1 基础概率

1.1 概率公理

1. **非负性** $\boldsymbol{P} \geq 0$
2. **规范性**对于必然事件 $\boldsymbol{P}(\boldsymbol{\Omega}) = 1$
3. **可列可加性**对于两两互不相容的事件 $\{\boldsymbol{A}_i\}, \boldsymbol{A_i A_j} = \emptyset$.

$$\boldsymbol{P}(\bigcup_{i=1}^{+\infty} \boldsymbol{A}_i) = \sum_{i=1}^{+\infty} \boldsymbol{P(A_i)} \tag{1}$$

1.2 公式

$$\begin{aligned}
&P(\bigcup A_i) = \sum P(A_i) - \sum P(A_i A_j) + \sum P(A_i A_j A_k) + \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n) \\
&|\bigcup A_i| = \sum |A_i| - \sum |A_i A_j| + \sum |A_i A_j A_k| + \ldots + (-1)^{n-1} |A_1 A_2 \ldots A_n| \\
&|A - B| = |A| - |AB|
\end{aligned}$$

1.3 条件概率

$$\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{P}(\boldsymbol{AB})}{\boldsymbol{P}(\boldsymbol{B})} \tag{2}$$

假设 $\boldsymbol{\Omega}$ 不变，则条件 $\boldsymbol{B}$ 说明 $\boldsymbol{B}$ 成为全空间。概率即为 $\boldsymbol{A}$ 的测度比上 $\boldsymbol{B}$ 的测度。

1.4 Bayes 公式

设 $\{\boldsymbol{B}_i\}$ 为 $\boldsymbol{\Omega}$ 的一个划分

$$\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{\Omega}) = \sum \boldsymbol{P}(\boldsymbol{B}_i)\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B_i}) \tag{3}$$

于是

$$\boldsymbol{P}(\boldsymbol{B_i}|\boldsymbol{A}) = \frac{\boldsymbol{P}(\boldsymbol{AB_i})}{\boldsymbol{P}(\boldsymbol{A})} = \frac{\boldsymbol{P}(\boldsymbol{B}_i)\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B_i})}{\sum \boldsymbol{P}(\boldsymbol{B}_i)\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B_i})} \tag{4}$$

1.5 独立性

设 n 个事件 $\boldsymbol{A_1, A_2, \ldots, A_n}$, 相互独立要求

$$\begin{aligned}
\boldsymbol{P}(\boldsymbol{A_i A_j}) &= \boldsymbol{P}(\boldsymbol{A}_i)\boldsymbol{P}(\boldsymbol{A}_j) \\
\boldsymbol{P}(\boldsymbol{A_i A_j A_k}) &= \boldsymbol{P}(\boldsymbol{A}_i)\boldsymbol{P}(\boldsymbol{A}_j)\boldsymbol{P}(\boldsymbol{A}_k) \\
&\cdots\cdots \\
\boldsymbol{P}(\boldsymbol{A_i A_j \ldots A_n}) &= \prod \boldsymbol{p}(\boldsymbol{A}_i)
\end{aligned} \tag{5}$$

互斥一定不独立；看成全空间下均匀散点比较好: 不均匀则有的地方” 更浓”

2 随机变量

2.1 随机变量 is 函数

随机变量是 $\boldsymbol{\Omega}$ 上的实单值函数 $\boldsymbol{X}(\boldsymbol{\omega})$

2.2 分布函数

$$\boldsymbol{F}(x)=\boldsymbol{P}(\boldsymbol{X}\leq x),\qquad -\infty<x<+\infty \tag{6}$$

性质：

$$\begin{aligned} \boldsymbol{P}(a<x\leq b)&=\boldsymbol{F}(b)-\boldsymbol{F}(a)\\ \boldsymbol{P}(\boldsymbol{X}=x_0)&=\boldsymbol{F}(x_0)-\boldsymbol{F}(x_0-0)\\ \text{左极限存在，右连续}\ \lim_{t\to x^+}\boldsymbol{F}(t)&=\boldsymbol{F}(x)\end{aligned} \tag{7}$$

