ECEN 649 Pattern Recognition – Spring 2018
Problem Set 1
Due on: Feb 8

Problems:

1. This problem demonstrates nicely subtle issues regarding partial information and prediction. A certain show host has placed a case with US$1,000,000 behind one of three identical doors. Behind each of the other two doors he placed a donkey. The host asks the contestant to pick one door but not to open it. The host then opens one of the other two doors to reveal a donkey. He then asks the contestant if he wants to stay with his door or switch to the other unopened door. Assume that the host is honest and that if the contestant initially picked the correct door, the host randomly picks one the two donkey doors to open. Which of the following strategies is rationally justifiable:

   (a) The contestant should never switch to the other door.

   (b) The contestant should always switch to the other door.

   (c) There is not enough information or the choice between (a) and (b) is indifferent.

   To get full credit, you must argue this by correctly computing the probabilities of success.

   Solution: Let us define the events:

   $$A = \{\text{The door first opened by the contestant has the prize}\}$$
   $$B = \{\text{The last unopened door has the prize}\}$$

   There are three possibilities: the contestant should never switch, always switch, or it does not matter, if $P(A) > P(B)$, $P(A) < P(B)$, or $P(A) = P(B)$, respectively. It is obvious that $P(A) = \frac{1}{3}$. As for $P(B)$, we have:

   $$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

   Clearly, $P(B|A) = 0$ and $P(B|A^c) = 1$ (the latter is so because the host is forced to open the remaining donkey door), so that

   $$P(B) = 0 + 1.P(A^c) = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}$$

   Therefore, $P(B) > P(A)$, and the contestant should always switch.

   The solution was easy because of the appropriate definition of the events. The solution can be much more complicated if the events are defined differently — for example, take a look at the solution in the Wikipedia entry on this problem (live link):

   http://en.wikipedia.org/wiki/Monty_Hall_problem

   Note also that the answer would be different if the host was dishonest. For example, if the host would only open a door and offer a switch if the contestant picked the correct door first, then the probability of winning by switching would be zero.

Note: This problem is sometimes called the "Monty Hall Problem" (Monty Hall was a popular TV game show host). It is a "paradox" in the sense that many people intuitively expect that, since there are two options available after the host opens a door, one should have $P(A) = P(B) = \frac{1}{2}$, so it should not matter whether the contestant switches or not. On the other hand, other people feel that the host may be trying to mislead the contestant, who thus should never switch doors (this is of course not allowed in the above formulation of the problem). These wrong perceptions do not stand up to the probabilistic analysis of the problem. This problem is completely equivalent to another (and older) paradox called "The Three Prisoners Problem," proposed by Martin Gardner, in which there are three prisoners, one of which is going to be executed and the rest will be pardoned. The prison guard reveals to the first prisoner which of the other two will be freed, which effectively makes the first prisoner not want to switch his fate with the remaining prisoner. For those who are curious about the Monty Hall paradox, I recommend the Wikipedia entry mentioned earlier. In addition, you can actually play the game here: `http://math.ucsd.edu/~crypto/Monty/monty.html`.

2. The random experiment consists of throwing two fair dice. Let us define the events:

   $D = \{\text{the sum of the dice equals 6}\}$

   $E = \{\text{the sum of the dice equals 7}\}$

   $F = \{\text{the first die lands 4}\}$

   $G = \{\text{the second die lands 3}\}$

   Show the following, both by arguing and by computing probabilities:

   (a) D is not independent of F and D is not independent of G.

   Solution: Intuitively, D is not independent of either F or G because if we are interested in throwing a combined 6 on the sum of the dice, it is necessary *not* to throw a 6 on any individual die. So the occurrence of D depends on the outcomes of each of the dice, and thus D cannot be independent of either F or G. In probabilistic terms, we have:

   $$P(D) = P(\{1,5\} \cup \{2,4\} \cup \{3,3\} \cup \{4,2\} \cup \{5,1\}) = \frac{5}{36}$$

   so that

   $$P(D,F) = P(\{4,2\}) = \frac{1}{36} \neq \frac{5}{216} = \frac{5}{36} \times \frac{1}{6} = P(D)P(F)$$

   and similarly

   $$P(D,G) = P(\{3,3\}) = \frac{1}{36} \neq \frac{5}{216} = \frac{5}{36} \times \frac{1}{6} = P(D)P(G)$$

   (b) E is independent of F and E is independent of G.

