数模笔记

Noflowerzzk

目　录

第一部分 概率论

第一章 随机变量

§ 1.1 随机变量的数值性质

定义 1.1.1

- 协方差 $\operatorname{cov}(X,Y)=E((X-E(X))(Y-E(Y)))$
- 相关系数/标准化协方差 $\rho(X,Y)=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$
- 变异系数 $\delta_X=\dfrac{\sqrt{D(X)}}{|E(X)|}$
- k 阶原点矩 $E(X^k)$
- k 阶中心矩 $E\left((X-E(X))^k\right)$

§ 1.2 大数定律

定理 1.2.1 (Chebyshev 不等式)

$$P(|X-E(X)|\geqslant\varepsilon)\leqslant\frac{D(X)}{\varepsilon^2}$$

定义 1.2.1 (依概率收敛) $\exists c,\forall\varepsilon>0,\ \lim\limits_{n\to+\infty}P\left(|X_n-c|\leqslant\varepsilon\right)=1$, 则称随机变量序列 $\{X_n\}$ 依概率收敛于 c, 记作 $X_n\xrightarrow{P}c$

定理 1.2.2 (Chebyshev 大数定律) 随机变量序列 $\{X_n\}$ 两两不 (线性) 相关，且 $D(X_i)$ 有一致上界 c (即 $D(X_i)<c$), 则有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\frac{1}{n}\sum_{i=1}^{n}E(X_i)$$

定理 1.2.3 (相互独立同分布 (辛钦) 大数定律)　$\{X_i\}$ 相互独立同分布，$E(X_i)=\mu$ 有限，则

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu$$

§ 1.3　中心极限定理

定理 1.3.1 (列维-林德伯格中心极限定理)　$\{X_i\}$ 相互独立同分布，$D(X_i)=\sigma^2$ 有限，$E(X_i)=\mu$，则

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

即当 n 充分大时，可以认为 $\sum\limits_{i=1}^{n}X_i\overset{\text{近似}}{\sim}N(n\mu,n\sigma^2)$ 或者 $\overline{X}\overset{\text{近似}}{\sim}N\left(\mu,\frac{\sigma^2}{n}\right)$

第二部分　数理统计

第二章　统计量

§ 2.1　无偏估计

一、样本方差

定义 2.1.1　$(X_1,X_2,\cdots,X_n)$ 是取自总体的一个样本，称

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

为样本均值，

$$S^2=\frac{1}{n-1}\sum_{i=1}^{m}\left(X_i-\overline{X}\right)^2$$

为样本方差.

§ 2.2 三大分布

一、χ^2 分布

定义 2.2.1 设 $\{X_i\}_{i=1}^n$ 为相互独立的标准正态分布随机变量，称随机变量 $Y = \sum_{i=1}^{n} X_i^2$ 服从自由度为 n 的 χ^2 分布，记为 $Y \sim \chi^2(n)$.

χ^2 分布的密度函数为

$$f(y) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{y}{2}}, & y > 0 \\ 0, & \text{其它} \end{cases}$$

定理 2.2.1 $Y \sim \chi^2(n)$ 有以下性质

- $E(Y) = n, D(Y) = 2n$
- 可加性，$X \sim \chi^2(m), Y \sim \chi^2(n), X, Y$ 相互独立，则 $X + Y \sim \chi^2(m+n)$

二、t 分布 (学生氏分布)

定义 2.2.2 设 X, Y 相互独立，$X \sim N(0,1), Y \sim \chi^2(n)$, 则称 $T = \dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布.

t 分布的密度函数为

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$t(n)$ 的密度函数与标准正态分布 $N(0,1)$ 密度很相似, 它们都是关于原点对称, 单峰偶函数, 在 $x = 0$ 处达到极大. 但 $t(n)$ 的峰值低于 $N(0,1)$ 的峰值, $t(n)$ 的密度函数尾部都要比 $N(0,1)$ 的两侧尾部粗一些. 容易证明:

$$\lim_{n\to\infty} f(x) = \Phi(x)$$

三、F 分布

定义 2.2.3 设 X,Y 相互独立，$X\sim\chi^2(m),Y\sim\chi^2(n)$, 则称 $F=\dfrac{X/m}{Y/n}$ 服从 F 分布，记为 $F\sim F(m,n)$ 其中 m 称为第一自由度，n 称为第二自由度.
$F(m,n)$ 分布的概率密度函数为

$$f(y)=\begin{cases}\dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\dfrac{m}{n}\right)^{\frac{m}{2}}y^{\frac{m}{2}-1}\left(1+\dfrac{m}{n}y\right)^{-\frac{m+n}{2}}, & y>0\\ 0, & \text{其它}\end{cases}$$

定理 2.2.2 记 $F_\alpha(m,n)$ 为 F 分布的第 α 分位数 (即 $P(F\leqslant F_\alpha(m,n))=\alpha$) 有性质：

$$F_\alpha(m,n)=\frac{1}{F_{1-\alpha}}(n,m)$$

§ 2.3 正态总体的抽样分布

暂时略.