2.3 离散型随机变量

离散型随机变量的可能取值是可列个，设可能取值 $\boldsymbol{X}=x_k\ (k=1,2,\ldots)$，不妨设 $x_1<x_2<\ldots$，称 $\boldsymbol{P}(\boldsymbol{X}=x_k)=p_k$，$k=1,2\ldots$ 为 $\boldsymbol{X}$ 的分布律，且有：

$$\begin{aligned} &p_k\geq 0\qquad k=1,2,\ldots\\ &\sum_{k=1}^{+\infty}p_k=1\end{aligned} \tag{8}$$

2.3.1 离散型随机变量分布律

二项分布 $\boldsymbol{X}$~$\boldsymbol{B}(n,p)$ 进行 n 次成功 k 次 $\boldsymbol{P}(\boldsymbol{X}=k)=C_n^k p^k(1-p)^{n-k}$

帕斯卡分布 成功 r 次恰需要 k 次，k 成功且前 k-1 成功 r-1

$$\boldsymbol{P}(\boldsymbol{X}=k)=pC_{k-1}^{r-1}p^{r-1}(1-p)^{k-r}=C_{k-1}^{r-1}p^r(1-p)^{k-r} \tag{9}$$

Poisson 分布

定理： 若 $\lim_{n\to+\infty}np_n=\lambda>0$,

$$\lim_{n\to+\infty}C_n^k p_n^k(1-p_n)^{n-k}=e^{-\lambda}\frac{\lambda^k}{k!} \tag{10}$$

Poisson 分布 X~P(λ)： 均匀概率密度，λ，发生次数。

$$\boldsymbol{P}(\boldsymbol{X}=k)=e^{-\lambda}\frac{\lambda^k}{k!} \tag{11}$$

2.4 连续型随机变量

设 $\boldsymbol{X}$ 是一随机变量，$\boldsymbol{F}(x)$ 是他的分布函数，若存在一个非负可积函数 $f(X)$，使得

$$\boldsymbol{F}(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t,\qquad -\infty<x<+\infty \tag{12}$$

则 X 是连续型随机变量，f(x) 是概率密度函数，且积上去不一定可导（似乎一般可导）：

$$\begin{aligned} f(x) &\geq 0 \\ \int_{-\infty}^{+\infty} f(x)\mathrm{d}x &= \boldsymbol{F}(+\infty) = 1 \\ f(x) &= \boldsymbol{F}'(x) \qquad \textit{where } f(x) \textit{ is continuous} \end{aligned} \tag{13}$$

2.4.1 连续性随机变量分布律

指数分布 $\boldsymbol{X}$~$\boldsymbol{E}(\lambda)$： 无记忆性 or 均匀概率密度，相邻两次发生的间隔。

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 \quad , & x \leq 0 \end{cases} \qquad (\lambda > 0) \tag{14}$$

正态分布 $\boldsymbol{X}$~$\boldsymbol{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in R \tag{15}$$

参数 $\mu = 0$, $\sigma = 1$ 的正态分布称为标准正态分布，记为 $\boldsymbol{X}^*$~$\boldsymbol{N}(0,1)$。其概率密度 $\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}}$. 分布函数 $\boldsymbol{\Phi}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t$

3 二维随机变量及分布律

3.1 二维随机变量

对随机试验 E 的样本空间 Ω, 若对 Ω 中任一样本点 ω, 存在实数 $(X(\omega), Y(\omega))$ 与之对应，则称 (X, Y) 为二维随机变量

3.2 （联合）分布函数

定义：

$$F(x,y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x, Y \leq y) \tag{16}$$

联合分布函数的性质：

1. $0 \leq F(x,y) \leq 1$
$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0, F(+\infty, +\infty) = 1$

2. 固定其他变量，对某一特定变量是单调不减函数。
3. 固定其他变量，关于某特定变量右连续

4. 对于任意的 a, b, c, d, 满足 $a < b$, $c < d$，有：
$F(b,d) - F(a,d) - F(b,c) + F(a,c) = P(a < X \leq b, c < Y \leq d) \geq 0$
NOTE：不像一维那样，这里存在只满足前三点不满足第四点的函数（但第四点是用定义推出来的）

3.3 边缘分布函数

定义：边缘分布函数为 $F_X(x)$ 和 $F_Y(y)$.

