Problem 1

Let A and B be two events, and $P(B) \neq 0$. Let X and Y be two discrete random variables, then

i) Bayes' rule for events is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)};$$

for discrete r.v. is

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)},$$

for $p_Y(y) \neq 0$.

ii) Total probability law for events is

$$P(A) = \sum_n P(A|B_n)P(B_n),$$

where (B_n) is a partition of the whole sample space; for discrete r.v. is

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

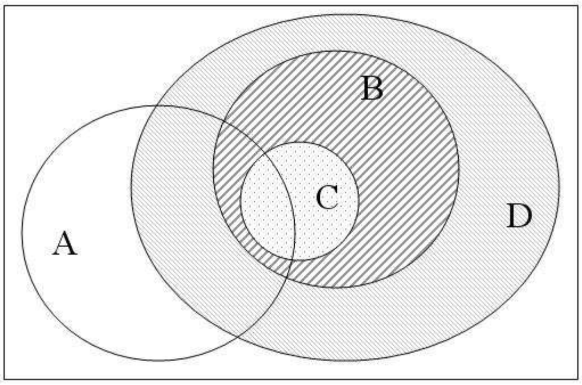
Problem 2

a) none.

b) A and B; A and C; A and D.

c) $C \subset B \subset D$.

d)



Problem 3

a) False. A counter example is as follows. Let X and Y be independent r.v.s, and $Z = X + Y$. Then X and Y are independent by our setting, but they are not independent conditional on Z. In fact, conditional on Z, X and Y are determined by each other.

(b) False. A counter example is as follows. Let Z, N_1 and N_2 be independent r.v.s. Define $X = Z + N_1$, and $Y = Z + N_2$. Then X and Y are obviously not independent by our setting. However they are independent conditional on Z, since then X and Y are determined by N_1 and N_2 respectively and N_1 and N_2 are independent.

Problem 4

a)

$$\begin{aligned} Var(X) = E[(X - E(X))^2] &= E[X^2 - 2XE(X) + E(X)^2] \\ &= E(X^2) - 2E(X)E(X)] + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

b) Let X be a geometric r.v. with parameter p, Y be indicator r.v. of the result of the first trial, i.e. $Y = 1$ represents the first trial successes and $Y = 0$ represents the failure. We use "divide and conquer" method to compute $E(X)$:

$$E(X) = E(X|Y = 1)P(Y = 1) + E(X|Y = 0)P(Y = 0).$$

Conditional on $Y = 1$, the experiment successes in the first trial, thus the conditional expectation is 1; conditional on $Y = 0$, the first trial fails, but the later trials are independent of the first one. Thus the experiment 'restarts' again after the first trial, thus the conditional expectation is $1 + E(X)$. So we have

$$\begin{aligned} & E(X) = 1 \cdot p + (1 + E(X))\,(1 - p); \\ \Rightarrow\quad & E(X) = \frac{1}{p}. \end{aligned}$$

c) From part a) we know we need to compute $E(X^2)$ to get $Var(X)$. To compute $E(X^2)$, we use the same method in part b):

$$\begin{aligned} & E(X^2) = 1 \cdot p + E\left((1 + X)^2\right)(1 - p); \\ \Rightarrow\quad & E(X^2) = p + E(X^2 + 2X + 1)(1 - p); \\ \Rightarrow\quad & E(X^2) = \frac{2 - p}{p^2}. \end{aligned}$$

Finally, we get

$$Var(X) = E(X^2) - E(X)^2 = \frac{1 - p}{p^2}.$$

Problem 5

Part I:
a) Let W represent a packet is transmitted successfully, and L represent a packet is erased. $\Omega = \{WW, WL, LW, LL\}$

b)

$$P(WW) = (1-p)^2, \quad P(WL) = P(LW) = p(1-p), \quad P(LL) = p^2.$$

c) Yes, it is enough to specify the probability law, as follows:

$$P(WW) = 1 - p - p(1-q), \quad P(WL) = P(LW) = p(1-q), \quad P(LL) = pq.$$

d) Let S be the event that packet A is successfully communicated. Then for the model in b), $S = \{LL\}^c$ and thus $P(S) = 1 - p^2$. For the model in c), $P(S) = 1 - pq$.

e) From description, the pair far apart in time can be considered to be transmitted independently. By part d), we know that if $q > p$, i.e. the loss of successive packets are highly positively dependent, then scheme (ii) is better, resulting in higher successfully communication probability.

f) A natural decoding scheme for packet A (and so for packet B) is as follows. If packet A is successfully transmitted, then A is already there and no need for further decoding. If A is erased but B and C are successfully transmitted, then we decode $A = B \bigoplus C$. If none of above events happens, we declare A can not be decoded, and the communications of A fails.
Thus the probability of successfully decoding packet A is just the summation of the probability of two events that result in successful decoding:

$$P(S) = 1 - p + p(1-p)^2 = 1 - p^2 - (p^2 - p^3).$$

Compared to the duplication scheme, we can see the duplication scheme has a higher successful transmission probability. More generally, this result is of two folds. First, for small p, the successful transmission probability is much better than no coding but not as good as the duplication scheme; but the data rate of this scheme is 1.5 times of that of duplication scheme.

Part II:
a) In this part, we assume the feedback channel is perfect, i.e. no feedback packet gets lost. Define a r.v. X as the number of packet transmissions, then X follows the geometric distribution. Then the PMF is

$$p_X(x) = p^{x-1}(1-p), \quad x \in \mathcal{N},$$

and the expectation is

$$E(X) = \frac{1}{1-p}.$$

b) By setting,

$$\begin{aligned} P(X = x) &= P(\text{1st packet lost, 2nd packet lost, } \ldots \text{, xth packet is received}) \\ &= P(\text{1st packet lost})P(\text{2nd packet lost}|\text{1st packet lost}) \cdots \\ &\quad\; P(\text{xth packet is received}|\text{1st, 2nd, } \ldots \text{, packets lost}) \\ &= \begin{cases} 1-p, & x = 1; \\ pq^{x-2}(1-q), & x \geq 2. \end{cases} \end{aligned}$$

Thus the expectation is

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} x\, p_X(x) \\ &= 1-p+\left(\frac{p}{q}\sum_{x=1}^{\infty} x\, q^{x-1}(1-q)\right) - 1\cdot\frac{p}{q}(1-q) \\ &= 1-p+\frac{p}{q}\cdot\frac{1}{1-q}-\frac{p}{q}(1-q) \\ &= 1+\frac{p}{1-q}. \end{aligned}$$

Compared to the expectation in a), we conclude if we want the expected time to be longer, the following must be true:

$$1+\frac{p}{1-q} > \frac{1}{1-p} \quad\Leftrightarrow\quad q > p.$$

c) Under this situation, the number of packet transmissions X still follows geometric distribution, but with parameter $(1-p)(1-r)$ since a transmission is successful if and only if both forward transmission and feedback are successful. Let $p_1 = (1-p)(1-r)$, then we have PMF as

$$p_X(x) = (1-p_1)^{x-1}p_1, \quad x \in \mathcal{N},$$

and expectation as

$$E(X) = \frac{1}{p_1} = \frac{1}{(1-p)(1-r)}.$$