

概率论与数理统计

xyfJASON

1 随机变量的函数

1.1 一元函数

1.1.1 一般情形

1.1.2 线性函数情形

1.1.3 单调函数情形

1.2 二元函数

1.2.1 一般情形

1.2.2 独立随机变量之和

1.2.3 极值分布

1.2.4 瑞利分布

1.3 多元函数

2 独立性与相关性

2.1 概览

2.2 协方差与相关系数

3 多元随机变量

3.1 概览

3.2 二元正态分布相关

3.2.1 定义

3.2.2 密度分解

3.2.3 边缘分布

3.2.4 协方差与相关系数

3.2.5 独立性

4 条件

4.1 条件概率

4.1.1 条件概率

4.1.2 全概率公式

4.1.3 贝叶斯公式

4.1.4 条件独立

4.2 条件分布

4.2.1 条件分布列

4.2.2 条件概率密度函数

4.3 条件期望

4.3.1 条件期望

4.3.2 全期望定理

5 极限理论

5.1 马尔可夫不等式

5.2 切比雪夫不等式

5.3 弱大数定律

5.4 中心极限定理

6 数理统计

6.1 正态分布的三个导出分布

6.1.1 χ^2 分布

6.1.2 t 分布

6.1.3 F 分布

6.1.4 性质

6.2 点估计

6.2.1 矩估计

6.2.2 极大似然估计

6.2.3 无偏性

1 随机变量的函数

1.1 一元函数

1.1.1 一般情形

PMF/PDF: 设 X 是一离散随机变量, 则 $Y = g(X)$ 也是一个离散随机变量, 且其分布列为:

$$p_Y(y) = \sum_{\{x|y=g(x)\}} p_X(x)$$

设 X 是一连续随机变量, 则 $Y = g(X)$ 也是一个连续随机变量, 且其分布函数为:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{\{x|g(x) \leq y\}} f_X(x) dx$$

于是 Y 的概率密度函数为:

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

期望: 设 X 是一离散随机变量, 则 $Y = g(X)$ 的期望为:

$$\mathbb{E}Y = \mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$$

设 X 是一连续随机变量, 则 $Y = g(X)$ 的期望为:

$$\mathbb{E}Y = \mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx$$

也就是说, 我们不必先求出 Y 的分布, 只需知道 X 的分布就能求出 Y 的期望。

1.1.2 线性函数情形

设 X 是一连续随机变量, $a, b \in \mathbb{R}$ 且 $a \neq 0$, 设 $Y = aX + b$, 则:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

证: 使用1.1节中描述的方法, 先求 Y 的分布函数:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \begin{cases} \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) & a > 0 \\ \mathbb{P}\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right) & a < 0 \end{cases}$$

然后求导得到 Y 的概率密度函数:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right) & a > 0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) & a < 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

证毕。

例子：正态分布的线性变换仍然是正态分布，且：

$$X \sim N(\mu, \sigma^2) \implies aX + b \sim N(a\mu + b, a^2\sigma^2)$$

证：运用上述定理，

$$\begin{aligned} f_{aX+b}(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma|a|} \exp - \frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma|a|} \exp - \frac{(y - a\mu - b)^2}{2a^2\sigma^2} \end{aligned}$$

所以， $aX + b \sim N(a\mu + b, a^2\sigma^2)$. 证毕。

1.1.3 单调函数情形

设 X 是一连续随机变量， g 是一严格单调的可逆函数，且其反函数 h 可微，则 $Y = g(X)$ 在 $\{y | f_Y(y) > 0\}$ 内的概率密度函数是：

$$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$$

证：仍然先求 Y 的分布函数：

$$F_Y(y) = \mathbb{P}(Y \leq y) = \begin{cases} \mathbb{P}(X \leq h(y)) = F_X(h(y)) & g \text{ 单调递增} \\ \mathbb{P}(X \geq h(y)) = 1 - F_X(h(y)) & g \text{ 单调递减} \end{cases}$$

