EECS 126: Probability and Random Processes

Solutions to Problem Set 4 (mid-term 1)

Note: Please send your score to cchang@eecs.berkeley.edu

Problem 1.1

a) True.
proof:

$$\begin{aligned}
P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cap C)) && (1)\\
&= P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C)) && (2)\\
&= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) && (3)\\
&= P(A)P(B) + P(A)P(C) - P(A)P(B)P(C) && (4)\\
&= P(A)P(B) + P(A)P(C) - P(A)P(B \cap C) && (5)\\
&= P(A)(P(B) + P(C) - P(B \cap C)) && (6)\\
&= P(A)P(B \cup C) && (7)
\end{aligned}$$

(2): De Morgan law, (4), (6) : independence, (7) De Morgan law.

b) False.
First,

$$\begin{aligned}
& Var(X+Y) - (Var(X) + Var(Y)) && (8)\\
= \; & E((X+Y)^2) - E(X+Y)^2 - (E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2) && (9)\\
= \; & 2E(XY) - 2E(X)E(Y) && (10)
\end{aligned}$$

So $Var(X+Y) = Var(X) + Var(Y)$ is equivalent to $E(XY) = E(X)E(Y)$. A counterexample is as follows. The joint PMF of P_{XY} is

	X=-1	X=0	X=1
Y=-1	0.2	0.2	0.2
Y=1	0.2	0	0.2

It's easy to verify that $E(XY) = 0$ and $E(X) = 0$ thus $E(XY) = E(X)E(Y)$, but X, Y are not independent.

c) False.

A counterexample: X is uniformly distributed in the interval $[-0.1, 0.1]$, then $E(X) = 0$, and $f_X(x) = 5$, $x \in [-0.1, 0.1]$.

Problem 1.2

a) 10

b) $100p + 0 \times (1-p) = 100p$

c)

$$\begin{aligned}
&1 + 10P(\text{positive}) + 100P(\text{negative and defective })\\
&= 1 + 10(P(\text{positive}|\text{defective})P(\text{defective}) + P(\text{positive}|\text{non-defective})P(\text{non-defective}))\\
&+100P(\text{non-negative}|\text{defective})P(\text{defective})\\
&= 1 + 10(qp + \frac{q}{5}(1-p)) + 100(1-q)p\\
&= 1 + 100p + 2q - 92pq
\end{aligned} \tag{11}$$

d) We pick the strategy (a) (b) or (c) with the minimum cost expectation. $0 \le p \le 1$, $0 \le q \le 1$ we want to determine the following three regions inside $[0,1] \times [0,1]$:

Region R_a is $\{(p,q) \in [0,1] \times [0,1] | 10 \le 100p, 10 \le 1 + 100p + 2q - 92pq\}$
Region R_b is $\{(p,q) \in [0,1] \times [0,1] | 100p \le 10, 100p \le 1 + 100p + 2q - 92pq\}$
Region R_c is $\{(p,q) \in [0,1] \times [0,1] | 1 + 100p + 2q - 92pq \le 10, 1 + 100p + 2q - 92pq \le 100p\}$
In region $R_i, i = a, b, c$, we apply strategy i.

For $p \le 0.1$, $100p \le 10$ and for $p \ge 0.1$, $10 \le 100p$. So only need to compare $100p$ and $1 + 100p + 2q - 92pq$ when $p \le 0.1$; 10 and $1 + 100p + 2q - 92pq$ when $p \ge 0.1$

d.1)$p \le 0.1$, compare $100p$ and $1 + 100p + 2q - 92pq$.

