = E\xi\eta - E\xi E\eta$

性质 2 设 a, b 是常数, 则

$$Cov(a\xi, b\eta) = abCov(\xi,\eta)$$

性质 3 $Cov\left(\sum_{i=1}^{n}\xi_i, \eta\right) = \sum_{i=1}^{n} Cov(\xi_i, \eta)$.

协方差阵

对于 n 维随机向量 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)'$, 可定义它的协方差阵如

$$\boldsymbol{B} = E(\boldsymbol{\xi} - E\boldsymbol{\xi})(\boldsymbol{\xi} - E\boldsymbol{\xi})^T = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

其中

$$b_{ij} = \begin{cases} Cov(\xi_i, \xi_j), & i \ne j \\ Var(\xi_i), & i = j \end{cases}$$

由上面的性质可知 $\boldsymbol{B}$ 是一个对称阵, 且对任何实数 $t_j, j=1,2,\cdots,n$, 二次型

$$[t_1,t_2,\cdots,t_n]\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}=\sum_{j,k} b_{jk}t_jt_k$$

$$\sum_{j,k} b_{jk}t_jt_k=\sum_{j,k} t_jt_kE\left(\xi_j-E\xi_j\right)\left(\xi_k-E\xi_k\right) \tag{6.12}$$

$$=E\left(\sum_{j=1}^{n} t_j\left(\xi_j-E\xi_j\right)\right)^2\geq 0 \tag{6.13}$$

即随机向量 ξ 的协方差阵 $\boldsymbol{B}$ 是非负定的.

性质 4 设

$$\xi=(\xi_1,\xi_2,\cdots,\xi_n)^T,\quad \boldsymbol{C}=\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

则 $C\xi$ 的协方差阵为 CBC^T, 其中 B 是 ξ 的协方差阵。因为

$$EC(\xi-E\xi)(C(\xi-E\xi))^T=CBC^T$$

性质 5 设 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)^T\sim N(\mu,B)$, 其中 μ 为 n 维向量,B 为 $n\times n$ 正定对称矩阵, 则 ξ 的数学期望为 μ, 协方差矩阵为 B。

特别地, 当 $\mu=0, B=I$ 时, $\xi_1,\xi_2,\cdots,\xi_n$ 为相互独立的标准正态随机变量,

有 $E\xi_i=0, Var\left(\xi_i\right)=1, Cov\left(\xi_i,\xi_j\right)=0, i\neq j$。即 $E\xi=0_{n\times 1}$, 协方差矩阵为 I。

一般地，设 $T:V\to V,\mathcal{M}(T)=B$，$V$ 是所有 n 元服从正态分布的随机向量构成的线性空间。由于 B 为正定对称矩阵，$\exists$ 正的实特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 和正交矩阵 Q 使得

$$B=Q^T\mathrm{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_n\}Q \tag{6.14}$$

$$=Q^TD^2Q(\text{令}D=\mathrm{diag}\{\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n}\}) \tag{6.15}$$

$$=Q^TD(QQ^T)DQ \tag{6.16}$$

$$=(Q^TDQ)(Q^TDQ) \tag{6.17}$$

令 $L=Q^TDQ$，则 $B=L^2$。考虑 L 的特征多项式 $f(\lambda)$：

$$f(\lambda)=|\lambda E-L|=|\lambda Q^TQ-Q^TDQ|=|Q^T|\cdot|\lambda E-D|\cdot|Q|=\prod_{i=1}^{n}(\lambda-\sqrt{\lambda_i})$$

因此 $\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_n}$ 是 L 的 n 个正的实特征值，可知 L 也是正交对称矩阵。

所以 $B=L^2=LL^T=LIL^T$。令 $\eta=L^{-1}(\xi-\mu)$ ，则 ξ 可以分解为 $\xi=L\eta+\mu$。

可以证明,这样分解得到的 η 服从标准正态分布,即 $\eta=L^{-1}(\xi-\mu)\sim N(0,I)$,$E\eta=0$,协方差矩阵为单位矩阵 I.

6.3 Pearson 相关系数及其性质

6.3.1 Pearson 相关系数

令 $\xi^*=\dfrac{\xi-E\xi}{\sqrt{Var\xi}},\eta^*=\dfrac{\eta-E\eta}{\sqrt{Var\eta}}$. 称

$$r_{\xi\eta}=Cov\,(\xi^*,\eta^*)=\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{Var\xi Var\eta}}$$

为 ξ,η 的 Pearson 相关系数 (correlation coefficient)。

6.3.2 Pearson 相关系数的性质

上一节的末尾，已经证明得到

$$E(X-EX)(Y-EY)\leq\left(E(X-EX)^2E(Y-EY)^2\right)^{1/2}$$

因此显然可以得到性质 1。

性质 1 对相关系数 $r_{\xi n}$ 有

$$|r_{\xi\eta}|\leq 1$$

结合空间向量几何相关知识，$r_{\xi n}=1$ 当且仅当

$$P\left(\frac{\xi-E\xi}{\sqrt{\operatorname{Var}\xi}}=\frac{\eta-E\eta}{\sqrt{\operatorname{Var}\eta}}\right)=1$$

$r_{\xi n}=-1$ 当且仅当

$$P\left(\frac{\xi-E\xi}{\sqrt{\operatorname{Var}\xi}}=-\frac{\eta-E\eta}{\sqrt{\operatorname{Var}\eta}}\right)=1$$

由性质 1，$r_{\xi\eta}=\pm1$ 时, ξ 与 η 存在线性关系。

另一个极端情形，定义 $r_{\xi\eta}=0$ 时，ξ 与 η **不相关** (uncorrelated).