于是求导得：

$$f_Y(y) = \begin{cases} f_X(h(y))h'(y) & g \text{ 单调递增} \\ -f_X(h(y))h'(y) & g \text{ 单调递减} \end{cases} = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$$

证毕。

1.2 二元函数

1.2.1 一般情形

PMF/PDF: 设 X, Y 是离散随机变量，则 $Z = g(X, Y)$ 也是一个离散随机变量，且其分布列为：

$$p_Z(z) = \sum_{\{(x,y) | z=g(x,y)\}} p_{X,Y}(x, y)$$

设 X, Y 是连续随机变量，则 $Z = g(X, Y)$ 也是一个连续随机变量，且其分布函数为：

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X, Y) \leq z) = \iint_{\{(x,y)|g(x,y)\leq z\}} f_{X,Y}(x,y) dx dy$$

于是 Z 的概率密度函数为:

$$f_Z(z) = \frac{dF_Z(z)}{dz}$$

期望: 设 X 和 Y 是联合离散随机变量, 则 $Z = g(X, Y)$ 的期望为:

$$\mathbb{E}Z = \mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) f_{X,Y}(x, y)$$

设 X 和 Y 是联合连续随机变量, 则 $Z = g(X, Y)$ 的期望为:

$$\mathbb{E}Z = \mathbb{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy$$

同样地, 我们不必先求出 Z 的分布, 只需知道 X 和 Y 的联合分布就能求出 Z 的期望。

1.2.2 独立随机变量之和

设 X, Y 是独立的离散随机变量, $Z = X + Y$, 则:

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x)$$

设 X, Y 是独立的连续随机变量, $Z = X + Y$, 则:

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

称上述两个式子为卷积 (convolution)。

例子: 相互独立的正态随机变量之和仍服从正态分布。

设 $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ 且相互独立, $Z = X + Y$, 则 $Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ 。

证: 运用上述定理,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(z-x-\mu_y)^2}{2\sigma_y^2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\frac{u^2}{\sigma_x^2} + \frac{(v-u)^2}{\sigma_y^2}\right]} du \quad u = x - \mu_x, v = z - \mu_x - \mu_y \end{aligned}$$

由于

$$\frac{u^2}{\sigma_x^2} + \frac{(v-u)^2}{\sigma_y^2} = \frac{u^2}{\sigma_x^2} + \frac{u^2}{\sigma_y^2} + \frac{v^2}{\sigma_y^2} - \frac{2uv}{\sigma_y^2} = \left(\frac{\sqrt{\sigma_x^2 + \sigma_y^2}}{\sigma_x\sigma_y} u - \frac{\sigma_x v}{\sigma_y\sqrt{\sigma_x^2 + \sigma_y^2}} \right)^2 + \frac{v^2}{\sigma_x^2 + \sigma_y^2}$$

令 $t = \frac{\sqrt{\sigma_x^2 + \sigma_y^2}}{\sigma_x\sigma_y} u - \frac{\sigma_x v}{\sigma_y\sqrt{\sigma_x^2 + \sigma_y^2}}$, 则:

$$f_Z(z) = \frac{1}{2\pi\sigma_x\sigma_y} \frac{\sigma_x\sigma_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} e^{-\frac{z^2}{\sigma_x^2 + \sigma_y^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2 + \sigma_y^2}} e^{-\frac{(z-\mu_x-\mu_y)^2}{\sigma_x^2 + \sigma_y^2}}$$

所以 $Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$. 证毕。

推论：设 $X_i \sim N(\mu_i, \sigma_i^2)$ 且相互独立，则 $\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

1.2.3 极值分布

设 X, Y 是独立的随机变量， $M = \max(X, Y)$, $N = \min(X, Y)$ ，则：

$$F_M(z) = \mathbb{P}(M \leq z) = \mathbb{P}(X \leq z \wedge Y \leq z) = \mathbb{P}(X \leq z)\mathbb{P}(Y \leq z) = F_X(z)F_Y(z)$$

$$F_N(z) = \mathbb{P}(N \leq z) = 1 - \mathbb{P}(N > z) = 1 - \mathbb{P}(X > z \wedge Y > z) = 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

