2022Fall Probability & Mathematical Statistics 2022/12/01

Homework 10

Professor: Ziyu Shao *Due: 2022/12/11 10:59pm*

1. Let $X \sim \text{Pois}(\lambda)$. The conditional distribution of X, given that $X \geq 1$, is called a truncated Poisson distribution.
   (a) Find $E(X|X \geq 1)$.
   (b) Find $\text{Var}(X|X \geq 1)$.

2. Let X and Y be independent, positive r.v.s. with finite expected values.
   (a) Give an example where $E(\frac{X}{X+Y}) \neq \frac{E(X)}{E(X+Y)}$, computing both sides exactly.
   Hint: Start by thinking about the simplest examples you can think of!
   (b) If X and Y are i.i.d., then is it necessarily true that $E(\frac{X}{X+Y}) = \frac{E(X)}{E(X+Y)}$?
   (c) Now let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$. Show *without using calculus* that
   $$E\left(\frac{X^c}{(X+Y)^c}\right) = \frac{E(X^c)}{E((X+Y)^c)},$$
   for every real $c > 0$.

3. Alice walks into a post office with 2 clerks. Both clerks are in the midst of serving customers, but Alice is next in line. The clerk on the left takes an $\text{Expo}(\lambda_1)$ time to serve a customer, and the clerk on the right takes an $\text{Expo}(\lambda_2)$ time to serve a customer. Let T_1 be the time until the clerk on the left is done serving his or her current customer, and define T_2 likewise for the clerk on the right.
   (a) If $\lambda_1 = \lambda_2$, is T_1/T_2 independent of $T_1 + T_2$?
   Hint: $T_1/T_2 = (T_1/(T_1 + T_2))/(T_2/(T_1 + T_2))$.
   (b) Find $P(T_1 < T_2)$ (do not assume $\lambda_1 = \lambda_2$ here or in the next part, but do check that your answers make sense in the that special case).
   (c) Find the expected total amount of time that Alice spends in the post office (assuming that she leaves immediately after she is done being served).

4. A DNA sequence can be represented as a sequence of letters, where the "alphabet" has 4 letters: A,C,T,G. Suppose such a sequence is generated randomly, where the letters are independent and the probabilities of A,C,T,G are p_1, p_2, p_3, p_4, respectively.

(a) In a DNA sequence of length 115, what is the expected number of occurrences of the expression "CATCAT" (in terms of the p_j)? (Note that, for example, the expression "CATCATCAT" counts as 2 occurrences.)

(b) What is the probability that the first A appears earlier than the first C appears, as letters are generated one by one (in terms of the p_j)?

(c) For this part, assume that the p_j are unknown. Suppose we treat p_2 as a $\text{Unif}(0, 1)$ r.v. before observing any data, and that then the first 3 letters observed are "CAT". Given this information, what is the probability that the next letter is C?

5. A coin with probability p of Heads is flipped repeatedly. For (a) and (b), suppose that p is a known constant, with $0 < p < 1$.

(a) What is the expected number of flips until the pattern HT is observed?

(b) What is the expected number of flips until the pattern HH is observed?

(c) Now suppose that p is unknown, and that we use a $\text{Beta}(a, b)$ prior to reflect our uncertainty about p (where a and b are known constants and are greater than 2). In terms of a and b, find the corresponding answers to (a) and (b) in this setting.