   Solution: Here, a curious thing happens. No single outcome on any of the die can invalidate the possibility of throwing a combined 7 on the sum of the dice. So we cannot conclude as before that E is not independent of either F or G. Let us examine the probabilities:

   $$P(E) = P(\{1,6\} \cup \{2,5\} \cup \{3,4\} \cup \{4,3\} \cup \{5,2\} \cup \{6,1\}) = \frac{6}{36} = \frac{1}{6}$$

so that

$$P(E,F) = P(\{4,3\}) = \frac{1}{36} \;=\; \frac{1}{6} \times \frac{1}{6} = P(E)P(F)$$

and similarly

$$P(E,G) = P(\{4,3\}) = \frac{1}{36} \;=\; \frac{1}{6} \times \frac{1}{6} = P(E)P(G)$$

Therefore, E is indeed independent of both F and G (separately). This is not true for any combined sum other than 7.

(c) E is not independent of (F, G), in fact, E is completely determined by (F, G). (Here is an example where an event is independent of each of two other events but is not independent of the joint occurrence of these events.)

Solution: If we consider the outcomes of both dice together, then obviously the sum will depend on that. Furthermore, it will be completely determined in the sense that the conditional probability of the sum given the individual outcomes will be either one or zero. In the present case, we have:

$$P(E|F,G) = \frac{P(E,F,G)}{P(F,G)} = \frac{P(\{4,3\})}{P(\{4,3\})} = 1 \;\neq\; \frac{1}{6} = P(E)$$

3. Suppose that a typist monkey is typing randomly, but that each time he types the "wrong character," it is discarded from the output. Assume also that the monkey types 24-7 at the rate of one character per second, and that each character can be one of 27 symbols (the alphabet without punctuation plus space). Given that *Hamlet* has about 130,000 characters, what is the average number of days that it would take the typist monkey to compose the famous play?

   Solution: Let T_i be the amount of tries the monkey takes to get the i-th character correct. The variables T_i are independent and identically-distributed. Furthermore, each T_i is a geometric random variable with parameter $p = 1/27$, so that $E[T_i] = 27$. The average total time to complete *Hamlet* is simply $130,000 \times E[T_1] = 130,000 \times 27 = 3510000$ seconds (since each try takes one second). As each day contains 86400 seconds, this corresponds to 40.625 days, or precisely 40 days and 15 hours.

4. Suppose that 3 balls are selected without replacement from an urn containing 4 white balls, 6 red balls, and 2 black balls. Let $X_i = 1$ if the i-th ball selected is white, and let $X_i = 0$ otherwise, for $i = 1, 2, 3$. Give the joint PMF of

   (a) X_1, X_2

   Solution: By conditioning on the first draw,

$$\begin{aligned}
P(X_1 = 0, X_2 = 0) &= P(X_2 = 0 \mid X_1 = 0)P(X_1 = 0) = \frac{7}{11} \times \frac{8}{12} = \frac{14}{33} \\
P(X_1 = 0, X_2 = 1) &= P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) = \frac{4}{11} \times \frac{8}{12} = \frac{8}{33} \\
P(X_1 = 1, X_2 = 0) &= P(X_2 = 0 \mid X_1 = 1)P(X_1 = 1) = \frac{8}{11} \times \frac{4}{12} = \frac{8}{33} \\
P(X_1 = 1, X_2 = 1) &= P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) = \frac{3}{11} \times \frac{4}{12} = \frac{1}{11}
\end{aligned}$$

   Another way of solving this is to imagine that the balls are all numbered from 1 to 12. Then there is clearly a total of 12×11 possible outcomes for the first two draws. Of these, 8×7 consist of two non-white balls, 8×4 consist of one white ball and one non-white ball, and 4×3 consists of two white balls.