更一般的，设 X_1, X_2, \dots, X_n 是 n 个相互独立的随机变量， $M = \max_{1 \leq i \leq n} X_i$ ， $N = \min_{1 \leq i \leq n} X_i$ ，则：

$$F_M(z) = \prod_{i=1}^n F_{X_i}(z)$$

$$F_N(z) = 1 - \prod_{i=1}^n [1 - F_{X_i}(z)]$$

1.2.4 瑞利分布

设 $X, Y \sim N(0, \sigma)$ 且相互独立， $R = \sqrt{X^2 + Y^2}$ ，称 R 服从瑞利分布，其分布函数为：

$$\begin{aligned} F_R(r) &= \mathbb{P}(R \leq r) = \mathbb{P}(X^2 + Y^2 \leq r^2) \\ &= \iint_{\{(x,y)|x^2+y^2 \leq r^2\}} p_{X,Y}(x,y) dx dy \\ &= \iint_{x^2+y^2 \leq r^2} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy \\ &= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} d\theta \int_0^r e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho \\ &= e^{-\frac{\rho^2}{2\sigma^2}} \Big|_r^0 \\ &= 1 - e^{-\frac{r^2}{2\sigma^2}} \end{aligned}$$

故概率密度函数为：

$$f_R(r) = \frac{dF_R(r)}{dr} = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad (r > 0)$$

1.3 多元函数

设 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 是一个随机向量, T 是 \mathbb{R}^n 上的一可逆映射, $\mathbf{Y} = T(\mathbf{X})$, 则 \mathbf{Y} 的概率密度函数为:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(T^{-1}(\mathbf{y}))|J|$$

其中, J 表示 $T^{-1}: \mathbf{y} \mapsto \mathbf{x}$, 即 $\mathbf{x} = T^{-1}(\mathbf{y})$ 的雅各比行列式:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

证: 设 $D \subset \mathbb{R}^n$ 是一个性质好的集合, 则:

$$\begin{aligned} \mathbb{P}(\mathbf{Y} \in D) &= \mathbb{P}(T(\mathbf{X}) \in D) \\ &= \mathbb{P}(\mathbf{X} \in T^{-1}(D)) && \text{两边同时施以 } T^{-1} \\ &= \int_{T^{-1}(D)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_D f_{\mathbf{X}}(T^{-1}(\mathbf{y}))|J| d\mathbf{y} && \text{变量代换 } \mathbf{y} = T(\mathbf{x}) \end{aligned}$$

又

$$\mathbb{P}(\mathbf{Y} \in D) = \int_D f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$$

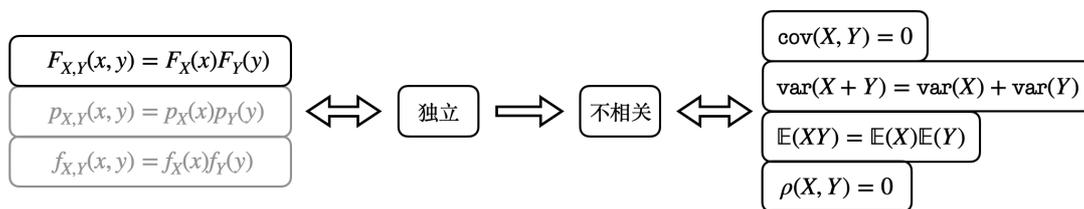
根据 D 一定的任意性, 可知:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(T^{-1}(\mathbf{y}))|J|$$

