大学公共数学系列教材

概率论与数理统计

第二版

武汉大学数学与统计学院

齐民友 主编

刘禄勤 王文祥 龚小庆 编

高等教育出版社
HIGHER EDUCATION PRESS

日期: /

随机事件

随机试验: 用 E 表示.

样本点/基本事件: 用 ω 表示.

样本空间: 所有可能的结果形成的集合, 用 Ω 表示

必然事件: 样本空间 Ω.

不可能事件: 空集 $\varnothing$

① 事件的包含与相等 $A \subset B$

② 并: $A \cup B$. 和事件

③ 交. $A \cap B$. 积事件. AB

④ 差: $A-B$、A发生 B不发生.

⑤ 逆、互斥: $\bar{A}$, A不发生的概率.

⑥ 互斥事件: A与B不能同时发生.

⑦ 不相容: $AB=\varnothing$. 两事件没有交集.

运算规律:

- 交换律: $A \cup B = B \cup A$
- 结合律: $A \cup (B \cup C) = (A \cup B) \cup C$
- 分配律: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ → 只要符号不同拿出来的时候都要分配
- 德摩根律. $\overline{A \cup B} = \bar{A} \cap \bar{B}$. → 并(交)集一起的非 可以转为先非再交(并)
- 不相容分解: $A \cup B = A \cup \bar{A}B$、$A = AB \cup A\bar{B}$

日期: /

所有事件的并集在其中

事件域: ①包含Ω与ϕ ②若A在其中则 $\bar{A}$也在其中 ③ $A_n \in F$ 则 $\bigcup_{n=1}^{\infty} A_n \in F$

↘F是概率的定义域，不在F中的事件没有意义.

频率 → 随实验次数发生变化. $f_n(A)$

- 非负性: $f_n(A) \geq 0$
- 规范性: $f_n(\Omega) = 1$.
- 可列可加性: $f_n\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} f_n(A_k)$

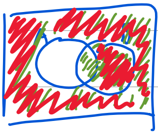
事件并的频率 = 所有事件的加和.

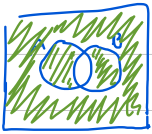
概率空间:

概率: P ⇒ 公理化定义:

- 非负性: $f_n(A) \geq 0$
- 规范性: $f_n(\Omega) = 1$.
- 可列可加性: $f_n\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} f_n(A_k)$

性质





① $P(\phi) = 0$ ✱反过来不对. → 在数轴上的长度表示概率的话 某一点P=0 但可能发生.

② $P(A) > 0 \Rightarrow A \neq \phi$.

③ 互不相容的事件 A_n: $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{n} P(A_k)$

④ 可减性: $A \subset B \Rightarrow P(B-A) = P(B) - P(A)$

⑤ 单调性: $A \subset B \Rightarrow P(A) \leq P(B)$

⑥ 逆事件: $P(\bar{A}) = 1 - P(A)$

⑦ $P(A \cup B) = P(A) + P(B) - P(AB)$

⑧ $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{1 \leq i < j \leq n} P(A_iA_j) + \sum_{1 \leq i < j < k \leq n} P(A_iA_jA_k) - \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n)$

日期: /

条件概率

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

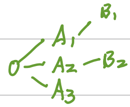
乘法公式：$P(AB)=P(B)P(A|B)$

一般乘法公式：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_{n-1}A_{n-2}\cdots A_1)$$

全概率公式

$$P(A)=\sum_{i=1}^{n}\underbrace{P(B_i)}_{\text{选到}A_i\text{概率}}\underbrace{P(A|B_i)}_{\text{在}A\text{中发生}B_i\text{概率}}$$



贝叶斯公式

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$$

$$\Rightarrow P(\overset{\text{原因}}{B}|\overset{\text{结果}}{A})=\frac{P(B)P(A|B)}{P(B)P(A|B)+P(\bar{B})P(A|\bar{B})}$$

由结果来分析原因，检测阳性后真患病概率

日期:　　/

事件独立性.

$P(AB) = P(A)P(B) \Longleftrightarrow$ A、B互不影响 $\Longleftrightarrow$ A、B相互独立

☆注. 3个事件相互独立 $\Leftrightarrow$ 两两独立且 3个独立.

伯努利试验

$P(B_k) = C_n^k p^k q^{n-k}$.

