ENEE620. Midterm examination, 4/25/2019. Instructor: A. Barg

READ THIS:
- The exam consists of four problems (Problem 1 to Problem 4). Each problem is 10 points. Max score=40 points
- Your answers should be justified. Giving just the answer may result in no credit for the problem.
- You do not have to copy problem statements into your paper.
- Please write legibly. If applicable, clearly identify the answer to the question.

Problem 1. (a) Let X, Y, Z be i.i.d. RVs. Show that X and Y are conditionally independent given Z. Show that X and $X+Y+Z$ are conditionally independent given $X+Y$.
(b) Let $Y_0, Y_1, \dots$ be a sequence of i.i.d. RV's defined on $\mathbb{Z}$. Show that the sequence $(X_n, n \ge 0)$ given by

$$X_n = \sum_{k=0}^{n} Y_k$$

forms a Markov chain.

SOLUTION: Part (a) is straightforward: Let $E_1, E_2, E_3 \in \mathcal{B}(\mathbb{R})$, then

$$P(X \in E_1, Y \in E_2 | Z \in E_3) = \frac{P(X \in E_1, Y \in E_2, Z \in E_3)}{P(Z \in E_3)}$$

$$= P(X \in E_1)P(Y \in E_2) = P(X \in E_1 | Z \in E_3)P(Y \in E_2 | Z \in E_3)$$

because $P(Y \in E_2) = P(Y \in E_2 | Z \in E_3)$ by independence

Similarly

$$P(X \in E_1, X+Y+Z \in E_2 | X+Y \in E_3) = \frac{P(X \in E_1, X+Y+Z \in E_2, X+Y \in E_3)}{P(X+Y \in E_3)}$$

$$= \frac{P(X \in E_1, Z \in E_2 - E_3, X+Y \in E_3)}{P(X+Y \in E_3)} = \frac{P(X \in E_1, X+Y \in E_3)P(Z \in E_2 - E_3)}{P(X+Y \in E_3)}$$

$$= P(X \in E_1 | X+Y \in E_3)P(Z \in E_2 - E_3) = P(X \in E_1 | X+Y \in E_3)P(Z \in E_2 - E_3 | X+Y \in E_3)$$

$$= P(X \in E_1 | X+Y \in E_3)P(X+Y+Z \in E_2 | X+Y \in E_3). \quad (1)$$

It is an interesting question (i), what does it mean to subtract sets, and (ii), whether $E_2 - E_3$ is defined. To answer these, note that $A - B = \{x - y | x \in A, y \in B\}$, and Borel sets are generated by open intervals, and we can subtract intervals as described, and thus, we can also subtract Borel sets.

(No, you cannot write $P(X, Y|Z)$ because this expression is not defined, and no, it is not enough to write $P(X = a, Y = b|Z = c)$ because this may very well be zero).

(b) For $n = 2$ this is proved in (1); after that extend by induction.

Problem 2. Consider a Markov chain with transition matrix

$$P = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where $\sum_{i\ge 1} p_i = 1$ and $p_i > 0$ for all i.
(a) Show that the chain is irreducible and recurrent.
(b) Find a necessary and sufficient condition for the chain to be positive recurrent.
(c) Find the stationary distribution if one exists.

SOLUTION:

(a) The chain is clearly irreducible because every pair of states communicates. It is recurrent because the *probability of first return in n steps* equals $\bar{p}_{11}^{(n)} = p_n$ and thus $\sum_{n\ge 1} \bar{p}_{11}^{(n)} = 1$.

(b) Positive recurrence we need that $m_1 = \sum_{n\ge 1} np_{11}^{(n)} = \sum_{n\ge 1} np_n < \infty$, equivalently, that the pmf p has a finite expectation. A necessary condition for this is that $np_n \to 0$, and there are multiple sufficient conditions, e.g., $\lim_{n\to\infty} \frac{(n+1)p_{n+1}}{np_n} < 1$, etc.

(c) The stationary distribution is obtained from

$$\pi_1 = \pi_1 p_1 + \pi_2$$
$$\pi_2 = \pi_1 p_2 + \pi_3$$
$$\dots$$

Or

$$\pi_2 = \pi_1(1 - p_1)\pi_3 = \pi_1(1 - p_1 - p_2)$$
$$\dots$$
$$\pi_n = \pi_1(1 - \sum_{i=1}^{n-1} p_i) = \pi_1 \sum_{i\ge n} p_i$$

whence, summing the above equations for $n = 1, 2, \dots,$

$$\pi_1 = \frac{1}{\sum_{n=1}^{\infty} \sum_{i\ge n} p_i} = \frac{1}{Ep}; \quad \pi_n = \frac{1}{Ep} \sum_{i\ge n} p_i, n \ge 2.$$

The stationary distribution exists under the assumption in Part (b).

