知识点归纳 · 概率论与数理统计

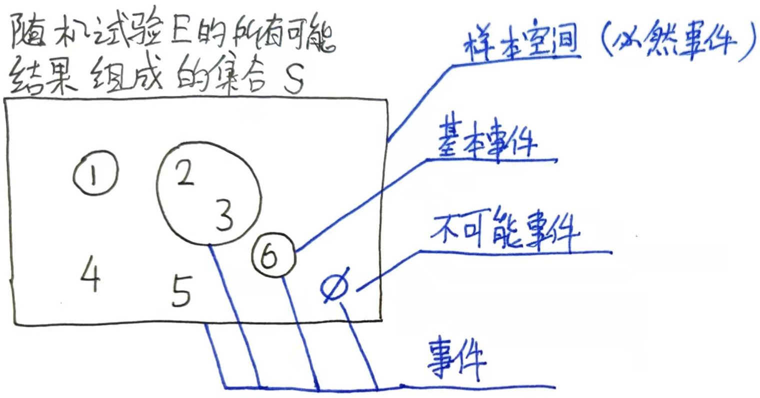
21305337 *Betelgeuxe*

仅供学习参考，勿作商业用途

目录

1　概率论的基本概念

1.1　样本空间



1.2　常用公式

交换律：$A\cup B = B\cup A$; $A\cap B = b\cap A$

结合律：$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$

$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$

对偶律：$\overline{A\cup B} = \overline{A}\cap\overline{B}$; $\overline{A\cap B} = \overline{A}\cup\overline{B}$

其他：$P(A\overline{B}) = P(A) - P(AB)$

$P(A\cup B) = P(A) + P(B) - P(AB)$(加法公式)

1.3　等可能概型（古典概型）

条件：1. 有穷性；2. 等可能性

其中有放回抽样和不放回抽样

1.4　实际推断原理

概率很小的事件在一次试验中实际上几乎是不发生的

1.5　条件概率（乘法公式）

$$\boxed{P(B|A) = \frac{P(AB)}{P(A)}}$$

1.6 全概率公式

划分

设 S 为试验 E 的样本空间，$B_1, B_2, \cdots, B_n$ 为 E 的一组事件。若 $B_i \cap B_j = \varnothing$，$i, j = 1, 2, \cdots, n$；且 $\bigcap_{k=1}^{n} B_k = S$ 则 $B_1, B_2, \cdots, B_n$ 为 S 的一个**划分**.

若 $B_1, B_2, \cdots, B_n S$ 的一个划分，则

$$\boxed{P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k)}$$

贝叶斯 (Bayes) 公式

若 $B_1, B_2, \cdots, B_n$ 为 S 的一个划分，则 $P(B_i|A) = \dfrac{P(A|B_i)P(B_i)}{\sum_{k=1}^{n} P(A|B_k)P(B_k)}, i = 1, 2, \cdots, n$.

1.7 独立性

二元情形 def: 若 $P(AB) = P(A)P(B)$，则 A, B **互相独立**（**独立**）

- 若 A, B 独立,$P(A) > 0$, 则 $P(B|A) = P(B)$.
- 若 A, B 独立, 则 A 与 $\overline{B}, \overline{A}$ 与 $B, \overline{A}$ 与 $\overline{B}$ 也独立.

两两独立 $n > 2$ 时，对于事件 $A_1, A_2, \cdots, A_n$，若 $\forall i \neq j$ 有 $P(A_iA_j) = P(A_i)P(A_j)$，则称 $A_1, A_2, \cdots, A_n$ **两两独立**.

互相独立 $n > 2$ 时，对于事件 $A_1, A_2, \cdots, A_n$，若任意 $2, 3, \cdots, n$ 个事件的积事件的概率都等于各事件概率之积，则称 $A_1, A_2, \cdots, A_n$ **互相独立**

2 随机变量

2.1 离散型随机变量

离散型随机变量 X 的分布律 表示 X 的所有可能取值及其概率的**通式**或**表格**

伯努利试验 试验 E 只有 A 及 $\overline{A}$ 两种结果

n 重伯努利试验 将 E 独立重复进行 n 次

2.1.1 二项分布

设随机变量 X 为 n 次试验 E 中 A 发生的次数，则 $P(X=k)=\binom{n}{k}p^kq^{n-k}$，并记 $X\sim B(n,p)$.