第三章 参数估计

§ 3.1 点估计

一、矩估计

用样本原点矩估计总体原点矩.
设总体的 k 阶原点矩为 $\mu_k=E(X^k)$, 样本的 k 阶原点矩为 $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$, 用 A_k 估计 μ_k, 对某个依赖 $\mu_1,\mu_2,\cdots,\mu_n$ 的分布参数 $\theta=\theta(\mu_1,\mu_2,\cdots,\mu_n)$, 有 θ 的估计

$$\hat{\theta}=\theta(A_1,A_2,\cdots,A_n)$$

二、极大似然估计

定义设总体 X 有分布律 $P(X=x;\theta)$ 或密度函数 $f(x;\theta)$（其中 θ 为一个未知参数或几个未知参数组成的向量 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$），已知 $\theta\in\Theta$，Θ 是参数空间. $(x_1,x_2,\cdots,x_n)$ 为取自总体 X 的一个样本 $(X_1,X_2,\cdots,X_n)$ 的观测值，将样本的联合分布律或联合密度函数看成 θ 的函数，用 $L(\theta)$ 表示，又称为 θ 的似然函数，则似然函数

$$L(\theta)=\prod_{i=1}^{n}P\left(X_i=x_i;\theta\right),\ \text{或}L(\theta)=\prod_{i=1}^{n}f\left(x_i;\theta\right),$$

称满足关系式 $L(\hat{\theta})=\max\limits_{\theta\in\Theta}L(\theta)$ 的解 $\hat{\theta}$ 为 θ 的极大似然估计量.

§ 3.2 点估计的优良性判断标准

一、无偏性

定义 3.2.1　设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的一个估计量，θ 取值的参数空间为 Θ，若对任意的 $\theta\in\Theta$，有

$$E_\theta\left(\hat{\theta}(X_1,X_2,\cdots,X_n)\right)=\theta,$$

则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的一个无偏估计（量），否则称为有偏估计（量）. 如果有

$$\lim_{n\to\infty}E_\theta\left(\hat{\theta}(X_1,X_2,\cdots,X_n)\right)=\theta,$$

则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是 θ 的一个渐近无偏估计（量）.

估计量的无偏性是指，由估计量得到的估计值相对于未知参数真值来说，取某些样本观测值时偏大，取另一些样本观测值时偏小。反复将这个估计量使用多次，就平均来说其偏差为 0。如果估计量不具有无偏性，则无论使用多少次，其平均值也与真值有一定的距离，这个距离就是系统误差了。

二、有效性

定义 3.2.2　设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是 θ 的两个无偏估计，若对任意的 $\theta \in \Theta$，有 $D(\hat{\theta}_1) \leqslant D(\hat{\theta}_2)$，且至少有一个 $\theta \in \Theta$ 使得上述不等式严格成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

三、相合性 (一致性)

定义 3.2.3　设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量，若对 $\forall \varepsilon > 0$，

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| \geqslant \varepsilon) = 0,$$

则称估计量 $\hat{\theta}$ 具有相合性（一致性），即 $\hat{\theta} \xrightarrow{P} \theta$，或称 $\hat{\theta}$ 是 θ 的相合（一致）估计量.

相合性被视为对估计的一个很基本的要求，如果一个估计量，在样本量不断增大时，它不能把被估参数估计到任意指定的精度内，那么这个估计是不好的. 通常，不满足相合性的估计一般不予考虑.

§ 3.3 区间估计

定义 3.3.1　设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的一个样本，总体 $X \sim f(x;\theta), \theta \in \Theta$ 未知，对于 $\forall 0 < \alpha < 1$，若统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n) < \overline{\theta}(X_1, X_2, \cdots, X_n) = \overline{\theta}$，使得

$$P(\underline{\theta} \leqslant \theta \leqslant \overline{\theta}) = 1 - \alpha, \theta \in \Theta,$$

则称 $[\underline{\theta}, \overline{\theta}]$ 为 θ 的双侧 $1-\alpha$ 置信区间，$\underline{\theta}, \overline{\theta}$ 分别称为 θ 的双侧 $1-\alpha$ 置信区间的置信下限和置信上限，$1-\alpha$ 为置信水平，一旦样本有观测值 $(x_1, x_2, \cdots, x_n)$，则称相应的 $[\underline{\theta}(x_1, x_2, \cdots, x_n), \overline{\theta}(x_1, x_2, \cdots, x_n)]$ 为置信区间的观测值。

第三部分　数学建模