概率

一、随机事件和概率

1、事件的描述

有事件A，B，C三个事件：
至少有一个发生：$A \cup B \cup C$，
不多于一个发生：$\bar{A}\bar{B}\bar{C} \cup A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C$

2、古典概率/几何概率

$$C_n^k = \frac{n!}{k!(n-k!)} = \frac{A_n^k}{A_k^k} \qquad A_n^k = \frac{n!}{(n-k)!}$$

3、概率的性质

$P(A \cup B) = P(A) + P(B) - P(AB)$
$P(\bar{A}) = 1 - P(A)$
$P(A - B) = P(A) - P(AB)$
若事件 A, B 相互独立 则有$P(AB) = P(A) \cdot P(B)$
概率为 0 的事件不一定为不可能事件，同理，概率为 1 的事件不一定为必然事件

4、条件概率

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

5、全概率和贝叶斯

全概率

设A_i 是样本空间 S 的一个划分，因此 $B = BS = B \cap (\cup_{i=1}^n A_i) = \cup_{i=1}^n BA_i$,且BA_i 与BA_j 互不相容，$i \neq j$ 所以

$$P(B) = P(\cup_{i=1}^n BA_i) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

例1.21 设某人有三个不同的电子邮件账户，有70%的邮件进入账户1，另有20%的邮件进入账户2，其余10%的邮件进入账户3. 根据以往经验，三个账户垃圾邮件的比例分别为1%，2%，5%，问某天随机收到的一封邮件为垃圾邮件的概率.0.016

解：

设$A_i (i=1,2,3)$表示事件“产品分别来自账户1、2、3”，以B表示事件“邮件为垃圾邮件”

贝叶斯

设实验 E 的样本空间为 S, B 为E 的事件, $A_1, A_2, \cdots A_n$ 为样本空间 S 的一个划分且 $P(B)>0, P(A_i)>0 (i=1,2,\cdots,n)$,则

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_j)P(A_j)} (i=1,2,\cdots,n)$$

补例3 有三只箱子：第一个箱子中有4个黑球和1个白球；第二个箱子中有3个黑球和3个白球；第三个箱子中有3个黑球和5个白球. 求已知取到的是白球，则这个白球属于第二个箱子的概率。

解：

设$A_i (i=1,2,3)$取到第i个箱子，B表示取到的是白球

4.设一批产品中一、二、三等品各占60%，35%，5%.从中任意取一件，结果不是三等品，求取到一等品的概率？**12/19**

5.一盒晶体管中有8只合格品、2只不合格品.从中不返回地一只一只取出，试求第二次取出的是合格品的概率**4/5**

6.钥匙掉了，掉在宿舍里、教室里、路上的概率分别是50%、30%和20%而掉在上述三处地方被找到的概率分别是0.8、0.3和0.1.试求钥匙找到的概率 **0.51**

7.学生在做一道4个选项的单选题，如果他不知道问题的答案就作随机猜测，现从卷面上看是答对了，试在以下情况求学生确实知道正确答案的概率.(1)知道答案和胡乱猜测的概率均为1/2； **0.8**(2)知道正确答案的概率是0.2. **0.5**

二、随机变量及其分布

随机变量

通常用 X, Y, Z 表示

分布函数

$F(x)=P(X \leq x)$

性质

(1)单调性:若 $x_1<x_2$ 则$F(x_1) \leq F(x_2)$

(2)有界性:

$$0 \leq F(x) \leq 1, F(+\infty)=\lim_{x \rightarrow+\infty} F(X)=1, F(-\infty)=\lim_{x \rightarrow-\infty} F(x)=0$$

(3)右连续:$F(x+0)=F(x)$

离散型

$x_1, x_2, \cdots, x_n \leq x$,则$F(x)=P(x_1)+P(x_2)+\cdots+P(x_n)$

连续型

$F(x)=P(X \leq x)=\int_{-\infty}^x f(x) \mathrm{d} x$

性质:

$P(a<x \leq b)=F(b)-F(a)$

$P(x>a)=1-p(x \leq a)=1-F(a)$

分布律

$P_{X=x}=F(x)-F(x-0)$

```
 _______________________
| x  |   ......   |   ......   |   ......   |
|____|______|______|______|
| P  |   ......   |   ......   |   ......   |
|____|______|______|______|
```

密度函数f(x)

求随机变量函数的分布函数步骤

1. 确定$Y = g(x)$ 的取值范围$[a, b]$
2. 分区间讨论:

$$F_Y(y) = P\{Y \leq y\} = P\{g(x) \leq y\} = P\{y \leq h(x)\} = \int_a^b f_X x \mathrm{d}x$$

3. 求导:$f_Y y = [F_Y y]'$

常见分布

1. 二项分布: $X \sim B(n, p)$

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}$$

2. 均匀分布:$X \sim U(a, b)$

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

3. 指数分布:$X \sim E(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0 \\ 0, 其他, \end{cases}$$

4. 正态分布:$x \sim N(\mu, \sigma^2)$

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

正态分布求"概率"标准化

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= P\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\} \\ &= PZ \leq \frac{x-\mu}{\sigma} \\ &= \phi(\frac{x-\mu}{\sigma}) \end{aligned}$$

多维随机变量及其分布

求边缘分布,判断独立性

$$\begin{aligned}&f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\\&f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\\&\text{若}X,Y\text{相互独立则有}f(x,y)=f_X(x)\cdot f_Y(y)\end{aligned}$$