2.1.2 泊松分布

若离散型随机变量 X 的分布律为

$$\boxed{P\{X=k\}=\frac{\lambda^k}{e^\lambda\ k!},\ (\lambda>0)}$$

则称 X 服从参数为 λ 的**泊松分布**，记为 $X\sim\pi(\lambda)$或$Po(\lambda)$.

泊松定理 对常数 $\lambda,\forall n\in\mathbf{Z}$，且有 $np_n=\lambda$，则对 $\forall k\in\mathbf{Z}$ 有

$$\boxed{\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{e^\lambda\ k!}.}$$

当 n 很大，p 很小 $(np=\lambda)$，有

$$\boxed{\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k}{e^\lambda\ k!}.}$$

2.2 连续型随机变量

分布函数 $F(x)=P\{X\leq x\}$, 性质：1. 单调非降 2. 右连续性 3. 左极限存在 4. 规范性.

概率密度 $f(x)$, 其中 $F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t, f(x)=\frac{\partial}{\partial x}F(x)$.

小结论 X 为连续型随机变量$\Leftrightarrow F(x)$ 无间断点.

2.2.1 均匀分布

若 X 在 (a,b) 上服从均匀分布，记为 $X\sim U(a,b)$, 密度函数为

$$\boxed{f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b,\\ 0, & \text{其他}.\end{cases}}$$

2.2.2 指数分布

若 X 服从参数为 $\theta(\theta>0)$ 的指数分布, 密度函数为

$$\boxed{f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}}$$

无记忆性 $P\{X>s+t|X>s\}=PX>t.$

2.2.3 正态分布

X 服从参数为 μ,θ 的正态分布, 记为 $X\sim N(\mu,\sigma^2)$, 概率密度为

$$\boxed{f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty.}$$

标准正态分布 $X\sim N(0,1),\begin{cases}\text{概率密度}\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}},\\ \text{分布函数}\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{t^2}{2}}\mathrm{d}t.\end{cases}$

2.3 [方法] 求一元函数的分布函数或密度函数

分布函数方法 若已知 $F_X(x)$, 且 $Y=g(X)$, 则

$$\boxed{F_Y(y)=P\{Y\le y\}=P\{g(X)\le Y\}}$$

密度函数方法 (次选) 若已知 $F_X(x)$, 有 $Y=g(X)$,$h=g^{-1}$, 即 $X=h(Y)$(可数多段), 则

$$\boxed{f_Y(y)=\sum f_X(x)\cdot|h'(y)|=\sum f_X[h(y)]\cdot|h'(y)|.}$$

3 多维随机变量及其分布

n 维随机变量 $X(\omega)=\Big(X_1(\omega),\cdots,X_n(\omega)\Big)$, 其中 $X_1,\cdots,X_n$ 是定义在 S 上的随机变量.

3.1 联合分布

联合分布函数 $F(x,y)=P\{(X\le x)\cap(Y\le y)\}$, 记为 $P\{X\le x,Y\le y\}$.

- 关于 x 与 y 均单调非降, 右连续.
- 对 $\forall x_1 < x_2, y_1 < y_2, F(x_2,y_2)+F(x_1,y_1) \geq F(x_1,y_2)+F(x_2,y_1)$.

联合概率密度 $f(x,y), F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\mathrm{d}x\mathrm{d}y$, $\left(f(x,y)=\dfrac{\partial^2}{\partial x\partial y}F(x,y)\right)$, 则 (X,Y) 为二维连续型随机变量, $f(x,y)(x,y)$ 的概率密度（X 和 Y 的联合概率密度）

3.2 边缘分布

边缘分布律（离散） (X,Y) 关于 X 的边缘分布律 $p_{i\bullet}=\sum_j p_{ij}=P\{X=x_i\}$.

边缘概率密度（连续） (X,Y) 关于 X 的边缘概率密度为

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\frac{\partial F(x,+\infty)}{\partial x}.$$

3.3 二维正态分布（要背）

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
$$\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

其中 $\sigma_1,\sigma_2>0, -1\leq\rho\leq 1$, 称 (X,Y) 服从参数为 $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$ 的二维正态分布, 记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

- $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho) \overset{\Longrightarrow}{\not\Leftarrow} \begin{cases} X\sim(\mu_1,\sigma_1) \\ Y\sim(\mu_2,\sigma_2)\end{cases}$
- n 维正态随机变量的每个分量都是正态随机变量; **相互独立的** n 个正态随机变量组成 n 维正态随机变量.
- $\rho=0 \implies f(x,y)=f_X(x)\cdot f_Y(y)$.
- 若 $(X_1,\cdots,X_n)$ 服从 n 维正态分布, 则 $X_1,\cdots,X_n$ **相互独立**等价于**两两不相关**.
- n 维正态分布的线性组合仍服从正态分布 (包括一维和多维).(P114)