性质 2 对随机变量 ξ 和 η, 如果它们的方差有限, 则下列事实等价:

(1) $\mathrm{Cov}(\xi,\eta)=0$;

(2) ξ 与 η 不相关;

(3) $E\xi\eta=E\xi E\eta$;

(4) $\mathrm{Var}(\xi+\eta)=\mathrm{Var}\,\xi+\mathrm{Var}\,\eta$.

证明：显然 (1) 与 (2) 等价. 又由协方差的性质 1 得 (1) 与 (3) 等价.

$Var(\xi+\eta)=Var(\xi)+Var(\eta)+Cov(\xi,\eta)$(方差性质 4), 得 (1) 与 (4) 等价.

性质 3 若 ξ 与 η 独立, 且它们的方差有限, 则 ξ 与 η 不相关.

显然, 由 ξ 与 η 独立知 (3) 成立, 从而 ξ 与 η 不相关. 但其逆不真。

例设随机变量 θ 服从均匀分布 $U(0,2\pi)$.$\xi=\cos\theta,\eta=\sin\theta$. 显然 $\xi^2+\eta^2=1$, 故 ξ 与 η 不独立. 但

$$E\xi=E\cos\theta=\int_0^{2\pi}\frac{1}{2\pi}\cos\varphi d\varphi=0$$
$$E\eta=E\sin\theta=\int_0^{2\pi}\frac{1}{2\pi}\sin\varphi d\varphi=0$$
$$E\xi\eta=E\sin\theta\cos\theta=\int_0^{2\pi}\frac{1}{2\pi}\sin\varphi\cos\varphi d\varphi=0$$

即 $E\xi\eta=E\xi E\eta$，ξ 与 η 不相关。因此独立条件强于不相关，独立一定不相关，不相关不一定独立。

性质 4 对二元正态随机向量, 两个分量不相关与独立是等价的

习题 *：证明性质 4。详解可见附录 C。

第七章　矩

定义 k 阶原点矩

$$m_k = E\xi^k$$

数学期望就是一阶原点矩，另外在方差公式 $Var\xi = E\xi^2 - (E\xi)^2$ 中我们经常用到的 $E\xi^2$ 就是二阶原点矩。

原点矩简称为矩，可以对比力学中计算力矩时参考点选在原点时的情况，不过力学的力矩仅是一阶原点矩，二阶原点矩或许要用能量进行类比。但我们这里并不尝试直接阐明其应用（因为我目前也还不知道），仅先讲清楚这些抽象概念。

相对应地，参考点可以不选在原点，这样参考点和原点就会有偏移。此时就需要定义 k 阶中心距

$$c_k = E(\xi - E\xi)^k$$

从定义可以知道，一阶中心距总是 0，二阶中心矩就是方差。其他常用的中心距有三阶中心距和四阶中心距，可以用来表示随机变量分布函数的形状。

如偏态系数 (Coefficient of Skewness)，衡量随机变量分布的对称性。大于 0 表示正偏态，小于 0 表示负偏态。

$$Cs = \frac{c_3}{c_2^{1.5}}$$

峰态系数 (Coefficient of Kurtosis)，衡量均值处峰值高低，若大于 0 表明比正态分布更尖峭。

$$Ck = \frac{c_4}{c_2^2} - 3$$

对于正态分布 $\xi \sim N(\mu, \sigma^2)$，其偏态系数和峰态系数都是 0。其 k 阶中心矩都存在。考虑到正态分布 $\xi \sim N(\mu, \sigma^2)$ 的 k 阶中心距其实就是 $\eta \sim N(0, \sigma^2)$ 的 k 阶原点矩，因此有

$$E(\xi - \mu)^{2k} = E\eta^{2k} = (2k-1)!!\sigma^2, E(\xi - \mu)^{2k+1} == E\eta^{2k+1} = 0$$

第八章　特征函数

8.1　常见特征函数

8.1.1　退化分布

$P(\xi = c) = 1$

$$\varphi(t) = e^{ict}$$

8.1.2　二项分布

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

$$\varphi(t) = (pe^{it} + q)^n$$

证明：

$$\begin{aligned}
\varphi(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{itk} && (8.1) \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(pe^{it}\right)^k q^{n-k} && (8.2) \\
&= (pe^{it} + q)^n && (8.3)
\end{aligned}$$

