本学期主要的学习内容

概率统计

- 了解概率统计的发展历程
- **概率与统计**关系
- 随机现象：**二重性**
- 随机试验：三特点（可重复、多结果、不确定）
- 样本空间、样本点
- 随机事件：基本事件、不可能事件、必然事件
- 事件关系：$\subset$ 、$=$、$\cup$、 $-$ 、$\cap$ 、$\bar{A}$ **互斥与对立事件**的关系
- 事件运算：幂等、交换、结合、**分配**、**对偶**

概率公理化

- 频率、频率的稳定性、频率与概率的关系

- 概率的公理化

- $P(\emptyset) = 0$

- $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots P(A_n)$

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

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

- Union bound

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant P(A_1) + P(A_2) + \cdots + P(A_n)$$

古典概型、几何概型

- 古典概型：试验结果只有有限种可能、每种结果发生的可能性相同
  - 计数原理、排列组合
  - 各种例题：产品抽样、生日驳论、抽签、matching、超几何
- 几何概型：样本空间无限可测、基本事件等可能性
  - 概率计算就是长度、面积、体积的计算
  - 例题：规划公交车发车时间、三角形、见面问题
- 十二重计数/组合计数（不考）

条件概率

条件概率：$P(B|A) = P(AB)/P(A)$

乘法公式：$P(A_1 A_2 \cdots A_n) =$
$P(A_1)P(A_2|A_1)\cdots P(A_n|A_1 A_2 \cdots A_{n-1})$

全概率公式:若事件$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 对任意事件B有 $P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$

贝叶斯公式：设$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 且事件B满足$P(B) > 0$. 对任意$1 \leq i \leq n$有

$$P(A_i|B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

独立性

独立性：两个事件、多个事件相互独立性

相互独立性　　　两两独立性

事件的独立　　　事件的互斥（互不相容）

独立性的性质，以及如何判断独立性

小概率原理：若事件A在一次试验中发生的概率非常小, 但经过多次独立地重复试验, 事件A的发生是必然的

离散型随机变量

随机变量、离散型随机变量

分布列：$p_k = P(X = x_k)$　　分布列性质

期望: $E(X) = \sum_k p_k x_k$ 反映随机变量的平均值

对随机变量X和常数$a, b \in R$, 有$E(aX + b) = aE(X) + b$

函数的期望：$E[g(X)] = \sum_{k \geq 1} g(x_k) p_k$

凸函数、Jensen不等式

方差：$Var(X) = E\big(X - E(X)\big)^2 = E(X^2) - [E(X)]^2$

性质：$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}\,(X)$，$E\big(X - E(X)\big)^2 \leq E(X - a)^2$

对$X \in [a, b]$有 $Var(X) \leq \big(b - E(X)\big)(E(X) - a) \leq (b - a)^2/4$

常用离散随机变量

- 0/1分布：X $\sim Ber(p)$，$E(X) = p$　　$Var(X) = p(1-p)$
- 二项分布：$X \sim B(n,p)$，$E(X) = np$　　$Var(X) = np(1-p)$
- 几何分布：$X \sim G(p)$，$\mathrm{E}(X) = \frac{1}{p}$　　$\mathrm{Var}(X) = \frac{1-p}{p^2}$

  - 无记忆性：$P(X > m+n \mid X > m) = P(X > n)$

- 负二项分布：X服从参数为r和p的负二项分布

$$\mathrm{E}(X) = \frac{r}{p} \qquad 和 \qquad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

- 泊松分布：$X \sim P(\lambda)$, $E(X) = \lambda$　　$\mathrm{Var}\ (X) = \lambda$

  - 泊松定理：$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$

连续型随机变量

分布函数：$F(x) = P(X \leq x)$

- 单调性: 若$x_1 < x_2$, 则$F(x_1) \leq F(x_2)$
- 规范性: $F(x) \in [0,1]$且$F(-\infty) = 0, F(+\infty) = 1$
- 右连续性: $F(x+0) = F(x)$

连续随机变量、概率密度函数：$F(x) = \int_{-\infty}^{x} f(t)dt$

- 非负性：$f(x) \geq 0$
- 规范性：$\int_{-\infty}^{+\infty} f(t)dt = 1$

连续随机变量分布函数$F(x)$与$f(x)$的连续、可导关系

连续型随机变量$P(X = x) = 0$

期望$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$及其性质

方差：$\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - \big(E(X)\big)^2$

常用的连续型随机变量

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

指数分布 $X \sim e(\lambda)$: $E(X) = \frac{1}{\lambda}$ $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

