信息学中的概率统计：作业四

截止日期：2025 年 11 月 7 日（周五）下午 1 点前。请务必通过教学网提交电子版。可期中考试后同时提交纸质版。

与本次作业相关的事宜,请发邮件给我(ruosongwang@pku.edu.cn),抄送研究生助教叶昊洋(yhyfhgs@gmail.com),以及负责本次作业的本科生助教陈思危（2300017751@stu.pku.edu.cn）。

第一题

(1) 对于正整数 r 和实数 $0 < p < 1$，给定 $X \sim NB(1,p)$，$Y \sim NB(r,p)$，也即随机变量 X 服从参数为 $1,p$ 的负二项分布，随机变量 Y 服从参数为 r,p 的负二项分布。若 X 和 Y 相互独立，证明 $X+Y \sim NB(r+1,p)$，也即 $X+Y$ 服从参数为 $r+1,p$ 的负二项分布。提示：使用恒等式

$$\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m}。$$

(2) 对于正整数 r，若 $X_1, X_2, \ldots, X_r$ 为独立同分布的随机变量，且均服从 $G(p)$，也即参数为 p 的几何分布。证明 $X_1 + X_2 + \cdots + X_r \sim NB(r,p)$。

第二题

在第一次作业第二题中，我们介绍了事件的条件独立。在本题中，我们将引入随机变量的条件独立。我们称两个离散随机变量 X 和 Y 在给定一个事件 A 的条件下是条件独立的，当且仅当对于 X 的全部取值 x_i 和 Y 的全部取值 y_j，有

$$P(X = x_i \cap Y = y_j \mid A) = P(X = x_i \mid A) \cdot P(Y = y_j \mid A)。$$

(1) 若 X 和 Y 在给定事件 A 的条件下是条件独立的，证明 $E(XY \mid A) = E(X \mid A) \cdot E(Y \mid A)$。

(2) 在随机图模型中，共有 n 个人，且每一对人等概率认识或不认识。令 X_i 为第 i 个人认识的人数，在课上，我们计算了 $E(X_i X_j)$。对于 $i \neq j$，令事件 $E_{i,j}$ 为第 i 个与第 j 个人认识。计算 $E(X_i X_j \mid E_{i,j})$。

(3) 利用上一问的方法和重期望公式，当 $i \neq j$ 时，重新计算 $E(X_i X_j)$。

第三题

(1) 令随机变量 $X \sim G(p)$，也即随机变量 X 服从参数为 p 的几何分布。证明

$$E(X^2) = p + E((X+1)^2)(1-p),$$
$$E(X^3) = p + E((X+1)^3)(1-p),$$

并计算 $E(X^2)$ 和 $E(X^3)$。

(2) 令随机变量 Y 为伯努利试验中首次满足结果 A 和结果 $\overline{A}$ 均至少出现过一次时的试验次数。在第二次作业第四题中，我们计算了 $E(Y)$ 和 $\mathrm{Var}(Y)$。使用上一问中的方法，重新计算 $E(Y)$ 和 $\mathrm{Var}(Y)$。这里同样不需要对结果进行化简。

第四题

在课上，我们介绍了二维离散随机变量 (X,Y) 的重期望公式：$E(X)=E(E(X\mid Y))$。本题中，我们将该公式推广到高维随机变量的情况。给定 $n+1$ 维离散随机变量 $(X,Y_1,Y_2,\ldots,Y_n)$，对于 $(Y_1,Y_2,\ldots,Y_n)$ 的一组取值 $(y_1,y_2,\ldots,y_n)$，定义

$$g(y_1,y_2,\ldots,y_n)=E(X\mid Y_1=y_1,Y_2=y_2,\ldots,Y_n=y_n),$$

并定义随机变量 $E(X\mid Y_1,Y_2,\ldots,Y_n)$ 为 $g(Y_1,Y_2,\ldots,Y_n)$。

(1) 证明 $E(X)=E(E(X\mid Y_1,Y_2,\ldots,Y_n))$。

(2) 考虑课上和第三题中介绍的随机图模型，在每一对人等概率认识或不认识的基础上，将 n 个人分为两组，每个人独立且等概率地被分到第一组或第二组，且分组过程与随机图的生成过程相互独立。对于 $1\le i<j\le n$，令随机变量 $X_{i,j}$ 为

$$X_{i,j}=\begin{cases}1 & i \text{ 与 } j \text{ 认识}\\ 0 & i \text{ 与 } j \text{ 不认识}\end{cases}。$$

令随机变量 Y 为属于不同组且相互认识的人的对数。计算 $E(Y\mid\{X_{i,j}\}_{1\le i<j\le n})$ 和 $E(Y^2\mid\{X_{i,j}\}_{1\le i<j\le n})$。

(3) 计算 $E(Y)$ 和 $\mathrm{Var}(Y)$。

第五题

令 $X_1,X_2,\ldots,X_n$ 为独立同分布的离散随机变量，且对于任意 $1\le i\le n$，X_i 服从拉德马赫分布，也即 $P(X_i=+1)=P(X_i=-1)=1/2$。令 $Y=\sum_{i=1}^n X_i$。

(1) 计算 $E(Y)$。

(2) 计算 $E(Y^2)$。

(3) 计算 $E(Y^4)$。

(4) 证明对于任意 $k\ge 1$，$P(|Y|\ge k\cdot\sqrt{n})\le 3/k^4$。

第六题

在课上，我们考虑了如下球与桶模型：有 $n\ge 1$ 个球，每个球都等可能被放到 $m\ge 1$ 个桶中的任一个。对于 $1\le i\le m$，令随机变量 X_i 表示第 i 个桶中球的数量。令随机变量 Y 表示至少有一个球的桶的数量。

(1) 对于 $1\le i\le m$，计算 $E(X_i)$。

(2) 计算 $E(Y)$ 和 $\mathrm{Var}(Y)$ 并表示为仅与 m 和 n 相关的量

(3) 对于任意 $1\le i,j\le m$，计算 $\mathrm{Cov}(X_i,X_j)$