Week 11 – Confidence Intervals

Juan Pablo Lewinger

Last Class - Maximum Likelihood

- Maximum likelihood is a general method for estimating parameters of interest
- Idea: choose parameter the maximizes the probability of observing the data actually observed
- The probability of observing the data actually observed as a function of parameter = **likelihood**
- Previous examples of estimators were sample analogs of population parameters (e.g.sample mean and sample variance) or intuitive (e.g. estimator based on $\max_{1 \le i \le n} X_i$ for $U[0, \theta]$)
- Maximum likelihood can be used in cases when there is no 'natural' estimator like sample mean or variance

Maximum likelihood estimation

- The maximum likelihood **estimate** of θ is the value $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \ldots, x_n)$ that maximizes the log-likelihood function $l(\theta)$.
- The corresponding random variable $\widehat{\theta} = \widehat{\theta}(X_1, X_2, \ldots, X_n)$ is called the maximum likelihood **estimator** of θ.
- (We often use the same notation, $\widehat{\theta}$, for both the estimate and the estimator. We understand which one we are referring to based on context)

Properties of Maximum likelihood estimators

The maximum likelihood estimator (MLE) is consistent:

- $\widehat{\theta}_n \xrightarrow{P} \theta$ as $n \longrightarrow +\infty$.
- i.e. as the sample size gets larger the MLE gets closer and closer to the true parameter θ.
- Consistency is stronger than asymptotic unbiasedness (which MLEs also have)

Invariance:

- If $\widehat{\theta}$ is the MLE of θ then $g(\widehat{\theta})$ is the MLE of $g(\theta)$.
- E.g. MLE of θ^2 is $\widehat{\theta}$

Asymptotic minimum variance -lowest asymptotic mean squared error

Asymptotic normality - The distribution of the MLE gets closer and closer to a normal distribution as $n \longrightarrow \infty$

MLE's are generally biased estimators - But bias disappears as sample size increases (asymptotically unbiased)

Asymptotic variance of the MLE

- Let $X_1, \ldots, X_n \sim f(x;\theta)$, and the log-likelihood function: $l(\theta) = \sum_{i=1}^{n} \log f(X_i;\theta)$
- The Fisher's information of θ (for one sample) is defined as:

$$I(\theta) = -E\left[\frac{\partial^2 l(\theta, X)}{\partial \theta^2}\right]$$

- If $f(x;\theta)$ is 'well behaved' (e.g. its support does not depend on θ)

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

- $\frac{1}{nI(\theta)}$ is the asymptotic variance of $\widehat{\theta}$
- The standard error of the MLE is $se(\widehat{\theta}) = \frac{1}{\sqrt{nI(\theta)}}$
- We can estimate the standard error of the MLE by $\widehat{se}(\widehat{\theta}) = \frac{1}{\sqrt{nI(\widehat{\theta})}}$

Confidence intervals

- So far we've considered what are called point estimators.
- A point estimator by itself not that useful; need some measure of how variable the estimator is
- To supplement point estimators with information about their variability we can also report their SD, called the standard error (SE)
- A better option is to report a whole interval of plausible values for the parameter of interest, i.e. a **confidence interval**.
- Based on Chebysheb's inequality, if T is an estimator for θ, and σ_T is the standard deviation of T:

    $P(|T - \theta| \geq 2\sigma_T) \leq \frac{3}{4}$

    $P\left(\theta \in (T - 2\sigma_T, T + 2\sigma_T)\right) \geq \frac{3}{4}$

    $(T - 2\sigma_T, T + 2\sigma_T)$ is an interval estimator for θ

Confidence intervals

- Dataset $x_1, \ldots, x_n$ modeled as a realization of random variables $X_1, \ldots, X_n$.
- θ the parameter of interest (e.g. mean, variance, probability, etc.), and $0 < \alpha < 1$
- A $(1-\alpha)$-level interval estimator for θ is a a pair of sample statistics $L_n = g(X_1, \ldots, X_n)$ and $U_n = h(X_1, \ldots, X_n)$ such that:

  $$P(L_n < \theta < U_n) = 1 - \alpha \text{ for every value of } \theta$$

- The interval (l_n, u_n) where $l_n = g(x_1, \ldots, x_n)$ and $u_n = h(x_1, \ldots, x_n)$, is called a $100(1-\alpha)\%$ confidence interval for θ.
- The number α is called the confidence level.
- Often, we only require $P(L_n < \theta < U_n) > 1 - \alpha$ or $P(L_n < \theta < U_n) \approx 1 - \alpha$
  - These are conservative and approximate interval estimators respectively

