Inference for sample means

Lecture 12b (STAT 24400 F24)

Confidence intervals for μ

- Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$

  We've derived

$$\mathbb{P}\left(|\bar{X} - \mu| > z_{\alpha/2} \cdot \tfrac{\sigma}{\sqrt{n}}\right) = \alpha$$

  where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, recall $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$.

  We may state the equation as

$$\mathbb{P}\left(|\bar{X} - \mu| \le z_{\alpha/2} \cdot \tfrac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

  Equivalently,

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2} \cdot \tfrac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \cdot \tfrac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

  which can be written as

$$\mathbb{P}\Bigg(\mu \;\in\; \underbrace{\bar{X} \;\pm\; z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{this is a } (1-\alpha) \text{ confidence interval for } \mu}\Bigg) = 1 - \alpha$$

  (when σ^2 is known).

Confidence intervals for μ

- If $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2 (not necessarily normal),

  then

$$\bar{X} \pm z_{\alpha/2} \cdot \tfrac{\sigma}{\sqrt{n}}$$

  is an *approximate* $1 - \alpha$ confidence interval for μ, i.e.,

$$\mathbb{P}\left(\mu \;\in\; \bar{X} \;\pm\; z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \;\approx\; 1 - \alpha$$

  (when σ^2 is known).

Confidence intervals for μ

In practice, generally cannot compute $\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ since σ^2 is unknown.

Can we use sample variance S^2 in place of σ^2?

- Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$

  We just learned

  $F_{t_{n-1}}^{-1}(1 - \alpha/2)$, where $F_{t_{n-1}}$ = CDF of t_{n-1}

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad \Rightarrow \quad \mathbb{P}\left(|\bar{X} - \mu| > t_{n-1,\alpha/2} \cdot \frac{S}{\sqrt{n}}\right) = \alpha$$

$$\Rightarrow \quad \mathbb{P}\Bigg(\mu \in \underbrace{\bar{X} \pm t_{n-1,\alpha/2} \cdot \tfrac{S}{\sqrt{n}}}_{\text{this is a } (1-\alpha) \text{ confidence interval for } \mu}\Bigg) = 1 - \alpha$$

Remarks: t- vs z-confidence intervals for μ

- Note that the t distribution has heavier tails than the normal,

$$t_{n-1,\alpha/2} > z_{\alpha/2}$$

- Therefore, in general, the confidence interval for μ is wider if σ^2 unknown.
- For large n, they will be similar: $t_{n-1,\alpha/2} \searrow z_{\alpha/2}$ as $n \to \infty$.

Confidence intervals for μ: overview

- If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$,

  these confidence intervals have $1-\alpha$ coverage for μ:

  - If σ^2 is known, use $\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
  - If σ^2 is unknown, use $\bar{X} \pm t_{n-1,\alpha/2} \cdot \frac{S}{\sqrt{n}}$

- If $X_1, \dots, X_n$ are i.i.d. with mean μ,

  then the confidence intervals above have $\approx 1-\alpha$ coverage

  (as long as n is not too small)

Why "confidence"?

If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$, then we've calculated

$$\mathbb{P}\left(\mu \in \ \bar{X} \pm 1.96 \cdot \tfrac{\sigma}{\sqrt{n}}\right) = 95\%$$

(& a similar calculation for σ^2 unknown with t distr.)

Caution: Suppose $n = 100$ and $\sigma^2 = 1$, and we observe data $\bar{X} = 5.5$.
Is it correct to write

$$\mathbb{P}\left(\mu \in \ 5.5 \pm 1.96 \cdot \tfrac{1}{\sqrt{100}}\right) = 95\% \ ?$$

No. This is incorrect — the parameter μ is not random!

Analogy: if $X = \#$ Heads out of 4 coin tosses, $\mathbb{P}(X = 0) = \frac{1}{2^4}$.
But after observing $X = 3$ Heads, we can't write $\mathbb{P}(3 = 0) = \frac{1}{2^4}$.

Why "confidence" (rather than probability)

We say we have 95% confidence that μ lies in $5.5 \pm 1.96 \cdot \frac{1}{\sqrt{100}}$

Interpretation: if 1000 researchers run the same experiment,
and each researcher computes a conf. int. $\bar{X} \pm 1.96 \cdot \frac{1}{\sqrt{100}}$,
then $\approx 95\%$ of these intervals will contain μ

the value of $\bar{X}$ will be different for each researcher

Comparison: Bayesian inference for μ

Recall the Bayesian inference framework:

- The unknown parameter is random, drawn from the *prior distribution*
- After observing data, compute the parameter's *posterior distribution*

For data $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$, to perform inference on μ:

- If σ^2 is known, we should choose a prior distribution for μ
- If σ^2 is unknown, we should choose a prior distribution for (μ, σ^2) (even if we are only interested in estimating μ)

Bayesian inference for μ (σ^2 is known)

Simple case: σ^2 is known

$$\begin{cases} \mu & \sim \mathrm{N}(\mu_0, \sigma_0^2) \quad \leftarrow \text{the prior distribution} \\ X_1, \dots, X_n \mid \mu & \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2) \end{cases}$$