   (b) X_1, X_2, X_3

   Solution: By conditioning on the first two draws, and using the result in a),

$$\begin{aligned}
P(X_1 = 0, X_2 = 0, X_3 = 0) &= P(X_3 = 0 \mid X_1 = 0, X_2 = 0)P(X_1 = 0, X_2 = 0) \\
&= \frac{6}{10} \times \frac{14}{33} = \frac{14}{55} \\
P(X_1 = 0, X_2 = 0, X_3 = 1) &= P(X_3 = 1 \mid X_1 = 0, X_2 = 0)P(X_1 = 0, X_2 = 0) \\
&= \frac{4}{10} \times \frac{14}{33} = \frac{28}{165} \\
P(X_1 = 0, X_2 = 1, X_3 = 0) &= P(X_3 = 0 \mid X_1 = 0, X_2 = 1)P(X_1 = 0, X_2 = 1) \\
&= \frac{7}{10} \times \frac{8}{33} = \frac{28}{165} \\
P(X_1 = 0, X_2 = 1, X_3 = 1) &= P(X_3 = 1 \mid X_1 = 0, X_2 = 1)P(X_1 = 0, X_2 = 1) \\
&= \frac{3}{10} \times \frac{8}{33} = \frac{12}{165} \\
P(X_1 = 1, X_2 = 0, X_3 = 0) &= P(X_3 = 0 \mid X_1 = 1, X_2 = 0)P(X_1 = 1, X_2 = 0) \\
&= \frac{7}{10} \times \frac{8}{33} = \frac{28}{165}
\end{aligned}$$

$$
\begin{aligned}
P(X_1 = 1, X_2 = 0, X_3 = 1) &= P(X_3 = 1 \mid X_1 = 1, X_2 = 0)P(X_1 = 1, X_2 = 0) \\
&= \frac{3}{10} \times \frac{8}{33} = \frac{12}{165} \\
P(X_1 = 1, X_2 = 1, X_3 = 0) &= P(X_3 = 0 \mid X_1 = 1, X_2 = 1)P(X_1 = 1, X_2 = 1) \\
&= \frac{8}{10} \times \frac{1}{11} = \frac{4}{55} \\
P(X_1 = 1, X_2 = 1, X_3 = 1) &= P(X_3 = 1 \mid X_1 = 1, X_2 = 1)P(X_1 = 1, X_2 = 1) \\
&= \frac{2}{10} \times \frac{1}{11} = \frac{1}{55}
\end{aligned}
$$

As before, this could be solved by realizing there are $12 \times 11 \times 10$ ways of performing the first three draws. Of these, $8 \times 7 \times 6$ consist of three non-white balls, etc.

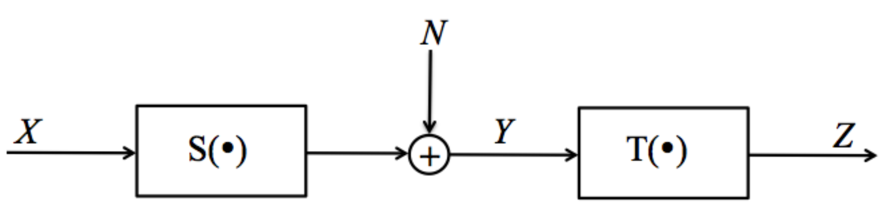
5. Consider 12 independent rolls of a 6-sided die. Let X be the number of 1's and let Y be the number of 2's obtained. Compute $E[X]$, $E[Y]$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $E[X+Y]$, $\mathrm{Var}(X+Y)$, $\mathrm{Cov}(X,Y)$, and $\rho(X,Y)$. (Hint: You may want to compute these in the order given.)

   Solution: Note that X and Y are binomial r.v.s with parameters $(n = 12, p = \frac{1}{6})$, whereas $X + Y$ is a binomial r.v. with parameters $(n = 12, p = \frac{1}{3})$. Therefore,

$$
\begin{aligned}
E[X] &= E[Y] = 12 \times \frac{1}{6} = 2 \\
\mathrm{Var}(X) &= \mathrm{Var}(Y) = 12 \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{3} \\
E[X+Y] &= 12 \times \frac{1}{3} = 4 \quad (= E[X] + E[Y]) \\
\mathrm{Var}(X+Y) &= 12 \times \frac{1}{3} \times \frac{2}{3} = \frac{8}{3} \quad (\neq \mathrm{Var}(X) + \mathrm{Var}(Y)!) \\
\mathrm{Cov}(X,Y) &= \frac{1}{2}(\mathrm{Var}(X+Y) - \mathrm{Var}(X) - \mathrm{Var}(Y)) = \frac{1}{2}\left(\frac{8}{3} - \frac{5}{3} - \frac{5}{3}\right) = -\frac{1}{3} \\
\rho(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-\frac{1}{3}}{\frac{5}{3}} = -0.2\,.
\end{aligned}
$$

   Notice that X and Y are negatively correlated (since the more 1's there are, the fewer 2's there must be, and vice-versa).

