1. Let X be a continuous r.v. X with pdf

$$f_X(x) = \begin{cases} kx & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

where k is a constant.

(a) Determine the value of k and sketch $f_X(x)$

(b) Find and sketch the corresponding cdf $F_X(x)$

(c) Find $P(\frac{1}{4} < X \le 2)$.

Sol)

(a)

By $f_X(x) \ge 0$, we must have $k > 0$ and by $\int_{-\infty}^{\infty} f_X(x)dx = 1$

$$\int_0^1 kxdx = \frac{k}{2} = 1$$

Thus, $k = 2$ and

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

Which is sketched in figure 1.

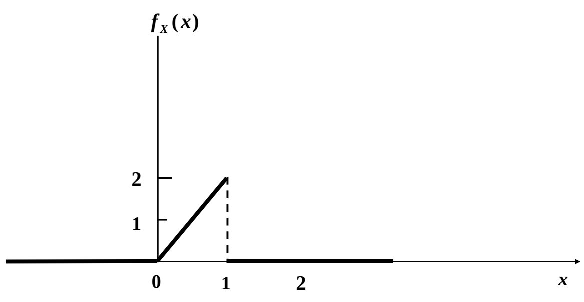


Figure 1

(b) $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(\xi)d\xi$,the cdf of X is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_0^x 2\xi d\xi = x^2 & 0 \le x < 1 \\ \int_0^1 2\xi d\xi = 1 & 1 \le x \end{cases}$$

Which is sketched in figure 2.

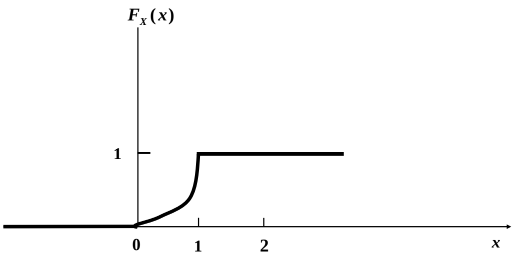


Figure 2

(C) $P(\frac{1}{4} < X \le 2) = F_X(2) - F_X(\frac{1}{4}) = 1 - (\frac{1}{4})^2 = \frac{15}{16}$

2. Let X = $N(0;\sigma^2)$. Find $E(X \mid X > 0)$ and $Var(X \mid X > 0)$

Sol)

The pdf of X = $N(0;\sigma^2)$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}$$

Then

$$f_X(x \mid X > 0) = \begin{cases} 0 & x < 0 \\ \dfrac{f_X(x)}{\int_0^\infty f_X(\xi)d\xi} = 2\dfrac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} & x \geq 0 \end{cases}$$

Hence,

$$E(X \mid X > 0) = 2\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty xe^{-x^2/(2\sigma^2)}dx$$

Let $y = x^2/(2\sigma^2)$. Then $dy = x\,dx/\sigma^2$ and we get

$$E(X \mid X > 0) = \frac{2\sigma}{\sqrt{2\pi}\sigma}\int_0^\infty e^{-y}dy = \sigma\sqrt{\frac{2}{\pi}}$$

Next

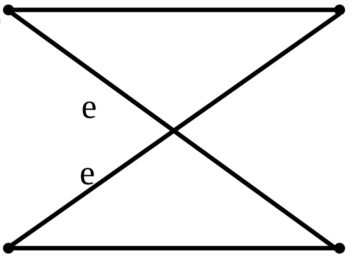
$$E(X^2 \mid X > 0) = 2\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty x^2e^{-x^2/(2\sigma^2)}dx$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty x^2e^{-x^2/(2\sigma^2)}dx = Var(X) = \sigma^2$$

Then we obtain

$$Var(X \mid X > 0) = E(X^2 \mid X > 0) - [E(X \mid X > 0)]^2$$
$$= \sigma^2(1 - \frac{2}{\pi}) \approx 0.363\sigma^2$$

3. A Communication Example During lecture, we looked at a simple point-to-point communication example. The task was to transmit messages from one point to another, by using a noisy channel. We modeled the problem probabilistically as follows:

- A binary message source generates independent successive messages M1, M2, · · · : each message is a discrete random variable that takes the value 1 with probability p and the value 0 with probability $1 - p$. For a single message, we write the PMF as:
- A binary symmetric channel acts on an input bit to produce an output bit, by flipping the input with "crossover" probability e, or transmitting it correctly with probability $1 - e$.