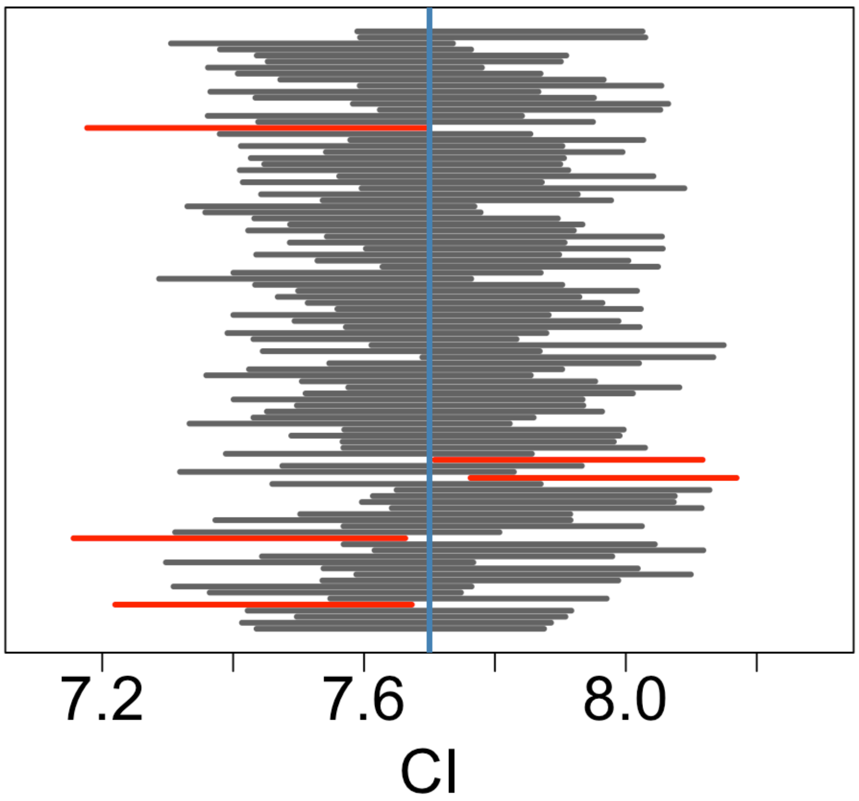
Interval estimator for the normal mean – variance known

$X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, with σ^2 known. We want to construct an interval estimator for μ.

- $\overline{X_n}$ is an unbiased estimator for μ. It is also consistent and it's the MLE (with all its properties)
- We know that $\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- Then, $P(-z_{\alpha/2} \leq \frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}} \leq -z_{\alpha/2}) = 1-\alpha$
- Equivalently, $P\left(\mu \in (\overline{X_n} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X_n} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})\right) = 1-\alpha$
- $(\overline{x_n} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x_n} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$ is a $100(1-\alpha)\%$ confidence interval for μ
- Common confidence levels are 95% and 99%

