概率论与数理统计笔记

made by LaTeX

NP_123

2022 年 5 月 30 日

概率论的基本概念

公式

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \qquad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

$$P(A-B) = P(A) - P(AB) = P(A\bar{B})$$

$$P(A_1+A_2+A_3) = P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)$$

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

$$P(AB) = P(A)P(B) \Longleftrightarrow \text{A 与 B 独立}$$

全概率公式： $$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

贝叶斯公式： $$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} = \frac{P(B_i)P(A|B_i)}{P(A)}$$

概念

两两独立 $\neq$ 相互独立
超几何分布与是否计序无关
抽签原理：与抽签的次序无关

模型

- 会面问题
- 蒲丰投针
- 不放回抽样与二项分布
- 放回抽样与超几何分布
- 匹配问题

随机变量的概率分布

基本概念

$$P(a<\xi\leq b)=F(b)-F(a)\qquad P(a\leq\xi<b)=F(b-0)-F(a=0)$$

定理：若 $g(x)$ 处处可导，恒有 $g'(x)>0$或$g'(x)<0$，则 $Y=g(X)$ 是连续型随机变量，概率密度

$$f_Y(y)=\begin{cases}f_X[h(y)]\,|h'(y)|, & \alpha<y<\beta,\\ 0, & \text{其他},\end{cases}$$

普通情况/证明

$$\begin{aligned}F_Y(y)&=P\{Y\leq y\}=P\{g(X)\leq y\}\\&=P\{X\leq h(y)\}=F_X[h(y)]\end{aligned}$$

离散型随机变量

- 单点分布 $\quad\xi\sim(x-c)$

$$F(x)=\begin{cases}0 & ,x<c\\ 1 & ,x\geq c\end{cases}$$

  $E\xi=c\qquad D\xi=0$

- (0-1) 分布

$$P\{X=k\}=p^k(1-p)^{1-k}\qquad k=0,1$$

  $E\xi=p\qquad D\xi=p(1-p)$

- 二项分布 $\quad\xi\sim B(n,p)$

$$P\{X=k\}=\binom{n}{k}p^k(1-p)^{1-k}\qquad k=0,1,2\cdots n$$

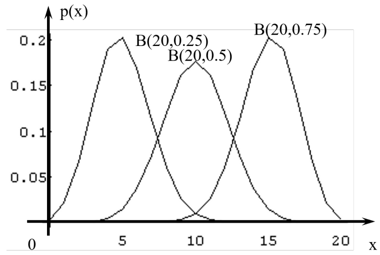

图 1: 二项分布的概率密度函数

$E\xi = np \qquad D\xi = np(1-p)$

- 泊松分布 $\quad \xi \sim P(\lambda)$

$$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0,1,2\cdots$$

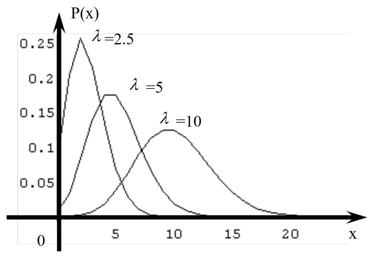

图 2: 泊松分布的概率密度函数

注意：泊松分布是非对称的，越大，非对称性越不明显。

$E\xi = \lambda \qquad D\xi = \lambda$

- 超几何分布 $\quad \xi \sim H(n,M,N)$

$$P\{X=k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \qquad k = 0,1,2\cdots r = min(n,M)$$

- 几何分布 $\quad \xi \sim Ge(p) \qquad$ 首次成功发生在第 k 次

$$P\{X=k\} = (1-p)^{k-1}p \qquad k = 1,2,3\cdots$$

p:A 发生的概率 $\qquad \xi$:A 首次发生时的试验次数

$E\xi = \frac{1}{p} \qquad D\xi = \frac{q}{p^2}$

连续型随机变量

- 均匀分布 $\quad \xi \sim U(a,b)$

$$p(x)=\begin{cases}\frac{1}{b-a}, & x\in(a,b)\\ 0 & \text{其他}\end{cases}\qquad F(x)=\begin{cases}0, & x\le a\\ \frac{x-a}{b-a}, & a<x\le b\\ 1, & x>b\end{cases}$$

$E\xi=\frac{a+b}{2} \qquad D\xi=\frac{(b-a)^2}{12}$

- 指数分布 $\quad \xi\sim E(\lambda)$ 　　寿命、某种服务的等待时间

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & a<x\le 0\\ 0, & x<0\end{cases}$$