For a single transmission, this channel can be modeled using a conditional PMF $p_{Y|X}(y|x)$, where X and Y are random variables for the channel input and output bits respectively. We then have:

$$p_{Y/X}(y \mid x) = \begin{cases} 1 - e & \text{If } y = x \\ e & \text{If } y \neq x \end{cases}$$

Next, we make a series of assumptions that make the above single-transmission description enough to describe multiple transmissions as well. We say that transmissions are:

(a) Independent: Outputs are conditionally independent from each other, given the inputs.

(b) Memoryless: Only the current input affects the current output.

(c) Time-invariant: We always have the same conditional PMF.

We write:

$$p_{Y_1,Y_2,\cdots|X_1,X_2,\cdots}(y_1, y_2, \cdots |x_1, x_2, \cdots)$$

$$\overset{(a)}{=} p_{Y_1|X_1,X_2,\cdots}(y_1|x_1, x_2, \cdots) \cdot p_{Y_2|X_1,X_2,\cdots}(y_2|x_1, x_2, \cdots) \cdots$$

$$\overset{(b)}{=} p_{Y_1|X_1}(y_1|x_1) \cdot p_{Y_2|X_2}(y_2|x_2) \cdots$$

$$\overset{(c)}{=} p_{Y|X}(y_1|x_1) \cdot p_{Y|X}(y_2|x_2) \cdots$$

Any transmission scheme through this channel must transform messages into channel inputs (encoding), and transform channel outputs into estimates of the transmitted messages (decoding).

For this problem, we will encode each message separately (more elaborate schemes look at blocks and trails of messages), and generate a sequence of n channel input bits. The encoder is therefore a map:

$$\{0, 1\} \rightarrow \{0, 1\}^n$$
$$M \rightarrow \quad X_1, X_2, \cdot\cdot\cdot, X_n$$

The decoder is simply a reverse map:

$$\{0, 1\}^n \rightarrow \{0, 1\}$$
$$Y_1, Y_2, \cdot\cdot\cdot, Y_n \quad \rightarrow \hat{M}$$

Note that we use the "hat" notation to indicate an estimated quantity, but bare in mind that M and $\hat{M}$ are two distinct random variables. The complete communication problem looks as follows:

Finally, to measure the performance of any transmission scheme, we look at the probability of error, i.e. the event that the estimated message is different than the transmitted message:

$$P(error) = P(\hat{M} \neq M)$$

(a) No encoding:

The simplest encoder sends each message bit directly through the channel, i.e. $X_1 = X = M$. Then, a reasonable decoder is to use the output channel directly as message estimate: $\hat{M} = Y_1 = Y$. What is the probability of error in this case?

(b) Repetition code with majority decoding rule:

The next thing we attempt to do is to send each message n > 1 times through this channel. On the decoder end, we do what seems natural: decide 0 when there are more 0s than 1s, and decide 1 otherwise. This is a "majority" decoding rule, and we can write it as follows (making the dependence of $\hat{M}$ on the channel outputs explicit):

$$\hat{M}(y_1, \dots, y_n) = \begin{cases} 0 & \text{If } y_1 + \cdots + y_n < n/2 \\ 1 & \text{If } y_1 + \cdots + y_n \geq n/2 \end{cases}$$

Analytical results:

i. Find an expression of the probability of error as a function of e, p and n. [Hint: First use the total probability rule to divide the problem into solving **P**(error|M = 0) and **P**(error|M = 1), as we did in the lecture.]

ii. Choose $p = 0.5$, $e = 0.3$ and use your favorite computational method to make a plot of P(error) versus n, for $n = 1, \cdot\cdot\cdot, 15$.

iii. Is majority decoding rule optimal (lowest P(error)) for all p?
[Hint: What is the best decoding rule if $p = 0$?].