二、随机变量

随机变量定义：在Ω中对于任意的x，$\{X\leq x\}$是随机事件即

$$\{\omega: X(\omega)\leq x\}\in F.$$

随机变量表示：

$$\{X<x\}=\bigcup_{k=1}^{\infty}\left\{X\leq x-\frac{1}{k}\right\}\in F$$

$$\{X=x\}=\{X\leq x\}-\{X<x\}\in F$$

$$\{a<X\leq b\}=\{X\leq b\}-\{X\leq a\}\in F$$

$F(x)=P(X\leq x)$ 为X的**分布函数**

$$\begin{cases} P(a<X\leq b)=P(X\leq b)-P(X\leq a)=F(b)-F(a). \\ P(X>x)=1-F(x) \\ P(X<x)=F(x-0) \\ P(X=x)=F(x)-F(x-0) \end{cases}$$

$\rightarrow$ 分布函数在x点左极限.

定理：X的分布函数为$F(x)$则有：

$$\begin{cases} F(x)\text{是}x\text{的单调增函数} \\ F(x)\text{是右连续函数} \\ F(-\infty)=0,\quad F(\infty)=1 \end{cases}$$

$\rightarrow$ 多一部分事件多一份概率.

退化分布(单点分布)：$F(x)=\begin{cases} 0 & x<a \\ 1 & x\geq a. \end{cases}$

两点分布(伯努利分布)：$F(x)=\begin{cases} 0 & x<0 \\ 1-p & 0\leq x<1 \\ 1 & x\geq 1 \end{cases}$

日期: /

离散型随机变量

$\boxed{\text{离散型随机变量}}$：X只取有限个或可列无穷多个 $p_k = P(X = x_k)$ 分布律

分布律：表格

分布函数：画数轴按区间分"[ )"。

$\boxed{\text{二项分布}}$：X为随机变量且取值 $0,1,\cdots n$ 分布律为 $P_k = P(X=k) = C_n^k p^k q^{n-k}$

记为 $X \sim B(n,p)$，$b(k;n,p)$

$\boxed{\text{泊松定理}}$：若 $\lim\limits_{n\to\infty} n p_n = \lambda$ 则有 $\lim\limits_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$。

注 p 要 ≤ 0.05，$n \geq 20$。

$\boxed{\text{泊松分布}}$：X为随机变量且取值 $0,1,\cdots n$ 分布律为 $P(k;\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$。

日期: /

连续型随机变量

概率密度:

其中的 $F(x)$ 为连续型随机变量

- $F(x)$ 为事件的分布函数, 则 $F(x)=\int_{-\infty}^{x} f(y)dy$, 称 $f(y)$ 为概率密度.
- $f(x)\geq 0$, $-\infty < x < \infty$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$ → 与 X 轴围成面积为 1.
- $P(a<X\leq b)=\int_a^b f(x)dx$
- $F(x)$ 是连续函数 且当 $f(x)$ 在 $x=x_0$ 点连续时 $F'(x_0)=f(x_0)$
- 对任意常数 c, $P(X=c)=0$ → 无须在意某(个)点的情况

均匀分布

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

记为 $X\sim U[a,b]$

指数分布

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0\\ 0 & \text{其他}\end{cases}$$

记为 $X\sim \mathcal{E}(\lambda)$

一般用作各种"寿命"的分布的近似,

"无记忆性" $\Rightarrow P(X>s+t\mid X>s)=P(X>t)$

日期:　　/

正态分布（高斯分布）

1. $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$　记作 $X\sim N(\mu,\sigma^2)$

正态密度

2. 性质：
① 最大值点在 $(\mu,\frac{1}{\sqrt{2\pi}\sigma})$ 处
② 图形关于 $x=\mu$ 对称
③ 图形拐点在 $x=\mu\pm\sigma$ 处
④ $x\to\infty$ 时，曲线以 x 轴为渐近线.
⑤ σ 越小，图形越尖.