证毕。

2 独立性与相关性

2.1 概览



注意，我们所说的不相关默认指非线性相关，即 X 与 Y 没有线性关系，但是可能具有其他关系而导致不独立。

2.2 协方差与相关系数

定义：

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}X\text{var}Y}}$$

当 $\text{cov}(X, Y) = 0$ 时 ($\rho(X, Y) = 0$)，称 X 和 Y 不相关。

性质：

- $\text{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$
- $\text{cov}(X, X) = \text{var}(X)$
- $\text{cov}(X, aY + b) = a \cdot \text{cov}(X, Y)$
- $\text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z)$
- $\text{var}(cX) = c^2\text{var}X$
- $\text{var}(X + Y) = \text{var}X + 2\text{cov}(X, Y) + \text{var}Y$
- $\text{var}(aX + bY) = a^2\text{var}X + 2abc\text{cov}(X, Y) + b^2\text{var}Y = [a \quad b] \begin{bmatrix} \text{var}X & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{var}Y \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$
- $\text{var}(X) = 0 \implies \mathbb{P}(X = \mathbb{E}X) = 1$

证：根据切比雪夫不等式， $\mathbb{P}(|X - \mathbb{E}X| \geq \frac{1}{n}) \leq n^2\text{var}(X) = 0$ 对 $\forall n$ 都成立，故 $\mathbb{P}(X = \mathbb{E}X) = 1$. 证毕。

- $|\rho| \leq 1$ ，且 $\rho = 1$ 当且仅当 $\mathbb{P}(Y = aX + b) = 1$.

证：对 $\forall t \in \mathbb{R}$ ，由于 $\text{var}(Y - tX) = t^2\text{var}X - 2t\text{cov}(X, Y) + \text{var}Y \geq 0$ ，所以：

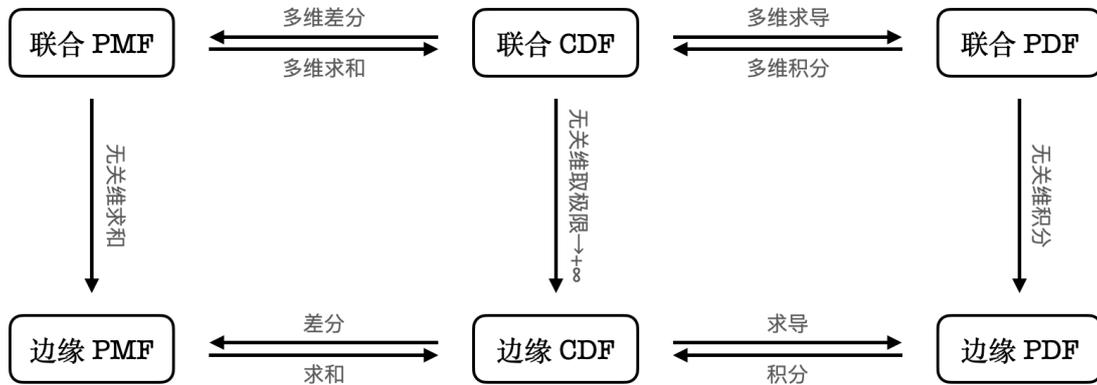
$$\Delta = 4\text{cov}^2(X, Y) - 4\text{var}X\text{var}Y \leq 0$$

即有： $|\rho| \leq 1$ ，且 $\rho = 1$ 当且仅当存在某个 b 使得 $\text{var}(Y - bX) = 0$ ，于是根据方差的性质有：

$\mathbb{P}(Y - bX = \mathbb{E}[Y - bX] = a) = 1$ ，其中 a, b 均为常数，证毕。

3 多元随机变量

3.1 概览



3.2 二元正态分布相关

3.2.1 定义

若随机变量 X, Y 有如下联合概率密度函数：

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

称 X, Y 服从参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二元正态分布。

矩阵形式：设 $\mathbf{X} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, 则：

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right)$$

这一形式可以推广到多元正态分布。

3.2.2 密度分解

对定义式进行变形可以得到：

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

注意到，前一部分是 $N(\mu_1, \sigma_1)$ 的概率密度函数，后一部分是 $N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$ 的概率密度函数。又由于：

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

所以事实上后一部分就是 $p_{Y|X}(y|x)$.