三项分布 * $P\{(X,Y)=(i,j)\}=C_n^iC_{n-1}^jp^iq^jr^{n-i-j}=\dfrac{n!}{i!j!(n-i-j)!}p^iq^jr^{n-i-j}$, 其中 $p+q+r=1$, $i,j,n\in Z^+$, 则称 (X,Y) 服从三项分布.

3.4 条件分布

$Y\leq y$ 条件下 X 的条件分布函数:

$$F(x|Y\le y)=P(X\le x|Y\le y)=\frac{P(X\le x,Y\le y)}{P(Y\le y)}=\frac{F(x,y)}{F_Y(y)}$$

•$Y=y$ 条件下 X 的条件分布函数:

$$F_{X|Y}(x|y)=P(X\le x|Y=y)$$
$$=\begin{cases}\dfrac{P(X\le x,\ Y=y)}{P(Y=y)}\\ \lim\limits_{\Delta y\to 0}\dfrac{P(X\le x,\ y-\Delta y<Y\le y)}{P(y-\Delta y<Y\le y)}=\dfrac{\int_{-\infty}^{x}f(u,y)\mathrm{d}u}{f_Y(y)}=\displaystyle\int_{-\infty}^{x}\dfrac{f(u,y)\mathrm{d}u}{f_Y(y)}\end{cases}$$

3.4.1 随机变量的独立性

若 $\forall x,y$, 均有

$$\begin{aligned}&P(X\in I,Y\in J)=P(X\in I)\cdot P(Y\in J),\forall I,J\\ \iff\quad & F(x,y)=F_X(x)\cdot F_Y(y)\\ \iff\quad & f(x,y)=f_X(x)\cdot f_Y(y)\text{在除去“面积”为 0 的集合成立}\\ \implies\quad & F_X(x)=F(x|Y\le y),F_Y(y)=F(y|X\le x)\text{(代替条件概率)}\end{aligned}$$

则称随机变量 X,Y 相互独立.

- A 与 B 互相独立 $\Leftrightarrow$ A 与 $\overline{B}$ 互相独立 $\Leftrightarrow$ $\overline{A}$ 与 B 互相独立 $\Leftrightarrow$ $\overline{A}$ 与 $\overline{B}$ 互相独立
- 随机变量 $X_1,X_2,\cdots,X_n$ 互相独立, 对 $\forall g_1(t),g_2(t),\cdots,g_n(t)$, 随机变量 $g_1(X_1),g_2(X_2),\cdots,g_n(X_n)$ 互相独立

3.4.2 条件分布律（离散情形）

在 $Y=y_j$ 条件下随机变量 X 的**条件分布律**为:

$$P\{X=x_i|Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\bullet y}}$$

3.4.3 条件概率密度（连续情形）

$Y=y$ 条件下 X 的**条件概率密度**为:

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{\text{概率密度}}{\text{边缘概率密度}}$$

3.5 二元随机变量的一些函数的分布

3.5.1 离散型

二项分布 若 $X\sim B(n,p),Y\sim B(m,p)$, X,Y 互相独立, 则 $X+Y\sim B(n+m,p)$.

泊松分布 若 $X\sim Po(\lambda_1),Y\sim Po(\lambda_2)$, X,Y 互相独立, 则 $X+Y\sim Po(\lambda_1+\lambda_2)$.

3.5.2 连续型

(一) $Z=X+Y$ **的分布**

卷积公式

设 (X,Y) 具有概率密度 $f(x,y)$，且 X,Y 独立，则 Z=X+Y 的概率密度为:

$$F_{X+Y}(z)=\begin{cases}\int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\mathrm{d}y \\ \int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}y\end{cases} \equiv f_X * f_Y$$

正态分布 若 $X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$，$X,Y$ 互相独立，那么 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