Interval estimator for the normal mean - variance known

One hundred 95% confidence intervals for $\mu = 7.7$

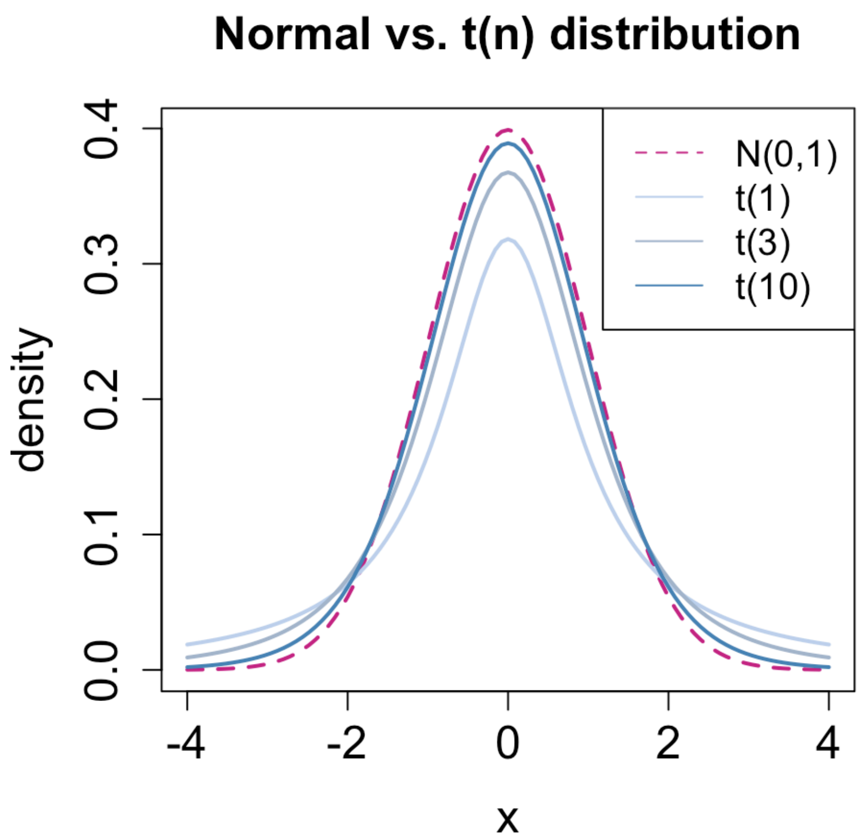


Confidence interval for the normal mean - variance unknown

$X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, with σ^2 unknown. We want an interval estimator for μ.

- In most situations the σ^2 is not known, so we cannot use the interval estimator above
- Idea: estimate σ^2 by the sample variance S_n^2 (and σ by S_n)
- $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ is not normal; it has a t-distribution $n-1$ degrees of freedom, $t(n-1)$
- The pdf of a $t(n)$ distribution is $f(x) = k_n(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$, for $-\infty < x < +\infty$, where k_n is a constant.

t-distribution



Confidence interval for the normal mean – variance unknown

- We know that $\frac{\overline{X}_n-\mu}{S_n/\sqrt{n}} \sim t(n-1)$
- So, $P\left(\mu \in (\overline{X}_n - t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}})\right) = 1-\alpha$
- $\left(\overline{x_n} - t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{x_n} + t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}\right)$ is a $100(1-\alpha)\%$ confidence interval for μ

Asymptotic interval estimator for the mean

$X_1, \ldots, X_n \overset{i.i.d}{\sim} F$ with $E[X_i] = \mu$. We want an interval estimator for μ.

- Now we don't know the distribution of X_i (it could be normal or anything else)
- The CLT states that: $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$
- A variant of the CLT applies if we replace σ with the estimate S_n: $\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{D} N(0, 1)$
- So, $P\left(\mu \in (\overline{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}})\right) \approx 1 - \alpha$
- $\left(\overline{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right)$ is an **approximate** $100(1 - \alpha)\%$ interval estimator for μ

Confidence interval for the mean - example

Let $x_1, \dots, x_n$ be a dataset modeled as a realization of i.i.d random variables $X_1, \dots, X_n$ with $E[X_i] = \mu$

Which interval for μ would you choose in each of these scenarios?

- $X_i \sim N(\mu, 2)$
- $X_i \sim N(\mu, \sigma^2)$
- $X_i \sim F$

Confidence interval for a proportion

$X_1, \dots, X_n \overset{i.i.d}{\sim} Bernoulli(p)$ or equivalently $X = X_1 + \dots + X_n \sim Binomial(p)$

- $\hat{p} = \overline{X_n} = \frac{X}{n}$ is the MLE of p
- $Var(\hat{p}) = \frac{p(1-p)}{n}$; $SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$
- A natural estimator of $SD(\hat{p})$ is $\widehat{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$; standard error of $\hat{p}$
- By the CLT we have $\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}} \approx N(0,1)$.
- By a variant of the CLT we have $\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \approx N(0,1)$
- So, $\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ is an **approximate** $100(1-\alpha)\%$ interval estimator for p
- Can also write as: $\left(\hat{p} - z_{\alpha/2}\widehat{se}(\hat{p}), \quad \hat{p} + z_{\alpha/2}\widehat{se}(\hat{p})\right)$