6. Consider the system represented by the block diagram below.



The functionals are given by $S(X) = aX + b$, and $T(Y) = Y^2$. The additive noise is $N \sim N(0, \sigma_N^2)$. Assuming that the input signal is $X \sim N(\mu_X, \sigma_X^2)$:

(a) Find the pdf of Y.

Solution: First recall that if $X \sim N(\mu_X, \sigma_X^2)$ then $S(X) = aX + b$ is again Gaussian, with $S(X) \sim N(a\mu_X + b, a^2\sigma_X^2)$. In addition, recall that the sum of two independent Gaussian random variables is again Gaussian, the mean and variance of which are simply the sum of the means and variances, respectively, of the original variables. Since $S(X)$ and N are independent, we have $Y = S(X) + N \sim N(a\mu_X + b, a^2\sigma_X^2 + \sigma_N^2)$. The PDF of Y is therefore

$$f_Y(y) = \frac{1}{\sqrt{2\pi(a^2\sigma_X^2 + \sigma_N^2)}} \exp\left\{-\frac{(y - a\mu_X - b)^2}{2(a^2\sigma_X^2 + \sigma_N^2)}\right\}.$$

(b) Find the pdf of Z.

Solution: First we find the PDF of Z:

$$F_Z(z) = P(Z \le z) = P(Y^2 \le z) = P(-\sqrt{z} \le Y \le \sqrt{z}) = F_Y(\sqrt{z}) - F_Y(-\sqrt{z}).$$

Differentiation gives the pdf of Z:

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{1}{2\sqrt{z}}(f_Y(\sqrt{z}) + f_Y(-\sqrt{z})).$$

Using the result of the previous item, one obtains:

$$f_Z(z) = \frac{1}{2\sqrt{2\pi(a^2\sigma_X^2 + \sigma_N^2)z}} \left(\exp\left\{-\frac{(\sqrt{z} - a\mu_X - b)^2}{2(a^2\sigma_X^2 + \sigma_N^2)}\right\} + \exp\left\{-\frac{(\sqrt{z} + a\mu_X + b)^2}{2(a^2\sigma_X^2 + \sigma_N^2)}\right\}\right).$$

(c) Compute the probability that the output is bounded by a constant $k > 0$, i.e., find $P(Z \le k)$.

Solution: From the previous item:

$$\begin{aligned}
P(Z \le k) &= P(-\sqrt{k} \le Y \le \sqrt{k}) = P\left(\frac{-\sqrt{k} - \mu_Y}{\sigma_Y} \le \frac{Y - \mu_Y}{\sigma_Y} \le \frac{\sqrt{k} - \mu_Y}{\sigma_Y}\right) \\
&= \Phi\left(\frac{\sqrt{k} - \mu_Y}{\sigma_Y}\right) - \Phi\left(\frac{-\sqrt{k} - \mu_Y}{\sigma_Y}\right) \\
&= \Phi\left(\frac{\sqrt{k} - a\mu_X - b}{\sqrt{a^2\sigma_X^2 + \sigma_N^2}}\right) - \Phi\left(\frac{-\sqrt{k} - a\mu_X - b}{\sqrt{a^2\sigma_X^2 + \sigma_N^2}}\right).
\end{aligned}$$

7. (Bi-variate Gaussian Distribution) Suppose (X, Y) are jointly Gaussian.

(a) Show that the joint pdf is given by:

$$\begin{aligned}
f_{X,Y}(x, y) = &\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \\
&\times \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x - \mu_x}{\sigma_x}\right)^2 + \left(\frac{y - \mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y}\right]\right\}
\end{aligned}$$

where $E[X] = \mu_X$, $\mathrm{Var}(X) = \sigma_X^2$, $E[Y] = \mu_Y$, $\mathrm{Var}(Y) = \sigma_Y^2$, and ρ is the correlation coefficient between X and Y.

Solution: The multivariate Gaussian density is given by:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \tag{1}$$

In the bivariate case, one has

$$d = 2, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & \mathrm{cov}(x,y) \\ \mathrm{cov}(x,y) & \sigma_y^2 \end{bmatrix}$$

$$\det(\Sigma) = \sigma_x^2\sigma_y^2 - \mathrm{cov}^2(x,y) = \sigma_x^2\sigma_y^2(1-\rho^2)$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$$

$$-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = -\frac{1}{2(1-\rho^2)} \begin{bmatrix} x-\mu_x & y-\mu_y \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}$$

$$= -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]$$

Substituting these into (1) yields

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

which is the required expression.