(8.2) 到 (8.3) 使用了二项式定理。

8.1.3　泊松分布

$\xi \sim \mathcal{P}(\lambda)$

$$\varphi(t) = e^{\lambda(e^{it}-1)}$$

证明：

$$\varphi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} \tag{8.4}$$

$$= \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} e^{-\lambda} \tag{8.5}$$

$$= e^{\lambda e^{it}} \cdot e^{-\lambda} = e^{\lambda(e^{it}-1)} \tag{8.6}$$

8.1.4　均匀分布

$\xi \sim U(a,b)$

$$\varphi(t) = \frac{e^{ibt} - e^{iat}}{i(b-a)t}$$

证明：

$$\varphi(t) = \int_a^b e^{itx} \frac{1}{b-a} \mathrm{d}x \tag{8.7}$$

$$= \frac{1}{i(b-a)t} e^{itx} \bigg|_a^b \tag{8.8}$$

$$= \frac{e^{ibt} - e^{iat}}{i(b-a)t} \tag{8.9}$$

8.1.5　正态分布

$\xi \sim N(\mu, \sigma^2)$

$$\varphi(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

证明：

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \tag{8.10}$$

$$\xlongequal{y=\frac{x-\mu}{\sigma}} \sigma e^{it\mu} \int_{-\infty}^{+\infty} e^{i(\sigma t)y} e^{-\frac{y^2}{2}} \mathrm{d}y \tag{8.11}$$

考虑标准正态分布需要的积分

$$\int_{-\infty}^{+\infty} e^{itx} e^{-\frac{x^2}{2}} \mathrm{d}x = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cos tx \mathrm{d}x + i \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \sin tx \mathrm{d}x \tag{8.12}$$

$$= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cos tx \mathrm{d}x (\text{第二个积分因奇函数为0}) \tag{8.13}$$

设 $g(t) = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cos tx \mathrm{d}x$ ，考虑求导

$$g'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cos tx \mathrm{d}x\right) \tag{8.14}$$

$$= -\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} \sin tx \mathrm{d}x \tag{8.15}$$

$$= \int_{-\infty}^{+\infty} \sin tx \mathrm{d}e^{-\frac{x^2}{2}} \tag{8.16}$$

$$= e^{-\frac{x^2}{2}} \sin tx \Big|_{-\infty}^{+\infty} - t\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \cos tx \mathrm{d}x = -tg(t) \tag{8.17}$$

解微分方程

$$g'(t) = -tg(t) \tag{8.18}$$

$$\frac{\mathrm{d}g}{g} = -t\mathrm{d}t \tag{8.19}$$

$$g = Ae^{-\frac{t^2}{2}} \tag{8.20}$$

$g(0) = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x = 1$，定得 $A = 1$
因此有

$$\int_{-\infty}^{+\infty} e^{itx} e^{-\frac{x^2}{2}} \mathrm{d}x = e^{-\frac{t^2}{2}}$$

则原 $\varphi(t)$ 有

$$\varphi(t) = e^{it\mu} \cdot e^{-\frac{\sigma^2 t^2}{2}} = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

8.2　可微性

8.2.1　预备知识

记

$$F(t) = \int_R f(x,t)p(x)dx$$

假定其存在, 然后以下的 $g \geqslant 0$ 需满足要求

$$\int_R g(x)p(x)dx < \infty$$

(1) $\exists g, s.t.\forall x, t$

$$|f(x,t)| < g(x)$$

对某个 x, 若

$$\lim_{t\to t_0} f(x,t) = f(x,t_0)$$

则

$$\lim_{t\to t_0} F(t) = F(t_0)$$

即 f 连续 $\Rightarrow F$ 关于 t 连续

(2)$\exists g, s.t.\forall x, t$

$$\left|\frac{\partial f(x,t)}{\partial t}\right| < g(x)$$

则

$$F'(t) = \int_R \frac{\partial f(x,t)}{\partial t} p(x)dx$$

8.2.2 可微性

现在令 $g(x) = |x|$，由于预设 X 期望存在，则

$$\int_{-\infty}^{+\infty} g(x)dF(x) = \int_{-\infty}^{+\infty} |x|dF(x) < \infty$$

对于特征函数

$$\varphi(t) = Ee^{itX} = \int_{-\infty}^{+\infty} e^{itx}dF(x)$$

令 $f(x,t) = e^{itx}$，则

$$\left|\frac{\partial f(x,t)}{\partial t}\right| = \left|ixe^{itx}\right| \leqslant |x| = g(x)$$

那么就有

$$\varphi'(t) = \int_{-\infty}^{+\infty} \frac{\partial f(x,t)}{\partial t} dF(x) = i\int_{-\infty}^{+\infty} xe^{itx}dF(x)$$

