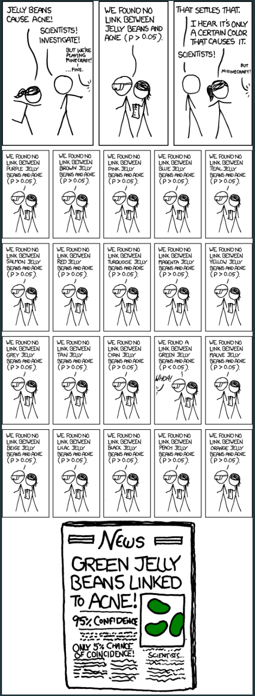


class poll: does this xkcd comic make sense?

(a) always did . . .
(b) not really - is it that jelly beans don't come in 20 different colors?
(c) ooh I get it now!

confidence intervals: definition

Each random expt results in 3 outputs
1) sample mean $\bar{X}_n$, 2) lower CI A, 3) upper CI B

(ideally, smallest)

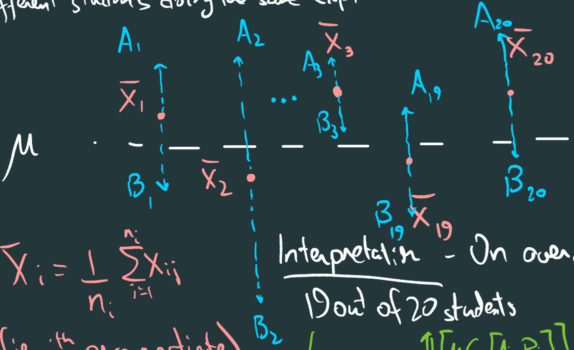
an random interval $[A, B]$ (computed from data/experiments) is called a 95% confidence interval for some (deterministic) quantity μ if

ARE RANDOM

$$\mathbb{P}[A \le \mu \le B] \ge 0.95$$

IS NOT $\frac{1}{n}\sum x_i$, it is the 'truth'

20 different students doing the same expt



We will say $[A,B]$ is a valid $1-\alpha$ CI if $\mathbb{P}[A \le \mu \le B] \geq 1-\alpha$

$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ (ie, i^{th} persons estimate)

Interpretation - On average, $\mu \in [A_i, B_i]$ for 19 out of 20 students

(moreover, $\mathbb{1}[\mu \in [A_i, B_i]] \sim Ber(p \geq 0.95)$)

class poll: abusing CI (?)

(Explanation)

in a homework assignment in Simulation, students were asked to do a Monte Carlo simulation to compute π up to 2 decimal places

there were 100 homework submissions, and *all* submissions reported 95% CIs which included 3.14

the probability that this happened 'by chance' is approximately

(a) 0.4

(b) 0.05

(c) 0.006

(d) 0.0007

(e) 0

By defn of a valid CI

$\mathbb{P}[\text{student } i \text{ has } 3.14 \text{ in the CI}] \approx 0.95$

$\Rightarrow \mathbb{P}[\text{ALL students have } 3.14 \text{ in CI}] = (0.95)^{100} \approx 0.0059$

'Should be'

All of these are examples of 'p-hacking'

- student assignments are independent
- The CIs are 'overcalibrated' (i.e, >0.95 chance of success)
- Experimenter 'selectively reveals' the CI

confidence intervals as a social contract

FREQUENTIST CI

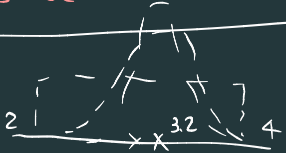
a random interval $[A, B]$ (computed from data/experiments) is a 95% confidence interval for some unknown μ if before the experiment is done

$$\mathbb{P}[A \leq \mu \leq B] \geq 0.95$$

Ideal world - Analyst $\boxed{\text{Code Data}} \longrightarrow$ Client runs code ONCE, and sees CIs

p-hacking - run code multiple times until you 'think CIs are correct'

alternate : BAYESIAN Interval.



confidence intervals for population mean

Define a CI taking the CLT to be literally true

$X_1, X_2, \ldots$ are i.i.d. rvs with $\mathbb{E}[X_1] = \mu$ and $Var(X_1) = \sigma^2 < \infty$; $\boxed{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$

Sample mean

- from the central limit theorem:

Truth

$$\sqrt{n}(\bar{X}_n - \mu) \overset{D}{\cancel{\approx}} \sigma\mathcal{N}(0,1) = \mathcal{N}(0,\sigma^2)$$

# of replications

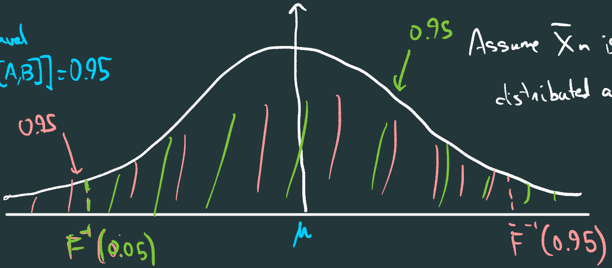
pretend

- from the inverse cdf of $\mathcal{N}(0,1)$, we can compute

$$\mathbb{P}[\quad \bar{X}_n + \bar{F}^{-1}(0.95) \le \mathcal{N}(0,1) \le \bar{X}_n + \bar{F}^{-1}(0.95) \quad] \ge 0.95.$$

Aim - output interval $[A,B]$ s.t. $\mathbb{P}[\mu \in [A,B]] = 0.95$

Assume $\bar{X}_n$ is EXACTLY distributed as $N(\mu, \sigma^2/n)$



confidence intervals for population mean

Define a CI taking the CLT to be literally true

$X_1, X_2, \ldots$ are i.i.d. rvs with $\mathbb{E}[X_1] = \mu$ and $Var(X_1) = \sigma^2 < \infty$; $\boxed{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$

Sample mean

- from the central limit theorem:

Truth

$$\sqrt{n}(\bar{X}_n - \mu) \underset{\text{pretend}}{\overset{D}{\cancel{\approx}} =} \sigma\mathcal{N}(0,1) = \mathcal{N}(0,\sigma^2)$$

\# of replications

- from the inverse cdf of $\mathcal{N}(0,1)$, we can compute the SMALLEST interval s.t

$$\mathbb{P}[\underbrace{-1.96}_{\Phi^{-1}(0.025)} \leq \mathcal{N}(0,1) \leq \underbrace{1.96}_{\approx 2}] \geq 0.95.$$

$\mathbb{P}[-2\sigma \leq N(0,\sigma^2) \leq 2\sigma]$

Aim - output interval $[A,B]$ s.t $\mathbb{P}[\mu \in [A,B]] = 0.95$

Assume $\bar{X}_n$ is EXACTLY distributed as $N(\mu, \sigma^2/n)$

pdf of $\bar{X}_n$

confidence intervals (Gaussian approximation for CI)

want to measure $\mu = \mathbb{E}[X_1]$ from simulations

Assumption - CLT is literally true

- from the central limit theorem: $\sqrt{n}(\bar{X}_n - \mu) \overset{D}{\approx} \sigma\mathcal{N}(0,1)$
- from the cdf of $\mathcal{N}(0,1)$, we have $\mathbb{P}[-1.96 \leq \mathcal{N}(0,1) \leq 1.96] \geq 0.95$

putting these together, we have:

$$\mathbb{P}\left[-1.96\sigma \leq \sqrt{n}(\bar{X}_n - \mu) \leq 1.96\sigma\right] = 0.95$$

$$\Rightarrow \quad \mathbb{P}\left[\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right] = 0.95$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\sigma^2 = \mathrm{Var}(X_i)$

confidence intervals: problems

- the confidence interval is approximate because

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \text{ is only approximately } N(0,1)$$

- the confidence interval is 'exact' when

$$\bar{X}_n \text{ is } \underline{\text{exactly}} \text{ Gaussian} \Leftarrow X_i \sim N(\mu, \sigma^2)$$

- the confidence interval above requires knowledge of σ^2

. In some cases, can upper bound σ^2 - Eg - If $X_i \sim Ber(\mu)$
Then $Var(X_i) = \mu(1-\mu) \leq \frac{1}{4}$

. In practice, often replace σ^2 by Empirical Variance $\equiv S_n = f(X_1, X_2, \ldots, X_n)$
s.t $\mathbb{E}[S_n] = \sigma^2$

confidence intervals: problems

- the confidence interval is approximate because

- ~~the confidence interval is 'exact' when~~

<u>Notes</u>

· S_n^2 is built in to most packages (np.random = std-dev, -1 specified via input)

