Part III-B: Statistic

Lecture by SongHao
Note by THF

2024 年 11 月 17 日

目录

Learn 5

11.15

0.1 总体与样本

Definition. 总体：全体，分作有限总体和无限总体

令总体为 X，总体的分布记为 X 的分布

在总体中抽样：样本

Definition. 抽样：在总体中抽一部分样本

假设抽了 n 次：得到样本 $X_1, X_2, \dots, X_n$ ，如果测出了这些样本的值，记为 $x_1, x_2, \dots, x_n$ ，即**观测值**

Notation. 抽样的方法：简单随机抽样，满足以下条件：

1. 同分布
2. 相互独立

Learn 6

11.16

1 大数定律

Corollary. 切比雪夫不等式：随机变量中任意 X 离期望的距离比 ϵ 大的概率比 $\frac{DX}{\epsilon^2}$ 小

证明. 假设 X 为连续型，由概率定义：

$$P\left(|X-EX|\geq\epsilon\right)=\int_{|x-EX|\geq\epsilon}f\left(x\right)\mathrm{d}x.$$

这里因为 $|X-EX|\geq\epsilon$ ，同时平方：$\left(X-EX\right)^2/\epsilon^2\geq1$ ，即原式可以转为：

$$P\left(|X-EX|\geq\epsilon\right)=\int_{|x-EX|\geq\epsilon}1\cdot f\left(x\right)\mathrm{d}x\leq\int_{|x-EX|\geq\epsilon}\frac{\left(x-EX\right)^2}{\epsilon^2}f\left(x\right)\mathrm{d}x.$$

这个积分有积分域，大小必然比从 $+\infty$ 到 $-\infty$ 更小（或等于）：因为密度函数 $f\left(x\right)\geq0$，因此不需要考虑被积函数正负问题，即：

$$P\left(|X-EX|\geq\epsilon\right)\leq\int_{-\infty}^{+\infty}\frac{\left(x-EX\right)^2}{\epsilon^2}f\left(x\right)\mathrm{d}x=\frac{\int_{-\infty}^{+\infty}\left(x-EX\right)^2f\left(x\right)\mathrm{d}x}{\epsilon^2}=\frac{DX}{\epsilon^2}.$$

□

该结论等价于：

1. $P\left(|X-EX|<\epsilon\right)\geq1-\frac{DX}{\epsilon^2}$
2. $P\left(|X-EX|\leq\epsilon\right)\geq1-\frac{DX}{\epsilon^2}$

当 DX 越小时 P 越大，即方差越小，X 落在外面的概率越小

Example. 一个电网有 10000 盏灯，每一盏灯打开的概率为 0.7，求一个晚上开着的灯数量在 6800 到 7200 盏中的概率

解：期望为 7000：$EX=7000$ ，即 $X\sim B\left(10000,0.7\right)$ ，由于二项分布 $EX=7000,DX=np\left(1-p\right)=10000\times0.7\times\left(1-0.7\right)=2100$

要求：$P\left\{X\in[6800,7200]\right\}$ ：转换后可得：$P\left\{X\in[6800,7200]\right\}=P\left\{|X-7000|\leq200\right\}$

由切比雪夫不等式：

$$P\left(|X-7000|\leq200\right)\geq1-\frac{2100}{200^2}\approx0.9475.$$

Example. 白细胞平均数为 7300，标准差为 700，求白细胞数在 5200 到 9400 的概率

解：$EX=7300,DX=700^2=490000$ ，由切比雪夫不等式：

$$P\left(5200\leq X\leq9400\right)=P\left(|X-7300|<2100\right)\geq1-\frac{DX}{2100^2}.$$

Notation. 用切比雪夫不等式做题：两边取到的距离应该相等

Example. 3σ 原则：求 $P\left(|X-\mu|\geq3\sigma\right)$

解：由切比雪夫不等式：

$$P\left(|X-\mu| \leq 3\sigma\right) \geq 1-\frac{\sigma^2}{(3\sigma)^2} \quad P\left(|X-\mu| \geq 3\sigma\right) \leq \frac{\sigma^2}{(3\sigma)^2}=\frac{1}{9}.$$

Notation. 当 $X \sim N\left(\mu, \sigma^2\right)$ 时，这个概率约等于 0.0027

1.1 切比雪夫大数定律

Definition. 依概率收敛：有一个随机变量序列，越到 n 大时 X_n 能取得某数 a 的概率越来越接近 1，即 X_n 的取值密集在 a 的附近，记为：

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