3.2.3 边缘分布

根据密度分解容易知道，二元正态分布的边缘分布就是正态分布，且：

$$X \sim N(\mu_1, \sigma_1^2), \quad Y \sim N(\mu_2, \sigma_2^2)$$

3.2.4 协方差与相关系数

运用密度分解，可以计算：

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\ &= \iint_{\mathbb{R}^2} (x - \mu_1)(y - \mu_2)f_{X,Y}(x, y)dx dy \\ &= \int_{-\infty}^{+\infty} (x - \mu_1)f_X(x)dx \int_{-\infty}^{+\infty} (y - \mu_2)f_{Y|X}(y|x)dy \\ &= \int_{-\infty}^{+\infty} (x - \mu_1)f_X(x)dx \left[\int_{-\infty}^{+\infty} yf_{Y|X}(y|x) - \mu_2 \right] \\ &= \int_{-\infty}^{+\infty} (x - \mu_1)f_X(x)dx \left[\mathbb{E}[Y|X = x] - \mu_2 \right] \\ &= \int_{-\infty}^{+\infty} (x - \mu_1)f_X(x)dx \left[\rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) \right] \\ &= \rho \frac{\sigma_2}{\sigma_1} \int_{-\infty}^{+\infty} (x - \mu_1)^2 f_X(x)dx \\ &= \rho \frac{\sigma_2}{\sigma_1} \text{var} X \\ &= \rho \sigma_1 \sigma_2 \end{aligned}$$

由此可得相关系数：

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var} X} \sqrt{\text{var} Y}} = \rho$$

也即二元正态分布定义中的 ρ 就是其相关系数。

3.2.5 独立性

定理：设 (X, Y) 服从二元正态分布，则 X, Y 独立当且仅当 $\rho = 0$.

证：由于 X, Y 独立蕴含着 X, Y 不相关，而后者等价于相关系数 $\rho = 0$ ，所以独立 $\implies \rho = 0$.

又设 $\rho = 0$ ，则：

$$p_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(y - \mu_2)^2}{2\sigma_2^2}\right) = p_X(x)p_Y(y)$$

所以 $\rho = 0 \implies$ 独立。证毕。

由该定理可知，对于二元正态分布而言，独立和不相关是等价的。

4 条件

4.1 条件概率

4.1.1 条件概率

定义： 设 A, B 为两个事件且 $\mathbb{P}(B) > 0$ ，则称 $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ 为 B 发生的条件下 A 发生的概率，记作 $\mathbb{P}(A | B)$ ，即：

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$$

公理化： 可以证明，条件概率是一个公理化定义下的概率：

1. 非负性：显然；
2. 规范性：

$$\mathbb{P}(\Omega | B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

3. 可数可加性：

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i | B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cap B\right)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}\left(\bigcup_{i=1}^n (A_i \cap B)\right)}{\mathbb{P}(B)} \\ &= \frac{\sum_{i=1}^n \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\ &= \sum_{i=1}^n \mathbb{P}(A_i | B) \end{aligned}$$

证毕。

因此，**概率的所有性质都适用于条件概率。**

乘法公式： 由条件概率定义有：

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A | B)$$

推广：

$$\mathbb{P}(A_1 A_2 \cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 A_2) \cdots \mathbb{P}(A_n | A_1 A_2 \cdots A_{n-1})$$

4.1.2 全概率公式

设 A_1, A_2, \dots, A_n 互不相容, $B \subset A_1 \cup A_2 \cup \dots \cup A_n$, 则:

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

证:

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}\left(B \cap \left(\bigcup_{i=1}^n A_i\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n (A_i \cap B)\right) \\ &= \sum_{i=1}^n \mathbb{P}(A_i \cap B) \\ &= \sum_{i=1}^n \mathbb{P}(A_i) \mathbb{P}(B | A_i) \end{aligned}$$

证毕。

4.1.3 贝叶斯公式

设 A_1, A_2, \dots, A_n 互不相容且 $\mathbb{P}(A_i) > 0$, $B \subset A_1 \cup A_2 \cup \dots \cup A_n$, 则:

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

关于贝叶斯公式的理解: 视事件 A_i 是导致事件 B 发生的原因, 我们对于事件 A_i 已有一个先验概率 $\mathbb{P}(A_i)$, 现在事件 B 发生了, 这必然给我们带了一定的信息, 于是我们可以由此修正 A_i 发生的概率, 得到 $\mathbb{P}(A_i | B)$, 即后验概率。

