第五次习题课讨论题目

1. 第 9、10 次作业题目选讲.

2. 假设总体 $X \sim U(a,b)$，$X_1,\cdots,X_n$ 为其随机样本，请分别给出 a,b 的矩估计和极大似然估计.

3. 设随机样本 $X_1,\cdots,X_n$ 来自具有概率密度函数为 $f(x)=\begin{cases}\theta x^{\theta-1}, & 0<x<1\\ 0, & \text{其他}\end{cases}$ 的分布，其中 $\theta>0$ 是未知参数.

   （1） 求 θ 的矩估计 $\hat{\theta}$.

   （2） 求 θ 的极大似然估计 θ^*.

   （3） 判断 $\frac{1}{\theta^*}$ 是否是 $\frac{1}{\theta}$ 的无偏估计并给出相应理由.

4. （预测区间）假设总体服从正态分布 $N(\mu,\sigma^2)$（参数未知），$X_1,\cdots,X_n$ 为其随机样本，$\overline{X}$ 和 S^2 分别样本均值和为样本方差，又设 X_{n+1} 服从正态分布 $N(\mu,\sigma^2)$ 且与 $X_1,\cdots,X_n$ 独立.

   （1） 给出 $\dfrac{X_{n+1}-\overline{X}}{S\sqrt{1+\frac{1}{n}}}$ 的分布.

   （2） 若视 X_{n+1} 为新观测，请利用（1）给出新观测值的 $1-\alpha$ 置信区间.

5. （简单线性回归）假定变量 Y（响应变量）与 X（预测变量）之间的关系可用如下的线性模型刻画：$Y=\beta_0+\beta_1 X+\varepsilon$，其中 β_0 是常数项，β_1 称为模型的回归系数，皆为常数，ε 为随机误差项. 假设有 (X,Y) 的观测：(x_i, y_i)（$i=1,\cdots,n$）.

（1）　（机器学习）用最小二乘法（即求$S(\beta_0,\beta_1)=\sum_{i=1}^{n}\left(y_i-(\beta_0+\beta_1 x_i)\right)^2$的最小值点）估计参数$\beta_0,\beta_1$（相应的估计分别表示为$\hat{\beta}_0,\hat{\beta}_1$，$Y=\hat{\beta}_0+\hat{\beta}_1 X$称为回归直线）.

（2）　（统计）若假设$\varepsilon\sim N(0,\sigma^2)$，$\sigma>0$是常数，观测之间独立，用极大似然方法估计参数$\beta_0,\beta_1$，并与（1）的结果相比较.

6. （简单线性回归续）考虑四对变量(X_i,Y_i)（$i=1,\cdots,4$），下表是这四对变量的观测数据.

（1）　分别画出四组数据的散点图.

（2）　分别计算四对变量观测数据的均值和相关系数（注：若(X,Y)有观测：(x_i,y_i)（$i=1,\cdots,n$），则(X,Y)相关系数的估计为

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

）.

（3）　计算每组数据相应的回归直线并在对应的散点图中将其画出来.

（4）　这个例子对你有什么启示？

X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04

6 7.24	6 6.13	6 6.08	8 5.25
4 4.26	4 3.10	4 5.39	19 12.50
12 10.84	12 9.13	12 8.15	8 5.56
7 4.82	7 7.26	7 6.42	8 7.91
5 5.68	5 4.74	5 5.73	8 6.89