特别地，有

$$\varphi'(0) = i\int_{-\infty}^{+\infty} xdF(x) = i\mu$$

同理，考虑 k 阶 (原点) 矩，则若 $E|X|^k < \infty$，则

$$\varphi^{(k)}(t) = i^k \int_{-\infty}^{+\infty} x^k e^{itx}dF(x)$$

那么原点处 $\varphi(x)$ 可做 k 次 Taylor 展开

$$
\begin{aligned}
\varphi(x) &= \varphi(0) + \varphi'(0)x + \frac{\varphi''(0)}{2!}x^2 + \cdots + \frac{\varphi^{(k)}(0)}{k!}x^n + o(t^k) \\
&= 1 + iEXt - \frac{1}{2}EX^2t^2 + \cdots + i^k\frac{EX^k}{k!}t^k + o(t^k)
\end{aligned}
$$

第九章　概率极限定理

9.1　Bernoulli 大数律

9.1.1　内容

$p \in (0,1)$,$S_n \sim B(n,p)$, 则

$$\frac{S_n}{n} \to p, \quad n \to \infty$$

即

$$P\left(\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right) \to 0, \quad n \to \infty$$

若引入依概率收敛的概念，那么其实就是

$$\frac{S_n}{n} \xrightarrow{P} p, \quad n \to \infty$$

9.2　De Moive-Laplace 中心极限定理

9.2.1　内容

$p \in (0,1)$,$S_n \sim B(n,p)$, 则

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

这里是依分布收敛的概念。

9.3　Poisson 极限定理

$p_n \in (0,1)$,$S_n \sim B(n,p_n), np_n \to \lambda, \lambda \in (0,1)$，则

$$P(S_n = k) \to \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots, \quad n \to \infty$$

我猜想这里应该是依概率收敛。不过我也猜想这里不是重点。

9.4 Chebyschev 大数律

9.4.1 内容

ξ_k 是一列随机变量。$E\xi_k = \mu$, $S_n = \sum_{k=1}^{n} \xi_k$，如果有

$$\frac{Var(S_n)}{n^2} \to 0, \quad n \to \infty$$

那么

$$\frac{S_n}{n} \xrightarrow{P} \mu, n \to \infty$$

9.4.2 推广

$E\xi_k = \mu_k$，则满足方差条件后有

$$\frac{S_n}{n} - \frac{\sum_{k=1}^{n} \mu_k}{n} \xrightarrow{P}, n \to \infty$$

9.4.3 回忆 Chebyschev 不等式

$\exists EX, EX^2$, 则 $\forall \varepsilon > 0$

$$P(|X - EX| > \varepsilon) \leqslant \frac{Var(X)}{\varepsilon^2}$$

可以用以证明 Bernoulli 大数律

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) &\leqslant \frac{Var(\frac{S_n}{n})}{\varepsilon^2} &(9.1)\\ &= \frac{Var(\sum_{k=1}^{n} \xi_k)}{n^2\varepsilon^2} &(9.2)\\ &= \frac{\sum_{k=1}^{n} Var(\xi_k)}{n^2\varepsilon^2} &(9.3)\\ &= \frac{p(1-p)}{n\varepsilon^2} \to 0, \quad n \to \infty &(9.4) \end{aligned}$$

9.5 Khinchine 大数律

9.5.1 内容

ξ_k 独立同分布，$E\xi_k = \mu$，则

$$\frac{S_n}{n} \to \mu, \quad n \to \infty$$

9.6 Levy-Feller 中心极限定理

9.6.1 内容

$\xi_k, k \geqslant 1$ 是一列 **独立同分布**随机变量，$E\xi_k = \mu, Var(\xi_k) = \sigma^2$. 记 $S_n = \sum_{k=1}^n \xi_k$，则 $\forall x$,

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right) \to \varphi(x)$$

即

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

9.6.2 意义

(1) 应用于**一般**随机变量，是 de Moivre-Laplace 中心极限定理的**推广**

(2) 说明测量误差可用**正态分布**描述

随机测量值 X_i，均值 μ，每次误差为 $X_i - \mu$，n 次观测叠加误差 (注意是离差不是方差，可以相消) 为 $\sum\limits_{i=1}^{n}(X_i - \mu)$，则

$$\sum_{i=1}^{n}(X_i - \mu) \sim N(0, n\sigma^2), \quad n \gg 1$$

9.6.3 证明

组合计算失效，使用特征函数方法。

9.7 Lyapunov 中心极限定理

9.7.1 内容

$\xi_k, k \geqslant 1$ 是一列 **独立**随机变量 (不一定同分布)，$E\xi_k = \mu_k, Var(\xi_k) = \sigma_k^2$. 记 $S_n = \sum_{k=1}^n \xi_k, B_k = \sum_{k=1}^n \sigma_k^2$ ，如果

(1)$B_n \to \infty$

(2)$E(|X_k|^3) < \infty$, 且

$$\frac{\sum\limits_{k=1}^n E|X_k|^3}{B_n^{\frac{3}{2}}} \to 0, \quad n \to \infty$$

