若总体 X 的概率密度为 $f(x)$, 则样本 $X_1, X_2, \cdots, X_n$ 的联合概率密度为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i).$$

若总体 X 的分布列 $\Pr(X = x_i)$, 则样本 $X_1, X_2, \cdots, X_n$ 的联合分布列为

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

8.2 常用统计量

为研究样本的特性, 我们引入统计量:

定义 8.2 设 $X_1, X_2, \cdots, X_n$ 是来自总体 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**.

由于 $X_1, X_2, \cdots, X_n$ 是随机变量, 因此统计量 $g(X_1, X_2, \cdots, X_n)$ 是一个随机变量. 而 $g(x_1, x_2, \cdots, x_n)$ 为 $g(X_1, X_2, \cdots, X_n)$ 的一次观察值. 下面研究一些常用统计量.

假设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 定义 **样本均值** 为

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

根据样本的独立同分布性质有

引理 8.1 设总体 X 的期望为 $E[X] = \mu$, 方差 $\mathrm{Var}(X) = \sigma^2$, 则有

$$E[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \sigma^2/n, \qquad \bar{X} \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n).$$

假设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 定义 **样本方差** 为

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

引理 8.2 设总体 X 的期望为 $E[X] = \mu$, 方差 $\mathrm{Var}(X) = \sigma^2$, 则有

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

证明 根据 $E[X_i^2] = \sigma^2 + \mu^2$ 有

$$E(\bar{X}^2) = E\left[\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = \frac{1}{n^2} E\left[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right] = \frac{\sigma^2}{n} + \mu^2,$$

于是有

$$E(S_0^2) = E(X_i^2) - E(\bar{X}^2) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2.$$

由此可知样本方差 S_0^2 与总体方差 σ^2 之间存在偏差.

进一步定义 **样本标准差** 为:

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

定义 **修正后的样本方差** 为:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \text{即} \qquad S^2 = \frac{n}{n-1}S_0^2,$$

引理 8.3 设总体 X 的期望为 $E[X] = \mu$, 方差 $\mathrm{Var}(X) = \sigma^2$, 则有

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

证明 根据期望的性质有

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

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 8.1 设总体 $X \sim \mathcal{N}(20, 3)$, 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.

解 设 $X_1, X_2, \ldots, X_{10}$ 和 $X_1', X_2', \ldots, X_{15}'$ 分别为来自总体 $X \sim \mathcal{N}(20, 3)$ 的两个独立样本. 根据正态分布的性质有

$$\bar{X}_1 = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}(20, 3/10), \qquad \bar{X}_2 = \frac{1}{15}\sum_{i=1}^{15} X_i' \sim \mathcal{N}(20, 1/5).$$

进一步根据正态分布的性质有 $\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}(0, 1/2)$, 于是可得

$$\Pr(|\bar{X}_1 - \bar{X}_2| > 0.3) = 2 - 2\Phi(0.3/\sqrt{1/2}).$$

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:

$$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\} \qquad \text{和} \qquad X_{(n)} = \max\{X_1, X_2, \cdots, X_n\},$$

以及定义 **样本极差** 为

$$R_n = X_{(n)} - X_{(1)}.$$

设总体 X 的分布函数为 $F(x)$, 则有

$$F_{X_{(1)}}(x) = \Pr(X_{(1)} \leqslant x) = 1 - \Pr(X_{(1)} > x) = 1 - (1 - F(x))^n, \qquad F_{X_{(n)}}(x) = F^n(x).$$

定理 8.1 设总体 X 的密度函数为 $f(x)$, 分布函数为 $F(x)$, $X_1, X_2, \cdots, X_n$是来自总体 X 的一个样本, 则第 k 次序统计量 $X_{(k)}$ 的分布函数和密度函数分别为

$$\begin{aligned} F_k(x) &= \sum_{r=k}^{n} \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r} \\ f_k(x) &= \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x). \end{aligned}$$

证明 根据题意有第 k 次序统计量 $X_{(k)}$ 的分布函数为

$$\begin{aligned} F_k(x) &= \Pr[X_{(k)} \leqslant x] = \Pr[X_1, X_2, \cdots, X_n \text{ 中至少有 } k \text{ 个随机变量 } \leqslant x] \\ &= \sum_{r=k}^{n} \Pr[X_1, X_2, \cdots, X_n \text{ 中恰有 } r \text{ 个随机变量 } \leqslant x,\ n-r \text{ 个随机变量 } > x] \\ &= \sum_{r=k}^{n} \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}. \end{aligned}$$

利用恒等式

$$\sum_{r=k}^{n} \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt \quad (r \in [n],\ p \in [0, 1])$$

由此可知

$$F_k(x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1} (1-t)^{n-k} dt,$$

根据积分函数求导完成证明.

8.3 Beta 分布、Γ 分布、Dirichlet 分布

首先介绍两积分函数.

定义 8.3 (Beta-函数) 对任意给定 $\alpha_1 > 0$ 和 $\alpha_2 > 0$, 定义 Beta 函数为

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

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

根据数学分析可知 $\mathrm{Beta}(\alpha_1, \alpha_2)$ 在定义域 $(0, +\infty) \times (0, +\infty)$ 连续. 利用变量替换 $t = 1 - x$,根据定义有

$$\begin{aligned}
\mathrm{Beta}(\alpha_1, \alpha_2) &= \int_0^1 t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1} dt = \int_1^0 (1-x)^{\alpha_1 - 1} x^{\alpha_2 - 1} d(1-x) \\
&= \int_0^1 x^{\alpha_2 - 1}(1-x)^{\alpha_1 - 1} dx = \mathrm{Beta}(\alpha_2, \alpha_1),
\end{aligned}$$

由此可知 Beta 函数的对称性: $\mathrm{Beta}(\alpha_1, \alpha_2) = \mathrm{Beta}(\alpha_2, \alpha_1)$.