3. 标准正态分布:

$$\begin{cases}\mu=0,\ \sigma=1 \quad N(0,1)\\ \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad \Phi=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy\end{cases}$$

4. 一般正态分布

$$F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

5. 对数正态分布

$$f_Y(y)=\begin{cases}\frac{1}{y\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}} & y>0\\ 0 & y\le 0\end{cases}$$

日期:　　　/

连续型随机变量函数的分布

已知 X 的概率密度，求 $Y=g(x)$ 的概率密度 $f_Y(y)$

1 分布函数微分法：

对 X 有 $f_X(x)$，$Y=aX+b$，求 $f_Y(y)$

$F_Y(y) = P(aX+b \le y)$ → $P(Y\le y)$ 分布函数定义

$$= \begin{cases} P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) & a>0 \\ P\left(X \ge \frac{y-b}{a}\right) = 1-F_X\left(\frac{y-b}{a}\right) & a<0 \end{cases}$$

考虑 a 的正负

→ 对两边同时求导

$$f_Y(y) = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right) & a>0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) & a<0 \end{cases}$$

$$= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

2. 积分转化法. 设 $h(x)$

令 $y=g(x)$ 用 $x=\cdots y$ 代换掉 x，只剩 y

$$\int_{-\infty}^{\infty} h[\underbrace{g(x)}_{\text{已知}}] \underbrace{f_X(x)}_{\text{已知}} dx = \int_{\alpha}^{\beta} h(y) p(y) dy$$

得出的 $f_Y(y) = \begin{cases} p(y) & \alpha<y<\beta \\ 0 & 其他 \end{cases}$

日期: /

3.1 二维随机变量及其联合分布函数

1.定义: $\underbrace{F(x,y)}_{\text{二元函数}} = P(X \le x, Y \le y)$

二元函数 ⇓ 二维联合分布函数

二维随机变量

2. 定理
① $P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1) \ge 0$
② $F(x,y)$ 关于 x 和 y 都是右连续的.
③ $F(-\infty,y) = 0$, $F(x,-\infty) = 0$, $F(\infty,\infty) = 1$

3. $F_X(x) = F(x,\infty)$, $F_Y(y) = F(\infty,y)$ 为边缘分布函数

日期:　　/

二维连续型随机变量

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\underbrace{f(u,v)}_{\text{概率密度}}\,du\,dv$$

对面积为零的区域概率为0

$$P(x_1<X\leqslant x_2,\ y_1<Y\leqslant y_2)=\int_{x_1}^{x_2}\int_{y_1}^{y_2}f(x,y)\,dx\,dy$$

$$P((X,Y)\in G)=\iint_G f(x,y)\,dx\,dy$$

- 已知 $f(x,y)$ 求 $P(G)$ $F(x,y)$ $\Rightarrow$ 直接对区域积分

  $F_X(x)$ $\Rightarrow$ $y\to\infty$ 取极限 / 用 $f(x,y)$ 积分，y 用 $(-\infty,\infty)$

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx \qquad f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy$$

二维均匀分布

$$f(x,y)=\begin{cases}\dfrac{1}{S(D)} & (x,y)\in D\\ 0 & \text{其他}\end{cases}$$

$$P((X,Y)\in G)=\iint_{S(G)}\frac{1}{S(D)}\,dx\,dy=\frac{S(G)}{S(D)}$$

二维正态分布

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

$|\rho|<1$　　记作 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

二维正态分布的两个边缘分布都是一维正态分布. 反之不成立

日期: /

随机变量的独立性

$F(x,y)=F_X(x)\cdot F_Y(y)$ 则称X与Y相互独立

$\Downarrow$

在一切公共连续点上成立
若除有限个点或可列无穷个点外成立 $F(x,y)=F_X(x)\cdot F_Y(y) \Longleftrightarrow$ X与Y独立
存在G(面积非零)使 $F(x,y)\neq F_X(x)\cdot F_Y(y)$ 则X与Y不独立

X与Y相互独立 $\Longleftrightarrow \rho=0$

日期: /

3.6 随机变量函数的分布

已知 $X\ Y$ 的联合分布 和 $g(x,y)$ 求 $Z=g(X,Y)$ 的概率分布

①离散型：
- 1> 确定 $Z=g(x,y)$ 的取值范围
- 2> 求 Z 取每个可能值的概率

$$P(Z=k)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)$$

- $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$ 且 X 与 Y 相互独立 $\longrightarrow X+Y\sim P(\lambda_1+\lambda_2)$
- $X\sim B(n_1,p)$，$Y\sim B(n_2,p)$ 且 X 与 Y 相互独立 $\longrightarrow X+Y\sim B(n_1+n_2,p)$