性质：$F_X(x)=F(x,+\infty)$，$F_Y(y)=F(+\infty,y)$

3.4 二维离散随机变量

$$\textbf{定义:} \text{样本点可列，} p_{ij} \text{ 为分布律。} F(x,y)=\sum_{i=-\infty}^{x}\sum_{j=-\infty}^{y}p_{ij}.$$
$$P((X,Y)\in D)=\sum\sum_{(i,j)\in D}p_{i,j}$$

3.5 二维连续随机变量

定义：若存在非负可积函数 $f(x,y)$, s.t. $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v$

$$P((X,Y)\in D)=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$

3.6 条件分布

离散：$P(X=x_i|Y=y_j)=\dfrac{p_{i,j}}{p_{\cdot j}}$

连续：在 {Y=y} 条件下 X 的条件概率密度: $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

3.7 独立性

$$F(x,y)=F_X(x)F_Y(y).$$
$$p_{i,j}=p_{i,\cdot}p_{\cdot,j}$$
$$f(x,y)=f_X(x)f_Y(y)$$

3.7.1 独立性的定理

1. 设 X,Y 为相互独立的随机变量，u(x),v(y) 为连续函数。则 U，V 也相互独立
2. 若 X,Y 在相互独立的定义域上，有 $f(x,y)=r(x)g(y)$，则 X,Y 独立。

3.8 二维随机变量函数的分布

$$P(Z=z_k)=\sum_{g(x_{i_k},y_{j_k})=z_k}P(X=x_{i_k},Y=y_{j_k})$$
$$Poisson\text{分布、二项分布可加（X,Y 独立）} \tag{17}$$
$$F_Z(z)=\iint_{g(x,y)\leq z}f(x,y)\mathrm{d}x\mathrm{d}y$$

3.8.1 卷积，一般线性函数的分布

$f(z)\mathrm{d}z=f(z)\mathrm{d}z\wedge R$ （不太对）

3.8.2 正态分布可加

3.8.3 m 个独立指数分布的最小 $\min\{X_1, X_2, \ldots, X_m\} \sim E(m\lambda)$

4 随机变量的数字特征

4.1 期望

4.1.1 定义：绝对收敛

4.1.2 公式

（对于同一个样本空间下，在同样的随机分布中不同的随机变量 X_i 有：）
线性：$E((\sum A_i X_i + B_i)) = \sum A_i E X_i + b_i$,
独立时 $E(\prod X_i) = \prod E(X_i)$
多维时总可以把联合分布合起来，这样就是同一个分布函数下的变量问题。

4.2 方差

4.2.1 定义

$$\text{方差：} D(X) = E[X - EX]^2 = E(X^2) - (EX)^2$$
$$\text{标准差：} \sigma_X = \sqrt{D(X)}$$

多维则用联合概率分布。

4.2.2 性质

$$\text{D 存在} \iff E(X^2) < +\infty.$$
$$D(X - Y) = D(X) + D(Y) - 2E[(X - EX)(Y - EY)]$$
$$D(\sum a_i X_i + b) = \sum a_i^2 D(X_i)$$

4.3 标准化随机变量

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}} \Longrightarrow EX^* = 0, DX^* = 1$$

4.4 协方差和相关系数

4.4.1 定义

$$\operatorname{cov}(X, Y) = E[(X - EX)(Y - EY)]$$
$$\rho_{XY} = \frac{E[(X - EX)(Y - EY)]}{\sqrt{D(X)}\sqrt{D(Y)}} = E[(X^* - EX^*)(Y^* - RY^*)]$$

4.4.2 公式

$$
\rho_{XY} = \frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = cov(X^*, Y^*)
$$
$$
cov(X,Y) = E(XY) - E(X)E(Y)
$$
$$
D(X) = cov(X,X)
$$

4.4.3 性质

独立一定不相关，相关一定不独立。
不独立不一定相关，不相关不一定独立
线性：

$$
cov(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum\sum a_i b_j cov(X_i, Y_j)
$$