定义 8.4 (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.

性质 8.1 对 Γ-函数, 有 $\Gamma(1) = 1$ 和 $\Gamma(1/2) = \sqrt{\pi}$, 以及对 $\alpha > 1$ 有 $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

证明 根据定义有

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

利用变量替换 $x = t^{1/2}$ 有

$$\Gamma(1/2) = \int_0^{+\infty} t^{-\frac{1}{2}} e^{-t} dt = \int_0^{+\infty} x^{-1} e^{-x^2} dx^2 = 2\int_0^{+\infty} e^{-x^2} dx = \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

进一步有

$$\Gamma(\alpha) = -\int_0^{\infty} x^{\alpha - 1} de^{-x} = -[x^{\alpha - 1} e^{-x}]_0^{+\infty} + (\alpha - 1)\int_0^{+\infty} x^{\alpha - 2} e^{-x} dx = (\alpha - 1)\Gamma(\alpha - 1)$$

对任意正整数 n, 根据上面的性质有

$$\Gamma(n) = (n-1)!$$

关于 Beta 函数和 Γ-函数, 有如下关系:

定理 8.2 对任意给定 $\alpha_1>0$ 和 $\alpha_2>0$, 有

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

证明 根据 Γ-函数的定义有

$$\Gamma(\alpha_1)\Gamma(\alpha_2)=\int_0^{+\infty}t^{\alpha_1-1}e^{-t}dt\int_0^{+\infty}s^{\alpha_2-1}e^{-s}ds=\int_0^{+\infty}\int_0^{+\infty}e^{-(t+s)}t^{\alpha_1-1}s^{\alpha_2-1}dtds.$$

引入变量替换 $x=t+s$ 和 $y=t/(t+s)$, 反解可得 $t=xy$ 和 $s=x-xy$, 计算雅可比行列式有

$$\begin{vmatrix}\frac{\partial t}{\partial x} & \frac{\partial t}{\partial y}\\ \frac{\partial s}{\partial x} & \frac{\partial s}{\partial y}\end{vmatrix}=\begin{vmatrix}y & x\\ 1-y & -x\end{vmatrix}=-x.$$

同时有 $x\in(0,+\infty)$ 和 $y\in(0,1)$ 成立, 由此可得

$$\begin{aligned}\Gamma(\alpha_1)\Gamma(\alpha_2) &= \int_0^1\int_0^{+\infty}e^{-x}x^{\alpha_1-1}y^{\alpha_1-1}x^{\alpha_2-1}(1-y)^{\alpha_2-1}|x|dxdy\\ &= \int_0^1\int_0^{+\infty}e^{-x}x^{\alpha_1+\alpha_2-1}y^{\alpha_1-1}(1-y)^{\alpha_2-1}dxdy\\ &= \int_0^{+\infty}e^{-x}x^{\alpha_1+\alpha_2-1}dx\int_0^1 y^{\alpha_1-1}(1-y)^{\alpha_2-1}dy\\ &= \Gamma(\alpha_1+\alpha_2)\mathrm{Beta}(\alpha_1,\alpha_2)\end{aligned}$$

定理得证.

根据上述定理可知

推论 8.1 对任意 $\alpha_1>1$ 和 $\alpha_2>0$, 有

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

证明 根据前面的定理有

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

定义 8.5 对任意 $\alpha_1,\alpha_2,\cdots,\alpha_k>0$, 定义多维 Beta 函数为

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

下面介绍三种分布:

定义 8.6 (Beta 分布) 给定 $\alpha_1 > 0$ 和 $\alpha_2 > 0$, 若随机变量 X 的概率密度为

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

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

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

证明 根据期望的定义有

$$\begin{aligned} E[X] &= \frac{1}{B(\alpha_1,\alpha_2)}\int_0^1 x \cdot x^{\alpha_1-1}(1-x)^{\alpha_2-1}dx = \frac{B(\alpha_1+1,\alpha_2)}{B(\alpha_1,\alpha_2)} = \frac{\alpha_1}{\alpha_1+\alpha_2}, \\ E[X^2] &= \frac{1}{B(\alpha_1,\alpha_2)}\int_0^1 x^{\alpha_1+1}(1-x)^{\alpha_2-1}dx = \frac{B(\alpha_1+2,\alpha_2)}{B(\alpha_1,\alpha_2)} = \frac{\alpha_1+1}{\alpha_1+\alpha_2+1}\frac{\alpha_1}{\alpha_1+\alpha_2}, \end{aligned}$$

由此可得

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

例 8.2 设独立同分布随机变量 $X_1, X_2, \cdots, X_n$ 服从均匀分布 $\mathcal{U}(0,1)$, 记 $X_{(k)}$ 为其顺序统计量, 则

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

证明 若随机变量 $X_i \sim U(0,1)$ $(i \in [n])$, 则当 $x \in (0,1)$ 时其分布函数 $F(x) = x$. 由此可得到第 k 个统计量 $X_{(k)}$ 的概率密度函数

$$\begin{aligned} f(x) &= \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k} = \frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k} \\ &= \frac{1}{B(k,n-k+1)}x^{k-1}(1-x)^{n-k}. \end{aligned}$$

下面定义 Γ 分布:

定义 8.7 如果随机变量 X 的概率密度

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$