指数分布的无记忆性: $P(X > s + t | X > t) = P(X > s)$

正态分布 $X \sim N(\mu, \sigma^2)$ 定义 、图像

- 若$X \sim \mathrm{N}(\mu, \sigma^2)$, 则 $Y = (X - \mu)/\sigma \sim N(0,1)$

- 若$X \sim N(0,1)$, 则 $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$

已知连续随机变量X的概率密度为$f_X(x)$，求随机变量$Y = g(X)$的概率密度$f_Y(y)$？

- 求解Y=g(X)的分布函数$F_Y(y) = P(Y \le y) = \int_{g(x)\le y} f_X(x)dx$
- 利用分布函数和概率密度关系求解密度函数$f_Y(y) = F_Y'(y)$

二维随机变量

二维联合分布函数 $F(x,y)=P(X\leq x,Y\leq y)$、性质

边缘分布函数$F_X(x)=\lim\limits_{y\to+\infty}F(x,y)$

随机变量X和Y的独立性

二维离散随机变量：联合分布列、边缘分布列、独立性

二维连续随机变量：联合概率密度、边缘概率密度

二维连续随机变量的独立性

二维正态分布$N(\mu,\Sigma)$, $N(\mu_x,\mu_x,\sigma_x^2,\sigma_y^2,\rho)$

$$f(x,y)=(2\pi)^{-2/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(\xi-\mu)^{\top}\Sigma^{-1}(\xi-\mu)}$$

正态分布的边缘概率密度，独立性

多维正态分布的定义、边缘分布、独立性、标准化

已知(X,Y)的分布, 求$Z=g(X,Y)$的分布（离散、连续）

多维随机变量函数的分布

多维随机变量函数$\max(X,Y)$ 和 $\min(X,Y)$

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

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

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

$X\sim B(n_1,p)$和$Y\sim B(n_2,p)$独立, $X+Y\sim B(n_1+n_2,p)$

$X\sim P(\lambda_1)$ 和$Y\sim P(\lambda_2)$ 独立, 则$X+Y\sim P(\lambda_1+\lambda_2)$

$X\sim N(\mu_1,\sigma_1^2)$和$Y\sim N(\mu_2,\sigma_2^2)$ 独立, 则

$$X+Y\sim \mathrm{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

多维随机变量的数字特征

若已知 X和Y的联合分布，计算$Z = g(X, Y)$的期望$E[Z]$

期望的性质$E[aX + bY] = aE[X] + bE[Y]$

独立: $\boldsymbol{E[XY] = E[X]E[Y]}$**非独立:** $\boldsymbol{E[XY] \leq \sqrt{E[X^2]E[Y^2]}}$

随机变量X和Y的协方差为

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

协方差的性质：$\mathrm{Cov}(X, c) = 0$, $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$$

相关系数

X与Y的相关系数 $\rho_{XY} = \dfrac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$、性质

二维正态分布$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$， ρ为X与Y的相关系数、**X与Y独立$\Leftrightarrow$ X与Y不相关**

随机向量$\boldsymbol{X}$的协方差矩阵

$$Cov(X) = \Sigma = \begin{pmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & \cdots & Cov(X_2, X_n) \\ \vdots & & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_n) \end{pmatrix}$$

半正定

正态分布相关的结论$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathsf{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

条件分布与条件期望

条件概率密度: $f_{X|Y}(x|y) = f(x,y)/f_Y(y)$

条件分布函数: $F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(v|y)dv$

乘法公式：$f(x,y) = f_X(x)f_{Y|X}(y|x), \quad (f_X(x) > 0)$

独立性：$f_{Y|X}(y|x) = f_Y(y)$

多维正态分布的条件分布是正态分布

随机变量X的期望为$E[X|Y=y] = \int_{-\infty}^{+\infty} x\, f(x|y)dx$

全期望公式：$E[X] = E[X|A]P(A) + E[X|\bar{A}]\big(1 - P(A)\big)$

集中不等式

Markov不等式: $P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$

Chebyshev不等式: $P(|X-\mu| > \epsilon) \leq \frac{Var(X)}{\epsilon^2}$

单边Chebyshev不等式: $P(X-\mu \geq \epsilon) \leq \frac{\sigma^2}{\sigma^2+\epsilon^2}$

Hölder不等式: $E(|XY|) \leq \left(E(|X|^p)\right)^{\frac{1}{p}}\left(E(|Y|^q)\right)^{\frac{1}{q}}$

集中不等式

随机变量X的矩生成函数为$M_X(t) = E[e^{tX}]$、性质

Chernoff方法： $P[X \geq \epsilon] \leq \min_{t>0}\{e^{-t\epsilon}E[e^{tX}]\}$.