$$100p \le 1 + 100p + 2q - 92pq \tag{12}$$
$$(92p - 2)q \le 1 \tag{13}$$

If $p \leq \frac{2}{92} = 0.0217$, i.e. $92p - 2 \leq 0$ then the above equality is always true since $q \geq 0$. If $\frac{2}{92} = 0.0217p \leq \frac{3}{92} = 0.0326$, i.e. $92p - 2 \leq 1$ then the above equality is always true because $0 \leq q \leq 1$. If $0.0326 \leq p \leq 0.1$, then $100p \leq 1 + 100p + 2q - 92pq$ if $0 \leq q \leq \frac{1}{92p-2}$ and $1 + 100p + 2q - 92pq \leq 100p$ if $\frac{1}{92p-2} \leq q$

d.2) $p \geq 0.1$, compare 10 and $1 + 100p + 2q - 92pq$

$$10 \leq 1 + 100p + 2q - 92pq \tag{14}$$

$$(92p - 2)q \leq 100p - 9 \tag{15}$$

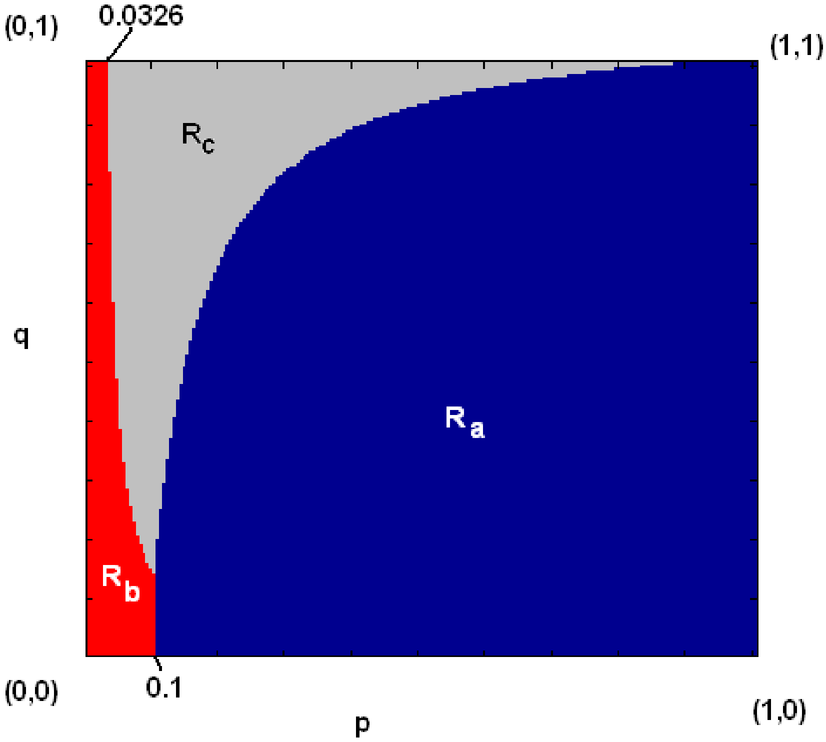
$$q \leq \frac{100p - 9}{92p - 2} \tag{16}$$

Summarize d.1) and d.2) we have

$$R_a = \{(p, q) \in [0, 1] \times [0, 1] | p \geq 0.1 \text{ and } q \leq \frac{100p - 9}{92p - 2}\}$$

$$R_b = \{(p, q) \in [0, 1] \times [0, 1] | p \leq 0.0326 \text{ and } 0.0326 \leq p, \quad q \leq \frac{1}{92p - 2}\}$$

$$R_c = \{(p, q) \in [0, 1] \times [0, 1] | 0.0326 \leq p \leq 0.1, \quad q \geq \frac{1}{92p - 2} \text{ and } p \geq 0.1, \quad q \leq \frac{100p - 9}{92p - 2}\}$$



If the defective probability is too high (p very big), you should just throw it away. If the defective probability is too low, you should just send it to the customers. Otherwise, you should test the chip if the test is good enough (q reasonably big).