Confidence interval for a proportion

Example: $n = 1{,}000$ poll showed that 557 voters intend to vote for candidate A and 443 for candidate B.

p proportion of all voters that intend to vote for A

$\widehat{p} = 557/1000 = 0.557 = 55.7\%$

$\widehat{se}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{0.557(1-0.557)/1000)} = 0.0157$

95% CI for p: $\left(\widehat{p} - 1.96 \times \widehat{se}(\widehat{p}), \quad \widehat{p} + 1.96 \times \widehat{se}(\widehat{p})\right) = (0.526, 0.588)$

Margin of error = $1.96 \times \widehat{se}(\widehat{p}) = 0.03$

“This poll is accurate to plus or minus 3% points 19 times out of 20”

Asymptotic variance of the MLE

- Let $X_1, \ldots, X_n \sim f(x;\theta)$, and the log-likelihood function: $l(\theta) = \sum_{i=1}^{n} \log f(X_i;\theta)$
- The Fisher's information of θ (for one observation) is defined as:

$$I_1(\theta) = -E\left[\frac{\partial^2 l_1(\theta, X)}{\partial \theta^2}\right]$$

  where $l_1(\theta) = \log f(X;\theta)$ is the log-likelihood for one observation

- If $f(x;\theta)$ is 'well behaved' (e.g. its support does not depend on θ)

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{1}{I_1(\theta)}\right)$$

- $\frac{1}{nI_1(\theta))}$ is the asymptotic variance of $\widehat{\theta}$
- The standard error of the MLE is $se(\widehat{\theta}) = \frac{1}{\sqrt{nI_1(\theta))}}$
- We can estimate the standard error of the MLE by $\widehat{se}(\widehat{\theta}) = \frac{1}{\sqrt{nI_1(\widehat{\theta})}}$

Asymptotic confidence intervals based on MLEs

Estimate of the standard error (i.e. $sd(\hat{\theta})$):

$$\widehat{se}(\hat{\theta}) = \frac{1}{\sqrt{nI(\hat{\theta})}}$$

Based on the asymptotic normality of the MLE an approximate $100(1-\alpha)\%$ confidence interval can be computed as:

$$\left(\hat{\theta} - z_{\alpha/2}\widehat{se}(\hat{\theta}),\quad \hat{\theta} + z_{\alpha/2}\widehat{se}(\hat{\theta})\right) = \left(\hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}},\quad \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}}\right)$$

Asymptotic confidence intervals based on MLEs

Example: $X_1, \ldots, X_n \sim Geometric(p)$; $\quad P(X = x) = p(1-p)^{x-1}$

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1} = p^n (1-p)^{\sum_{i=1}^{n} X_i - n}$$

