信息学中的概率统计：期中考试答案

第一题

(1) $P(\overline{A}) = 1/4$，$P(\overline{B}) = 2/3$。$P(C) = P((A \cap \overline{B}) \cup (\overline{A} \cup B)) = 3/4 \cdot 2/3 + 1/4 \cdot 1/3 = 7/12$。$P(A \cap C) = P(A) = 3/4$。$P(A \mid C) = 6/7$。

(2) $E(2^{1_A}) = 3/4 \cdot 2 + 1/4 \cdot 1 = 7/4$。$E(2^{-1_B}) = 1/3 \cdot 1/2 + 2/3 \cdot 1 = 5/6$。$E(2^{1_A - 1_B}) = 35/24$。

(3) 事件 $1_A + 1_B \le 1$ 等价于 $\overline{A} \cup \overline{B}$。由于 $\overline{B} \subseteq \overline{A} \cup \overline{B}$，$P(\overline{A} \cup \overline{B}) \ge P(\overline{B}) = 2/3$。由 Union Bound，$P(\overline{A} \cup \overline{B}) \le P(\overline{A}) + P(\overline{B}) = 11/12$。

第二题

(1) 对于任意 $b < y < a + b$，$P(Y \le y) = P(a\sqrt{X} + b \le y) = P(X \le (y-b)^2/a^2) = \frac{(y-b)^2}{a^2}$。求导，得 $f_Y(y) = 2(y-b)/a^2$。

(2) 当 $a = 1$，$b = 0$，$E(\sqrt{X}) = \int_0^1 2y^2 dy = 2/3$。因此，$E(Y) = aE(\sqrt{X}) + b = 2a/3 + b$。$\mathrm{Var}(Y) = \mathrm{Var}(a\sqrt{X} + b) = a^2 \mathrm{Var}(\sqrt{X}) = a^2(E(X) - (E(\sqrt{X})^2)) = a^2(1/2 - 4/9) = a^2/18$。

第三题

(1) 当 $k \ge 1$，$P(X = k) = (1/3)^{k-1} \cdot 2/3$。$X \sim G(2/3)$。$E(X) = 3/2$。$\mathrm{Var}(X) = 3/4$。

(2) 当 $k \ge 1$，$P(Y = k) = (1/3)^{k-1} \cdot 1/3 + (1/3)^{k-1} \cdot 2/3 \cdot (k-1)$。这里第一项对应前 $k-1$ 次得 0 分，第 k 次得 2 分的情况。第二项对应前 $k-1$ 次中有一次为 1 分，其余为 0 分，最后一次为 1 分或 2 分。因此 $P(Y = k) = (2k-1)/3^k$。

(3) $E(Y) = \sum_{k \ge 1} (2k^2 - k)/3^k$。注意到，$E(X^2) = \sum_{k \ge 1} 2k^2/3^k = 3$。另外，$E(X) = \sum_{k \ge 1} 2k/3^k = 3/2$。因此，$E(Y) = E(X^2) - E(X)/2 = 3 - 3/4 = 9/4$。也可以使用作业四第三题中基于条件期望的做法。

第四题

令

$$\delta_{ij} = \begin{cases} 1 & \text{球 } j \text{ 在桶 } i \text{ 中} \\ 0 & \text{球 } j \text{ 不在桶 } i \text{ 中} \end{cases}，\sigma_j = \begin{cases} 1 & \text{球 } j \text{ 为蓝色} \\ -1 & \text{球 } j \text{ 为红色} \end{cases}。$$

则有 $Y_i = \sum_{j=1}^n \delta_{ij}\sigma_j$。令 $X_i = \sum_{j=1}^n \delta_{ij}$，也即桶 i 中球的数量。

(1) $E(Y_i) = \sum_{j=1}^n E(\delta_{ij}\sigma_j) = \sum_{j=1}^n E(\delta_{ij})E(\delta_j) = 0$。