Cauchy-Schwarz:

$$
|cov(X,Y)| \leq \sqrt{D(X)D(Y)}
$$

证 cov(x,y) 是内积: 线性，可交换

正定：$cov(x,x) \geq 0$

4.5 高阶矩

k 阶原点矩 $E(X^k)$

k 阶中心矩 $E([X - E(X)]^k)$

X 和 Y 的 k+l 阶混合原点矩 $E(X^k Y^l)$

X 和 Y 的 k+l 阶混合中心矩 $E([X - EX]^K [Y - EY]^l)$

4.5.1 协方差矩阵

$$
c_{ij} = cov(X_i, X_j) \tag{18}
$$
$$
\boldsymbol{C} = \left(c_{ij} \right) \tag{19}
$$

性质：（内积）
n 维正态分布：

$$
f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{C}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T C^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \tag{20}
$$

5 大数定理和中心极限

5.1 Chebyshev Inequality

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}, \quad \forall \epsilon > 0$$

5.2 依概率收敛

已知 $\{Y_n\} = Y_1, Y_2, \ldots, Y_n$, X 是一个随机变量。依概率收敛定义为：$\forall \epsilon > 0$,

$$\lim_{n \to +\infty} P(|Y_n - X| \geq \epsilon) = 0 \text{ or } \lim_{n \to +\infty} P(|Y_n - X| < \epsilon) = 1$$

这里 X 一般单值，也可以 $Y_n = X + X_n$，但可能独立不太行

5.3 大数定律——次数多起来后频率趋于期望

5.3.1 Bernoulli 大数定理——二项分布的频率收敛于概率

5.3.2 Chebyshev 大数定律

设 $\{X_n\}$ 两两不相关，方差存在且一致有界，则

$$\lim_{n \to +\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} a_i| < \epsilon) = 1$$

但需要注意普遍来说这里需要方差，极限只要求期望。这不好

5.3.3 Khintchine 大数定律

$\{X_n\}$: **独立同分布 I,I,D** 且期望存在，则

$$\lim_{n \to +\infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k - E(X)| \geq \epsilon) = 0$$

变化一下，$X^k \to E(X_k)$

5.4 中心极限定理

5.4.1 独立同分布的中心极限定理

设 $\{X_n\}$ 为独立同分布的随机变量序列，且他们的期望方差都存在。
$E(X_k) = \mu, D(X_k) = \sigma^2.$，则对任意实数 x，有：

$$\lim_{n \to +\infty} P(\frac{\sum_{i=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq x) = \Phi(x) \tag{21}$$

若记 $\sum_{k=1}^{n} X_k$ 的标准化随机变量为 $Y_n = \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}\sigma}$, 即 $Y_n \sim N(0,1)$

5.4.2 De Moivre-Laplace 中心极限定理

设随机变量 $Y_n \sim B(n,p))$，$0 < p < 1, n = 1, 2$，则对任一实数 x，有：

$$\lim_{n \to +\infty} P(\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x) = \Phi(x) \tag{22}$$

6 统计

6.1 基本概念

总体： 研究对象的某个（或某些）数量指标的全体称为**总体**

个体： 总体的每个元素（数量指标）

样本： 总体的部分

样本观测值： 样本值

样本空间： 样本 $(X_1, X_2, \cdots, X_n))$ 的所有可能取值的集合 $\chi = \{(x_1, x_2, \cdots, x_n)\}$

简单随机样本： 独立同分布

6.2 简单随机样本

6.2.1 统计量

设 $(X_1, X_2, \ldots, X_n)$ 为总体 X 的简单随机样本，$g(r_1, r_2, \cdots, r_n)$ 是一个实值连续函数，且不含除自变量之外的未知参数.
称随机变量 $g(X_1, X_2, \ldots, X_n)$ 为**统计量**，$g(x_1, x_2, \ldots, x_n)$ 为样本值

6.2.2 常用统计量

样本均值 $\bar{X}$,$\bar{x}$

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, s^2 $\bar{X}$ 是一个变量，导致不能提出去。

样本标准差 $S = \sqrt{S^2}$, s

样本 k 阶原点矩 $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$,$m_k$

样本 k 阶中心矩 $(CM)_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$, $(cm)_k$

顺序统计量 $\mathrm{sort}((X_1, X_2, \cdots, X_n))$，极差

NOTE!! X（or $\bar{X}$）不是一个数，而是一个随机变量（分布）
$E(\bar{X}) = \mu, \bar{X} \neq E(X), D(\bar{X}) = \frac{\sigma^2}{n}, E(S^2) = \sigma^2$