Simulation:

i. Set $p = 0.5$, $e = 0.3$, generate a message of length 20.

ii. Encode your message with $n = 3$.

iii. Transmit the message, decode it, and write down the value of the message error rate (the ratio of bits in error, over the total number of message bits).

iv. Repeat Step 3 several times, and average out all the message error rates that you obtain. How does this compare to your analytical expression of P(error) for this value of n?

v. Repeat Steps 3 and 4 for $n = 5, 10, 15$, and compare the respective average message error rates to the corresponding analytical P(error).

(c) Repetition code with maximum a posteriori (MAP) rule:

In Part b, majority decoding was chosen almost arbitrarily, and we have alluded to the fact that it might not be the best choice in some cases. Here, we claim that the probability of error is in fact minimized when the decoding rule is as follows:

$$\hat{M}(y_1,\ldots,y_n) = \begin{cases} 0 & \text{If } P(M = 0 \mid Y_1 = y_1,\cdots,Y_n < y_n) \\ & \quad > P(M = 1 \mid Y_1 = y_1,\cdots,Y_n = y_n) \\ 1 & \text{otherwise.} \end{cases}$$

This decoding rule is called "maximum a posterior" or MAP, because it chooses the value of M which maximizes the posterior probability of M given the channel ouput bits $Y_1 = y_1, \cdots, Y_n = y_n$ (another term for Bayes' rule).

i. Denote by N0 the number of 0s in $y_1, \cdots, y_n$, and by N_1 the number of 1s. Express the MAP rule as an inequality in terms of N_0 and N_1, and as a function of e and p. [Hint: Use Bayes' rule to decompose the posterior probability. Note that $p_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n)$ is a constant during one decoding.]

ii. Show that the MAP rule reduces to the majority rule if $p = 0.5$.

iii. Give an interpretation of how, for $p \neq 0.5$, the MAP rule deviates from the majority rule. [Hint: Use the simplification steps of your previous answer, but keep p arbitrary.]

iv. (Optional) Prove that the MAP rule minimizes **P**(error).

Sol)

Analytical Results

(a) No encoding:

$$\begin{aligned} P(error) &= P(\hat{M} = 1, M = 0) + P(\hat{M} = 1, M = 0) \\ &= P(Y = 1, X = 0) + P(Y = 0, X = 1) \\ &= P(Y = 1 \mid X = 0)P(X = 0) + P(Y = 0 \mid X = 1)P(X = 1) \\ &= e(1 - p) + ep \end{aligned}$$

(b) Repetition encoding:

i. Similar to the previous case, the errors can be attributed to the conditions $\hat{M}$ =1, M = 0 and $\hat{M}$ =0, M = 1. Assuming we encode each bit by repeating it n =2m+ 1, m = 0, 1… times, we get

$$\begin{aligned} P(error) &= P(\hat{M} = 1, M = 0) + P(\hat{M} = 1, M = 0) \\ &= P(Y_1 + Y_2 \ldots Y_n \geq m + 1 \mid M = 0)(1 - p) + P(Y_1 + Y_2 \ldots Y_n \leq m \mid M = 1)p \\ &= P(atleast(m + 1)\, errors \mid M = 0)(1 - P) + P(atleast(m + 1)\, errors \mid M = 0)p \\ &= \sum_{k=m+1}^{n} \binom{n}{k} e^{k} (1 - e)^{(n-k)} \end{aligned}$$

Note: It can be seen that the probability of error is independent of the a priori probabilities *p* and (1 −*p*). The error varies only as a function of

ii. The variation of the error probability as a function of n is shown in Figure 3

iii. Even though, the error is independent of p, in degenerate cases such as p=1, p =0, the majority decoding does not give us any advantage. It is easier to always output $\hat{M}$ =1, $\hat{M}$ =0 respectively.

