

Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫

Note by THF

2024 年 10 月 30 日

目录

0.1	协方差的性质	3
0.2	相关系数	5
0.2.1	标准化	5
0.2.2	性质	5
1	大数定律和中心极限定理	6
1.1	大数定律	7
1.2	中心极限定理	8

Lecture 11

10.24

Review:

Notation. 数学期望的性质:

1. $E(c) = c$
2. $E(cX) = cE(X)$
3. $E(X + Y) = E(X) + E(Y)$
- 3.1 $E(E(Y)X) = E(Y)E(X)$
4. X, Y 相互独立, $E(XY) = E(X)E(Y)$

协方差: $\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - E(X)E(Y)$

若 X, Y 独立则 $\text{cov}(X, Y) = 0$

Notation. 方差的性质:

1. $D(c) = 0$

2. $D(cX) = c^2 D(X)$

- 2.1. $D(X) = E(X - EX)^2 = E(X^2) - E(X)^2$

3. X, Y 相互独立, $D(X + Y) = D(X) + D(Y)$

$$\text{cov}(X, Y) = E(X - EX)(Y - EY)$$

当 $X = Y$, $\text{cov}(X, Y) = \text{cov}(X, X) = E(X - EX)^2 = D(X)$

或: $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^2) - E(X)^2$

Example. $D(aX + bY + c) = D(aX + bY)$

$$\begin{aligned} D(aX + bY) &= E((aX + bY) - E(aX + bY))^2 \\ &= E(a(X - EX) + b(Y - EY))^2 \\ &= E(a^2(X - EX)^2 + 2ab(X - EX)(Y - EY) + b^2(Y - EY)^2) \\ &= a^2 D(X) + b^2 D(Y) + 2ab \text{cov}(X, Y). \end{aligned}$$

◦ 切比雪夫不等式: 已知一个随机变量的方差可以估算出数学期望

Question. 一个随机变量 X 分布未知, 已知 $\mu = 18, \sigma = 2.5$, 求 $P\{X \in (8, 28)\}$

解: 由切比雪夫不等式:

$$\begin{aligned} P\{X \in (8, 28)\} &= P\{X - 18 \in (-10, 10)\} \\ &= P\{|X - 18| < 10\} \\ &= P\{|X - \mu| < \varepsilon\} \\ &\geq 1 - \frac{\sigma^2}{\varepsilon^2} \\ &= 1 - \frac{2.5^2}{10^2} = 0.9375. \end{aligned}$$

◦ 马尔可夫不等式

Example. $X_1, X_2, \dots, X_n : i.i.d, X \sim N(\mu, \sigma^2)$, 证明:

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

$$2. \text{ 设 } Y_i = \frac{X_i - \mu}{\sigma}, i = 1, 2, \dots, n \text{ 则 } E\left(\sum_{i=1}^n Y_i^2\right) = n$$

证明. 1. 由线性性:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(E\bar{X}, D\bar{X}).$$

由于 X 之间相互独立, 有 $D(X_1 + X_2) = D(X_1) + D(X_2)$

$$E\bar{X} = \frac{1}{n} \sum_{i=1}^n EX_i = \mu, \quad D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i = \frac{\sigma^2}{n}.$$

2. 由题: $EY_i = 0, DY_i = 1$

$$E\left(\sum_{i=1}^n Y_i^2\right) = \sum_{i=1}^n EY_i^2.$$

Notation. Y_i^2 符合自由度为 1 的卡方分布: $Y_i^2 \sim X^2(1)$

$$\text{即: } \sum_{i=1}^n E(Y_i^2) = nE(Y_i^2)$$

由方差的定义: $D(Y_i) = E(Y_i^2) - E(Y_i)^2$:

$$EY_i^2 = D(Y_i) + E(Y_i)^2 = 1 + 0^2 = 1$$

$$\sum_{i=1}^n E(Y_i^2) = nE(Y_i^2) = n.$$

□

0.1 协方差的性质

- $\text{cov}(X, Y) = \text{cov}(Y, X)$ (对称性)
- $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$