条件概率密度

$$\begin{aligned}&f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)},f_X(x)>0\\&f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)},f_Y(y)>0\\&P(Y\le a|x=c)=\int_{-\infty}^{a}f_{Y|x}(y|x=c)\mathrm{d}y\end{aligned}$$

求Z=X+Y的密度函数

$$\begin{aligned}&z=x+y\\&\text{用}x\text{和}z\text{表示}y,f(x,y)\Rightarrow f(x,z-x)\\&\text{对}z\text{区间进行分类讨论}\\&\text{最后解得}f_Z(z)\end{aligned}$$

例题：

$$\begin{aligned}&f(x,y)=\begin{cases}2-x-y,0<x<1,0<y<1\\0,\text{其他}\end{cases}\\&\text{求}Z=X+Y\text{的该路密度}f_Z(z)\end{aligned}$$

解：

$$\begin{aligned}&f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)\mathrm{d}x,\text{其中}\\&f(x,z-x)=\begin{cases}2-z,0<x<1,0<z-x<1,\\0,\text{其他},\end{cases}\\&\text{当}z\le 0\text{或}z\ge 2\text{时},f_Z(z)=0\text{，}\\&\text{当}0<z<1\text{时},f_Z(z)=\int_0^z(2-z)\mathrm{d}x=z(2-z)\\&\text{当}1\le z<2\text{时},f_z(z)=\int_{z-1}^{1}\mathrm{d}x=(2-z)^2\\&\text{因此},Z=X+Y\text{的概率密度为}f_Z(z)=\begin{cases}z(2-z),0<z<1,\\(2-z)^2,1\le z<2\\0,\text{其他}\end{cases}\end{aligned}$$

数字特征与极限定理

求期望

$$
\begin{aligned}
&E(x)=\sum_{k=1}^{n} kP_k\\
&E(x)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x\\
&E(g(x))=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x\\
&E(ax+b)=aE(x)+b
\end{aligned}
$$

求方差

$$
\begin{aligned}
&D(x)=E(x^2)-E(x)^2\\
&D(ax\pm b)=a^2D(x)
\end{aligned}
$$

中心极限定理（近似正态分布）

$$
\lim_{n\to+\infty} P\left\{\frac{X-np}{\sqrt{np(1-p)}}\le x\right\}=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t=\phi(x), X\sim B(n,p)
$$

协方差

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

常见分布的期望与方差

1. 二项分布: $X\sim B(n,p)$

$$
P\{X=k\}=C_n^k p^k(1-p)^{n-k}\qquad E(x)=np,\quad D(x)=np(1-p)
$$

2. 均匀分布:$X\sim U(a,b)$

$$
f(x)=\begin{cases}\frac{1}{b-a}, a<x<b,\\ 0, 其他,\end{cases}\qquad E(x)=\frac{a+b}{2},\quad D(x)=\frac{(b-a)^2}{12}
$$

3. 指数分布:$X\sim E(\lambda)$

$$
f(x)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0, 其他,\end{cases}\qquad E(x)=\frac{1}{\lambda},\quad D(x)=\frac{1}{\lambda^2}
$$

4. 正态分布:$x\sim N(\mu,\sigma^2)$

$$
f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},(-\infty<x<+\infty)\qquad E(x)=\mu,\quad D(x)=\sigma^2
$$

统计量及其分布

统计量

不含有位置参数

常见统计量

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

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

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

$$E(\bar{X}) = 总体期望 \qquad D(\bar{X}) = \frac{总体方差}{n}$$

$$E(S^2) = 总体方差$$

三大抽样分布

$\mathcal{X}^2$分布

设 $X_1, X_2, X_3, \cdots, X_n$ 是来自标准正态总体 $N(0,1)$ 的样本，则称统计量

$$\mathcal{X}^2 = X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2 \sim \mathcal{X}^2(n)$$

t分布

设 $X \sim N(0,1), Y \sim \mathcal{X}^2(n)$,且 X, Y 相互独立，则

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

F分布

设$U \sim \mathcal{X}^2(n_1), V \sim \mathcal{X}^2(n_2)$，且 U，V 相互独立，则

$$F = \frac{U \setminus n_1}{V \setminus n_2} \sim F(n_1, n_2)$$

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

上侧α分位数（点）

设有随机变量X，对给定的$\alpha(0 < \alpha < 1)$，若存在实数x_α满足$P\{X > x_\alpha\} = \alpha$,则称$x_\alpha$为$X$的上侧$\alpha$分位点

$$u_{1-\alpha} = -u_\alpha \qquad t_{1-\alpha} = -t_\alpha \qquad F_{1-\alpha} = \frac{1}{F_\alpha(n_2, n_1)}$$

参数估计

矩估计步骤

1. 求$E(x)$
2. 反解θ

3. 令$\bar{X} = E(x)$

最大似然估计步骤

1. 构造$L(\theta)$

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

2. 求出$\ln L(\theta)$
3. 令$\frac{\mathrm{d} \ln L(\theta)}{\mathrm{d}\theta} = 0$
4. 解出$\hat{\theta}$

无偏估计量

若$f(x)$是θ的无偏估计量，则$E(f(x)) = \theta$

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

区间估计

存在统计量$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$和$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$，使$P\{\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2\} = 1 - \alpha$,则随机变量区间$[\hat{\theta}_1, \hat{\theta}_2]$为参数$\theta$的置信度为$1 - \alpha$的置信区间。
$[\bar{X} \pm \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)]$