Ch 8 统计的基本概念

回顾前一次课

- 林德贝格-勒维中心极限定理：独立同分布随机变量，若$E[X_k]=\mu$和$\mathrm{Var}(X_k)=\sigma^2$，则$\sum_{k=1}^{n} X_k \xrightarrow{d} N(n\mu, n\sigma^2)$
- 棣莫弗-拉普拉斯中心极限定理：若$X_n \sim B(n,p)$，则$X_n \xrightarrow{d} N\big(np, np(1-p)\big)$
- 李雅普诺夫定理：独立不同分布中心极限定理

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n} E\left[|X_k-\mu_k|^{2+\delta}\right] \to 0$$

回顾前一次课

总体: 研究对象的全体，用随机变量X表示(分布未知)

样本：从总体中随机抽取一些个体，表示为$X_1, X_2, \cdots, X_n$，称$X_1, X_2, \cdots, X_n$为取自总体X的随机样本，其样本容量为n

抽样、**样本值**、**样本的二重性**、**简单样本**

样本的分布：$F(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n)$

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

统计量

设$X_1, X_2, \cdots, X_n$来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$是关于$X_1, X_2, \cdots, X_n$的一个连续、且不含任意参数的函数, 称$g(X_1, X_2, \cdots, X_n)$是一个**统计量**

- $g(X_1, X_2, \cdots, X_n)$是随机变量: $X_1, X_2, \cdots, X_n$是随机变量
- $g(x_1, x_2, \cdots, x_n)$ 为$g(X_1, X_2, \cdots, X_n)$的一次观察值

样本均值

设$X_1, X_2, \cdots, X_n$是总体X的一个样本, 定义**样本均值**为

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

引理：总体X的期望为$E[X] = \mu$, 方差$\mathrm{Var}(X) = \sigma^2$, 有

$$E[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

样本方差

设$X_1, X_2, \cdots, X_n$是总体X的一个样本, 定义**样本方差**为

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - \bar{X}^2$$

定义**样本标准差**为 $S_0 = \sqrt{S_0^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

引理： 总体X的期望为$E[X] = \mu$, 方差$\mathrm{Var}(X) = \sigma^2$, 有

$$E[S_0^2] = \frac{n-1}{n}\sigma^2$$

修正后的样本方差

设$X_1, X_2, \cdots, X_n$是总体X的一个样本，定义**修正后的样本方差**为

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

引理： 总体X的期望为$E[X] = \mu$, 方差$\mathrm{Var}(X) = \sigma^2$, 有

$$E[S^2] = \sigma^2$$

原点矩和中心矩

假设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本, 定义样本k阶原点矩为:

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad k = 1,2,\cdots.$$

定义样本k阶中心矩为:

$$B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k \qquad k = 1,2,\cdots$$

例题

设总体$X \sim N(20,3)$，从总体中抽取两独立样本，容量分别为10和15．求这两个样本均值之差的绝对值大于0.3的概率

次序统计量

假设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本，定义最小次序统计量和最大次序统计量分别为:

$$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}$$

$$X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$$

以及定义样本极差: $R_n = X_{(n)} - X_{(1)}$.

设总体X的分布函数为$F(x)$, 则有

$$F_{X_{(1)}}(x) = P\left(X_{(1)} \leq x\right) = 1 - \left(1 - F(x)\right)^n$$

$$F_{X_{(n)}}(x) = F^n(x)$$

k次序统计量

设$X_1, X_2, \cdots, X_n$是总体X的一个样本，总体的密度函数为$f(x)$，分布函数为$F(x)$，则第k次序统计量$X_{(k)}$的分布函数和密度函数分别为

$$F_k(x) = \sum_{r=k}^{n} \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}$$

$$f_k(x) = \frac{n!}{(k-1)!\,(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

Beta-函数

对任意给定$\alpha_1 > 0$和$\alpha_2 > 0$, 定义Beta函数为

$$\mathrm{Beta}\left(\alpha_1, \alpha_2\right) = \int_0^1 x^{\alpha_1 - 1} \left(1 - x\right)^{\alpha_2 - 1} dx$$

有些书简记为$\mathrm{B}(\alpha_1, \alpha_2)$, 被称为第一类欧拉积分函数.

定理： $\mathrm{Beta}\left(\alpha_1, \alpha_2\right)$在定义域$(0, +\infty) \times (0, +\infty)$连续.

利用变量替换$t = 1 - x$有

$$\mathrm{Beta}(\alpha_1, \alpha_2) = \mathrm{Beta}(\alpha_2, \alpha_1)$$

Γ-函数

对任意给定$\alpha > 0$, 定义Γ-函数为

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

又被称为第二类欧拉积分函数.

定理：对Γ-函数有$\Gamma(1) = 1$和$\Gamma(1/2) = \sqrt{\pi}$, 对$\alpha > 1$有

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

定理：对任意给定$\alpha_1 > 0$和$\alpha_2 > 0$有

$$\mathrm{Beta}(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

第一与第二类欧拉积分函数的关系

推论：对任意$\alpha_1 > 1$和$\alpha_2 > 0$有

$$\mathrm{Beta}(\alpha_1, \alpha_2) = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 1}\mathrm{Beta}(\alpha_1 - 1, \alpha_2)$$

对任意$\alpha_1, \alpha_2, \cdots, \alpha_k > 0$, 定义多维Beta函数为

$$\mathrm{Beta}(\alpha_1, \alpha_2, \cdots, \alpha_k) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)}{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}$$

Beta分布

给定$\alpha_1 > 0$和$\alpha_2 > 0$, 若随机变量X的概率密度为

$$f(x) = \begin{cases} \dfrac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}}{\mathrm{B}(\alpha_1, \alpha_2)} & x \in (0,1) \\ 0 & \text{其它} \end{cases}$$

称X服从参数为α_1和α_2的Beta分布，记$X \sim B(\alpha_1, \alpha_2)$

定理：若随机变量$X \sim B(\alpha_1, \alpha_2)$, 则有

$$E[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2} \qquad \mathrm{Var}(X) = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}$$

例题

独立同分布随机变量$X_1, X_2, \cdots, X_n$服从均匀分布$U(0,1)$，记$X_{(k)}$为其顺序统计量，则

$$X_{(k)} \sim B(k, n-k+1).$$