分位数
对于连续随机变量 X，给定 $0 < \alpha < 1$.
若 $P(X > x_\alpha) = \alpha$, 则 x_α 为 X 所服从分布的上侧 α 分位数。
若 $f(x)$ 为偶函数，则如果 $P(|X| > x_{\alpha/2}) = \alpha$，则 $x_{\alpha/2}$ 为 X 服从分布的双侧 α 分位数
$x_{\alpha/2} = x_{1-\alpha/2}$

6.3 常用分布

正态分布

χ^2 分布 设 $\{X_n\}$ 独立且 $X_i \sim N(0,1)$，则称统计量$\chi^2=\sum\limits_{i=1}^{n} X_i^2$: 服从自由度为 n 的 χ^2 分布，记为 $\sum_{i=1}^{n} X_i \sim \chi^2(n)$.

$$f_{\chi^2}(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}e^{-\frac{x}{2}}x^{\frac{n}{2}}-1, & x>0\\ 0, & x\leq 0\end{cases} \tag{23}$$

其中 $\Gamma(x)=\displaystyle\int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t$

性质：

1. $E(\chi^2)=n,\ D(\chi^2)=2n$
2. $X_1\sim\chi^2(n_1),\ X_2\sim\chi^2(n_2)$, 且 X_1,X_2 独立，则 $X_1+X_2\sim\chi^2(n_1+n_2)$
3. n 很大时，χ^2 近似服从 $N(n,2n)$.

t 分布 $X\sim N(0,1), Y\sim\chi^2(n)$ 且 X,Y 独立，则$T=\dfrac{X}{\sqrt{Y/n}}=\dfrac{X}{\sqrt{\chi^2(n)/n}}$: 服从自由度为 n 的 t 分布，记为 $T\sim t(n)$.

$$f(t)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}, \qquad t\in R \tag{24}$$

性质

1. $f(t)=-f(t)$
2. $n\longrightarrow+\infty,\ f(t)\longrightarrow\varphi(t)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$
3. 上侧 α 分位数 $t_\alpha(n)$: $t_{1-\alpha}(n)=-t_\alpha(n)$

F 分布 设 $U\sim\chi^2(m),\ V\sim\chi^2(n)$, 且 U 与 V 独立，则称随机变量$F=\dfrac{U/m}{V/n}=\dfrac{\chi^2(m)/m}{\chi^2(n)/n}$: 服从第一自由度为 m, 第二自由度为 n 的 F 分布，记为 $F\sim F(m,n)$

$$f_F(t)=\begin{cases}\dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}(\dfrac{m}{n})^{\frac{m}{2}}t^{\frac{m}{2}-1}(1+\dfrac{m}{n}t)^{-\frac{m+n}{2}} & t>0,\\ 0 & t\leq 0\end{cases} \tag{25}$$

性质

1. $F\sim F(m,n),\ \frac{1}{F}\sim F(n,m)$
2. $F(m,n)$ 的上侧 α 分位数 $F_\alpha(m,n)$: $F_{1-\alpha}(m,n)=\frac{1}{F_\alpha(n,m)}$

6.4 正态总体的抽样分布

$X \sim N(\mu, \sigma^2), \{X_i\} \sim N(\mu, \sigma^2)$ 是来自总体的 $i.i.d.$ 独立同分布。

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n) \tag{26}$$

$$\frac{(n-1)S^2}{\sigma^2} = \sum \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1) \tag{27}$$

$$\bar{X}\text{与}\frac{(n-1)S^2}{\sigma^2}\text{独立} \tag{28}$$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{29}$$

$$\frac{\bar{x}-\mu}{S/\sqrt{n}} \sim t(n-1) \tag{30}$$

两个正态总体的情形

$\{X_n\} \sim N(\mu_1, \sigma_1^2), \{Y_m\} \sim N(\mu_2, \sigma_2^2)$. 且 X, Y 独立

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1). \tag{31}$$

若 $\sigma_1 = \sigma_2 = \sigma$:

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{n}) \tag{32}$$

$$\text{由可加性}\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2) \tag{33}$$

$$\bar{X} - \bar{Y}\text{与}\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2}\text{独立} \tag{34}$$

$$\frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{m+n-2}}} \sim T(m+n-2) \tag{35}$$

