2022Fall Probability & Mathematical Statistics 2022/11/17

Homework 8

Professor: Ziyu Shao *Due: 2022/11/27 10:59pm*

1. Consider a setting where a Poisson approximation should work well: let $A_1, \ldots, A_n$ be independent, rare events, with n large and $p_j = P(A_j)$ small for all j. Let $X = I(A_1) + \cdots + I(A_n)$ count how many of the rare events occur, and let $\lambda = E(X)$.

   (a) Find the MGF of X.

   (b) If the approximation $1 + x \approx e^x$ (this is a good approximation when x is very close to 0 but terrible when x is not close to 0) is used to write each factor in the MGF of X as e to a power, what happens to the MGF? Explain why the result makes sense intuitively.

2. Let X and Y be i.i.d. Geom(p), $L = \min(X, Y)$, and $M = \max(X, Y)$.

   (a) Find the joint PMF of L and M. Are they independent?

   (b) Find the marginal distribution of L in two ways: using the joint PMF, and using a story.

   (c) Find $E[M]$. Hint: A quick way is to use (b) and the fact that $L + M = X + Y$.

   (d) Find the joint PMF of L and $M - L$. Are they independent?

3. Let X and Y be i.i.d. Expo(λ), and $T = X + Y$.

   (a) Find the conditional CDF of T given $X = x$. Be sure to specify where it is zero.

   (b) Find the conditional PDF $f_{T|X}(t \mid x)$, and verify that it is a valid PDF.

   (c) Find the conditional PDF $f_{X|T}(x \mid t)$, and verify that it is a valid PDF.

   (d) In class we have shown that the marginal PDF of T is $f_T(t) = \lambda^2 t e^{-\lambda t}$, for $t > 0$. Give a short alternative proof of this fact, based on the previous parts and Bayes' rule.

4. Let U_1, U_2, U_3 be i.i.d. Unif(0, 1), and let $L = \min(U_1, U_2, U_3)$, $M = \max(U_1, U_2, U_3)$.

   (a) Find the marginal CDF and marginal PDF of M, and the joint CDF and joint PDF of L, M. Hint: For the latter, start by considering $P(L \geq l, M \leq m)$.

   (b) Find the conditional PDF of M given L.

5. In humans (and many other organisms), genes come in pairs. Consider a gene of interest, which comes in two types (alleles): type a and type A. The genotype of a person for that gene is the types of the two genes in the pair: AA, Aa, or aa (aA is equivalent to Aa). According to the Hardy-Weinberg law, for a population in equilibrium the frequencies of AA, Aa, aa will be $p^2, 2p(1-p), (1-p)^2$, respectively, for some p with $0 < p < 1$. Suppose that the Hardy-Weinberg law holds, and that n people are drawn randomly from the population, independently. Let X_1, X_2, X_3 be the number of people in the sample with genotypes AA, Aa, aa, respectively.

   (a) What is the joint PMF of X_1, X_2, X_3?

   (b) What is the distribution of the number of people in the sample who have an A ?

   (c) What is the distribution of how many of the $2n$ genes among the people are A's?

   (d) Now suppose that p is unknown, and must be estimated using the observed data X_1, X_2, X_3. The maximum likelihood estimator (MLE) of p is the value of p for which the observed data are as likely as possible. Find the MLE of p.

   (e) Now suppose that p is unknown, and that our observations can't distinguish between AA and Aa. So for each person in the sample, we just know whether or not that person is an aa (in genetics terms, AA and Aa have the same phenotype, and we only get to observe the phenotypes, not the genotypes). Find the MLE of p.