Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫
Note by THF

2024 年 11 月 1 日

目录

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. $f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \le 1 \\ 0, & x^2+y^2 > 1 \end{cases}$，求：
1. X,Y 的相关性；2. X,Y 的独立性

解：1.

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

同理 $EY = 0$，即不相关

2.

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

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：

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

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 $X_1, X_2, \ldots X_i$ 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

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

如何判断随机变量的敛散性：

Corollary. 依概率收敛：

对 $\forall \varepsilon$ 有：

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

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

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:

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

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

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:

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

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 $DX_i \le C$ （方差有界），则对 $\forall \varepsilon > 0$ 当:

$$\lim_{n\to\infty} P\left\{\left|\bar{X_n} - E\bar{X_n}\right| < \varepsilon\right\} = 1.$$

时:

$$\bar{X_n} \xrightarrow[n\to+\infty]{P} E\bar{X_n}.$$

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

$$\begin{aligned} E\bar{X_n} &= \frac{1}{n}\sum_{i=1}^{n} EX_i \\ D\bar{X_i} &= \frac{1}{n^2}\sum_{i=1}^{n} DX_i \\ &\le \frac{C}{n}. \end{aligned}$$

由切比雪夫不等式:

$$\begin{aligned} P\left\{\left|\bar{X_n} - E\bar{X_n}\right| < \varepsilon\right\} &\ge 1 - \frac{D\bar{X_n}}{\varepsilon^2} \\ &\ge 1 - \frac{C}{n\varepsilon^2}. \end{aligned}$$

当 $n\to\infty$ 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：

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

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：

$$\lim_{n\to\infty} P\left\{\frac{\bar{X_n} - \mu}{\sigma/\sqrt{n}} \le x\right\} = \Phi(x).$$

Corollary. 棣莫弗-拉普拉斯定理：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：

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

Lecture 13

10.31

Notation. 偏度 r_1：三阶标准化随机变量的矩，用于描述对称性

峰度 r_2：四阶标准化随机变量的矩，一般使正态分布的峰度 $r_2 = 0$，描述分布的陡峭程度

表 1: 常见分布的数字特征

分布	EX	DX	r_1	r_2
$B(1,p)$	p	$p(1-p)$	$\frac{1-2p}{\sqrt{p(1-p)}}$	$\frac{1}{p(1-p)-6}$
$B(n,p)$	np	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$	$\frac{1-6p(1-p)}{np(1-p)}$
$P(\lambda)$	λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$
$G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$	$6 + \frac{p^2}{1-p}$
$U[a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0	$\frac{9}{5} - 3$
$\Gamma(1,\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	6
$N(\mu,\sigma^2)$	μ	σ^2	0	0

2 数理统计基本概念

随机变量引入：使样本空间映射到实数轴上

分布函数：任意随机变量的概率

大数定律和中心极限定理：由概率论过渡到数理统计

$$\begin{cases} \text{描述统计学：过去的实验数据/相关分析图} \\ \text{推断统计学：根据现有的实验数据决策} \begin{cases} \text{参数估计：第七章} \\ \text{假设检验：第八章} \\ \text{回归分析：第九章} \end{cases} \end{cases}.$$

Definition. 总体：全部研究对象，可以用分布描述（随机变量组）

Definition. 个体：组成总体的成员，符合总体分布（每一个个体都是一个随机变量）

Example. 从总体中抽取 n 个样本

对数据记录：$x_1, x_2, \ldots, x_n$ 称为 n 维随机变量 $X_1, X_2, \ldots, X_n$ 对应的观测值，$X_1, X_2, \ldots, X_n$ 为来自总体 X 的一个样本

Notation. 简单样本：$X_1, X_2, \ldots, X_n$ $i.i.d$，且与总体分布相符

特点：

◦ 独立性

◦ 代表性

Definition. 样本空间：$\boldsymbol{\Omega} = \{(x_1, x_2, \ldots, x_n) \,|x_i \in \mathbb{R}, \mathrm{i} = 1, 2, \ldots, n\}$

Notation. 样本联合分布和总体分布的关系 $(i.i.d)$：

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\} \\ &= \prod_{i=1}^{n} P\{X_i \le x_i\} \\ &= \prod_{i=1}^{n} F(x_i). \end{aligned}$$

扩展：X 为连续型，密度函数的关系：

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} f_{X_i}(x_i) \\ &= \prod_{i=1}^{n} f(x_i) \quad x_i \in \mathbb{R}, i = 1, 2, \ldots, n. \end{aligned}$$

2.1 经验分布函数

经验分布函数：$F_n(x)$

将样本观测值 $x_1, x_2, \ldots, x_n$ 按大小分类为 $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

$$\begin{aligned} F_n(x) &= f_n\{X \le x\} \\ &= \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x \in [x_{(k)}, x_{(k+1)}) \\ 1, & x \ge x_{(n)} \end{cases} \\ &\approx F(X). \end{aligned}$$

Corollary. 格利文科定理:

$$P\left\{\limsup_{n\to\infty, x\in\mathbb{R}} |F(x) - F_n(x) = 0|\right\} = 1.$$

根据格利文科定理：可以使用经验分布函数来估计理论分布函数