(二) $Z=\dfrac{Y}{X}, Z=XY$ **的分布**

设 (X,Y) 具有概率密度 $f(x,y)$，且 X,Y 独立，则 $Z=\dfrac{Y}{X}$，$Z=XY$ 的概率密度分别为:

$$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|f(x,xz)\mathrm{d}x = \int_{-\infty}^{+\infty}|x|f_X(x)f_Y(xz)\mathrm{d}x$$

$$f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f(x,\frac{z}{x})\mathrm{d}x = \int_{-\infty}^{+\infty}\frac{1}{|x|}f_X(x)f_Y(\frac{z}{x})\mathrm{d}x$$

(三) $\max\{X,Y\}$ **和** $\min\{X,Y\}$ **的分布**

设 $U\equiv\max\{X_1,\cdots,X_n\}$, 则有

$$F_U(u)=P\{X_1,\cdots,X_n\le u\} \xlongequal{X_1,\cdots,X_n\text{相互独立}} \prod_{i=1}^{n}F_{X_i}(x_i).$$

设 $V\equiv\min\{X_1,\cdots,X_n\}$, 则有

$$F_V(v)=P\{X_1,\cdots,X_n\le u\}=1-P\{X_1,\cdots,X_n>u\} \xlongequal{X_1,\cdots,X_n\text{相互独立}} 1-\prod_{i=1}^{n}[1-F_{X_i}(x_i)].$$

(四) $g_1(x,y)$ **和** $g_2(x,y)$ **的联合分布**

$$F(g_1,g_2)=P(g_1\le z_1,g_2\le z_2)=\iint\limits_{\substack{g_1(x,y)\le z_1\\ g_2(x,y)\le z_2}} f(x,y)\mathrm{d}x\mathrm{d}y \xlongequal[y=\psi(g_1,g_2)]{\text{设}\,x=\varphi(g_1,g_2)} \iint\limits_{\substack{g_1\le z_1\\ g_2\le z_2}} f(\varphi,\psi)\cdot|J|\mathrm{d}g_1\mathrm{d}g_2$$

$$=\iint\limits_{\substack{g_1\le z_1\\ g_2\le z_2}} f(\varphi,\psi)\cdot\frac{\mathrm{D}(\varphi,\psi)}{\mathrm{D}(\mathrm{g}_1,\mathrm{g}_2)}\mathrm{d}g_1\mathrm{d}g_2$$

4 随机变量的数字特征

4.1 数学期望的定义及一些性质

定义（离散情形） 若级数 $\sum\limits_{k}^{+\infty} x_k P(X = x_k)$ **绝对收敛**则数学期望存在，（注意：条件收敛不保证期望存在!）则记

$$E(X) = \sum_k x_k P(x = x_k) = \sum_k x_k p_k$$

并称 $E(X)$ 为 X 的**数学期望**，简称**期望**，又称**均值**.

定义（连续情形） 若 $\left|\int_{-\infty}^{+\infty} x\,|f(x)|\,\mathrm{d}x\right| < +\infty$ **(绝对收敛)**，则记

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x} \left(= \int_0^1 x\mathrm{d}F(x)\right) < +\infty$$

并称 $E(X)$ 为 X 的**期望**.

4.1.1 一元函数期望的计算公式

$g(x)$ 是实函数,

1) 若 X 为离散型,

$$E\left(g(X)\right) = \sum_k g(x_k)P\{X = x_k\}$$

2) 若 X 为连续型, 密度函数为 $f(x)$,

$$E\left(g(X)\right) = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x.$$

以上两式都可写为

$$E\left(g(X)\right) = \int_0^1 g(x)\mathrm{d}F(x) \qquad \text{(勒贝格积分)}$$

4.1.2 二元函数期望的计算公式

线性函数的期望

$$\boxed{E(aX + bY + c) = aE(X) + bE(Y) + c}$$

X, Y 独立的情形（线性性） 若 X, Y 互相独立, 则

$$E(XY) = E(X)E(Y)$$

一般二元函数的期望 $g(x,y)$ 是实函数,

1) 若 (X,Y) 为离散型,

$$E\left(g(X,Y)\right)=\sum_{i,j}g(x_i,y_j)P\{X=x_i,Y=y_j\}$$

2) 若 (X,Y) 为连续型, 密度函数为 $f(x,y)$,

$$E(g(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y.$$

以上两式都可写为

$$E\left(g(X,Y)\right)=\iint g(x,y)\mathrm{d}F(x,y)\qquad \text{(勒贝格积分)}$$

4.2 方差

$$\boxed{D(X)\equiv E\{[X-E(X)]^2\}=E(X^2)-[E(X)]^2.}$$

- $D(X)=0\Longleftrightarrow P\{X=E(X)\}=1.$
- $D(aX+c)=a^2D(X).$
- 如果X,Y独立, 则$D(X\pm Y)=D(X)+D(Y).$

4.2.1 切比雪夫不等式

设随机变量 X 具有数学期望 $E(X)$, 方差 $D(X)$, 则对 $\forall\varepsilon>0$,

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}.$$

4.3 协方差及相关系数

4.3.1 协方差

$$\operatorname{cov}(X,Y)=\sigma_{XY}\equiv E\{[X-E(X)][Y-E(Y)]\}$$

上式称为 X 和 Y 的协方差, 即 X 和 Y 的二阶混合中心矩.