②连续型
- 分布函数微分法：$Z=aX+bY$，$Z=\frac{X}{Y}$，$M=\max\{X,Y\}$，$N=\min\{X,Y\}$
- 积分转化法

日期: /

★分布函数微分法

1> $Z=X+Y$ $\Rightarrow$ $x=u-y$, $u=z$

$$F(z)=\int_{-\infty}^{z} f(u)\,du$$

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx \quad XY独立$$

$$\parallel \int_{-\infty}^{\infty} f(x,z-x)\,dx$$ →把y换掉

$\left.\begin{array}{l} X\sim N(\mu_1,\sigma_1^2),\ Y\sim N(\mu_2,\sigma_2^2) \\ X与Y独立 \end{array}\right\} \Rightarrow X+Y\sim N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$

2> $Z=aX+bY$ →找出该直线从下到上移动改变线与y轴交点从而改变到情况与积分区域有交集后积分

3> $Z=\frac{X}{Y}$: $f_Z(z)=\int_{-\infty}^{\infty}|y|f(yz,y)\,dy$

4> $M=\max\{X,Y\}$ $N=\min\{X,Y\}$ X与Y独立

$M \Rightarrow F_{max}(z)=F_X(z)F_Y(z)$

$N \Rightarrow F_{min}(z)=1-[1-F_X(z)][1-F_Y(z)]$

日期: /

积分转化法:

(X,Y) 的联合概率密度 $f(x,y)$ $\quad z=g(x,y)$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h[g(x,y)]f(x,y)\,dx\,dy=\int_{\alpha}^{\beta} h(z)\,p(z)\,dz$$

$$f(z)=\begin{cases} p(z) & \alpha<z<\beta \\ 0 & \text{其他} \end{cases}$$

$dy=\sim dz$

把dy换掉

日期: /

4.1 数学期望

离散型随机变量的数学期望

已知分布律

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

当 $\sum_{n\geq1}|x_n|p_n<\infty$ 时 数学期望存在

$\sum_{n\geq1}x_np_n$ 为X的期望 $E(X)$

若 $X\sim B(n,p)$ 则 $E(X)=np$ → 二项分布

$X\sim P(\lambda)$ 则 $E(X)=\lambda$ → 泊松分布

连续型随机变量的期望

对X的取值范围

已知概率密度 $f(x)$ 如果 $\int_{-\infty}^{\infty}|x|f(x)dx<\infty$ 则期望存在，

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx$$

均匀分布: $X\sim U(a,b)=\frac{a+b}{2}$

指数分布: $X\sim E(\lambda)=\frac{1}{\lambda}$

正态分布: $X\sim N(\mu,\sigma^2)=\mu$

$Y=g(X)$ 的数学期望

注意判断是否存在

一维

- 若X为离散型 $E(Y)=E(g(X))=\sum_{k\geq1}g(x_k)p_k$
- 若X为连续型 $E(Y)=E(g(X))=\int_{-\infty}^{\infty}g(x)f(x)dx$

二维

- 已知 $f(x,y)$　$g(x,y)=Z$，　$E(Z)=E(g(X,Y))=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)dxdy$

日期: /

期望的性质

$E(cX) = cE(X)$

$E(X_1+X_2) = E(X_1)+E(X_2)$

若 $X \leq Y$ 则 $E(X) \leq E(Y)$

若 $a \leq X \leq b$ 则 $a \leq E(X) \leq b$

若 $X_1, X_2 \cdots X_n$ 相互独立 $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$

$|E(X)| \leq E(|X|)$

马尔可夫不等式 $P(|X| \geq c) \leq \frac{E(|X|)}{c}$

$E(|X|)=0$ ，则 $P(X=0)=1$

日期: /

方差

定义: $D(X)=E([X-E(X)]^2)=E(X^2)-[E(X)]^2$

$\sigma_X=\sqrt{D(X)}$ 标准差, 均方差

离散型: $D(X)=\sum_{k\geq1}[X_k-E(X)]^2 p_k$

连续型: $D(X)=\int_{-\infty}^{\infty} f(x)\cdot[x-E(X)]^2 dx$

两点分布: $E(X)=P$, $D(X)=P(1-P)$

泊松分布: $X\sim P(\lambda)$ $E(X)=\lambda$, $D(X)=\lambda$ $\rightarrow$ $\frac{\lambda^k}{k!}e^{-\lambda}$

均匀分布: $X\sim U(a,b)$ $E(X)=\frac{a+b}{2}$, $D(X)=\frac{(b-a)^2}{12}$

指数分布: $X\sim E(\lambda)$ $E(X)=\frac{1}{\lambda}$, $D(X)=\frac{1}{\lambda^2}$ $\rightarrow$ $\lambda e^{-\lambda x}$