Simulations the matlab code for simulating the communication channel is attached with the solutions. Figure 4 shows the close agreement between the analytical results (indicated by ◦) and the simulated error rates (indicated by ∗).

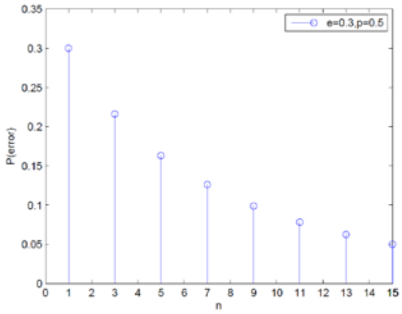


Figure 3 Analytical Results

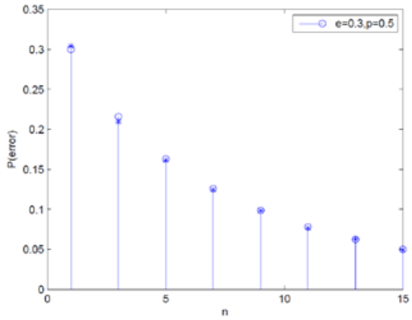


Figure 4 simulation results

(c) Repetition code with maximum a posteriori (MAP) rule:

i. Using Baye's rule to compute the posterior probabilities, we get

$$P(M=0|Y_1=y_1,\cdots,Y_n=y_n)=\frac{P(Y_1=y_1,\cdots,Y_n=y_n|M=0)P(M=0)}{P(Y_1=y_1,\cdots,Y_n=y_n)}$$

$$(numerator)=>e^{N_1}(1-e)^{N_0}(1-p)$$

$$similarly, P(M=1|Y_1=y_1,\cdots,Y_n=y_n)=\frac{P(Y_1=y_1,\cdots,Y_n=y_n|M=1)P(M=1)}{P(Y_1=y_1,\cdots,Y_n=y_n)}$$

$$(numerator)=>e^{N_0}(1-e)^{N_1}(p)$$

Since, the denominators are the same, the MAP decision rule reduces to

$$\hat{M}(y_1,\cdots,y_n) = \begin{cases} 0 & If \;\; e^{N_1}(1-e)^{N_0}(1-p) \geq e^{N_0}(1-e)^{N_1}(p) \\ 1 & otherwise \end{cases}$$

ii. When, p=0.5, the decision rule reduces to

$$\hat{M}(y_1,\cdots,y_n) = \begin{cases} 0 & If \;\; e^{N_1}(1-e)^{N_0} \geq e^{N_0}(1-e)^{N_1} \\ 1 & otherwise \end{cases}$$

The test expression can be rewritten as $(1-e)^{N_0-N_1} \geq e^{N_0-N_1}$ if we assume that $e \leq 0.5$, then $(1-e) \geq e$. Therefore, we can re-write the test as,

$$\hat{M}(y_1,\cdots,y_n) = \begin{cases} 0 & If \;\; N_0 > N_1 \\ 1 & otherwise \end{cases}$$

iii. In this case, we see that the decision rule and hence the probability of error depends oupon the *a-priori* probabilities p and $1-p$. In particular, if we consider the degenerate cases again p =0, p = 1, we see from the rule that we always decide in favor of $\hat{M}$ =0, $\hat{M}$ = 1 irrespective of the received bits. This is in contrast to the majority decoding that still has to count the number of 1s in the output.