(b) Show that the conditional pdf of Y, given $X = x$, is a univariate Gaussian density with parameters:

$$\mu_{Y|X} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x) \quad \text{and} \quad \sigma_{Y|X}^2 = \sigma_y^2(1-\rho^2)$$

Solution: From the definition of conditional pdf:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

We know that the marginal density $f_X(x)$ is a univariate Gaussian with parameters μ_x and σ_x^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)$$

whereas the joint density $f_{XY}(x, y)$ was calculated in the previous item. Substituting these into (1) yields, after some algebraic manipulation:

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y - \mu_y - \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right)^2\right]$$

By direct inspection, we can see that this is a univariate Gaussian density with parameters:

$$\mu_{Y|X} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x) \quad \text{and} \quad \sigma_{Y|X}^2 = \sigma_y^2(1-\rho^2)$$

as required.

(c) Conclude that the conditional expectation $E[Y|X]$ (which can be shown to be the "best" predictor of Y given X), is in the Gaussian case a linear function of X. This is the foundation of optimal linear filtering in Signal Processing. Plot the regression line for the case $\sigma_x = \sigma_y$, $\mu_x = 0$, fixed μ_y and a few values of ρ. What do you observe as the correlation ρ changes? What happens for the case $\rho = 0$?

Solution: From the previous item, we conclude that the

$$E[Y|X = x] = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x) = \rho\frac{\sigma_y}{\sigma_x}\,x + \left(\mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x\right)$$

so that $E[Y|X = x] = ax + b$ is a linear function of x. In the case $\sigma_x = \sigma_y$, $\mu_x = 0$, this reduces to a line with slope $a = \rho$ and intercept $b = \mu_y$, and the prediction will deviate from the mean μ_y by an amount proportional to the value $X = x$, with sensitivity given by the correlation coefficient ρ. If additionally $\rho = 0$, then the regression line is horizontal. In this case, the variable Y is uncorrelated from X (and thus independent, since they are jointly Gaussian) and there is no change in prediction as the value $X = x$ varies; the best predictor reduces to the no-information constant estimator μ_y.

8. Consider the example of a random sequence $X(n)$ of 0-1 binary r.v.'s given in class:

- Set X(0) = 1
- From the next 2 points, pick one randomly and set to 1, the other to zero.
- From the next 3 points, pick one randomly and set to 1, the rest to zero.
- From the next 4 points, pick one randomly and set to 1, the rest to zero.
- ...

Show that $X(n)$:

(a) converges to 0 in probability

Solution: The sequence is composed of blocks, where the first block has length 1, the second block has length 2, and so on. Let $B(n)$ be the ordinal number of the block to which n belongs. Clearly,

$$P(X_n = 1) = \frac{1}{B(n)}$$

Note that $B(n) \to \infty$ as $n \to \infty$. Therefore, for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - 0| > \epsilon) = \lim_{n\to\infty} P(X_n = 1) = \lim_{n\to\infty} \frac{1}{B(n)} = 0$$

so that $X_n \to 0$ in probability.

(b) converges to 0 in the mean-square sense

Solution: Each X_n is a Bernoulli random variable with parameter $P(X_n = 1)$. It is easy to see then that

$$E[X_n^2] = 1^2 \times P(X_n = 1) + 0^2 \times P(X_n = 0) \; = \; \frac{1}{B(n)}$$

Therefore, we have that

$$\lim_{n\to\infty} E[|X_n - 0|^2] \; = \; \lim_{n\to\infty} E[X_n^2] \; = \; \lim_{k\to\infty} \frac{1}{B(n)} = 0$$

so that $X_n \to 0$ in the mean-square sense. Of course, we know that this implies that $X_n \to 0$ in probability, so showing (b) automatically shows (a).

(c) does not converge to 0 with probability 1. In fact, show that

$$P\left(\lim_{n\to\infty} X(n) = 0\right) = 0$$

Solution: Within each block, the probability of getting a 1 is one and so is the probability of getting a 0. By the 2nd Borel-Cantelli lemma, we conclude that

$$P([X_n = 1 \;\; i.o.]) = P([X_n = 0 \;\; i.o.]) = 1$$

It follows that

$$P\left(\lim_{n\to\infty} X(n) = 0\right) = 0$$

and thus X_n does not converge to 0 with probability one.