- X 和 Y 成线性关系 $(Y=aX+b, a\neq 0)\iff \operatorname{cov}(X,Y)=\sqrt{D(X)D(Y)}$
- X 和 Y 不相关 $\iff \operatorname{cov}(X,Y)=0$
- 若 X 和 Y 独立, 则 $\operatorname{cov}(X,Y)=0$(反之不成立)
- $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)$
- $\operatorname{cov}(X,aY+bZ+c)=a\operatorname{cov}(X,Y)+b\operatorname{cov}(X,Z)$
- $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$
- $D(\sum_{i=1}^{n}a_iX_i+b)=\sum_{i=1}^{n}a_i^2D(X_i)+\sum_{i<j}2a_ia_j\operatorname{cov}(X_i,X_j)$

4.3.2 相关系数

若 $D(X)\neq 0, D(Y)\neq 0$, 则 X,Y 的相关系数为

$$\rho_{XY}\equiv\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}$$

上式称为 X 和 Y 的协方差, 即 X 和 Y 的二阶混合中心矩.

- $|\rho_{XY}|\leq 1$
- $Y=aX+b(a>0)\iff \rho_{XY}=1$
- $Y=aX+b(a<0)\iff \rho_{XY}=-1$
- X 和 Y 不相关 $\iff \rho_{XY}=0$
- 若 X 和 Y 独立, 则 $\rho_{XY}=0$(反之不成立)
- X,Y 为二元正态分布或都是二值随机变量, 则不相关性与独立性等价.

4.4 矩、协方差矩阵

X 的 K 阶原点矩（K 阶矩） 若 $E(X^k)$ 存在, 则称为 X 的 K 阶原点矩（K 阶矩）.

X 的 K 阶中心矩 若 $E\{[X-E(X)]^k\}$ 存在, 则称为 X 的 K 阶中心矩.

X 和 Y 的 $k+l$ 阶混合矩 若 $E(X^kY^l)$ 存在, 则称为 X 和 Y 的 $k+l$ 阶混合矩.

X 和 Y 的 $k+l$ 阶混合中心矩 若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$ 存在，则称为 X 和 Y 的 $k+l$ 阶混合中心矩.

4.4.1 协方差矩阵

$(X_1, X_2, \cdots, X_n)$ 中分量的两两的协方差都存在, 则协方差矩阵为

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

---

n 维正态分布 n 维正态随机变量 $(X_1, X_2, \cdots, X_n)$ 的概率密度定义为

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\boldsymbol{C}|}} \exp[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})],$$

其中 $\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$

5 大数定律及中心极限定理

5.1 弱大数定律 (辛钦大数定律)

依概率收敛

若 $X_1, X_2, \cdots$ 是一个无穷的随机变量列, 若 $\forall \varepsilon > 0$ 都有

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1,$$

则称序列 $X_1, X_2, \cdots$ 依概率收敛于随机变量 X(退化情况为常数 a),
记为 $X_n \xrightarrow{P} X$.

若 $X_1, X_2, \cdots, X_n$ 相互独立且服从同一分布, 且有 $E(X_k) = \mu$, 则 $\boxed{\overline{X} \xrightarrow{P} \mu}$.
n 充分大时, 频率接近概率.

依分布收敛 *

若 $X_1, X_2, \cdots$ 是一个无穷的随机变量列, 分布函数为 $F_1(x), F_2(x), \cdots$, 如果对 X 的分布函数 $F(x)$ 的任意连续点 x, 都有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称随机变量列 $X_1, X_2, \cdots$ 依分布收敛于 X, 记为 $X_n \xrightarrow{L} X$.

5.1.1 伯努利大数定律

若 $f_A \sim B(n, p)$, 则 $\dfrac{f_A}{n} \xrightarrow{L} p$.

