Answers to Stochastic Process Pre-req Test

BA 2022, MSE, July 25th, 3:45-4:15 PM, 15 marks

1. Write your name and roll number on the top right.
2. Answer all the questions on the paper and return the paper. You can draw a margin and do rough work on the side. **Put a box around the final answer.**
3. For Problems $1, 2, 3, 4, 5, 6$: **Just write the final answer**. Use the space for rough work.
4. The symbol $X \sim D(f)$ means that the random variable X is distributed according to the distribution D with parameters f.
5. A Bernoulli(p) random variable X takes only two values $\{0, 1\}$ for the purposes of this test. By definition,

$$\Pr(X = 1) = p.$$

Questions

1. **[6 marks]** Write down the range *of the random variable*, pdf/pmf, the mean and the variance of the following distributions. I have written the first one as an example.
   (a) Exponential(λ):
      i. Range is the set of all non-negative real numbers.
      ii. Probability density function $f(x) = \lambda e^{-\lambda x}$ in the range.
      iii. Mean is $\frac{1}{\lambda}$ and variance is $\frac{1}{\lambda^2}$.

   (b) Poisson(λ)
      i. Range is the set of <u>all non negative integers</u>.
      ii. Probability mass function

$$\underline{f(k) = \frac{\lambda^k e^{-\lambda}}{k!}}$$

      in the range.
      iii. Mean is $\underline{\lambda}$ and variance is $\underline{\lambda}$.

   (c) Normal(μ, σ^2)
      i. Range is the set of <u>all real numbers</u>.
      ii. Probability density function

$$\underline{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

      in the range.
      iii. Mean is $\underline{\mu}$ and variance is $\underline{\sigma^2}$.

2. **[1 mark]** Let $N \sim \text{Binomial}(5, \frac{1}{2})$. What is the maximum value of the pmf of N? [In other words, let $f_N(k) = \Pr(N = k)$ denote the pmf of N where k is in the range of N. Compute the maximum value of $f_N(k)$ as k varies].

   <u>Solution:</u> The p.m.f of N is

$$f_N(k) = \frac{\binom{5}{k}}{2^5}, \quad \text{for } k = 0, 1, 2, 3, 4, 5.$$

   If you compute all 6 values of the p.m.f., then we get

$$\left(\frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32}\right).$$

   Clearly the maximum value is $\boxed{\frac{5}{16}}$.

   This question checks if one knows the definition of Binomial pmf and is willing to list out numbers to check for maximum value ('getting your hands dirty').

3. **[1 mark]** Let $X, Y \sim \text{Bernoulli}(0.314)$ and suppose they are independent random variables. Compute

$$\Pr(X = 0 \mid X + Y = 1)$$

Solution: Let $p = 0.314$. Applying the definition of conditional probability

$$\Pr(X = 0 \mid X + Y = 1) = \frac{\Pr(X = 0 \text{ and } X + Y = 1)}{\Pr(X + Y = 1)}$$

We compute both the numerator and denominator separately.

(a) The numerator computation:

$$\Pr(X = 0 \text{ and } X + Y = 1) = \Pr(X = 0 \text{ and } Y = 1) = \Pr(X = 0)\Pr(Y = 1) = p(1 - p).$$

The second equality above follows from independence of X, Y and the third from $X, Y \sim \text{Bernoulli}(p)$.

(b) Now for the denominator computation:

$$\Pr(X+Y = 1) = \Pr(X = 0 \text{ and } Y = 1)+\Pr(X = 1 \text{ and } Y = 0) = \Pr(X = 0)\Pr(Y = 1)+\Pr(X = 1)\Pr(X = 1) = 2p(1-p).$$

So we see that

$$\Pr(X = 0 \mid X + Y = 1) = \frac{\Pr(X = 0 \text{ and } X + Y = 1)}{\Pr(X + Y = 1)} = \frac{p(1 - p)}{2p(1 - p)} = \boxed{\frac{1}{2}}.$$

This question checks if people can recall and apply the definition of conditional probability and independence.

4. **[1 mark]** A fair coin is tossed 5 times. Let S_n denote the number of tails in the nth toss for $n = 1, 2, 3, 4, 5$. Compute the probability

$$\Pr(S_3 = 2 \mid S_1 = 1, S_2 = 2, S_5 = 2)$$

Solution: $S_5 = 2, S_3 = 2$ implies that between the second and fifth tosses the number of tails has not increased. This means that $S_3 = 2, S_4 = 2$ for sure. Thus $S_1 = 1, S_5 = 2, S_2 = 2 \implies S_3 = 2$.Therefore the required probability is $\boxed{1}$.

At this point, if you want a rigorous proof for the above argument used in the conclusion, let us state it precisely and prove it.

Claim. If events A, B are events such that $A \implies B$, then $\Pr(B|A) = 1$.

Proof. If $A \implies B$, then the event $(A \text{ and } B) = A$. Thus if $A \implies B$, we have

$$\Pr(B|A) = \frac{\Pr(A \text{ and } B)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1.$$

□

This question checks if people understand the meaning of conditional probability.