则 $\forall x$,

$$P\left(\frac{\sum\limits_{k=1}^n (\xi_k - \mu_k)}{\sqrt{B_n}} \leqslant x\right) \to \phi(x)$$

则

$$\frac{\sum\limits_{k=1}^n (\xi_k - \mu_k)}{\sqrt{B_n}} \xrightarrow{d} N(0,1)$$

第十章　概率论的收敛

10.1　依概率收敛

10.1.1　定义

在概率空间 (Ω, Σ, P) 中

10.1.2　敛散性判别法则

10.1.3　性质

4. 连续映射保持依概率收敛性
设 $f : \mathbf{R} \to \mathbf{R}$ 是连续映射，则

$$X_n \xrightarrow{P} X \Rightarrow f(X_n) \xrightarrow{P} f(X)$$

10.2　依分布收敛

10.2.1　定义

$X, X_n, n \geqslant 1$ 是一列随机变量，其分布函数分别为 $F, F_n, n \geqslant 1$
若对 $\forall F$ 的连续点 x，

$$F_n(x) \to F(x), \quad n \to \infty$$

则有

$$X_n \xrightarrow{d} X$$

(1) 回顾 F 的基本条件：左极限存在，右连续

(2)F 若在 $\mathbf{R}$ 上连续，则 F_n 处处收敛于 F

(3) 如果 F 单调有界，则不连续点最多可数个。

$$D_F = \{x : F(x) - F(x-0) > 0\} = \cup_{n=1}^{\infty}\{x : F(x) - F(x-0) \geqslant \frac{1}{n}\}$$

因为有界，$\{x : F(x) - F(x-0) \geqslant \frac{1}{n}\}$ 就是有限集；n 从 1 数到 ∞，则可数。

(4)F 的连续点集在 $\mathbf{R}$ 上稠密

10.2.2 依概率收敛强于依分布收敛

(1) 依概率收敛 $\Rightarrow$ 依分布收敛

即

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

证明：$\forall x \in \mathbf{R}, \varepsilon > 0$,

$$P(X_n \leqslant x) = P(X_n \leqslant x, X_n - X \geqslant -\varepsilon) + P(X_n \leqslant x, X_n - X < -\varepsilon)$$

然而，

$$P(X_n \leqslant x, X_n - X < -\varepsilon) \leqslant P(X_n - X < -\varepsilon) \to 0$$

则

$$\begin{aligned}\lim_{n\to\infty} \sup P(X_n \leqslant x) =& P(X_n \leqslant x, X_n - X \geqslant -\varepsilon)\\ =& P(-\varepsilon \leqslant X_n - X \leqslant x - X)\\ \leqslant& P(X \leqslant x + \varepsilon) \to P(X \leqslant x), \varepsilon \to 0\end{aligned}$$

因为分布函数右连续所以可以直接这么趋向。然而分布函数只是左极限存在，并不一定左连续，所以有

同理，

$$\lim_{n\to\infty} \inf P(X_n \leqslant x) \geqslant P(X \leqslant x - \varepsilon) \to P(X < x), \varepsilon \to 0$$

但是在连续点，就也左连续了，那么就可以得到

$$\lim_{n\to\infty} P(X_n \leqslant x) = P(X \leqslant x)$$

也就得到了依分布收敛。

(2) 依概率收敛 $\nRightarrow$ 依分布收敛

有反例：
设 Y 为非退化对称随机变量，则显然有

$$Y \overset{d}{=} -Y$$

那么令 $X_n = Y, X = Y$，则 $X, X_n, n \geqslant 1$ 分布相同，有

$$X_n \xrightarrow{d} X$$

然而

$$P(|X_n - X| - \varepsilon) = P(2|Y| > \varepsilon)$$

Y 不恒等于 0（否则就是退化分布了），那么 X_n 不依概率收敛到 X
对于退化分布，特别地，有

$$X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{P} c$$

即退化分布情况下，依概率收敛和依分布收敛是等价的。
证明：

10.2.3　敛散性判别法则

Levy 连续性定理

$X, X_n, n \geqslant 1$ 是一列随机变量，其特征函数分别为 $\phi, \phi_n, n \geqslant 1$，则

$$X_n \xrightarrow{d} X \Rightarrow \phi_n(t) \to \phi(t), \quad t \in \mathbf{R}$$

另一种形式，如果

$$\phi_n(t) \to \phi(t), \quad t \in \mathbf{R}$$

且 ϕ 在 0 处连续，那么 ϕ 一定是特征函数，其对应的随机变量 X 满足

$$X_n \xrightarrow{d} X$$

可以认为是等价条件。

证明 Khinchine 大数律

分析：频数除以次数依概率收敛于期望，由于期望可以看成退化分布的随机变量，所以和按分布收敛于期望是等价的。

$\xi_k, k \geqslant 1$ 独立同分布，$E\xi_k = \mu$ 时，令 $X_n = \dfrac{1}{n}\sum_{k=1}^{n}\xi_n$，由独立性有

$$\begin{aligned}\phi_n(t) &= E\exp\left\{i\frac{1}{n}\sum_{k=1}^{n}\xi_n t\right\}\\ &= \prod_{k=1}^{n}\left[E\exp\left\{i\frac{1}{n}\xi_k t\right\}\right] \text{(独立性)}\\ &= \left[E\exp\left\{i\frac{1}{n}\xi_1 t\right\}\right]^n \text{(同分布)}\end{aligned}$$