5.2 中心极限定理 (独立同分布)

若 $X_1, X_2, \cdots$ 相互独立且服从同一分布, 且有 $E(X_k) = \mu, \sqrt{D(X_k)} = \sigma$, 设 $S_n = X_1 + \cdots + X_n$, 则

$$\boxed{\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{L} X, \text{其中} X \sim N(0,1).}$$

Tips

大数定律 定性
切比雪夫不等式 定量
中心极限定理 较精确的定量

6 样本及抽样分布

基本概念 总体 $\supseteq$ 样本 $\supseteq$ 个体

总体的容量 $>$ 样本容量

样本的性质 独立性、代表性.

样本值 样本 $(X_1, \cdots, X_n)$ 的值 $(x_1, \cdots, x_n)$.

样本分布 $F(x_1, \cdots, x_n) = \prod_{i=1}^{n} F(x_i)$.

样本密度函数 $f(x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$.

样本的概率分布 $p(x_1, \cdots, x_n) = \prod_{i=1}^{n} p(x_i)$.

统计量 (statistic) 样本 $(X_1, \cdots, X_n)$ 的函数 $g(X_1, \cdots, X_n)$, 函数 g 必须是确定的, 不含未知参数,g 的取值可以是一维、多维的.

6.1 常用的统计量

样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

未修正样本方差 $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

修正样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

样本标准差 $S = \sqrt{S^2}$.

k **阶样本原点矩** $a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.

k **阶样本中心矩** $m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{(}X))^k$.

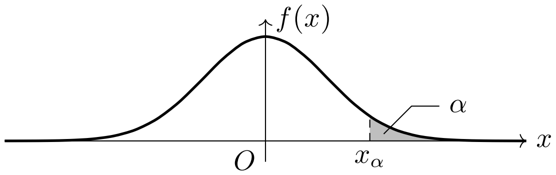
顺序统计量 随机变量 $X_1, \cdots, X_n$ 互相独立且同分布, 分布函数为 $F(x)$, 将其从小到大排序:$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, 则 $X_{(k)}$ 的分布函数为 $F_k(x) = \sum_{i=1}^{n} C_n^i F^i(x)[1 - f(x)]^{n-i}$.

样本极小、极大值 $X_{(1)}$、$X_{(n)}$

样本极差 $X_{(n)} - X_{(1)}$

6.2 分位数

上侧 α 分位数 若 $\alpha \in (0,1)$, $x_\alpha \in \mathbb{R}$ 满足 $P\{X > x_\alpha\} = \alpha$, 则称 x_α 为 X 的上侧 α 分位数.



通常, 对于上 α 分位数,

标准正态分布的记作 z_α,

$\chi^2(n)$ 分布的记作 $\chi^2_\alpha(n)$,

$t(n)$ 分布的记作 $t_\alpha(n)$,

$F(n_1, n_2)$ 分布的记作 $F_\alpha(n_1, n_2)$.

双侧 α 分位数 设 X 是对称分布 (密度函数是偶函数) 的连续性随机变量, 若 $\alpha \in (0,1)$, $T_\alpha \in \mathbb{R}$ 满足 $P\{|X| > T_\alpha\} = \alpha$, 则称 T_α 为 X 的双侧 α 分位数.

6.3 χ^2 分布

$X = X_1^2 + \cdots + X_n^2$, $X_1^2 + \cdots + X_n^2 \sim N(0,1)$ 且相互独立, 则称 X 服从自由度为 n 的 χ^2 分布, 记为 $X \sim \chi^2(n)$.

- $E(X) = n, D(X) = 2n$.
- **可加性** 若 $X \sim \chi^2(m), Y \sim \chi^2(n)$, 则 $X + Y \sim \chi^2(m+n)$.
- $n = 1, 2$ 时单调递减;$n \geq 3$ 时先增后减, 极大值点为 $x = n - 2$.
- $\chi^2(2) = e\left(\frac{1}{2}\right)$.(指数分布)
- n 充分大时,(如 $n > 40$), $\chi^2_\alpha(n) \approx \frac{1}{2}\left(z_\alpha + \sqrt{2n-1}\right)$.

6.4 F 分布

$X = \dfrac{U/m}{V/n}$, $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ X, Y 相互独立, 则称 X 服从第一自由度为 m, 第二自由度为 n 的 F 分布, 记为 $X \sim F(m, n)$.