(2) $E(Y_i^2) = \sum_{j_1, j_2} E(\delta_{ij_1}\delta_{ij_2}\sigma_{j_1}\sigma_{j_2})$。注意到若 $j_1 \ne j_2$，$E(\sigma_{j_1}\sigma_{j_2}) = E(\sigma_{j_1})E(\sigma_{j_2}) = 0$。因此仅需考虑 $j_1 = j_2$ 的情况，此时 $E(\delta_{ij_1}\delta_{ij_2}\sigma_{j_1}\sigma_{j_2}) = 1/m$，因此 $\mathrm{Var}(Y_i) = E(Y_i^2) = n/m$。另一种做法是注意到 $E(Y_i^2 \mid X_i) = X_i$（作业四第五题第二问），而 $E(X_i) = n/m$。

(3) $E(Z)=\sum_{i=1}^m E(Y_i^2)=n$。

(4) 当 $m=1$，$Y_1=\sum_{j=1}^n \sigma_j$。根据作业四第五题第三问，$E(Z^2)=E(Y_1^4)=3n^2-2n$。因此 $\mathrm{Var}(Z)=E(Z^2)-(E(Z))^2=2n^2-2n$。

(5) 根据上一问，$E(Y_i^4\mid X_i)=3X_i^2-2X_i$。另外，$X_i\sim B(n,1/m)$，因此 $E(X_i)=n/m$，$E(X_i^2)=\mathrm{Var}(X_i)+E(X_i)^2=n(m-1)/m^2+n^2/m^2$。因此，$E(Y_i^4)=3E(X_i^2)-2E(X_i)=3n(m-1)/m^2+3n^2/m^2-2n/m=n/m-3n/m^2+3n^2/m^2$。也可以使用 $Y_i=\sum_{j=1}^n \delta_{ij}\sigma_j$ 进行计算，仿照作业四第五题第三问中的做法对求和中的下标进行分类讨论。

(6) $E(Y_i^2Y_j^2)=\sum_{k_1,k_2,k_3,k_4} E(\delta_{ik_1}\delta_{ik_2}\delta_{jk_3}\delta_{ik_1}\sigma_{k_1}\sigma_{k_2}\sigma_{k_3}\sigma_{k_4})$。由于 $i\neq j$，必然有 k_1 与 k_3 和 k_4 均不相同，且 k_2 与 k_3 和 k_4 均不相同。若 $k_1\neq k_2$，有 $E(\sigma_{k_1}\sigma_{k_2}\sigma_{k_3}\sigma_{k_4})=0$。因此，必然有 $k_1=k_2$，$k_3=k_4$，此时 $E(\delta_{ik_1}\delta_{ik_2}\delta_{jk_3}\delta_{ik_1}\sigma_{k_1}\sigma_{k_2}\sigma_{k_3}\sigma_{k_4})=1/m^2$。因此，$k_1,k_2,k_3,k_4$ 有 $n(n-1)$ 种取法。$E(Y_i^2Y_j^2)=n(n-1)/m^2$。

也可以注意到 $E(Y_i^2Y_j^2\mid X_i,X_j)=X_iX_j$（条件独立，以及作业四第五题第二问），并使用作业四第六题中的结论进行计算。

(7) $E(Z^2)=\sum_{i=1}^m E(Y_i^4)+\sum_{i\neq j}E(Y_i^2Y_j^2)=m\cdot(n/m-3n/m^2+3n^2/m^2)+m(m-1)\cdot n(n-1)/m^2=n^2+2n^2/m-2n/m$。$\mathrm{Var}(Z)=E(Z^2)-(E(Z))^2\leq 4n^2/m$。当 $m\geq 100/\epsilon^2$，$\mathrm{Var}(Z)\leq n^2\epsilon^2/25$。因此，由切比雪夫不等式，$P(|Z-E(Z)|\geq \epsilon n)\leq 1/5$。