对 $E\exp\left\{i\dfrac{1}{n}\xi_1 t\right\}$ 进行泰勒展开，有

$$E\exp\left\{i\frac{1}{n}\xi_1 t\right\} = 1 + i\frac{t}{n}\mu + o(\frac{t}{n})$$

那么，$\forall t \in \mathbf{R}$，

$$\phi_n(t) = \left[1 + \frac{it\mu}{n} + o(\frac{t}{n})\right]^n \to e^{it\mu}$$

显然 $e^{it\mu}$ 在 0 处连续，且对应常数为 μ 的退化分布，那么得证依分布收敛

证明 Levy-Feller 中心极限定理

$\xi_k, k \geqslant 1$ 独立同分布，$E\xi_k = \mu, Var(\xi_k) = \sigma^2$，那么

$$X_n = \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(\xi_k - \mu) \xrightarrow{d} N(0,1)$$

证明：只需证

$$\phi_n(t) = E\exp\{itX_n\} \to exp\{-\frac{t^2}{2}\}$$

易得

$$\phi_n(t) = \left[E\exp\left\{it\frac{\xi_1 - \mu}{\sigma\sqrt{n}}\right\}\right]^n$$

泰勒展开，有

$$E\exp\left\{i\frac{t}{\sigma\sqrt{n}}(\xi_1 - \mu)\right\} = 1 + i\frac{t}{\sigma\sqrt{n}}E(\xi_1 - \mu) - \frac{t^2}{\sigma^2 n}E(\xi_1 - \mu)^2 + o(\frac{1}{n})$$

第十一章　正态分布

11.1　一元正态分布密度函数的规范性

先考虑重要积分 $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x$ 的值

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x\right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} \mathrm{d}y \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x && (11.1)\\
&= \iint\limits_{R^2} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\mathrm{d}y && (11.2)\\
&= \iint\limits_{R^2} e^{-\frac{\rho^2}{2}} \rho\mathrm{d}\rho\mathrm{d}\theta && (11.3)\\
&= \int_0^{2\pi} \mathrm{d}\theta \int_0^{+\infty} e^{-\frac{\rho^2}{2}} \rho\mathrm{d}\rho && (11.4)\\
&= \int_0^{2\pi} \mathrm{d}\theta \int_0^{+\infty} -e^{-\frac{\rho^2}{2}} \mathrm{d}(-\frac{\rho^2}{2}) && (11.5)\\
&= 2\pi && (11.6)
\end{aligned}$$

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

下证一元正态分布密度函数 $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ 的规范性

$$\begin{aligned}
\int_{-\infty}^{+\infty} p(x)\mathrm{d}x &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \mathrm{d}x && (11.7)\\
&\xlongequal{t=\frac{x-\mu}{\sigma}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} \mathrm{d}t = 1 && (11.8)
\end{aligned}$$

得证。

11.2　二元正态分布的边际分布与线性变换

11.2.1　二元正态分布的边际分布

考虑 $(X,Y)\sim N(a,\sigma_1,b,\sigma_2;r)$，求 $p_X(x),p_Y(y)$

在二元正态分布的密度函数

$$p(x)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2\rho(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

的指数中对 y 配方, 可把 $p(x,y)$ 写成

$$p(x,y)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-a)^2}{2\sigma_1^2}\right\}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{\left[y-b-\frac{r\sigma_2}{\sigma_1}(x-a)\right]^2}{2\sigma_2^2\left(1-r^2\right)}\right\}$$

令

$$q(x,y)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{\left[y-b-\frac{r\sigma_2}{\sigma_1}(x-a)\right]^2}{2\sigma_2^2\left(1-r^2\right)}\right\}$$

$$\int_{-\infty}^{+\infty}q(x,y)\mathrm{d}y=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{\left[y-b-\frac{r\sigma_2}{\sigma_1}(x-a)\right]^2}{2\sigma_2^2\left(1-r^2\right)}\right\}\mathrm{d}y \tag{11.9}$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}\mathrm{d}t\cdots\cdots t=\frac{y-b-\frac{r\sigma_2}{\sigma_1}(x-a)}{\sigma_2\sqrt{1-r^2}}\qquad=1 \tag{11.10}$$

$$\tag{11.11}$$

则

$$p_X(x)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-a)^2}{2\sigma_1^2}\right\}q(x,y)\mathrm{d}y \tag{11.12}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-a)^2}{2\sigma_1^2}\right\} \tag{11.13}$$

同理

$$p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-b)^2}{2\sigma_2^2}\right\}$$