证明. 已知: $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned}\text{cov}(aX, bY) &= E(aXbY) - E(aX)E(bY) \\ &= abE(XY) - abE(X)E(Y) \\ &= ab\text{cov}(X, Y).\end{aligned}$$

□

$$\circ \text{cov}(c, X) = 0$$

Notation. 协方差用于衡量随机变量之间的线性关系, 常数和其他随机变量不在线性关系

证明.

$$\begin{aligned}\text{cov}(cX) &= E(cX) - E(c)E(X) \\ &= cE(X) - cE(X) \\ &= 0.\end{aligned}$$

□

Notation. $\text{cov}(c, c) = D(c) = 0$

$$\circ \text{cov}(aX + bY, cZ) = a\text{ccov}(X + Y) + b\text{ccov}(Y + Z) \quad (\text{分配律})$$

证明.

$$\begin{aligned}\text{cov}(aX + bY, cZ) &= E((aX + bY)cZ) - E(aX + bY)E(cZ) \\ &= E(acXZ + bcYZ) - cEZ(aEX + bEY) \\ &= acE(XZ) + bcE(YZ) - acEXEZ - bcEYEZ \\ &= a\text{ccov}(X, Z) + b\text{ccov}(Y, Z).\end{aligned}$$

□

$$\text{Notation. } \text{cov}\left(\sum_{i=1}^n a_i X_i, b_i Z\right) = \sum_{i=1}^n a_i b_i \text{cov}(X_i, Z)$$

$$\text{Notation. } D\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 DX_i + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{cov}(X_i, X_j)$$

0.2 相关系数

0.2.1 标准化

$$X^* = \frac{X - EX}{\sqrt{DX}}.$$

标准化后的变量 $EX^* = 0, DX^* = 1$

Definition. X^*, Y^* 的协方差 $\text{cov}(X^*, Y^*)$ 为 X, Y 的相关系数 $\rho(X, Y)$

$$\begin{aligned}\text{cov}(X^*, Y^*) &= \text{cov}\left(\frac{X - EX}{\sqrt{DX}}, \frac{Y - EY}{\sqrt{DY}}\right) \\ &= \frac{1}{\sqrt{DX}\sqrt{DY}} \text{cov}(X - EX, Y - EY).\end{aligned}$$

易得 $\text{cov}(X - EX, Y - EY) = \text{cov}(X, Y)$

$$\begin{aligned}\text{cov}(X^*, Y^*) &= \frac{\text{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} \\ &= \rho(X, Y).\end{aligned}$$

0.2.2 性质

- $|\rho(X, Y)| \leq 1$
- $P\{X^* = \pm Y^*\} = 1$ 是 $\rho(X, Y) = \pm 1$ 的充要条件

Lecture 12

10.29

Notation. 相关系数/Pearson 相关系数: 描述两个随机变量之间的线性相关性
只能描述数值性的变量

$|\rho(X, Y)| = 1$ 时: 正相关

$|\rho(X, Y)| > 0.8$: 强相关

$|\rho(X, Y)| \in (0, 0.5)$: 弱相关

$\rho = 0$: 不相关/非线性关系

Notation. 相关系数本质上描述:

$$P\{Y = aX + b\}.$$

Example. $f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & x^2 + y^2 > 1 \end{cases}$, 求:

1. X, Y 的相关性; 2. X, Y 的独立性

解: 1.

