Dis 7B: CLT + Confidence Intervals

Tuesday, 4 August 2020 7:33 PM

What's a confidence interval of an unknown value x?

e.g. You a biased coin and you wanted to **estimate** the bias x of the coin.

X = # of heads observed when 100 x

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%.

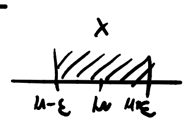
We will use a confidence interval of the form $(X-\varepsilon, X+\varepsilon)$, where $\varepsilon > 0$ is the width of the confidence interval. When ε is smaller, it means that the confidence interval is narrower, i.e., we are giving a more *precise* estimate of μ.

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

$$\leq \frac{Var(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

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

$X - \mu < \varepsilon$, $-|X -$

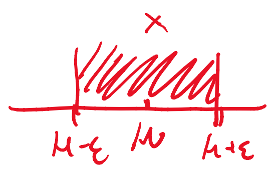


$\mu - \varepsilon \leq X \leq \mu + \varepsilon$

$-\varepsilon \leq X - \mu \leq \varepsilon$

$-\varepsilon - X \leq -\mu \leq -X + \varepsilon$ $\quad$ $+X + \varepsilon \geq \mu \geq X - \varepsilon$

$|X - \mu| < \varepsilon$



(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.]

$P(\mu \in (X-\varepsilon, X+\varepsilon)) \geq 0.95$

$\Leftrightarrow$ $\boxed{Pr(|X-\mu| < \varepsilon) \geq 0.95}$

$A \geq B$ $\quad$ 0.5 ≥ 0.2

$Pr(|X - \mu| \geq \varepsilon) \leq 0.05$

$\Updownarrow$

$\leq Var(X)$

$\Leftrightarrow \Pr(|X-\mu|<\varepsilon) \geq 0.95$

$\Leftrightarrow 1 - \Pr(|X-\mu|<\varepsilon)$

$\Pr(|X-\mu|\geq\varepsilon)$

$A \geq B$ $\quad 0.5 \geq 0.2$

$1-A \leq 1-B$

0.95

$\leq \dfrac{Var(X)}{\varepsilon^2}$

$\Pr(|X-\mu|\geq\varepsilon) \leq 0.05$

$\leq \dfrac{Var(X)}{\varepsilon^2}$

$\dfrac{\sigma^2}{\varepsilon^2} \leq 0.05$

$\sqrt{\dfrac{\sigma^2}{0.05}} \leq \varepsilon$

(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.]

2 Poisson Confidence Interval

You collect n samples (n is a positive integer) $X_1, \ldots, X_n$, which are i.i.d. and known to be drawn from a Poisson distribution (with unknown mean). However, you have a bound on the mean: from a confidential source, you know that $\lambda \leq 2$. Find a $1-\delta$ confidence interval ($\delta \in (0,1)$) for λ using Chebyshev's Inequality. (Hint: a good estimator for λ is the *sample mean* $\bar{X} := n^{-1}\sum_{i=1}^{n} X_i$)

$\Leftrightarrow$ Find interval (a,b) s.t.

$\Pr(\lambda \in (a,b)) \geq 1-\delta$

English to Math

Find an α confidence interval for a value x

$\Leftrightarrow$

Find an interval (a,b) where $Pr(x \in (a,b)) \geq \alpha$

Hint: use sample mean to make the interval + Chebyshev's $\mathbb{E}$

$(\mathbb{E}[\bar{X}] = \frac{1}{n}\,\mathbb{E}[\sum_{i=1}^{n} X_i] = \lambda)$

$\Pr(|\bar{X} - \mathbb{E}(\bar{X})| < \varepsilon) \geq 1-\delta$

$\Leftrightarrow \Pr(|\bar{X} - \lambda| \geq \varepsilon) \leq \delta$

$\Pr(|\bar{X} - \underset{\mathbb{E}[\bar{X}]}{\lambda}| \geq \varepsilon) \leq \delta$

cheby $\hookrightarrow \leq \dfrac{Var(\bar{X})}{\varepsilon^2} \leq \delta$

$Var(\bar{X}) = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$

$\Leftrightarrow \dfrac{\lambda}{n\varepsilon^2} \leq \delta \Rightarrow \sqrt{\dfrac{\lambda}{n\delta}} \leq \varepsilon$

$\sqrt{\dfrac{2}{n\delta}} \leq \varepsilon$

Central Limit Theorem

Let $X_1, \ldots, X_n$ be a bunch of RVs that are independent and identical to each other (i.i.d). Define the sum $S_n = X_1 + \ldots + X_n$. Say that each $X_i \sim Poi(\lambda)$ and let $A_n = \frac{S_n}{n}$.