11.2.2 二元正态分布的线性变换

假设 $(X,Y) \sim N\left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho\right)$ 。定义

$$\begin{pmatrix} U \\ V \end{pmatrix} = A \cdot \begin{pmatrix} X \\ Y \end{pmatrix}, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{可逆}$$

求 (U,V) 的分布密度

A 可逆，则

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A^{-1} \begin{pmatrix} U \\ V \end{pmatrix}$$

Jacobi 行列式

$$J = |A^{-1}| = |A|^{-1}$$

协方差矩阵 Σ，随机向量 $\vec{x}$，期望向量 $\vec{\mu}$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

则

$$p_{X,Y}(x,y) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right\}$$

随机向量 $\vec{u} = \begin{pmatrix} u \\ v \end{pmatrix}$

$$p_{U,V}(u,v) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(A^{-1}\vec{u}-\vec{\mu})^T\Sigma^{-1}(A^{-1}\vec{u}-\vec{\mu})\right\}|J| \tag{11.14}$$

$$= \frac{1}{2\pi|A^T\Sigma A|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{u}-A\vec{\mu})^T(A\Sigma A^T)^{-1}(\vec{u}-A\vec{\mu})\right\} \tag{11.15}$$

则

$$(U,V) \sim N(A\vec{\mu}, A\Sigma A^T)$$

$$A\vec{\mu} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} a\mu_1 + b\mu_2 \\ c\mu_1 + d\mu_2 \end{pmatrix}$$

$A\Sigma A^T$ 不再展开，那将会是个很长的可怕式子。

11.3　二元正态分布的条件分布和独立性

11.3.1　二元正态分布的条件分布

考虑 $(X,Y)\sim N(\mu_1,\sigma_1,\mu_2,\sigma_2;\rho)$，求 $p_{X|Y}(x|y),p_{Y|X}(y|x)$

当然，如果直接使用 C.2 中的配方结果，就不需要下面这么麻烦地化简了。因为 C.2 中没有给出配方过程，因此在这里稍微写得详细一点。

$$
\begin{aligned}
p_{X|Y}(x|y) &= \frac{p(x,y)}{p_Y(y)} &(11.16)\\
&= \frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\dfrac{(x-\mu_1)^2}{\sigma_1^2}-\dfrac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\dfrac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{\dfrac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\dfrac{(y-\mu_2)^2}{2\sigma_2^2}\right\}} &(11.17)\\
&= \frac{\exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\dfrac{(x-\mu_1)^2}{\sigma_1^2}-\dfrac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\dfrac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{\sqrt{2\pi(1-\rho^2)}\sigma_1\exp\left\{-\dfrac{(y-\mu_2)^2}{2\sigma_2^2}\right\}} &(11.18)\\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2(1-\rho^2)\sigma_1^2}+\frac{2\rho(x-\mu_1)(y-\mu_2)}{2(1-\rho^2)\sigma_1\sigma_2}-\frac{\rho^2(y-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}\right\} &(11.19)\\
&= A_1\exp\left\{A_2\left[(x-\mu_1)^2-2k\rho(x-\mu_1)(y-\mu_2)+k^2\rho^2(y-\mu_2)^2\right]\right\} &(11.20)\\
&\left(A_1=\frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1},A_2=-\frac{1}{2(1-\rho^2)\sigma_1^2},k=\frac{\sigma_1}{\sigma_2}\right) &(11.21)\\
&= A_1\exp\{A_2[(x-\mu_1)-\rho k(y-\mu_2)]^2\} &(11.22)\\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1}\exp\left\{-\frac{[(x-\mu_1)-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)]^2}{2(1-\rho^2)\sigma_1^2}\right\} &(11.23)
\end{aligned}
$$

同理

$$p_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2}\exp\left\{-\frac{[(y-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)]^2}{2(1-\rho^2)\sigma_2^2}\right\}$$

11.3.2　二元正态分布的独立性等价条件

考虑独立性，有结论

$$\rho=0\iff X,Y\text{相互独立}\iff p(x,y)=p_X(x)p_Y(y),\forall x,y$$

因为

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \tag{11.24}$$

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\} \tag{11.25}$$

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)}\cdot \tag{11.26}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \tag{11.27}$$

$$p(x,y) = p_X(x)p_Y(y) \iff \tag{11.28}$$

$$\frac{1}{(1-\rho^2)} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{\rho^2(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{\rho^2(y-\mu_2)^2}{\sigma_2^2}\right]\right\} = 1 \tag{11.29}$$

$$\iff \rho^2 a^2 - 2\rho ab + \rho^2 b^2 = -2\sigma_1^2(1-\rho^2)\ln(1-\rho^2), \forall a = (x-\mu_1), b = \frac{\sigma_1}{\sigma_2}(y-\mu_2) \tag{11.30}$$

$\rho = 0$ 时, 显然反推成立。则考虑正推，用反证法，若 $\rho \in (-1,0)\cup(0,1)$

有

$$\rho a^2 - 2ab + \rho b^2 = -2\frac{\sigma_1^2}{\rho}(1-\rho^2)\ln(1-\rho^2) = C$$