Problem 3. There are two independent Poison processes, with rates $\lambda_1 = 1$ and $\lambda_2 = 7$ per hour, respectively, passing through a registering device. Call them process of type 1 and process of type 2.
(a) What is the probability that the device registers exactly three arrivals (never mind which type) during the first hour?
(b) What is the probability that exactly three type-2 arrivals occur before the first arrival of type 1?
(c) Suppose that exactly 1/2 of all arrivals in each of the processes are diverted before reaching the registering device (in other words, those that reach are, deterministically, even-numbered arrivals in each of the two streams). What is the probability that the device registers no arrivals in 30 minutes?

SOLUTION: (a) The process of arrivals $N(t)$ is PP$(\lambda_1+\lambda_2)$, and thus

$$P(N(1)=3)=\frac{512}{6}e^{-8}.$$

(b) Conditional of the fact that the merged process registered n arrivals, the number of arrivals in the type-1 process is a binomial RV Binom(n,p), where $p=\frac{\lambda_1}{\lambda_1+\lambda_2}$. Thus, the answer is $p^3=(7/8)^3$.

(c) The question means that arrivals numbered 2 in each of the processes did not show up within the first 30 minutes. The time to the 2nd arrival in the Poisson process is Gamma-distributed; $P(T_2>t)=1-F(T_2<t)=1-\int_0^t \lambda^2 x e^{\lambda x}dx=(1+t\lambda)e^{-\lambda t}$. Thus, the required probability is

$$(1+0.5)(1+3.5)e^{-4}=6.75e^{-4}.$$

Or, a less fancy way: the event in question occurs exactly when each of the two processes registers at most one arrival, and thus, the answer is

$$(e^{-1/2}+(1/2)e^{-1/2})(e^{-7/2}+(7/2)e^{-7/2})=(1+\frac{7}{2}+\frac{1}{2}+\frac{7}{4})e^{-7/2},$$

as before.

Problem 4. Let $(X_k, k\geq 1)$ be a sequence of i.i.d. RVs with a finite mean. In this problem we study convergence of the averages $S_n:=\frac{1}{n}\sum_{k=1}^n X_k X_{k+1}$.
(a) True or False: SLLN implies that $S_n \stackrel{\text{a.s.}}{\to} (EX_1)^2$?
(b) Define $Y_k=X_{2(k-1)+1}X_{2k}, Z_k=X_{2k}X_{2k+1}, k\geq 1$. Do the sequences $\frac{1}{m}\sum_{k=1}^m Y_k$, $\frac{1}{m}\sum_{k=1}^m Z_k$ converge; if yes, then how, to which limits, and for what reason?
(c) Now write S_n as a sum of $\frac{1}{m}\sum_{k=1}^m Y_k$ and $\frac{1}{m}\sum_{k=1}^m Z_k$ and argue that S_n has a limit; also find this limit and characterize the type of convergence.

SOLUTION:

(a) False because X_1X_2 and X_2X_3 are not independent; thus SLLN cannot be used directly. At the same time, the dependence is weak and the statement of interest is still true, as shown below.

(b) Both sequences Y_k and Z_k are formed of iid RV's, and thus the sample averages $\frac{1}{n}\sum_{k=1}^n Y_k$ and $\frac{1}{n}\sum_{k=1}^n Z_k$ converge to $(EX)^2$ a.s.

(c) For $n=2m$ we have

$$S_n=\frac{1}{2m}\sum_{k=1}^m Y_k+\frac{1}{2m}\sum_{k=1}^m Z_k$$

and for $n=2m+1$ we have

$$S_n=\frac{m+1}{2m+1}\frac{1}{m+1}\sum_{k=1}^{m+1} Y_k+\frac{m}{2m+1}\frac{1}{m}\sum_{k=1}^m Z_k.$$

In both cases when $m\to\infty$ we obtain that $S_n \stackrel{\text{a.s.}}{\to} \frac{1}{2}(EX)^2+\frac{1}{2}(EX)^2=(EX)^2$.