Homework 6

Stat 345 - Spring 2020

Name: ________________________________

Problem 1

We load on a plane 100 packages whose weights are independent random variables that are uniformly distributed between 5 and 50 pounds. What is the probability that the total weight will exceed 3000 pounds? It is not easy to calculate the cdf of the total weight and the desired probability, but an approximate answer can be quickly obtained using the central limit theorem.

The distribution is $X_1, X_2, ..., X_{100} \sim Uniform(5, 50)$ with $E(X) = (5 + 50)/2 = 27.5$ and $Var = (50 - 5)^2/12 = 168.75$ based on the formulas for the mean and variance of the uniform pdf. We need to calculate $P(\sum_{i=1}^{100} > 3000)$.

$$P(\sum_{i=1}^{100} > 3000) = 1 - P(\sum_{i=1}^{100} \leq 3000)$$

Using the central limit theorem

$$P(\sum_{i=1}^{100} \leq 3000) = P(Z \leq \frac{3000 - (27.5)(100)}{\sqrt{168.75}\sqrt{100}}) = P(Z \leq 1.9245) = \Phi(1.9245)$$

Using R: $pnorm(1.9245) = 0.973$. And the probability is

$$P(\sum_{i=1}^{100} > 3000) = 1 - 0.973 = 0.027.$$

Problem 2

Let $X_1, X_2, ..., X_n$ be i.i.d. Gamma random variables with parameters α and λ. The likelihood function is difficult to differentiate because of the gamma function. Rather than finding the maximum likelihood estimators, what are the method of moments estimators of both parameters α and λ?

The expected value and variance for $Gamma(\alpha, \lambda)$ are

$$E(X) = \frac{\alpha}{\lambda}, Var(X) = \frac{\alpha}{\lambda^2}$$

To find the method of moments estimators we need to equate population moments to the sample moments:

$$E(X) = \bar{X}, \qquad E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\frac{\alpha}{\lambda} = \bar{X}, \quad \frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\alpha = \lambda\bar{X}, \quad \frac{\lambda\bar{X}}{\lambda^2} + \left(\frac{\lambda\bar{X}}{\lambda}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\alpha = \lambda\bar{X}, \quad \frac{\bar{X}}{\lambda} + \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\lambda = \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

$$\alpha = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Problem 3

Let $X_1, X_2, ..., X_n$ be i.i.d. Geometric(θ), $\theta = 1, 2, 3, ...$ random variables.

a) Find the maximum likelihood estimator of θ.

The pmf when $X_i \sim Geometric(\theta)$ is $p(x) = (1-\theta)^{x-1}\theta$. The likelihood is given by

$$L(\theta) = p(x_1)p(x_2)...p(x_n) = \prod_{i=1}^{n}(1-\theta)^{x_i-1}\theta$$

The log likelihood can be obtained by taking the natural logarithm of $L(\theta)$:

$$logL(\theta) = log\left(\prod_{i=1}^{n}(1-\theta)^{x_i-1}\theta\right) = \sum_{i=1}^{n} log\left((1-\theta)^{x_i-1}\theta\right)$$

$$logL(\theta) = \sum_{i=1}^{n}(x_i-1)log(1-\theta) + \sum_{i=1}^{n} log(\theta)$$

$$logL(\theta) = log(1-\theta)\sum_{i=1}^{n}(x_i-1) + nlog(\theta)$$

$$\frac{d\log L(\theta)}{d\theta} = -\frac{\sum_{i=1}^{n}(x_i-1)}{1-\theta} + \frac{n}{\theta} = -\frac{\sum_{i=1}^{n} x_i - n}{1-\theta} + \frac{n}{\theta}$$

$$\frac{\sum_{i=1}^{n} x_i - n}{\theta - 1} + \frac{n}{\theta} = 0$$

$$\theta\left(\sum_{i=1}^{n} x_i - n\right) + n(\theta - 1) = 0$$

$$\theta\sum_{i=1}^{n} x_i - n\theta + n\theta - n = 0$$

$$\theta\sum_{i=1}^{n} x_i - n\theta + n\theta - n = 0$$

$$\theta\sum_{i=1}^{n} x_i = n$$

$$\theta = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$\theta = \frac{1}{\bar{x}}$$

The maximum likelihood estimator of θ is

$$\hat{\theta} = \frac{1}{\bar{X}}$$

b) The nice property of maximum likelihood estimators is the *invariance property*. If $\hat{\theta}$ is the MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$. For example, if I have a random sample from a Bernoulli distribution $X_i \sim Bernoulli(\theta), i = 1, ..., n$ and I need to find the MLE of the function of θ, say $g(\theta) = (1 - \theta^2)$, then I can first find the MLE of θ which is $\hat{\theta} = \bar{X}$, and by the invariance property the MLE of $g(\theta)$ is $g(\hat{\theta}) = 1 - \hat{\theta}^2 = 1 - \bar{X}^2$.

In a certain hard video game a player is confronted with a series of AI opponents and has an θ probability of defeating each one. Success with any opponent is independent of previous encounters. Until first win, the player continues to AI contest opponents. Let X denote the number of opponents contested until the player's first win. Suppose that data of 10 players was collected:

$$7, 4, 3, 1, 12, 10, 2, 1, 4, 6$$

What is the MLE of the probability that a player contests five or more AI opponents in a game until the first win?