iv. Here we resort to a more intuitive proof of the statement that also illustrates the concept of 'risk' in decision. Consider the case when we have no received data and make the decision based entirely on prior probabilities P(M =0) =$1-p$ and P(M =1) = p. If we decide $\hat{M}$ =1, then we have the $1-p$ probability of being wrong. Similarly if we chose $\hat{M}$ =0, we have a probability p of being wrong. We choose $\hat{M}$ to minimize the risk of being wrong or maximize the probability of being right. Thus we choose

$$\hat{M} = arg \max_{M=(0,1)} P(X = M)$$

When we are given the data Y, we have to deal with the *a-posteriori* probabilities P(M =1|Y),P(M =0|Y) instead of *a-priori* probabilities P(M =0),P(M = 1) and the argument remains unchanged. Thus, to minimize probability or being wrong, or maximize the probability of being right, we choose

$$\hat{M} = arg \max_{M=(0,1)} P(X = M \mid Y)$$

4. Let X be uniformly distributed in the unit interval [0,1]. Consider the random variable Y=g(X), where

$$g(x) = \begin{cases} 1 & if\ \ x \le 1/3 \\ 2 & if\ \ x > 1/3 \end{cases}$$

Find the expected value of Y by first deriving its PMF. Verify the results using the expected value rule.

Sol)

The random variable Y = g(X) is discrete and its PMF is given by

$$p_Y(1) = P(X \le 1/3) = 1/3 \qquad\qquad p_Y(2) = 1 - p_Y(1) = 2/3$$

Thus,

$$E[Y] = \frac{1}{3} \times 1 + \frac{2}{3} \times 2 = \frac{5}{3}$$

The same result is obtained using the expected value rule:

$$E(Y) = \int_0^1 g(x) f_X(x) dx = \int_0^{1/3} dx + \int_{1/3}^1 2dx = \frac{5}{3}$$

5. Alvin throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Alvin's hit from the center.

(a) Find the PDF, the mean, and the variance of X.

(b) The target has an inner circle of radius t. If $X \leq t$, Alvin's gets a score of $S=1/X$. Otherwise his score is $S=0$. Find the CDF of S. Is S a continuous random variable?

Sol)

(a) We first calculate the CDF of X. For x $\in$ [0, r], we have

$$F_X(x) = P(X \leq x) = \frac{\pi x^2}{\pi r^2} = \left(\frac{x}{r}\right)^2$$

For $x < 0$, we have $F_X(x) = 0$, and for $x > r$, we have $F_X(x) = 1$. By differentiating, we obtain the PDF

$$f_X(x) = \begin{cases} \frac{2x}{r^2}, & \text{if } 0 \leq x \leq r \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$E(X) = \int_0^r \frac{2x}{r^2}\,dx = \frac{2r}{3}$$

Also

$$E(X^2) = \int_0^r \frac{2x^3}{r^2}\,dx = \frac{r^2}{2}$$

So

$$var(X) = E[X^2] - (E[X])^2 = \frac{r^2}{2} - \frac{4r^2}{9} = \frac{r^2}{18}$$

(b) Alvin gets a positive score in the range $[1/t, \infty)$ if and only if $X \le t$, and otherwise he gets a score of 0. Thus, for $s < 0$, the CDF of S is $F_S(s) = 0$. For $0 \le s < 1/t$, we have

$$F_S(s) = \mathbf{P}(S \le s) = \mathbf{P}\text{ (Alvin's hit is outside the inner circle)} = 1 - \mathbf{P}(X \le t) = 1 - \frac{t^2}{r^2}$$

For $1/t < s$, the CDF of S is given by

$$F_S(s) = \mathbf{P}(S \le s) = \mathbf{P}(X \le t)\mathbf{P}(S \le s \mid X \le t) + \mathbf{P}(X > t)\mathbf{P}(S \le s \mid X > t)$$

We have

$$\mathbf{P}(X \le t) = \frac{t^2}{r^2}, \mathbf{P}(X > t) = 1 - \frac{t^2}{r^2}$$

and since $S = 0$ when $X > t$,

$$\mathbf{P}(S \le s \mid X > t) = 1$$

Furthermore

$$\mathbf{P}(S \le s \mid X \le t) = \mathbf{P}(1/X \le s \mid X \le t) = \frac{\mathbf{P}(1/s \le X \le t)}{\mathbf{P}(X \le t)} = \frac{\frac{\pi t^2 - \pi(1/s)^2}{\pi r^2}}{\frac{\pi t^2}{\pi r^2}} = 1 - \frac{1}{s^2 t^2}$$