4.1.4 条件独立

我们知道, 两个事件 A, B 相互独立, 是指 $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$. 由于条件概率是符合概率公理化定义的概率律, 我们可以对条件独立有类似的定义。

定义: 给定事件 C , 若事件 A, B 满足:

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A | C) \mathbb{P}(B | C)$$

则称 A, B 在给定事件 C 下条件独立。

注意, A, B 条件独立并不能推出 A, B 独立, 反过来亦不成立。

4.2 条件分布

4.2.1 条件分布列

离散随机变量 Y 取定某个值 y 后，离散随机变量 X 的条件分布列为：

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

于是我们有联合分布列、边缘分布列、条件分布列之间的关系：

$$\begin{aligned} p_{X,Y}(x, y) &= p_Y(y)p_{X|Y}(x | y) \\ p_X(x) &= \sum_y p_{X,Y}(x, y) = \sum_y p_Y(y)p_{X|Y}(x | y) \end{aligned}$$

4.2.2 条件概率密度函数

连续随机变量 Y 取定某个值 y 后，连续随机变量 X 的条件概率密度函数为：

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

同样我们有联合概率密度函数、边缘概率密度函数、条件概率密度函数之间的关系：

$$\begin{aligned} f_{X,Y}(x, y) &= f_Y(y)f_{X|Y}(x | y) \\ f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy = \int_{-\infty}^{+\infty} f_Y(y)f_{X|Y}(x | y)dy \end{aligned}$$

4.3 条件期望

4.3.1 条件期望

对于离散随机变量，

$$\mathbb{E}[X | Y = y] = \sum_x xp_{X|Y}(x | y)$$

对于连续随机变量，

$$\mathbb{E}[X | Y = y] = \int_{-\infty}^{+\infty} xf_{X|Y}(x | y)dx$$

对于随机变量的函数，我们类似有：

$$\begin{aligned} \mathbb{E}[g(X) | Y = y] &= \sum_x g(x)p_{X|Y}(x | y) && \text{离散} \\ \mathbb{E}[g(X) | Y = y] &= \int_{-\infty}^{+\infty} g(x)f_{X|Y}(x | y)dx && \text{连续} \end{aligned}$$

4.3.2 全期望定理

$$\mathbb{E}X = \sum_y \mathbb{E}[X | Y = y]p_Y(y)$$

$$\mathbb{E}X = \int_{-\infty}^{+\infty} \mathbb{E}[X | Y = y]f_Y(y)dy$$

证：仅证明离散情形。根据全概率公式有：

$$p_X(x) = \sum_y p_{X|Y}(x | y)p_Y(y)$$

两边乘上 x 并求和：

$$\begin{aligned}\mathbb{E}X &= \sum_x xp_X(x) \\ &= \sum_x x \sum_y p_{X|Y}(x | y)p_Y(y) \\ &= \sum_y p_Y(y) \sum_x xp_{X|Y}(x | y) \\ &= \sum_y \mathbb{E}[X | Y = y]p_Y(y)\end{aligned}$$

证毕。

5 极限理论

5.1 马尔可夫不等式

设随机变量 X 只取非负值, 则对任意 $a > 0$, 有:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

粗略来讲, 马尔可夫不等式指出, 一个非负随机变量如果均值很小, 那么该随机变量取大值的概率也很小。

证: 这里假设 X 是连续随机变量, 离散类似。

$$\mathbb{E}X = \int_0^{+\infty} x f_X(x) dx \geq \int_a^{+\infty} x f_X(x) dx \geq a \int_a^{+\infty} f_X(x) dx = a \cdot \mathbb{P}(X \geq a)$$

证毕。

5.2 切比雪夫不等式

设随机变量 X 的均值为 μ , 方差为 σ^2 , 则对任意 $c > 0$, 有:

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

粗略来讲, 切比雪夫不等式指出, 如果一个随机变量的方差非常小, 那么该随机变量取远离均值的概率也非常小。

证: 利用马尔可夫不等式,

$$\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}((X - \mu)^2 \geq c^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

证毕。

5.3 弱大数定律

设 X_1, \dots, X_n 独立同分布, 其公共分布均值为 μ , 则对任意 $\epsilon > 0$, 有:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X} - \mu| \geq \epsilon) = 0$$

粗略来讲, 弱大数定律表明, 独立同分布的随机变量序列的样本均值, 在大样本的情况下, 以很大的概率与随机变量的均值接近, 或称作样本均值依概率收敛到真值 μ 。

证: 这里仅对方差有界的情形进行证明, 方差无界时弱大数定律依然成立, 但是证明较为精巧。

视 $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ 为我们要研究的随机变量, 其均值和方差为:

$$\begin{aligned} \mathbb{E}\bar{X} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \mu \\ \text{var}(\bar{X}) &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{\sigma^2}{n} \end{aligned}$$

于是根据切比雪夫不等式，有： $\forall \epsilon > 0$,

$$\mathbb{P}(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \quad (n \rightarrow \infty)$$

证毕。

一般情形的弱大数定理称为辛钦大数定律，而方差有界的情形称之为切比雪夫大数定律。

更特殊的，对于 X_1, \dots, X_n 独立同分布于 $B(1, p)$ 的情形而言，我们称之为伯努利大数定律：

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X} - p| \geq \epsilon) = 0$$

5.4 中心极限定理

设 X_1, \dots, X_n 是独立同分布的随机变量序列，序列的每一项的均值为 μ ，方差为 σ^2 ，则对 $\forall x \in \mathbb{R}$ ，有：

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x \right) = \Phi(x)$$

粗略来讲，中心极限定理表明，在大样本的情况下，独立同分布的随机变量序列的样本均值的标准化结果服从标准正态分布。

标准化：设一个随机变量 X 有均值和方差，称 $\frac{X - \mathbb{E}X}{\sqrt{\text{var}(X)}}$ 为 X 的标准化，因为这个结果的均值为 0 ，方差为 1 。

样本均值的标准化即为：

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

这个一般情形的定理被称作林德伯格-莱维中心极限定理。

对于 X_1, \dots, X_n 独立同分布于 $B(1, p)$ 的特殊情形而言，我们称之为棣莫弗-拉普拉斯中心极限定理：

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x)$$

6 数理统计

6.1 正态分布的三个导出分布

6.1.1 χ^2 分布

定义: 设 X_1, X_2, \dots, X_n 为 n 个独立的服从 $N(0, 1)$ 的随机变量, 则称

$$Z = \sum_{i=1}^n X_i^2$$

的分布为自由度为 n 的 χ^2 分布, 记作 $Z \sim \chi^2(n)$.

期望与方差:

$$\begin{aligned}\mathbb{E}Z &= n \\ \text{var}Z &= 2n\end{aligned}$$

证: 由于

$$\mathbb{E}X_i^2 = \text{var}X_i + (\mathbb{E}X_i)^2 = 1 + 0 = 1$$

故

$$\mathbb{E}Z = \mathbb{E}\left[\sum_{i=1}^n X_i^2\right] = \sum_{i=1}^n \mathbb{E}X_i^2 = n$$

又由于

$$\begin{aligned}\mathbb{E}X_i^4 &= \int_{-\infty}^{+\infty} x^4 \varphi(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^3 de^{-\frac{x^2}{2}} \\ &= \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = -\frac{3}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x de^{-\frac{x^2}{2}} \\ &= \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 3\end{aligned}$$

故

$$\text{var}X_i^2 = \mathbb{E}X_i^4 - (\mathbb{E}X_i^2)^2 = 3 - 1 = 2$$

故

$$\text{var}Z = \text{var}\sum_{i=1}^n X_i^2 = \sum_{i=1}^n \text{var}X_i^2 = 2n$$

证毕。

6.1.2 t 分布

定义: 设 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, X, Y 相互独立, 则称

$$t = \frac{X}{\sqrt{\frac{Y}{n}}}$$

的分布为自由度为 n 的 t 分布, 记作 $t \sim t(n)$.