The probability is

$$P(X \geq 5) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4)$$

This is the probability of failure on all of the first four contests (trials). The failures are independent events, so

$$P(X \geq 5) = (1-\theta)^4$$

To calculate this probability we need to know θ.

Note that $X_1, X_2, ..., X_{10} \sim Geometric(\theta)$. The MLE of θ is $\hat{\theta} = \frac{1}{\bar{X}}$.

From the data

$$\bar{x} = \frac{7+4+3+1+12+10+2+1+4+6}{10} = 5$$

Therefore,

$$\hat{\theta} = \frac{1}{\bar{x}} = \frac{1}{5} = 0.2$$

By the invariance property, the MLE of $g(\theta) = (1-\theta)$ is $g(\hat{\theta}) = (1-\hat{\theta})^4$. Since $\hat{\theta} = 0.2$ the maximum likelihood estimate of the probability that a player contests five or more AI opponents in a game until the first win is $(1-0.2)^4 = 0.41$.

Problem 4

Let $X_1, X_2, ..., X_n$ be a random sample from a uniform distribution on the interval $(0, a)$. Recall that the maximum likelihood estimator (MLE) of a is $\hat{a} = max(X_i)$.

a) Let $Y = max(X_i)$. Use the fact that $Y \leq y$ if and only if each $X_i \leq y$ to derive the cumulative distribution function of Y.

$$P(Y \leq y) = P(max(X_i) \leq y) = P(X_1 \leq y)P(X_2 \leq y) \cdot ... \cdot P(X_n \leq y) = \left(\frac{y-0}{a-0}\right)^n = \left(\frac{y}{a}\right)^n$$

b) Find the probability density function of Y from cdf.

By taking derivative of cdf with respect to y the pdf is

$$f(y) = \frac{n}{a}\left(\frac{y}{a}\right)^{n-1} = \frac{ny^{n-1}}{a^n}$$

c) Use the obtained pdf to show that MLE for a ($\hat{a} = max(X_i)$) is biased.

To show that the estimator $\hat{a}$ is unbiased for a we need to show that

$$E(\hat{a}) = a$$

$$E(\hat{a}) = E(max(X_i)) = E(Y) = \int_0^a yf(y)dy = \int_0^a y\frac{ny^{n-1}}{a^n}dy =$$

$$= \int_0^a \frac{ny^n}{a^n} dy = \frac{n}{a^n} \int_0^a y^n dy = \frac{n}{a^n} \frac{y^{n+1}}{n+1}\bigg|_0^a = \frac{n}{a^n} \frac{a^{n+1}}{n+1} = \frac{n}{n+1} a \neq a$$

d) Say I would like to consider another estimator for a, I will call it $\hat{b} = 2\bar{X}$. Is it unbiased estimator of a (show)? How you can explain someone without calculations why $\hat{b} = 2\bar{X}$ is a reasonable estimator of a?

$$E(\hat{b}) = E(2\bar{X}) = 2E(\bar{X}) = \frac{2}{n} \sum_{i=1}^{n} E(X_i) = \frac{2}{n} \sum_{i=1}^{n} \frac{0+a}{2} = \frac{2}{n} \frac{na}{2} = a$$

Therefore, $\hat{b}$ is the unbiased estimator of a.

Even without doing calculations we can conclude that $\hat{b} = 2\bar{X}$ is a good estimator of a. The range of values here is defined on the interval $(0, a)$. We expect the mean of the sample $\bar{X}$ to fall close to the midpoint of zero and a, hence $2\bar{X}$ should be close to a.

e) Based on the result in (c), I will propose to use unbiased estimator for a instead of $\hat{a} = max(X_i)$, say $\hat{c} = \frac{n+1}{n} max(X_i)$. Given that the relative efficiency of any two unbiased estimators $\hat{b}, \hat{c}$ is the ratio of their variances

$$\frac{Var(\hat{b})}{Var(\hat{c})},$$

explain which of these two unbiased estimators is more efficient. You can obtain the $Var(\hat{c}) = Var(\frac{n+1}{n} max(X_i))$ from $Var(\hat{a}) = Var(Y)$. The variance of the $Y = max(X_i)$ is

$$Var(Y) = \frac{n}{(n+1)^2(n+2)} a^2$$

Then

$$Var(\hat{c}) = Var(\frac{n+1}{n} max(X_i)) = \frac{(n+1)^2}{n^2} Var(Y) = \frac{(n+1)^2}{n^2} \frac{n}{(n+1)^2(n+2)} a^2 = \frac{a^2}{n(n+2)}$$

$$Var(\hat{b}) = Var(2\bar{X}) = 2^2 \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{4}{n} \sum_{i=1}^{n} \left(\frac{(a-0)^2}{12} \right) = \frac{a^2}{3n}$$

The relative efficiency is

$$\frac{Var(\hat{b})}{Var(\hat{c})} = \frac{\frac{a^2}{3n}}{\frac{a^2}{n(n+2)}} = \frac{n+2}{3}$$

indicating that for $n > 1$ the ratio is greater than 1. Therefore, $\hat{c} = \frac{n+1}{n} max(X_i)$ has a lower variance than $\hat{b} = 2\bar{X}$ and thus, $\hat{c}$ is more efficient than $\hat{b}$.