Combining the above equations, we obtain

$$\mathbf{P}(S \le s) = \frac{t^2}{r^2}(1 - \frac{1}{s^2 t^2}) + 1 - \frac{t^2}{r^2} = 1 - \frac{1}{s^2 r^2}$$

Collecting the results of the preceding calculations, the CDF of S is

$$F_S(s) = \begin{cases} 0 & \text{if } s < 0 \\ 1 - \frac{t^2}{r^2} & \text{if } 0 \le s < 1/t \\ 1 - \frac{1}{s^2 r^2} & \text{if } 1/t \le s \end{cases}$$

Because F_S has a discontinuity at $s = 0$, the random variable S is not continuous.

6. Let X be a normal random variable with zero mean and standard deviation σ. Use the normal tables to compute the probabilities of the events $\{X \geq k\sigma\}$ and $\{|X| \leq k\sigma\}$ for $k = 1,2,3$.

Sol)

The random variable Z = X/σ is a standard normal,so

$$\boldsymbol{P}(X \geq k\sigma) = \boldsymbol{P}(Z \geq k) = 1 - \Phi(k)$$

From the normal tables we have

$$\Phi(1) = 0.8413, \quad \Phi(2) = 0.9772, \quad \Phi(3) = 0.9986.$$

Thus $\boldsymbol{P}(X \geq \sigma) = 0.1587, \boldsymbol{P}(X \geq 2\sigma) = 0.0228, \boldsymbol{P}(X \geq 3\sigma) = 0.0014.$ *we also have*

$$\boldsymbol{P}(|X| \leq k\sigma) = \boldsymbol{P}(|Z| \leq k) = \Phi(k) - \boldsymbol{P}(Z \leq -k) = \Phi(k) - (1 - \Phi(k)) = 2\Phi(k) - 1.$$

Using the normal table values above, we obtain

$$\boldsymbol{P}(|X| \leq \sigma) = 0.6826, \boldsymbol{P}(|X| \leq 2\sigma) = 0.9544, \boldsymbol{P}(|X| \leq 3\sigma) = 0.9972$$

where *t* is a standard normal random variable.

7. An absent-minded professor schedules two student appointments for the same time. The appointment durations are independent and exponentially distributed with mean thirty minutes. The first student arrives on time, but the second student arrives five minutes late. What is the expected value of the time between the arrival of the first student and the departure of the second student?

Sol)

The expected value in question is

$$\boldsymbol{E}[Time] = (5 + \boldsymbol{E}[stay\ of\ 2nd\ student]) \cdot \boldsymbol{P}(1st\ stays\ no\ more\ tan\ 5\ minutes) +$$
$$(\boldsymbol{E}[stay\ of\ 1st \mid stay\ of\ 1st \geq 5] + \boldsymbol{E}[stay\ of\ 2nd]) \cdot \boldsymbol{P}(1st\ stays\ more\ than\ 5\ minutes)$$

We have **E**[stay of 2nd student] = 30, and, using the memorylessness property of the exponential distribution,

$$\boldsymbol{E}[stay\ of\ 1st \mid stay\ of\ 1st \geq 5] = 5 + \boldsymbol{E}[stay\ of\ 1st] = 35$$

Also

$$\boldsymbol{P}(1st\ stays\ no\ more\ tan\ 5\ minutes) = 1 - e^{-5/30}$$

$$\boldsymbol{P}(1st\ stays\ more\ tan\ 5\ minutes) = e^{-5/30}$$

By substitution we obtain

$$\boldsymbol{E}[Time] = (5 + 30) \cdot (1 - e^{-5/30}) + (35 + 30) \cdot e^{-5/30} = 35 + 30e^{-5/30} = 60.394$$