0/1-随机变量$P\left[\sum_{i=1}^{n} X_i \geq (1+\epsilon)\mu\right] \leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mu}$

Rademacher随机变量： $P\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geq \epsilon\right] \leq e^{-n\epsilon^2/2}$

有界：Chernoff引理

$$P\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \geq \epsilon\right] \leq e^{-2n\epsilon^2/(b-a)^2}$$

Sub-Gaussian随机变量: $E\left[e^{(X-E[X])t}\right] \leq e^{\frac{bt^2}{2}}$

有界随机变量、Gaussian随机变量

集中不等式（不考）

Bennet不等式: $P\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\ \geq \epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{2\sigma^2+2\epsilon/3}\right)$

Bernstein不等式: $P\ \left[\frac{1}{n}\sum_{i=1}^{n}X_i-\mu \geq \epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{2\sigma^2+2b\epsilon}\right)$

随机投影 JL-引理

大数定律

什么是大数定律、依概率收敛

Markov大数定律：若随机变量序列$\{X_i\}$满足$\mathrm{Var}(\sum_{i=1}^{n} X_n)/n^2 \to 0$, 则满足大数定律

Chebyshev大数定律：若独立随机变量序列$\{X_i\}$满足$\mathrm{Var}(X_i) \le c$, 则满足大数定律

Khintchine大数定律：若独立同分布随机变量序列$\{X_i\}$期望存在, 则满足大数定律;

Bernoulli大数定律: 对二项分布$X_n \sim \mathrm{B}(n, p)$, 有$X_n/n \xrightarrow{P} p$

中心极限定理

- 中心极限定理、依分布收敛
- **林德贝格-勒维中心极限定理**: 独立同分布随机变量，若 $E[X_k]=\mu$和$\mathrm{Var}(X_k)=\sigma^2$, 则$\sum_{k=1}^{n} X_k \xrightarrow{d} N(n\mu, n\sigma^2)$
- **棣莫弗-拉普拉斯中心极限定理**: 若 $X_n \sim B(n,p)$，则 $X_n \xrightarrow{d} N\big(np, np(1-p)\big)$
- **李雅普诺夫定理**: 独立不同分布中心极限定理

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n} E\left[|X_k-\mu_k|^{2+\delta}\right] \rightarrow 0$$

统计的基本概念

总体: 研究对象的全体，用随机变量X表示(分布未知)

样本：从总体中随机抽取一些个体，表示为$X_1, X_2, \cdots, X_n$，称$X_1, X_2, \cdots, X_n$为取自总体X的随机样本, 其样本容量为n

抽样、样本值、样本的二重性、简单样本

样本的分布：$F(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n)$

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

- 统计量: $g(X_1, X_2, \cdots, X_n)$ 关于$X_1, X_2, \cdots, X_n$连续、不含任意参数
- 样本均值、样本方差、样本标准差、修正后的样本方差、 k阶原点矩/中心矩、次序统计量

三大统计分布

两类积分函数与分布（不考）

- Beta函数、Γ-函数、两类函数的关系
- Beta分布、Γ-分布、性质、独立可加性
- 标准正态分布的平方$\Gamma(1/2,1/2)$
- Dirichlet分布、性质

统计三大分布

- 自由度为n的χ^2分布：$Y=X_1^2+X_2^2+\cdots+X_n^2$
- χ^2分布性质、独立可加性
- t分布$T=X/\sqrt{Y/n}$、性质

分布可加性

- 如果$X \sim \mathrm{N}(\mu_1,\sigma_1^2)$ 和$Y \sim \mathrm{N}(\mu_2,\sigma_2^2)$, 且$X$与$Y$独立, 那么$X \pm Y \sim \mathrm{N}(\mu_1 \pm \mu_2,\sigma_1^2+\sigma_2^2)$;
- 如果$X \sim B(n_1,p)$和$Y \sim B(n_2,p)$,且X与Y独立，那么$X+Y \sim B(n_1+n_2,p)$;
- 如果$X \sim P(\lambda_1)$和$Y \sim P(\lambda_2)$,且X与Y独立，那么$X+Y \sim P(\lambda_1+\lambda_2)$;
- 如果$X \sim \Gamma(\alpha_1,\lambda)$ 和$Y \sim \Gamma(\alpha_2,\lambda)$，且$X$与$Y$独立，那么$X+Y \sim \Gamma(\alpha_1+\alpha_2,\lambda)$
- 如果$X \sim \chi^2(m)$和$Y \sim \chi^2(n)$,且X与Y独立，那么$X+Y \sim \chi^2(m+n)$.

