PROBABILITY AND STATISTICS - HOMEWORK 8

DUE DATE: 25/NOV/2016 (IN TUTORIALS)

Homework problems are marked blue (but trying all problems is strongly recommended).

1. A large box contains 10000 marbles, of which some are red and the others are blue. To estimate the unknown proportion p of red balls, a sample of 100 marbles is drawn at random (with replacement) and it is observed that the number of red balls in the sample is 30. Construct a $1-\alpha$ confidence interval for p when (1) $\alpha = 0.01$, (2) $\alpha = 0.05$, (3) $\alpha = 0.10$. Repeat the same exercise when the number of red marbles in the sample is 40.

2. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where μ, σ^2 are both unknown. Summary statistics of the data obtained in an experiment are given as follows:

$$n = 20, \quad \sum_{i=1}^{n} X_i = 60, \quad \sum_{i=1}^{n} X_i^2 = 240.$$

(1) Find a two-sided confidence interval for μ with confidence level 0.90.
(2) Find an upper bound for σ^2 with confidence level 0.90.

3. Let $X_1, \ldots, X_n$ be i.i.d. Exp(λ). Let $\theta = \log \lambda$. Let $\gamma = \int_0^\infty \log t \, e^{-t} \, dt$.

(1) Show that $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(\gamma - \log X_i)$ is an unbiased estimate for θ.
(2) Compute the m.s.e of $\hat{\theta}$.
(3) Explain how you would give an $(1-\alpha)$-confidence interval for λ, based on $\hat{\theta}$. **[Hint:** If $X \sim$ Exp(λ), the distribution of $\log X + \log \lambda$ does not depend on λ.]

4. In each of the following cases, find the bias and m.s.e of the given estimate. The samples are $X_1, \ldots, X_n$, i.i.d. from the given distribution.

(1) Distribution is $N(\mu, \sigma^2)$, both parameters unknown. The estimate (for μ) if $\hat{\mu} = \overline{X}_n$.
(2) Distribution is Ber(p). The estimate for p is $\hat{p} = \overline{X}_n$.
(3) Distribution is Pois(λ). The estimate for λ is $\hat{\lambda} = \overline{X}_n$.

Date: November 15, 2016.

5. In the above problem, describe how you would construct a $1-\alpha$ confidence interval for the unknown parameter in terms of $\overline{X}_n$. You may assume that n is large enough that central limit approximation is valid.

6. In http://math.iisc.ernet.in/~manju/UGstatprob/newcomb_lightspeed.txt you will see the data from Simon Newcomb's experiment on the time taken (in nanoseconds) by light to travel 7442 meters at sea level.

(1) Compute the sample mean and sample standard deviation.

(2) Assuming normal distribution for the data, compute a confidence interval for the time taken. What confidence interval does it give for the speed of light (in meters per second)?

[**Note:** You are being asked to assume that the measured times have a normal distribution. It is different from assuming that the measured speeds (i.e., reciprocals of times essentially) are normally distributed.]

7. A box contains N marbles of which m are red in colour and $N-m$ are blue. We are interested in estimating the proportion $p=m/N$ of red balls. A sample of size k is drawn from the box and the number of red balls in the sample if observed, call it X. Then, X/k is a reasonable estimate for p. What are its bias and m.s.e if

(1) the sampling is done with replacement?

(2) the sampling is done without replacement?

Before you do the calculations, can you guess in which case would the mean squared error be smaller?