CS 70 Discrete Mathematics and Probability Theory
Summer 2021

DIS 07B

1 Confidence Interval Introduction

We observe a random variable X which has mean μ and standard deviation $\sigma \in (0, \infty)$. Assume that the mean μ is unknown, but σ is known.

We would like to give a 95% confidence interval for the unknown mean μ. In other words, we want to give a random interval (a,b) (it is random because it depends on the random observation X) such that the probability that μ lies in (a,b) is at least 95%.

We will use a confidence interval of the form $(X - \varepsilon, X + \varepsilon)$, where $\varepsilon > 0$ is the width of the confidence interval. When ε is smaller, it means that the confidence interval is narrower, i.e., we are giving a more *precise* estimate of μ.

(a) Using Chebyshev's Inequality, calculate an upper bound on $\mathbb{P}[|X - \mu| \geq \varepsilon]$.

(b) Explain why $\mathbb{P}(|X - \mu| < \varepsilon)$ is the same as $\mathbb{P}[\mu \in (X - \varepsilon, X + \varepsilon)]$.

(c) Using the previous two parts, choose the width of the confidence interval ε to be large enough so that $\mathbb{P}[\mu \in (X - \varepsilon, X + \varepsilon)]$ is guaranteed to exceed 95%. [Note: Your confidence interval is allowed to depend on X, which is observed, and σ, which is known. Your confidence interval is not allowed to depend on μ, which is unknown.]

(d) The previous three parts dealt with the case when you observe one sample X. Now, let n be a positive integer and let $X_1, \ldots, X_n$ be i.i.d. samples, each with mean μ and standard deviation $\sigma \in (0, \infty)$. As before, assume that μ is unknown but σ is known.

Here, a good estimator for μ is the *sample mean* $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$. Calculate the mean and variance of $\bar{X}$.

(e) We will now use a confidence interval of the form $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$ where $\varepsilon > 0$ again represents the width of the confidence interval. Imitate the steps of (a) through (c) to choose the width ε to be large enough so that $\mathbb{P}[\mu \in (\bar{X} - \varepsilon, \bar{X} + \varepsilon)]$ is guaranteed to exceed 95%.

To check your answer, your confidence interval should be *smaller* when n is larger. Intuitively, if you collect more samples, then you should be able to give a more *precise* estimate of μ.

2 Poisson Confidence Interval

For n a positive integer, you collect $X_1, \ldots, X_n$ i.i.d. samples drawn from a Poisson distribution (with unknown mean λ). However, you have a bound on the mean: from a confidential source, you know that $\lambda \leq 2$. For $0 < \delta < 1$, find a $1 - \delta$ confidence interval for λ using Chebyshev's Inequality.

3 Vegas

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip their coin. Assume that each person is independently likely to carry a fair or a trick coin.

(a) Let X be the proportion of people whose coin flip results in heads. Find $\mathbb{E}[X]$.

(b) Given the results of your experiment, how should you estimate p? (*Hint:* Construct an unbiased estimator for p using part (a))

(c) How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

(d) Suppose n is large. Construct an approximate 98% confidence interval for p.