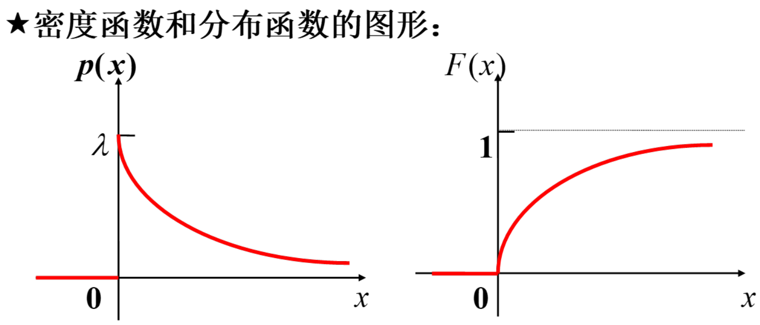

图 3: 指数分布

$E\xi=\frac{1}{\lambda} \qquad D\xi=\frac{1}{\lambda^2}$

$\star$ 模型：Poisson 分布与指数分布的关系 (电话，飞机) 　　无记忆性

- 正态分布 $\quad \xi\sim N(\mu,\sigma)$ 　　$\star 3\mu$ 原则

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\star\ E(|X|)=\sqrt{\frac{2}{\pi}} \qquad\qquad f_X(X^2)=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\sim\chi^2(1)$$

多元随机变量及其分布

二元正态分布:　$\sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

1. 联合正态 $\Longrightarrow$ 边际正态
2. 相同的边际正态，联合不一定正态
3. 相同的边际正态，联合即使是正态，也可能不一样

联合分布是均匀分布，但边缘分布都不是均匀分布。
二项分布具有可加性：$\zeta = \xi + \eta \sim B(n_1 + n_2, p)$
泊松分布具有可加性：$\zeta = \xi + \eta \sim P(\lambda_1 + \lambda_2)$
正态分布具有可加性：$\zeta = \xi + \eta \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
$\star$ 卷积公式：$p_\zeta(z) = \int_{-\infty}^{+\infty} p_\xi(x) p_\eta(y)|_{y=(z-x)} dx$
一般的如果 $Y = a\xi + b\eta$ 则

$$p_\zeta(z) = \frac{1}{|b|} \int_{-\infty}^{+\infty} p_{\xi\eta}(x, y)|_{y=\frac{z-ax}{b}} dx$$

商分布定理 ($Z = \frac{Y}{X}$、$Z = XY$)：

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

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

极值分布定理：

$$F_{max}(z) = [F(z)]^n$$

$$F_{min}(z) = 1 - [1 - F(z)]^n$$

$\star\Gamma$ 函数：

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
2. $\Gamma(n + 1) = n!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

随机变量的数字特征

期望的函数嵌套：	$E(Y)=E[g(X)]=\int_{-\infty}^{\infty}g(x)f(x)\,dx$
切比雪夫不等式：	$P\{\lvert X-\mu\rvert\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$ $\quad P\{\lvert X-\mu\rvert<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$
相关系数：	$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$
协方差：	$\mathrm{Cov}(\xi,\eta)=E(\xi\eta)-(E\xi)\cdot(E\eta)$ $\quad \mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$
	$\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$
期望与方差基本公式：	$E(CX)=CE(X)$ $\quad E(X+Y)=E(X)+E(Y)$
	$D(CX)=C^2D(X)$ $\quad D(X+C)=D(X)$
	$D(X)=E(X^2)-[E(X)]^2$

- 若独立则：
  $E(XY)=E(X)E(Y)$
  $D(X+Y)=D(X)+D(Y)$
- 非不独立则：
  $D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y)$

$\star\xi$ 与 η 独立是 ξ 与 η 不相关的充分不必要条件

$\star$ 下车问题：20 位旅客，十个车站可以下车，每个人每站下车的概率均等，设

$$X_i=\begin{cases}0,\text{在第 i 站没人下车}\\1,\text{在第 i 站有人下车}\end{cases}$$

所以 $X=X_1+X_2+\cdots+X_{10}$

$$E(X_i)=1-(\frac{9}{10})^{20}$$

$$E(X)=10[1-(\frac{9}{10})^{20}]=8.784$$

数理统计

基本概念

$$\text{卡方分布}\chi^2 \sim \chi^2(n) \qquad \chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \tag{1}$$

$$E(X) = n \qquad D(X) = 2n \tag{2}$$

$$\text{t 分布}t \sim t(n) \qquad t = \frac{X}{\sqrt{Y/n}} \tag{3}$$

$$\text{F 分布}F \sim F(n_1, n_2) \qquad F = \frac{U/n_1}{V/n_2} \tag{4}$$

定理：

1. $\overline{X} \sim N(\mu, \sigma^2/n)$
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-2)$
3. $\overline{X}$ 与 S^2 相互独立

区间估计评选标准：

1. 无偏性
$$E(\hat{\theta}) = \theta$$
2. 有效性
$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$$
3. 相和性
$$\lim_{n\to\infty} P\{|\hat{\theta} - \theta\| < \varepsilon = 1\}$$

枢轴量分布：

- σ^2 已知
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \text{置信区间：}(\overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$
- σ^2 未知
$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \qquad \text{置信区间：}(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1))$$

假设检验中的两类错误

真实情况	所作决策	
(未知)	接受 H_0	拒绝 H_0
H_0 为真	正确	犯第 I 类错误
H_0 不真	犯第 II 类错误	正确