· $\frac{1}{n}$ vs $\frac{1}{n-1}$ makes little diff for large n

- the confidence interval above requires knowledge of σ^2

can replace σ^2 with its sample estimator

$\frac{1}{n}\sum_{i=1}^{n} X_i$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \quad \Rightarrow \boxed{E}[S_n^2] = \sigma^2$$

Random Variable

(fn of random $X_1, \ldots, X_n$)

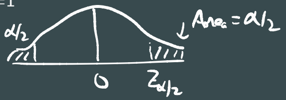
fixed sample-size: recipe for CI

approximate $100(1-\alpha)\%$ Gaussian CI for $\mathbb{E}X$

1. select a sample size N
2. generate N i.i.d. samples $X_1, X_2, \ldots, X_N$ of X
3. compute the estimators $\bar{X}_N, s_N^2$

$$\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n, \qquad S_N^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(X_n - \bar{X}_N\right)^2$$

4. look up the value of $z_{\alpha/2}$ such that



$$\mathbb{P}[-z_{\alpha/2} \le N(0,1) \le z_{\alpha/2}] = 1-\alpha$$

$\mathbb{E}[W_N^2] = \mathrm{Var}(\bar{X}_N)$

$\downarrow$

5. the approximate $100(1-\alpha)\%$ CI for $\mathbb{E}X$ is given by

$\mathbb{E}[S_N^2] = \mathbb{E}[X_i]$

$$\bar{X}_N \mp z_{\alpha/2}\frac{s_N}{\sqrt{N}} = \bar{X}_N + z_{\alpha/2} W_n$$

selecting the sample size

how large should N be so that the resulting $100(1-\alpha)\%$ confidence interval will have a $\boxed{\text{pre-specified width?}}$ ← interpretable measure of quality of measurement

- CI $\implies$ $\bar{X}_N \mp z_{\alpha/2}\left(\sigma/\sqrt{N}\right)$
- half-width $\implies$ $z_{\alpha/2}\dfrac{\sigma}{\sqrt{N}}$ $\approx z_{\alpha/2} \dfrac{S_n}{\sqrt{n}} \overset{\text{AIM}}{\leq} w$
- ℓ be the desired half-width
- Set $N = \dfrac{(z_{\alpha/2})^2 \, \sigma^2}{w^2}$ $\left(\text{ie, for } 95\%, \; N = \dfrac{4 \sigma^2}{w^2}\right)$ $1.96^2 \ldots$

. How do we know σ^2?
- If Ber, $\sigma^2 \leq 1/4 \Rightarrow$ choose $N = 1/w^2$
- more commonly - use 'pilot sims' to estimate σ^2

selecting the sample size

- problem: σ^2 is unknown!
- estimating σ^2 through s_N^2 requires simulation!
- solution: 'pilot runs'

  perform k simulation runs to get $[X_n' : n = 1, \ldots, k]$ as outcomes

  compute

$$\tilde{X}_k = \frac{1}{k}\sum_{n=1}^{k} X_n', \quad \underline{\tilde{s}_k^2} = \frac{1}{k-1}\sum_{n=1}^{k}\left(X_n' - \tilde{X}_k\right)^2$$

pilot estimate of σ^2

use $\tilde{s}_k^2$ to estimate σ^2

for confidence level α, half-width ℓ, set

$$N = \left\lceil \frac{z_{\alpha/2}^2 \tilde{s}_k^2}{\ell^2} \right\rceil$$

basic simulation workflow

- perform pilot run of k simulations (sufficient but not large k)
- compute required sample-size N for desired confidence interval
- run N additional simulations $\implies$ production runs
- form fixed-sample confidence intervals from these N samples
- note: final CI may be different than desired, because it is constructed by using s_N^2 (may be larger/smaller than $\tilde{s}_k^2$)
- for the final confidence interval, discard the information from the trial runs not a problem, since $N \gg k$ usually

(not important, but good habit to avoid 'contamination')

Issues

- S_n^2 is an estimate for σ^2
- $\bar{X}_n$ is not truly Gaussian

In reality, there is some c s.t.

$$\mathbb{P}\left[\mu \in \left[\bar{X}_n \mp c\frac{S_N}{\sqrt{N}}\right]\right] \leq 0.95$$