6.1.3 F 分布

定义: 设 $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$, X, Y 独立, 则称

$$Z = \frac{X/n}{Y/m}$$

的分布为自由度为 n, m 的 F 分布, 记作 $Z \sim F(n, m)$.

6.1.4 性质

设 $X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, 则:

性质 1: 样本均值服从期望相同、方差更小的正态分布:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

易证。

性质 2: 样本均值的标准化:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

易证。

性质 3: 记 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 为样本方差, 则

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

证明较为复杂, 此处略去。

性质 4: \bar{X} 与 S^2 独立, 证明略去。

性质 5:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$$

证: 首先, 由性质 2 知: $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$; 其次, 由性质 3 知: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$; 于是, 根据 t 分布的定义有:

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$$

证毕。

性质 6: 设 $X_i \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$, $i = 1, 2, \dots, n$, 样本方差为 S_1^2 ; $Y_i \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$, $i = 1, 2, \dots, m$, 样本方差为 S_2^2 , 且 X_i, Y_i 相互独立, 则:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$$

证: 由性质 3 知: $\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$, $\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$, 于是根据 F 分布的定义有:

$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)} \sim F(n-1, m-1)$$

证毕。

6.2 点估计

6.2.1 矩估计

基本思想: 对于总体矩 $\mu_k = \mathbb{E}X^k$, 直接用样本矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ 近似之, 得到:

$$\frac{1}{n} \sum_{i=1}^n X_i^k \approx \mathbb{E}X^k$$

左式包含已知的 X_i , 右式可用参数表达, 于是我们可以将未知的参数用已知的 X_i 表示出来, 这就是矩估计。

一般地, 有多少个参数, 就需要几阶矩, 这样才有足够的方程数。

6.2.2 极大似然估计

基本思想: 设 $X_i \stackrel{\text{iid}}{\sim} f(x; \vec{\theta})$, 则 (X_1, \dots, X_n) 的联合概率密度函数为 $\prod_{i=1}^n f(x_i; \vec{\theta})$. 对于一组已知的样本 (x_1, \dots, x_n) , 这是一个关于 $\vec{\theta}$ 的函数, 于是一个自然的想法是, 我们取这样的 $\vec{\theta}$, 它使得 $L(\vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\theta})$ 取到最大值。这就是极大似然估计。

于是问题转化成了一个多元微分学问题。我们知道, $L(\vec{\theta})$ 的最大值在它的驻点或边界上取到, 但直接令 $\frac{\partial L(\vec{\theta})}{\partial \theta_i} = 0$ 的计算量较大。注意到 $L(\vec{\theta})$ 和 $\ln L(\vec{\theta})$ 有相同的驻点, 于是我们令:

$$\frac{\partial \ln L(\vec{\theta})}{\partial \theta_i} = 0$$

就可以较为轻松的找到驻点, 随后在这些驻点以及边界上寻找最大值即可。

6.2.3 无偏性

定义： 对于我们的参数估计量 $\hat{\theta}$ ，若对 $\forall \theta$ 都有 $\mathbb{E}\hat{\theta} = \theta$ ，则称 $\hat{\theta}$ 是无偏的。

典型例子： 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是总体方差 σ^2 的无偏估计。

证：首先注意到： $\mathbb{E}\bar{X} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \mu$ ， $\text{var}\bar{X} = \frac{\sigma^2}{n}$ ，于是：

$$\begin{aligned}\mathbb{E}S^2 &= \mathbb{E} \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right] \\ &= \frac{1}{n-1} \mathbb{E} \left[\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \mathbb{E}X_i^2 - n\mathbb{E}\bar{X}^2 \right] \\ &= \frac{1}{n-1} \left[n\sigma^2 + n\mu^2 - n \cdot \frac{\sigma^2}{n} - n\mu^2 \right] \\ &= \sigma^2\end{aligned}$$

证毕。