CLT says:

For large enough n, A_n is approximately a ___N___ distribution, with parameter(s) (μ, σ^2)

$\mu = \lambda$

$\sigma^2 = \dfrac{\lambda}{n}$

$A_n \sim N(\lambda, \frac{\lambda}{n})$

____

Then, __$\dfrac{A_n - \lambda}{\sqrt{\frac{\lambda}{n}}}$__ $\sim N(0,1)$

Convert __________ into S_n form:

__________ $\sim N(0,1)$

$\dfrac{\frac{S_n}{n} - \lambda}{} = \dfrac{S_n - n\lambda}{n} \div \sqrt{\dfrac{\lambda}{n}}$

________________ $\sim N(0,1)$

$$\frac{\frac{S_n}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}} = \frac{S_n - n\lambda}{n} \div \sqrt{\frac{\lambda}{n}} = \frac{S_n - n\lambda}{\sqrt{n\lambda}} \sim N(0,1)$$

[Going to assume you know how to read Normal RV tables ... If you don't know, get up to the step of calculating the probability of a normal RV. Then, watch that Khan Academy walkthrough]

3 Hypothesis testing

We would like to test the hypothesis claiming that a coin is fair, i.e. $P(H) = P(T) = 0.5$. To do this, we flip the coin $n = 100$ times. Let Y be the number of heads in $n = 100$ flips of the coin. We decide to reject the hypothesis if we observe that the number of heads is less than $50 - c$ or larger than $50 + c$. However, we would like to avoid rejecting the hypothesis if it is true; we want to keep the probability of doing so less than 0.05. Please determine c. (*Hints: use the central limit theorem to estimate the probability of rejecting the hypothesis given it is actually true. Table is provided in the appendix.*)

What's the event of 'reject the hypothesis if it is true'?

$$\boxed{S_n \sim N(E[S_n], Var(S_n))}$$

$$P(Y < 50 - c \text{ or } Y > 50 + c) < 0.05$$

$X_i \sim Bern(0.5)$

$Y = X_1 + X_2 + \dots + X_{100}$

$Y \sim Bin(100, 0.5)$

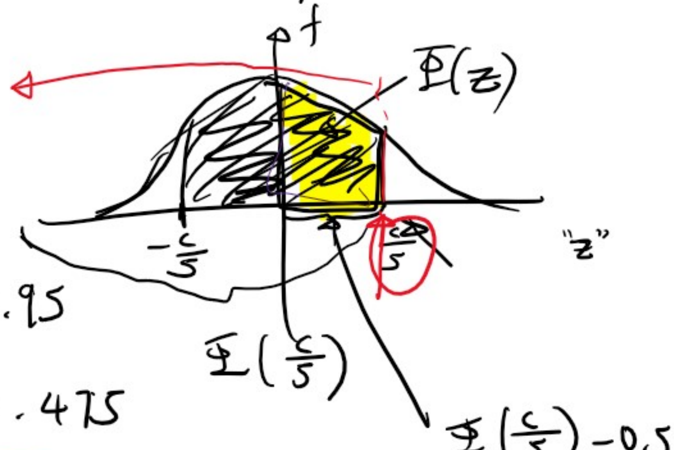
$$Pr(50 - c < Y < 50 + c) \geq 0.95$$

using CLT, $Y \sim N(E[Y], Var(Y))$

$Y \sim N(50, 25)$

$$Pr\left(\frac{50 - c - E(Y)}{\sqrt{Var(Y)}} < \frac{Y - E[Y]}{\sqrt{Var(Y)}} < \frac{50 + c - E(Y)}{\sqrt{Var(Y)}}\right) \geq 0.95$$

$$\Leftrightarrow \quad Pr\left(\frac{-c}{5} < N(0,1) < \frac{c}{5}\right) \geq 0.95$$



$$2\left(\Phi\left(\frac{c}{5}\right) - 0.5\right) \geq 0.95$$

$$\Phi\left(\frac{c}{5}\right) - 0.5 \geq 0.475$$

$$\boxed{\Phi\left(\frac{c}{5}\right) \geq 0.975}$$

0.975