第 7 次作业

1. *将线段$[0,1]$随机断开，将前半段（包含原线段左端点那部分）扔掉，将剩下线段再随机断开后扔掉前半段，求余下的那一段长度的期望值.
2. **一个矿工在井下迷了路，迷路的地方有三个门，假设从第一个门出来，经过 2 小时可达安全之处；若从第二个门出来，经过 3 小时后会回到原地；若从第三个门出来，经过 1 小时后会回到原地. 假定该人选择门是随机的，且始终无法区分这三个门，能够走到安全之处的平均用时是多少？
3. *证明：若$E(Y|X)=X$，则$\mathrm{Cov}(X,Y)=Var(X)$.
4. 定义条件方差为$Var(Y|X)\triangleq E\left[(Y-E(Y|X))^2|X\right]$. 证明：

   （1）$Var(Y|X)=E(Y^2|X)-E^2(Y|X)$

   （2）$Var(Y)=E(Var(Y|X))+Var(E(Y|X))$
5. 假设点(X,Y)服从单位圆盘上半部分上的均匀分布. 若观测到$X=0.5$，那么在均方误差意义下Y的最优预测值是什么？
6. *应用中常常基于随机变量X的观察值对随机变量Y的值进行预测，假设仅仅知道X和Y的期望分别为μ_1和μ_2，方差分别为σ_1^2和σ_2^2，相关系数为ρ.

   （1） 在均方误差意义下求Y的最优线性预测，即选择系数a,b使得$E\left[(Y-(aX+b))^2\right]$（均方误差）达到极小值.

   （2） 给出这个最优线性预测对应的均方误差，并指出其值何时接近 0.

   （3） 验证：若X,Y服从$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则$Y$的最优线性预测就是$E(Y|X)$.
7. *判断下述说法正确与否，并尝试给出说明：由概率的频率解释知，当试验次数$n\to\infty$时，事件 A 的频率$\frac{m}{n}$的极限就是概率，即有$\lim\limits_{n\to\infty}\frac{m}{n}=P(A)$.

8. *证明 Chebyshev 大数定律：设随机变量 X_i（$i=1,2,\cdots$）两两不相关，具有期望 $E(X_i)=\mu_i$，方差 $Var(X_i)=\sigma_i^2<c$，这里 c 为常数，则 $\forall\varepsilon>0$，有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}\mu_i\right|\geq\varepsilon\right)=0.$$

9. 利用中心极限定理证明（辛钦）弱大数定律.

10. **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，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 为样本均值，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 为样本方差. 证明：S^2 依概率收敛至 σ^2.

11. 以 X 表示抛 40 次均匀硬币出现的正面次数. 求 $P(X=20)$ 及其正态近似值.

12. *某保险公司向 10000 个投保人提供内容相同的汽车保险，假定这 10000 个投保人在一年内由于发生交通事故而造成的损失是独立同分布的，且一辆车在一年内发生事故的概率为 0.001，事故损失为 1000 元.
    （1） 如果忽略运营成本，保险公司每份保险卖 2 元合理吗？
    （2） 保险公司至少有 20%的毛利润（毛利润=保费-保险赔付-运营成本（忽略））的概率多大？
    （3） 以 95%的概率可以保证保险公司至少还有多少毛利润？

13. 假设某物理量的真值为 m，多次对其测量，每次测量产生一个随机误差，一个合理的假设是：在选择适当的单位下，随机误差服从 $U(-1,1)$.
    （1） 求 n 次测量的算术平均值与真值之差的绝对值低于微小正数 ε 的概率，并利用中心极限定理给出 $n=25$，$\varepsilon=0.2$ 情形的概率近似值；
    （2） 要使得算术平均值与真值之差的绝对值低于微小正数 ε 的概率超过 $1-\alpha$，应该进行多少次测量？请利用中心极限定理进行讨论，并给出 $\varepsilon=0.2$，$\alpha=0.05$ 情形的测量数.
    （3） 利用切比雪夫不等式求解（2），并比较两种方法得出的结果.

14. 假设某批次电子元件中合格品占 80%，从中任意购买 5000 个. 试问把误差限 ε 定为多少才能保证买到的元件的合格率与该批次的合格率之差不超过 ε 的概

率至少为 0.99？此时的合格品数的范围是多少？

15. （测量误差模型）测量误差由系统误差和随机误差构成. 测量 X 的模型如下：$X = x_0 + \beta + \varepsilon$. 其中 x_0 是测量对象的真实值；β 是常数，表示系统误差，也常称为测量过程的偏倚（bias）；ε 是随机误差，具有 $E(\varepsilon) = 0$，$\mathrm{Var}(\varepsilon) = \sigma^2$. 计算 $E(X)$，$\mathrm{Var}(X)$ 以及均方误差 $E\left[(X - x_0)^2\right]$.

16. （计算机实验）设 X_i（$i = 1, \cdots, n$）为独立同分布随机变量，$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

（1） 假设 $X_i \sim N(0,1)$，请分别对 $n = 1, 25, 100, 1000$ 模拟 $\overline{X}$ 的分布.

（2） 假设 $X_i \sim U(0,1)$，请分别对 $n = 1, 25, 100, 1000$ 模拟 $\overline{X}$ 的分布.

（3） 假设 X_i 的概率密度函数为 $f(x) = \dfrac{1}{\pi(1 + x^2)}, x \in \mathbb{R}$（Cauchy 分布），请分别对 $n = 1, 25, 100, 1000$ 模拟 $\overline{X}$ 的分布.

观察以上的实验结果是否与中心极限定理相符？