统计五大抽样定理

统计五大采样定理

1） 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，则有 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}(\mu, \sigma^2/n)$ $\qquad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$

2）有$\bar{X}$和S^2相互独立, 且$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

3） $\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \sim t(n-1)$

正态分布的抽样分布定理四和五

定理： 设$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$分别来自总体$N(\mu_X, \sigma_X^2)$和$N(\mu_Y, \sigma_Y^2)$的两个独立样本, 其修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

定理： 设$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$分别来自总体$N(\mu_X, \sigma^2)$和$N(\mu_Y, \sigma^2)$的两个独立样本，令其样本均值分别$\bar{X}$和$\bar{Y}$，修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

参数估计

参数估计：依据$X_1, X_2, \cdots, X_n$估计参数θ或函数$g(\theta)$

点估计

- 矩估计法：样本矩去估计总体矩求参数θ
- 最大似然估计: $\hat{\theta} = \operatorname{argmax}_\theta L(x_1, x_2, \cdots, x_m; \theta)$
- 最大似然估计的不变性

估计量的常用标准

- 无偏性：无系统偏差
- 有效性：估计量的方差越小越好，有效统计量
- 一致性：在数据足够多的情况下能有效估计统计量

Ch 9 参数估计

例题

设$X_1, X_2, \cdots, X_n$是来自总体$X \sim U(0, \theta)$的样本，证明：θ的最大似然估计量是一致估计量.

区间估计

区间估计问题: 设$X_1, X_2, \cdots, X_n$是来自总体X的样本, θ为总体X的分布函数$F(x,\theta)$的未知参数, 根据样本估计

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n) \text{ 和 } \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$$

使得以较大的概率保证有$\theta \in \left(\hat{\theta}_1, \hat{\theta}_2\right)$成立，称为参数估计问题

具体而言, 对任意给定$\alpha \in (0,1)$, 有

$$P\left[\hat{\theta}_1(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \cdots, X_n)\right] \geq 1 - \alpha$$

置信区间与置信度

设$X_1, X_2 \cdots, X_n$是来自总体X的样本，总体的分布函数含未知参数θ，找出统计量$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$使得

$$P\left[\hat{\theta}_1(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \cdots, X_n)\right] \geq 1-\alpha$$

成立，则称$1-\alpha$为置信度，$[\hat{\theta}_1, \hat{\theta}_2]$为$\theta$的置信度为$1-\alpha$的置信区间.

置信区间$[\hat{\theta}_1, \hat{\theta}_2]$是随机区间，$1-\alpha$为该区间包含$\theta$的概率或可靠程度. 若$\alpha = 0.05$, 则置信度为95%.

通常采用95%的置信度, 有时也可99%或90%等

置信区间与置信度

说明

- $\hat{\theta}_2 - \hat{\theta}_1$反映了估计精度: 长度越小精度越大.
- α反映了估计的可靠度: α越小可靠度越高.
- 给定α，区间$[\hat{\theta}_1, \hat{\theta}_2]$的选取并不唯一确定，通常选长度最小的一个区间.

置信区间的求解方法

枢轴变量法

- 先找一样本函数$W(X_1, X_2, \cdots, X_n; \theta)$包含待估参数$\theta$，但不含其它参数，且函数$W$分布已知，称为枢轴变量.
- 给定置信度$1-\alpha$，根据W的分布找出临界值a和b，使得$P[a < W < b] = 1-\alpha$成立.
- 根据$a < W < b$解出$\hat{\theta}_1 < \theta < \hat{\theta}_2$，则$[\hat{\theta}_1, \hat{\theta}_2]$为$\theta$的置信度为$1-\alpha$的置信区间.

正态总体, 方差已知, 求期望的区间估计

设$X_1, X_2, \cdots, X_n$是来自总体$X \sim N(\mu, \sigma^2)$的样本, 若方差σ^2已知. 给定$\alpha \in (0,1)$, 确定置信度为$1-\alpha$下μ的置信区间$\left[\hat{\theta}_1, \hat{\theta}_2\right]$.

正态总体, 方差未知, 求期望的区间估计

设$X_1, X_2, \cdots, X_n$是来自总体$X \sim N(\mu, \sigma^2)$的样本, 若方差σ^2未知. 给定$\alpha \in (0,1)$, 确定置信度为$1-\alpha$下μ的置信区间$\left[\hat{\theta}_1, \hat{\theta}_2\right]$.

正态总体, 求方差的置信区间

设$X_1, X_2, \cdots, X_n$是来自总体$X \sim N(\mu, \sigma^2)$的样本，考虑方差σ^2的置信度为$1-\alpha$的置信区间.