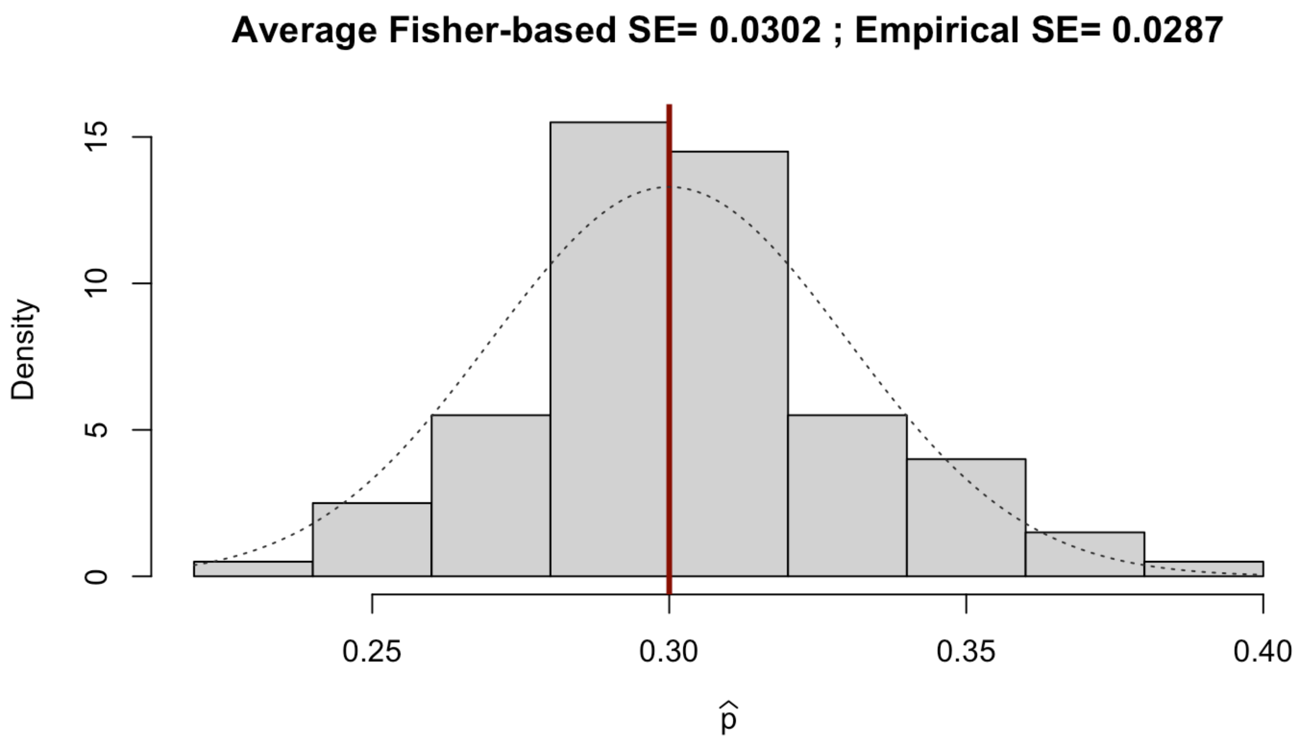
$$l(p) = n\log(p) + \left(\sum_{i=1}^{n} X_i - n\right)\log(1-p)$$

$$l'(p) = -\frac{n}{p} - \frac{\sum_{i=1}^{n} X_i - n}{1-p} = 0 \iff p = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}_n}$$

So, MLE of p is $\hat{p} = \frac{1}{\bar{X}_n}$

Asymptotic confidence intervals based on MLEs

```
set.seed(1909)
nsims = 100; n = 70; p = 0.3
p_hat = replicate(nsims, {X = rgeom(n, p) + 1; 1/mean(X)})
```



Asymptotic confidence intervals based on MLEs

$l_1''(p) = -\frac{1}{p^2} - \frac{X-1}{(1-p)^2}$ ($l_1(p)$ is log-likelihood for a single observation X)

Fisher information for one observation: $I_1(p) = -E[l_1''(p)] = \frac{1}{p^2} + \frac{\frac{1}{p}-1}{(1-p)^2} = \frac{1}{p^2(1-p)}$

$se(\hat{p}) = \sqrt{(nI_1(p))^{-1}} = \frac{p\sqrt{(1-p)}}{\sqrt{n}}$

$\widehat{se}(\hat{p}) = \frac{\hat{p}\sqrt{(1-\hat{p})}}{\sqrt{n}}$

$\left(\hat{p} - z_{\alpha/2}\frac{\hat{p}\sqrt{(1-\hat{p})}}{\sqrt{n}},\quad \hat{p} + z_{\alpha/2}\frac{\hat{p}\sqrt{(1-\hat{p})}}{\sqrt{n}}\right)$

Asymptotic confidence intervals based on MLEs

Example $X_1, \dots, X_{70} \sim Geometric(p), \bar{X} = 3.714$

$\widehat{p} = 0.269$

$\widehat{se}(\widehat{p}) = \frac{\widehat{p}\sqrt{(1-\widehat{p})}}{\sqrt{70}} = 0.0275$

$\left(\widehat{p} - z_{\alpha/2}\widehat{se}(\widehat{p}), \quad \widehat{p} + z_{\alpha/2}\widehat{se}(\widehat{\theta})\right) =$

$= (0.269 - 1.96 \times 0.0275, \quad 0.269 + 1.96 \times 0.0275) = (0.220, \quad 0.318)$

For next class (in two weeks)

- Hypothesis testing
- Read PSD 10.2

Happy Thanksgiving!!