正态分布: $X\sim N(\mu,\sigma^2)$. $E(X)=\mu$, $D(X)=\sigma^2$

性质:

- C 是常数: $D(C)=0$
- $D(CX)=C^2D(X)$, $D(-X)=D(X)$
- X,Y 相互独立则 $D(X\pm Y)=D(X)+D(Y)$
- $D(X)\leqslant E((X-C)^2)$ 当且仅当 $E(X)=C$ 时等号成立
- 若 $D(X)=0$ 则 $P(X=E(X))=1$

$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma^2)$, X,Y 独立. $aX+bY\sim N(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2)$

日期: /

协方差与相关系数

协方差：$cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

性质：

$$\begin{cases} X与Y相互独立 \rightarrow cov(X,Y)=0 \\ D(X\pm Y)=D(X)+D(Y)\pm 2cov(X,Y) \\ cov(X,Y)=E(XY)-E(X)E(Y) \\ cov(X,Y)=cov(Y,X) \\ cov(aX+c,bY+d)=ab\,cov(X,Y) \\ cov(X_1+X_2,Y)=cov(X_1,Y)+cov(X_2,Y) \end{cases}$$

相关系数：$\rho_{XY}=\dfrac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\dfrac{E\{[X-E(X)][Y-E(Y)]\}}{\sqrt{D(X)}\sqrt{D(Y)}}$

性质：

- $|E(XY)|^2\leq E(X^2)E(Y^2)$ 当且仅当 $P(Y=t_0X)=1$ 等式成立
- $(cov(X,Y))^2\leq D(X)D(Y)$ 当且仅当 $P(Y=aX+b)=1$ 等式成立
- $|cov(X,Y)|\leq\sqrt{D(X)}\sqrt{D(Y)}$ $\Rightarrow|\rho_{XY}|\leq 1$ 当且仅当 X与Y线性相关等式成立
- $|\rho_{XY}|=1\Rightarrow$ X与Y线性相关
- $|\rho_{XY}|<1$ 时 线性关联程度随 $\rho\downarrow$ 而 $\downarrow$
- $|\rho_{XY}|=0$ 时 X与Y不相关 但可能独立
- $\rho_{XY}>0$ 正相关，$\rho_{XY}<0$ 负相关

等价命题：

$$\begin{cases} X与Y不相关 \Leftrightarrow \rho_{XY}=0 \Leftrightarrow cov(X,Y)=0 \\ \Updownarrow \\ D(X+Y)=D(X)+D(Y) \Leftrightarrow E(XY)=E(X)E(Y) \end{cases}$$

日期:　　/

4.4

矩： $E(X^k)$ 称X的k阶原点矩，$E(x)$ 为X的一阶原点矩

$E((X-E(X))^k)$ 称X的k阶中心矩，$D(x)$ 为X的二阶中心矩

中位数与分位数

若 $F(x_\alpha)=P(X\leqslant x_\alpha)=\alpha$ 称 x_α 是X的下α分位数

$F(u_\alpha)=P(X\geqslant u_\alpha)=\alpha$ 称 u_α 是X的上α分位数

$\alpha=0.5$ 时为中位数　　$u_\alpha=1-x_\alpha$

切比雪夫不等式：

对任意 $\varepsilon>0$ 有 $\begin{cases} P(|X-E(x)|\geqslant\varepsilon)\leqslant\dfrac{D(x)}{\varepsilon^2} \\ P(|X-E(x)|<\varepsilon)\geqslant 1-\dfrac{D(x)}{\varepsilon^2} \end{cases}$

日期: /

5-1

依概率收敛

对任意 $\varepsilon>0$ 有 $\lim\limits_{n\to\infty} P(|X_n|>\varepsilon)=0$ 称 X_n 依概率收敛于0 $X_n \xrightarrow{P} 0$

n足够大时 $|X_n|>\varepsilon$ 发生的概率很小

大数定律:

若随机变量 $\{X_n: n=1,2\cdots\}$ 满足

$E(\bar{X})$

$$\lim_{n\to\infty} P\left(\left|\underbrace{\frac{1}{n}\sum_{k=1}^{n}X_k}_{\bar{X}} - \frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|>\varepsilon\right)=0$$

