第 6 次作业

1. 假设随机变量 $X_i \sim \mathrm{P}(\lambda_i)$（$i=1,2$）相互独立，请确定 $Y = X_1 + X_2$ 的分布. 尝试给出结果的一个直观解释.

2. 假设随机变量 X,Y 相互独立，且都服从标准正态分布.

   （1） 求 $Z = \dfrac{Y}{X}$ $(X \neq 0)$ 的概率密度函数.

   （2） 令 $X = R\cos\Theta$，$Y = R\sin\Theta$，计算 (R,Θ) 的概率密度函数，并确定 R,Θ 是否独立.

   （3） 令 $U = X+Y$，$V = X-Y$，求 (U,V) 的概率密度函数，并确定 U,V 是否独立.

3. 设随机变量 X_i（$i=1,\cdots,n$）独立同分布，其分布函数为 $F(x)$，令

$$Y=\max\{X_1,\cdots,X_n\}，\quad Z=\min\{X_1,\cdots,X_n\}$$

   分别求 Y,Z 的分布函数.

4. 了解统计上（与正态分布相关）的三大分布：卡方分布，t 和 F 分布（参阅 John Rice 的数理统计与数据分析或其他资料），给出其定义.

5. 假设有一场 3 匹马（分别记为 A，B，C）的比赛，在下注结束时假设有 500 元下注在 A 马，300 元下注在 B 马，200 元下注在 C 马. 如果投注站想确保每下注的 100 元中可赚取 5 元.

   （1） 投注站设定的每匹马的赔率应该是多少？

   （2） 此时赔率所隐含的每匹马获胜概率是多少？（对比作业题 1-5）

   （3） 比较 3 匹马的隐含获胜概率之和与 1 的大小，你对此有什么看法？（对比作业题 2-14（2））
   （注：A 马赔率若是 $20:5$ 意味着若 A 马获胜则下注 A 马者每投注 5 元可获得 5 元+20 元，额外收益为 20 元）

6. 判断下列结论对错并说明理由，这里假设所涉及的期望和方差皆存在.

（1） 若 X 和 Y 独立，则 $Var(XY)=Var(X)Var(Y)$.

（2） X 的中位数若存在则一定等于 $E(X)$.

7. 证明：对任何常数 c 有 $E((X-c)^2)\geq Var(X)$，且等号当且仅当 $c=E(X)$ 时成立.
8. *设 X 有概率密度函数 $f(x)$，其中位数为 m. 证明：对任何常数 c 都成立不等式 $E(|X-c|)\geq E(|X-m|)$.
9. 计算对数正态分布的期望和方差（对数正态分布定义可参见作业题 4-9）.
10. *设随机变量 X_i（$i=1,\cdots,n$）独立同分布（这样的序列也称为来自同一分布的随机样本)，其公共期望为 μ，公共方差为 σ^2，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 称为样本均值，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 称为样本方差. 求 $Var(\bar{X})$ 和 $E(S^2)$.
11. 下列叙述是否等价？请说明理由.

    (1) $Cov(X,Y)=0$；

    (2) X 与 Y 不相关；

    (3) $E(XY)=E(X)E(Y)$；

    (4) $Var(X+Y)=Var(X)+Var(Y)$
12. 验证：若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 $\rho=Corr(X,Y)$.
13. *设 X 为之前课上讨论的配对问题中拿到自己帽子的人数（假设总共有 n 个人)，求 $E(X)$ 和 $Var(X)$.
14. （1）证明：$E^2(UV)\leq E(U^2)E(V^2)$，且等号成立当且仅当存在常数 c 使得

$P(V=cU)=1$.（提示：考虑 $E\left[(U-tV)^2\right]\geq 0, \forall t\in R$）

（2）利用（1）证明：$|Corr(X,Y)|\leq 1$，且等号成立当且仅当存在常数 a,b（$a\neq 0$）使得 $P(Y=aX+b)=1$.

15. 设随机变量 X_i（$i=1,\cdots,n$）独立同分布，其公共期望为 μ，公共方差为 σ^2.

（1） 证明：$Cov(X_i-\bar{X},\bar{X})=0$.

（2） **判断 $X_i-\bar{X}$ 与 $\bar{X}$ 是否一定独立？尝试给出理由.

16. （计算机实验）利用 Q-Q 图验证正态性. 一个数据集的正态 Q-Q 图（分位数-分位数图）是一个散点图，其中横坐标（通常）是数据值，纵坐标（通常）是该数据值相应于标准正态分布的分位数. 具体对应关系为：将数据值由小到大按顺序排列，记为 $x_{(1)}\leq x_{(2)}\leq\cdots\leq x_{(n)}$（称为观测分位数），$x_{(i)}$ 为横坐标，对应的纵坐标为 $\Phi^{-1}(\frac{i-0.5}{n})$，这里 Φ^{-1} 为标准正态累积分布函数的反函数，“$\frac{i-0.5}{n}$”是作了所谓的连续性修正. 当数据集来自正态总体时，该散点图合理地接近为一条直线.

（1） 生成一组（100 个）标准正态随机数，画出其正态 Q-Q 图，看看是否近似为一条直线.

（2） 生成一组（100 个）服从指数分布（参数 $\lambda=2$）的随机数，画出其正态 Q-Q 图，看看是否近似为一条直线.

（3） 可否推广上述方法去验证给定的数据集是否服从假设的分布？

（4） 可否推广上述方法去验证给定的两个数据集是否来自同一未知总体？