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 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，证明：$Y = X_1 + X_2$ 服从正态分布 $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$。（可参考陈希孺书例 4.8 $\rho = 0$ 情形的讨论）

4. *设 X, Y 独立，概率密度函数分别为 $f(x)$ 和 $g(y)$，且 $X > 0$。 请分别用以下两种方法计算 $Z = XY$ 的概率密度：

   （1） 利用变换 $Z = XY$，$W = X$。

   （2） 把 XY 表示为 Y / X^{-1}，先算出 X^{-1} 的密度，再利用课上得到的两个随机变量商的概率密度结果。

5. 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。

6. 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（参阅陈希孺书或其他资料），给出其定义。

7. 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） 对任何常数c有$E((X-c)^2)\geq Var(X)$，等号当且仅当$c=E(X)$时成立。

（2） 若X和Y独立，则$Var(XY)=Var(X)Var(Y)$。

（3） X的中位数若存在则一定等于$E(X)$。

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均值和方差（对数正态分布定义可参见作业4）。
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样本），其公共期望为μ，公共方差为σ^2，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$称为样本均值，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$称为样本方差。 求$\mathrm{Var}(\overline{X})$和$E(S^2)$。
11. 下列叙述是否等价？请说明理由。

(1) $\mathrm{Cov}(X,Y)=0$;

(2) X与Y不相关;

(3) $E(XY)=E(X)E(Y)$;

(4) $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$

12. 完成课上二元正态分布ρ的计算，即验证：若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则$\rho=\mathrm{Corr}(X,Y)$。
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（总共有n个人），求$E(X)$和$\mathrm{Var}(X)$。

14. （1）*证明：$E^2(UV)\le E(U^2)E(V^2)$，且等号成立当且仅当存在常数c使得 $P(V=cU)=1$。

（2）利用（1）证明：$|\mathrm{Corr}(X,Y)|\le 1$，且等号成立当且仅当存在常数$a,b$（$a\ne 0$）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：$\mathrm{Cov}(X_i-\overline{X},\overline{X})=0$。

（2） $X_i-\overline{X}$与$\overline{X}$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$x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$（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$\Phi^{-1}$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）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 Cauchy 分布的随机数，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？