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y) dx dy \\ &= \int_{-1}^1 dx \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{x}{\pi} dy \\ &= 0. \end{aligned}$$

同理 $EY = 0$, 即不相关

2.

$$\begin{aligned} f_X(x) &= \int_D f(x, y) dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy \\ &= \frac{2}{\pi} \sqrt{1-x^2}. \end{aligned}$$

同理 $f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2}$, 易得 $f(x, y) \neq f_X(x) f_Y(y)$, 即不独立

数理统计部分

1 大数定律和中心极限定理

Definition. 大数定律:

$$\bar{X} \xrightarrow[P]{n \rightarrow +\infty} EX.$$

即: 以某事件发生的频率估计该事件的概率

Definition. 中心极限定理:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

其中 X_1, X_2, \dots, X_i 独立同分布

该随机变量序列存在分布, 中心极限定理提出不论 \bar{X} 的分布是什么, 该序列的分布为正态分布

$$\bar{X} \xrightarrow[n \rightarrow \infty]{L} N(E\bar{X}, D\bar{X}).$$

如何判断随机变量的敛散性:

Corollary. 依概率收敛:

对 $\forall \varepsilon$ 有:

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1.$$

代表序列 $\{X_n\}$ 收敛于随机变量 X , 记为 $X_n \xrightarrow[n \rightarrow \infty]{P} X$

Corollary. 依分布收敛:

序列的分布函数为 $F_n(x)$, 随机变量的分布函数 $F(x)$, 对 $\forall x$, 有:

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

则 $\{X_n\}$ 依分布收敛于 X , 记为 $X_n \xrightarrow[n \rightarrow \infty]{L} X$

Notation. 测度变换: 通过将问题映射到另一个空间简化计算

依分布收敛要求更弱, 即: 依概率收敛 \Rightarrow 依分布收敛

当收敛对象为常数时二者可互推

Notation. 撞骗: 只要发出的短信足够多, 成功率符合大数定律

三大大数定律:

$$\left\{ \begin{array}{l} \text{切比雪夫大数定律: 最根本} \\ \text{伯努利大数定律: 例子} \\ \text{辛钦大数定律} \end{array} \right. .$$

1.1 大数定律

Notation. 切比雪夫大数定律:

Definition. $\{X_i\}$ i.i.d, $\exists EX_i, DX_i$, 且 $\exists C$, 使得 $DX_i \leq C$ (方差有界), 则对 $\forall \varepsilon > 0$ 当:

$$\lim_{n \rightarrow \infty} P\{|\bar{X}_n - E\bar{X}_n| < \varepsilon\} = 1.$$

时:

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{P} E\bar{X}_n.$$

证明. $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, 有:

$$\begin{aligned} E\bar{X}_n &= \frac{1}{n} \sum_{i=1}^n EX_i \\ D\bar{X}_n &= \frac{1}{n^2} \sum_{i=1}^n DX_i \\ &\leq \frac{C}{n}. \end{aligned}$$

由切比雪夫不等式:

$$\begin{aligned} P\{|\bar{X}_n - E\bar{X}_n| < \varepsilon\} &\geq 1 - \frac{D\bar{X}_n}{\varepsilon^2} \\ &\geq 1 - \frac{C}{n\varepsilon^2}. \end{aligned}$$

当 $n \rightarrow \infty$ 时原式收敛于 1

□

Notation. 辛钦大数定律: 序列中的随机变量独立同分布

Notation. 伯努利大数定律: 序列中 $X_i \sim B(1, p)$ (已知分布), 记 μ_s 为随机变量序列之和, 有:

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{\mu_s}{n} - p\right| < \varepsilon\right\} = 1.$$

即: $\frac{\mu_s}{n}$ 依概率收敛于 p

1.2 中心极限定理

Example. 高尔顿钉板

Corollary. *i.i.d* 的中心极限定理:

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right\} = \Phi(x).$$

Corollary. 拉普拉斯中心极限定理: X_i 独立同分布, $X_i \sim B(1, p)$, 令 $Y = \sum_{i=1}^n X_i$, 对 $\forall x$ 有:

$$\lim_{n \rightarrow \infty} P \left\{ \frac{Y - np}{\sqrt{np(1-p)}} \leq x \right\} = \Phi(x).$$