5. **[1 mark]** Let X, Y, Z are all independent random variables and suppose all of them are $\text{Bernoulli}(0.25)$ random variables. Compute

$$\mathbb{E}((X + Y)(Y - Z))$$

Solution: For a *Bernoulli*(p) random variable

$$\mathbb{E}(X) = \mathbb{E}(X^2) = p.$$

Now we multiply out and apply linearity of expectation

$$\mathbb{E}((X + Y)(Y - Z)) = \mathbb{E}(XY - XZ + Y^2 - YZ) = \mathbb{E}(XY) - \mathbb{E}(XZ) + \mathbb{E}(Y^2) - \mathbb{E}(YZ).$$

For independent random variables: $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Using that we get:

$$\mathbb{E}(XY) - \mathbb{E}(XZ) + \mathbb{E}(Y^2) - \mathbb{E}(YZ) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Z) + \mathbb{E}(Y^2) - \mathbb{E}(Y)\mathbb{E}(Z) = p^2 - p^2 + p - p^2 = p - p^2 = \boxed{\frac{3}{16}}.$$

This problem was created to trap people who assume

$$\mathbb{E}((X + Y)(Y - Z)) = \mathbb{E}(X + Y)\mathbb{E}(Y - Z)$$

which is false since the two factors are not independent random variables.

6. **[1 mark]** If $X_1, X_2, \cdots, X_{10}$ are independent random variables then which of the following **options** are always true?

   (a) Three random variables $X_1 X_2^2$, $X_3 X_4^3 X_5^4$, $X_9 \sin X_8$ are independent.

   (b) If $Y_n = X_{n+1} - X_n$ for $1 \le n \le 9$ then the random variables Y_2, Y_5, Y_7, Y_9 are independent.

   (c) Three random variables $X_1 X_2^2 X_3$, $X_3 X_4^3 X_5^4$, $X_9 \sin X_8$ are independent.

   (d) If $Y_n = X_{n+1} - X_n$ for $1 \le n \le 9$ then the random variables Y_2, Y_3, Y_7, Y_9 are independent.

   Solutions: Only options a, b are true.

   Fact. *If we have a collection of independent random variables, then any collection of expressions made out of disjoint subsets of variables is again independent.*

   In option a, the expressions are made out of $\{X_1, X_2\}$,$\{X_3, X_4, X_5\}$, $\{X_8, X_9\}$ which are disjoint. In option b, the expressions are made out of $\{X_2, X_3\}$,$\{X_5, X_6\}\{X_7, X_8\}$, $\{X_9, X_{10}\}$ which are again disjoint. However in option c, X_3 is common to first two expressions and in option d, X_3 is common to the first two variables.

   This problem checks whether people know to spot independent random variables which is a fundamental requirement for the course.

7. **[2 marks]** A fair coin is tossed 5 times. Let S_n denote the number of tails in the first n tosses for $n = 1, 2, 3, 4, 5$. Compute the conditional expectation
$$\mathbb{E}(S_5 | S_3)$$
as a function of S_3.

   Solution: Let N_k denote the number of tosses in the kth toss, then we know that
$$S_k = N_1 + N_2 + \cdots + N_k.$$
We directly use the definition of conditional expectation:
$$g(y) = \mathbb{E}(X|Y = y) = \sum_x x \Pr(X = x|Y = y).$$
As y varies,we generally denote this random variable as $\mathbb{E}(X|Y) = g(Y)$. So applying definition, we compute
$$g(y) = \mathbb{E}(S_5|S_3 = y) = \mathbb{E}(S_3 + N_4 + N_5|S_3 = y) = \mathbb{E}(S_3|S_3 = y) + \mathbb{E}(N_4|S_3 = y) + \mathbb{E}(N_5|S_3 = y).$$
Now we can compute each term using the definition given above to continue our string of equalities:
$$\mathbb{E}(S_3|S_3 = y) + \mathbb{E}(N_4|S_3 = y) + \mathbb{E}(N_5|S_3 = y) = y + \mathbb{E}(N_4) + \mathbb{E}(N_5) = y + \frac{1}{2} + \frac{1}{2} = y + 1.$$
Thus $g(S_3) = S_3 + 1$. A short hand way to encode this entire calculation is by discarding the g function:
$$\mathbb{E}(S_5|S_3) = \mathbb{E}(S_3 + N_4 + N_5|S_3) = \mathbb{E}(S_3|S_3) + \mathbb{E}(N_4|S_3) + \mathbb{E}(N_5|S_3) = S_3 + \mathbb{E}(N_4) + \mathbb{E}(N_5) = S_3 + \frac{1}{2} + \frac{1}{2} = S_3 + 1.$$
The second equality follows from linearity of expectation. The third equality follows from the fact that N_4, N_5 are independent of S_3 (follows from Fact in the previous problem). The fourth equality follows from expectation of Bernoulli(0.5) random variable (see Problem 5 solution).

   This problem checks if you know, understand and can apply the definition of conditional expectation in a problem. This was the only problem which was supposed to be hard.