7 参数估计

7.1 点估计

7.1.1 频率替代法————大数定律

7.1.2 矩估计法

设总体 X 的分布函数为 $F(x; \theta_1, \cdots, \theta_k)$，并假设 k 阶原点矩存在。记

$$E(X^r) = \mu_r(\boldsymbol{\theta}) \quad r = 1, 2 \ldots, k \tag{36}$$

则 $\boldsymbol{\theta}$ 有 k 个未知数，需要 k 个方程。从理论上推导出 $\boldsymbol{\mu}$ 的 k 个表达式，将 μ_i 替换成样本矩 M^k，反解 $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$。称 $\hat{\boldsymbol{\theta}}(X_1, \ldots, X_n)$ 为矩估计量，$\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ 为矩估计量

7.1.3 极大似然估计

定义似然函数

$$L(\boldsymbol{\theta}) = P(\bigcap X_i = x_i) = \prod_{i=1}^{n} P(x_i; \boldsymbol{\theta})$$

连续情形下

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$$

L 极大，$L(\hat{\boldsymbol{\theta}}) = \max L(\boldsymbol{\theta})$. 称 $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ 为最大似然估计值，$\hat{\boldsymbol{\theta}}(X_1, \ldots, X_n)$ 为最大似然估计量

7.2 估计量的评价标准

7.2.1 无偏性

对于 $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$, 若

$$E(\hat{\theta}) = \theta$$

则称 $\hat{\theta}$ 是 θ 的无偏估计量，反之，称 $\epsilon = E(\hat{\theta}) - \theta$ 为估计量 $\hat{\theta}$ 的偏差

- 对任意分布 X，$\bar{X}$ 都是总体均值 $\mu = E(X)$ 的无偏估计量，S^2 是 σ^2 的无偏估计量
- $E(\hat{\theta}) = \theta$ 不一定得到 $E(f(\hat{\theta})) = f(\theta)$

7.2.2 有效性

设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 均为参数 θ 的**无偏估计量**，若

$$D(\hat{\theta}_1) < D(\hat{\theta}_2)$$

, 称 θ_1 比 θ_2 有效

· **Rao-Cramer 不等式**

设总体 X 为离散型随机变量，$P(X = x; \theta) = P(x; \theta)$. $(X_1, \ldots, X_n)$ 是简单随机样本

$$D(\hat{\theta}) \geq I(\theta) = \frac{1}{nE[(\frac{\partial \ln P(X;\theta)}{\partial \theta})^2]} \tag{37}$$

设总体 X 为连续性随机变量，概率密度为 $f(x; \theta)$，$\hat{\theta}$ 是**无偏估计量**，

$$D(\hat{\theta}) \geq I(\theta) = \frac{1}{nE[(\frac{\partial \ln f(X;\theta)}{\partial \theta})^2]} \tag{38}$$

$I(\theta)$ 称为无偏估计的方差下界。

7.2.3 有效估计量

设 $\hat{\theta}_0$ 是 θ 的无偏估计量，如果在所有 θ 的无偏估计量 $\hat{\theta}$ 中均有 $D(\hat{\theta}_0) \leq D(\hat{\theta})$. 称 $\hat{\theta}_0$ 是 θ 的有效估计量

7.2.4 一致性

设 $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ 是 θ 的估计量，如果 $\{\hat{\theta}_n\} \xrightarrow{p} \theta$, 则称 $\hat{\theta}_n$ 是 θ 的一致估计量。

- 不要求无偏。而无偏 + 方差 $\to 0 \Longrightarrow$ 一致

7.3 区间估计

设总体 X 的分布函数为 $F(x;\theta)$，$\{X_n\}$ 是一个样本，若

$$\forall\alpha,\exists\hat{\theta}_1=\hat{\theta}_1(X_1,\ldots,X_n),\hat{\theta}_2=\hat{\theta}_2(X_1,\ldots,X_n).\ s.t.\ P(\hat{\theta}_1<\theta<\hat{\theta}_2)=1-\alpha$$

成立，则称区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 是 θ 的置信度为 $1-\alpha$ 的置信区间，$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是置信下限和致信上限

枢轴量 $U=U(X_1\ldots X_n;\theta)$ 除 θ 没有其他未知参数，U 分布已知且不依赖于未知参数