- 若 $X \sim F(m,n)$, 则 $\dfrac{1}{X} \sim F(n,m)$; $F_{1-\alpha}(m,n) = \dfrac{1}{F_\alpha(n,m)}$

6.5 t 分布

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

- 当 n 充分大 (如 $n > 50$), $t(n)$ 近似于 $N(0,1)$.
- $E(X) = 0, D(X) = \dfrac{n}{n-2}$

6.6 正态总体的抽样分布 (小样本分布)

6.6.1 单正态总体的抽样分布

若 $X \sim N(\mu, \sigma^2)$, $X_1, \cdots, X_n$ 为样本, $\overline{X}$ 和 S^2 分别为样本均值和样本方差 (修正后), 有

$$\left\{\begin{array}{l} U = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ \boxed{\begin{array}{l}\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \\ (\mu\text{已知时用于推断}\sigma^2)\end{array}} \\ \overline{X}(\text{或}U)\text{ 与}S^2\text{相互独立} \end{array}\right. \Longrightarrow \boxed{\begin{array}{l} T = \dfrac{U}{S}\sigma = \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1). \\ (\sigma^2\text{已知时用于推断}\mu) \end{array}}$$

6.6.2 双正态总体的抽样分布

若 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $X_1, \cdots, X_{n_1}$ 与 $Y_1, \cdots, Y_{n_2}$ 为样本, $\overline{X}, \overline{Y}$ 和 S_1^2, S_2^2 分别为样本均值和样本方差 (修正后), 有

$$\left\{\begin{array}{l} U = \dfrac{(\overline{X}-\overline{Y})-(\mu_1,\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1) \\ \boxed{\begin{array}{l} F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1) \\ (\mu_1-\mu_2\text{已知时用于推断}\dfrac{\sigma_1^2}{\sigma_2^2}) \end{array}} \\ U\text{与}F\text{相互独立} \\ \text{设加权平均}S^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \end{array}\right. \Longrightarrow \boxed{\begin{array}{l} \text{当}\sigma_1=\sigma_2=\sigma, \\ T = \dfrac{U}{S}\sigma = \dfrac{(\overline{X}-\overline{Y})-(\mu_1,\mu_2)}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \\ \qquad \sim t(n_1+n_2-2). \\ (\dfrac{\sigma_1^2}{\sigma_2^2}\text{已知时用于推断}\mu_1-\mu_2) \end{array}}$$

6.7 一般总体的抽样分布的极限分布 *

若 $X_1, \cdots, X_n$ 为总体 X 的样本, $\overline{X}$ 和 S^2 分别为样本均值和样本方差 (修正后), 设 $E(X)=\mu, D(X)=\sigma^2$, 有

$$U_n = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{L} N(0,1) \quad (\sigma^2\textbf{已知}\text{时用于推断}\mu)$$

$$T_n = \frac{\overline{X}-\mu}{S/\sqrt{n}} \xrightarrow{L} N(0,1) \quad (\sigma^2\textbf{未知}\text{时用于推断}\mu)$$

7 参数估计

7.1 点估计

估计 $\hat{\theta}(X_1,\cdots,X_n)$ 为参数 θ 的**估计量**,$\hat{\theta}(x_1,\cdots,x_n)$ 为参数 θ 的**估计值**, 二者统称为**估计**.

7.1.1 矩估计法 (ME 法)

设总体 X 的分布依赖于参数 $\theta=(\theta_1,\cdots,\theta_m)$, $X_1,\cdots,X_n$ 是样本, 估计参数 θ 的过程如下:

Step1 由于各阶矩是参数的函数, 故可以列出方程组:

$$\begin{cases}\alpha_1=E(X)=g_1\left(\theta_1,\cdots,\theta_m\right),\\ \alpha_2=E(X^2)=g_2\left(\theta_1,\cdots,\theta_m\right),\\ \quad\vdots\\ \alpha_m=E(X^m)=g_m\left(\theta_1,\cdots,\theta_m\right).\end{cases}\text{一般情况下可解得}\begin{cases}\theta_1=h_1\left(\alpha_1,\cdots,\alpha_m\right)\\ \theta_2=h_1\left(\alpha_2,\cdots,\alpha_m\right)\\ \quad\vdots\\ \theta_m=h_m\left(\alpha_1,\cdots,\alpha_m\right).\end{cases}$$

注: 有时会出现某一个方程 (通常是期望的方程) 是恒等方程, 则需要把 $m+1$ 阶矩的方程也列出, 才能解出方程.(有效方程个数 = 未知参数个数)