即 $\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n}E(X_k)$

则称 $\{X_n: n=1,2\cdots\}$ 服从大数定律

切比雪夫大数定律:

$E(X_n)=\mu$, $D(X_n)=\sigma^2$ 则有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k - \frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|>\varepsilon\right)=0$$

即 $\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{P} \mu$

伯努利大数定律:

n_A 为事件A发生的次数, A在每次实验中概率为 p, 对任意 $\varepsilon>0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n}-p\right|<\varepsilon\right)=1$$

辛钦大数定律

$E(X_n)=\mu$ $\quad$ $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k - \frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|>\varepsilon\right)=0$

$$\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{P} \mu$$

日期:　　/

中心极限定理

无论X_n服从什么分布，只要n足够大，则

$$\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right) \text{ 服从 } N(0,1)$$

$$\sum_{k=1}^{n} X_k \text{ 服从 } N(n\mu, n\sigma^2)$$

n_A为n重伯努利实验中A发生的次数，对任意x

$$\lim_{n\to\infty} P\left(\frac{n_A - np}{\sqrt{np(1-p)}} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

日期: /

数理统计

X代表样本，x代表观察值

$g(X_1, X_2, \cdots, X_n)$为统计量，$g(x_1, x_2, \cdots, x_n)$为统计量的观察值

样本均值：$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{观察值}} \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$

样本方差：$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 \longrightarrow s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$

样本标准差：$S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2} \longrightarrow s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}$

样本k阶矩：$A_k = \frac{1}{n}\sum_{k=1}^{n} X_i^k \longrightarrow a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$

样本k阶中心矩：$B_k = \frac{1}{n}\sum_{k=1}^{n}(X_i - \bar{X})^k \longrightarrow b_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$

$E(X) = \mu$，$D(X) = \sigma^2$，$E(\bar{X}) = \mu$，$D(\bar{X}) = \frac{\sigma^2}{n}$，$E(S^2) = \sigma^2$

$E(Z^2) = E^2(Z) + D(Z)$

日期: 6.2 /

样本的顺序统计量：按样本的观察值排序的样本. $X_{(1)}, X_{(2)} \cdots$

$X_{(1)}$ 称最小统计量, $X_{(n)}$ 称最大统计量

$X_{(n)} - X_{(1)}$ 称为极差

$$\hat{F}_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{1}{n}, & x_{(1)} \le x < x_{(2)} \\ \frac{k}{n}, & x_{(k)} \le x < x_{(k+1)} \\ 1 & x_{(n)} < x \end{cases}$$

抽样分布

1. χ^2分布：$X_1, X_2 \cdots X_n$ 相互独立且服从 $N(0,1)$ 称 $\chi^2 = X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2$

称: $\chi^2 \sim \chi^2(n)$

自由度

若 $X_1^2 \sim \chi^2(m)$, $X_2^2 \sim \chi^2(n)$ 且 X_1^2 与 X_2^2 独立 则有 $X_1^2 + X_2^2 \sim \chi^2(m+n)$

若 $\chi^2 \sim \chi^2(n)$, 则 $E(\chi^2) = n$, $D(\chi^2) = 2n$

2. t分布: $X \sim N(0,1)$. $Y \sim \chi^2(n)$ 且X与Y独立

则 $T = \dfrac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)$

密度函数 $\lim\limits_{n\to\infty} h(t) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

3. $X \sim \chi^2(n_1)$. $Y \sim \chi^2(n_2)$ 且X与Y独立

则 $F = \dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

定理：①若 $F \sim F(n_1, n_2)$ 则 $\frac{1}{F} \sim F(n_2, n_1)$

②若 $T \sim t(n)$ 则 $T^2 \sim F(1, n)$

日期: /

上分位点:

对 $\chi^2 \Rightarrow P(X > \chi^2_\alpha(n)) = \alpha$ 

对 $t \Rightarrow P(X > t_\alpha(n)) = \alpha$ 

对 $F \Rightarrow P(X > F_\alpha(n_1, n_2)) = \alpha$ 

正态总体的抽样分布:

标准化

一个正态总体: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$\Rightarrow$ 根据变量公式, 标准化之后查表

$\bar{X}$与S^2相互独立 $\quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

两个正态总体: $X_1, X_2 \cdots X_m \sim N(\mu_1, \sigma^2)$

$Y_1, Y_2, \cdots Y_n \sim N(\mu_2, \sigma^2)$ 相互独立

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

日期:　　　/