Problem 1.3

a)Only on odd number bounces, the request can be served. So the bouncing always ends at odd numbers: $1, 3, 5....$ Write odd numbers as $2 \times 1 - 1, 2 \times 2 - 1, ..., 2 \times k - 1, ...$

$$P_X(X = 2k - 1) = (\frac{1}{2})^k, k = 1, 2, ...$$
$$P_X(X = 2k) = 0, k = 1, 2, ...$$

b) Notice that each time you have the request, the probability that you bounce the request to the neighbor with the file is $\frac{1}{3}$ and the probability that you bounce the request to the neighbors without the file is $\frac{2}{3}$, similarly to part a)

$$P_X(X = 2k - 1) = (\frac{2}{3})^{k-1}\frac{1}{3}, k = 1, 2, ...$$
$$P_X(X = 2k) = 0, k = 1, 2, ...$$

$$\begin{aligned}
E(X) &= \sum_{j=0}^{\infty} jP_X(X = j) && (17)\\
&= \sum_{k=1}^{\infty}(2k - 1)P_X(X = 2k - 1) && (18)\\
&= \sum_{k=1}^{\infty}(2k - 1)(\frac{2}{3})^{k-1}\frac{1}{3} && (19)\\
&= \sum_{k=0}^{\infty}(2k + 1)(\frac{2}{3})^{k}\frac{1}{3} && (20)\\
&= 2\sum_{k=0}^{\infty}k(\frac{2}{3})^{k}\frac{1}{3} + \sum_{k=0}^{\infty}(\frac{2}{3})^{k}\frac{1}{3} && (21)
\end{aligned}$$

The first part is twice of the expectation of a geometric random variable of parameter $\frac{1}{3}$, which is

$$2 \times \frac{1 - \frac{1}{3}}{\frac{1}{3}} = 4$$

The second part is simply the sum of the PMF of a geometric random variable which is 1.

$$E(X) = 4 + 1 = 5$$

c) Suppose that you have n neighbors, A has the subversive file and B is the spy. The probability that the request is bounced to A is equal to the probability that the request is bounced to B, which is $\frac{1}{n}$. So the probability that the request is served at bounce number $2k-1$ and the spy does *not* get the request is:

$$f_k = \frac{1}{n}(\frac{n-2}{n})^{k-1}$$

So the probability that the spy gets the request is:

$$1 - \sum_{k=1}^{\infty} f_k = 1 - \sum_{k=1}^{\infty} \frac{1}{n}(\frac{n-2}{n})^{k-1} \tag{22}$$

$$= 1 - \frac{1}{n}(\frac{1}{1-\frac{n-2}{n}}) \tag{23}$$

$$= \frac{1}{2} \tag{24}$$

It does not depend on n. Another interpretation is by symmetry. The probability that you bounce the request to A before you bounce it to B is equal to the probability that you bounce the request to B before you bounce it to A.

d) Let the $n-1$'th and the n'th neighbor have the file. And define $X_k = 1$ if the k'th neighbor gets the request before it's served, $X_k = 0$ if the k'th neighbor does not get the request before it's served, for $k = 1, 2, ..., n-2$.

$$Y = \sum_{k=1}^{n-2} X_k$$

So

$$E(Y) = \sum_{k=1}^{n-2} E(X_k) = (n-2)E(X_1)$$

The last equality is true because of the symmetry. Similar to c), the probability that the request is served at bounce number $2k-1$ and neighbor 1 does *not* get the request is:

$$f_k = \frac{2}{n}(\frac{n-3}{n})^{k-1}$$

So

$$
\begin{aligned}
E(X_1) &= 1 - \sum_{k=1}^{\infty} f_k & (25)\\
&= 1 - \sum_{k=1}^{\infty} \frac{2}{n}(\frac{n-3}{n})^{k-1} & (26)\\
&= 1 - \frac{2}{n}(\frac{1}{1-\frac{n-3}{n}}) & (27)\\
&= \frac{1}{3} & (28)
\end{aligned}
$$

$$E(Y) = (n-2)E(X_1) = \frac{n-2}{3}$$