Step2 计算

$$\begin{cases}A_1=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i\\ \quad\vdots\\ A_m=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^m\end{cases}\text{或}\begin{cases}a_1=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i\\ \quad\vdots\\ a_m=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i^m\end{cases}$$

再用$A_1,\cdots,A_m$或$a_1,\cdots,a_m$代替上式中的$\alpha_1,\cdots,\alpha_m$, 得到**矩估计量或矩估计值**

$$\begin{cases}\hat{\theta}_1=h_1(A_1,\cdots,A_m)\\ \quad\vdots\\ \hat{\theta}_m=h_m(A_1,\cdots,A_m).\end{cases}\qquad\text{或}\begin{cases}\hat{\theta}_1=h_1(a_1,\cdots,a_m)\\ \quad\vdots\\ \hat{\theta}_m=h_m(a_1,\cdots,a_m).\end{cases}$$

7.1.2 最大似然估计法 (MLE 法)

Step 1

1° **离散型** 若总体 X 属离散型, 其分布律 $P\{X=x\}=p(x;\theta_1,\cdots,\theta_m),\theta_i\in\Theta_i$ 的形式已知$\Big($如 $B(n,p)$, 未知参数$\theta_1=n,\theta_2=p\Big)$, 则事件 $\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$ 发生的概率为

$$L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_m)=\prod_{i=1}^{n}p(x_i;\theta_1,\cdots,\theta_m).$$

2° **连续型** 若总体 X 属离散型, 其密度函数 $f\{X=x\}=f(x;\theta_1,\cdots,\theta_m),\theta_i\in\Theta_i$ 的形式已知, 可设

$$L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_m)=\prod_{i=1}^{n}f(x_i;\theta_1,\cdots,\theta_m).$$

Step 2

取 $\hat{\theta}_1,\cdots,\hat{\theta}_m$ 使

$$L(x_1,x_2,\cdots,x_n;\hat{\theta}_1,\cdots,\hat{\theta}_m)=\max_{\theta_i\in\Theta_i}L(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_m).$$

这样得到的 $\hat{\theta}_i$ 与样本值 $x_1,x_2,\cdots,x_n$ 有关, 常记为 $\hat{\theta}_i(x_1,x_2,\cdots,x_n)$, 称为参数 θ_i 的**最大似然估计值**, 相应的统计量 $\hat{\theta}_i(X_1,X_2,\cdots,X_n)$ 称为参数 θ_i 的**最大似然估计量**.

Tips

连续型时, **Step 2** 通常是一个极值问题, 可以转化为求解以下方程

$$\frac{\mathrm{d}}{\mathrm{d}\theta_i}L(\theta_1,\cdots,\theta_m)=0,$$

若 $L>0$, 有时 $L(\theta_1,\cdots,\theta_m)$ 的对数更容易求导, 则可转化为**对数似然方程**, 即

$$\frac{\mathrm{d}}{\mathrm{d}\theta_i}\ln\Big(L(\theta_1,\cdots,\theta_m)\Big)=0.$$

注: 并不是所有连续型都可以通过以上两条方程解出极值点, 有时还要考虑定义域端点是否为最大值点.

7.2 估计量的评选标准

7.2.1 无偏性

设 $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ 是参数 θ 的估计量, 若 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的**无偏估计**.

若 $\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的**渐进无偏估计**.

7.2.2 有效性

设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是参数 θ 的**无偏估计量** $\forall \theta \in \Theta$, 都有 $D(\hat{\theta}_1) \le D(\hat{\theta}_1)$, 则一般情况下可称 $\hat{\theta}_1$ **较** $\hat{\theta}_2$ **有效**.

设 $\hat{\theta}_0$ 是参数 θ 的**无偏估计量**, 若对 $\forall \theta$ 的无偏估计量 $\hat{\theta}$, 有 $D(\hat{\theta}_0) \le D(\hat{\theta})$, 则称 $\hat{\theta}_0$ 为 θ 的**最有效无偏估计**.

7.2.3 相合性

若 $\hat{\theta} \xrightarrow{P} \theta$, 即对 $\forall \varepsilon > 0$, $\lim\limits_{n\to\infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1$, 则称 $\hat{\theta}$ 为 θ 的 (弱) **相合估计**.

若 $\lim\limits_{n\to\infty} P\{\hat{\theta} = \theta\} = 1$, 则称 $\hat{\theta}$ 为 θ 的强相合估计.