Calculate the joint distribution:

$$f_{\mu, X_1, \dots, X_n}(t, x_1, \dots, x_n) = f_\mu(t) \cdot f_{X_1, \dots, X_n \mid \mu}(x_1, \dots, x_n \mid t)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(t-\mu_0)^2/2\sigma_0^2}}_{\text{prior density}} \cdot \underbrace{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - t)^2/2\sigma^2}}_{\text{likelihood}}$$

$$= \begin{pmatrix} \text{terms not} \\ \text{depending on } t \end{pmatrix} \cdot \exp\left\{ -\frac{\left(t - \left[\bar{x} \cdot \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} + \mu_0 \cdot \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}\right]\right)^2}{2 \cdot \frac{1}{n/\sigma^2 + 1/\sigma_0^2}} \right\}$$

Bayesian inference for μ (σ^2 is known) cont.

$\Longrightarrow$ The conditional density (i.e., the posterior density) satisfies

$$f_{\mu \mid X_1, \dots, X_n}(t \mid x_1, \dots, x_n) = \frac{f_{\mu, X_1, \dots, X_n}(t, x_1, \dots, x_n)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}$$

$$= \begin{pmatrix} \text{terms not} \\ \text{depending on } t \end{pmatrix} \cdot \exp\left\{ -\frac{\left(t - \left[\bar{x} \cdot \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} + \mu_0 \cdot \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}\right]\right)^2}{2 \cdot \frac{1}{n/\sigma^2 + 1/\sigma_0^2}} \right\}$$

$\Longrightarrow$ The posterior distribution is:

$$\mu \mid X_1, \dots, X_n \sim \mathrm{N}\left(\underbrace{\bar{X} \cdot \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} + \mu_0 \cdot \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}}_{\approx \bar{X} \text{ if } n \text{ large}}, \underbrace{\frac{1}{n/\sigma^2 + 1/\sigma_0^2}}_{\approx \frac{\sigma^2}{n} \text{ if } n \text{ large}} \right)$$

Bayesian inference for μ (σ^2 is known) cont.

Using the posterior distribution, can also compute posterior probabilities, e.g., $\mathbb{P}(\mu > 0 \mid X_1, \dots, X_n)$.

We can also compute a **credible interval** —

$$\mathbb{P}\left(\mu \in \left(\underbrace{\bar{X} \cdot \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} + \mu_0 \cdot \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}}_{\approx \bar{X} \text{ if } n \text{ large}}\right) \pm z_{\alpha/2} \cdot \underbrace{\sqrt{\frac{1}{n/\sigma^2 + 1/\sigma_0^2}}}_{\approx \frac{\sigma}{\sqrt{n}} \text{ if } n \text{ large}} \;\middle|\; X_1, \dots, X_n\right) = 1 - \alpha$$

interval endpoints are $\approx$ equal

$\Rightarrow$ For large n, Bayesian credible interval $\approx$ frequentist confidence interval

interpretation: posterior prob. for a random μ

interpretation: "confidence" for a non-random μ

Bayesian inference for μ (Remarks)

Remarks

Although the Bayesian credible interval for μ and the frequentist confidence interval for μ have similar endpoints, their interpretations are very different.

Bayesian inference for μ (σ^2 unknown, outline)

If μ and σ^2 are both unknown...

$$\begin{cases} (\mu, \sigma^2) \sim \text{Prior density } f_{\mu,\sigma^2} \\ X_1, \dots, X_n \mid \mu, \sigma^2 \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2) \end{cases}$$

Joint distribution:

$$f_{\mu,\sigma^2,X_1,\dots,X_n}(t, z, x_1, \dots, x_n) = f_{\mu,\sigma^2}(t, z) \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z}} e^{-(x_i - t)^2/2z}$$

If only interested in the posterior distribution of μ, we marginalize over σ^2:

$$f_{\mu,X_1,\dots,X_n}(t, x_1, \dots, x_n) = \int_{z=0}^{\infty} f_{\mu,\sigma^2,X_1,\dots,X_n}(t, z, x_1, \dots, x_n)\, dz.$$

And then calculate the conditional density $f_{\mu|X_1,\dots,X_n}(t \mid x_1, \dots, x_n)$ to find the posterior distribution of μ.

Recap - population parameters and sample statistics

$X_1, \cdots, X_n$ are of a common distribution we want to learn.

- $\mu = \mathbb{E}(X_i)$ is called the population mean, often the quantity we want to estimate.
- $\sigma^2 = \mathbb{E}[(X_i - \mu)^2]$ is called the population variance, we may also want to learn.
- In general, $\mathbb{E}(X_i^r)$ is called the rth population moment.
- In general, the distribution may be parametrized by some population parameter θ.

The X_i's will be observed (as data) to make inference on population parameters.

- $\{X_1, \cdots, X_n\}$ is sometimes called the sample (realized into numbers once observed).
- n is called the sample size.
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean.
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is called the sample variance.
- $\frac{1}{n}\sum_{i=1}^{n} X_i^r$ is called the rth sample moment.

Look ahead

Next, we will look into parameter estimation systematically.

- Criteria of a good estimator.
- Method of moments for parameter estimation.
- Likelihood method for parameter estimation and beyond.