设

$$f(a,b) = \rho a^2 - 2ab + \rho b^2, f(a,b) = C \Rightarrow \frac{\partial f}{\partial a} = 0, \frac{\partial f}{\partial b} = 0$$

$$\therefore \begin{cases} \dfrac{\partial f}{\partial a} = 2\rho a - 2b = 0 \\ \\ \dfrac{\partial f}{\partial b} = 2\rho b - 2a = 0 \end{cases} \Rightarrow b = \rho a = \rho(\rho b) = \rho^2 b, (1-\rho^2)b = 0, b = 0$$

同理 $a = 0$

这与 a, b 任取，$f \equiv C$ 矛盾，则假设不成立，必有 $\rho = 0$

11.4　一元正态分布的期望与方差

设 $\xi \sim N(\mu, \sigma^2)$，则其密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

首先预备求两个积分：$\int_{-\infty}^{+\infty} xe^{-\frac{x^2}{2}}$ 和 $\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}}\mathrm{d}x$
对于第一个积分，由于被积函数是奇函数，有：

$$\int_{-\infty}^{+\infty} xe^{-\frac{x^2}{2}}\mathrm{d}x = 0$$

对于另外一个积分，有：

$$\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}}\mathrm{d}x = -\int_{-\infty}^{+\infty} x\mathrm{d}(e^{-\frac{x^2}{2}}) = -xe^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\mathrm{d}x = \sqrt{2\pi}$$

因为

$$\lim_{x\to\infty} xe^{-\frac{x^2}{2}} = \lim_{x\to\infty} \frac{x}{e^{\frac{x^2}{2}}} = \lim_{x\to\infty} \frac{1}{xe^{\frac{x^2}{2}}} = 0$$

考虑其期望，有

$$\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\mathrm{d}x && (11.31)\\
&\xlongequal{t=\frac{x-\mu}{\sigma}} \int_{-\infty}^{+\infty} (\sigma t+\mu)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t && (11.32)\\
&= \sigma\int_{-\infty}^{+\infty} t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t + \mu\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t && (11.33)\\
&= \mu && (11.34)
\end{aligned}$$

考虑方差，有

$$\begin{aligned}
EX^2 &= \int_{-\infty}^{+\infty} x^2\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\mathrm{d}x && (11.35)\\
&\xlongequal{t=\frac{x-\mu}{\sigma}} \int_{-\infty}^{+\infty} (\sigma t+\mu)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t && (11.36)\\
&= \frac{1}{\sqrt{2\pi}}\left(\sigma^2\int_{-\infty}^{+\infty} t^2e^{-\frac{t^2}{2}}\mathrm{d}t + 2\sigma\mu\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}\mathrm{d}t + \mu^2\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\mathrm{d}t\right) && (11.37)\\
&= \sigma^2+\mu^2 && (11.38)
\end{aligned}$$

则 $Var\xi = EX^2 - (EX)^2 = \sigma^2$

11.5　二元正态分布的协方差、Pearson 相关系数

$(\xi,\eta)\sim N\left(a,b,\sigma_1^2,\sigma_2^2,r\right)$, 求 $Cov(\xi,\eta)$ 和 $r_{\xi,\eta}$.

解

$$
\begin{aligned}
Cov(\xi,\eta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)(y-b)p(x,y)dxdy && (11.39)\\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)(y-b) && (11.40)\\
&\cdot\exp\left\{-\frac{1}{2\left(1-r^2\right)}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right)^2-\frac{(y-b)^2}{2\sigma_2^2}\right\}dxdy && (11.41)\\
& && (11.42)
\end{aligned}
$$

令

$$z=\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2},\quad t=\frac{y-b}{\sigma_2}$$

则

$$\frac{x-a}{\sigma_1}=z+rt,\quad J=\frac{\partial(x,y)}{\partial(z,t)}=\sigma_1\sigma_2$$

于是

$$
\begin{aligned}
Cov(\xi,\eta) =&\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(zt+rt^2\right)e^{-\frac{z^2}{2(1-r^2)}}e^{-\frac{t^2}{2}}dzdt && (11.43)\\
=&\sigma_1\sigma_2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}te^{-\frac{t^2}{2}}dt\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}\int_{-\infty}^{\infty}ze^{-\frac{z^2}{2(1-r^2)}dz} && (11.44)\\
&+\frac{r\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}\int_{-\infty}^{\infty}e^{-\frac{s^2}{2(1-r^2)}}dz && (11.45)\\
=&r\sigma_1\sigma_2 && (11.46)
\end{aligned}
$$

故得

$$r_{\xi\eta}=\frac{Cov(\xi,\eta)}{\sqrt{\operatorname{Var}\xi\operatorname{Var}\eta}}=r$$

因此，ξ 与 η 不相关等价于 $r=0$，也就等价于 ξ 与 η 相互独立。