STA 250/MTH 342 – Intro to Mathematical Statistics

Lecture 13

Interval estimation

- In the Bayesian framework, by treating unknown parameter θ as a random variable and place a prior distribution $\xi(\theta)$ on it, Bayes' theorem allows us to get a posterior on θ

$$\xi(\theta|x) \propto \xi(\theta)f(x|\theta).$$

- This is a probability distribution on θ that represents *all of* our knowledge about the parameter θ after observing the data.
- Based on this distribution we can make various kinds of inference about θ.
- One example is *point estimation*.
- Another example is *interval estimation*.

Interval estimation from Bayes' perspective

- Instead of giving a single "guess" of θ, can you give me a range of values such that *given the data*, θ is in that range with high probability, e.g. 95% chance.
- Given the posterior distribution $\xi(\theta|x)$ this is easy to do.
- Just find an interval $[\theta_0, \theta_1]$, such that given the data,

$$P(\theta \in [\theta_0, \theta_1]|x) = \int_{\theta_0}^{\theta_1} \xi(u|x) du = 95\%.$$

- Such an interval $[\theta_0, \theta_1]$ is called a 95% (posterior) *credible interval* of θ.

- In particular, if choose θ_0 and θ_1 to be the 2.5% and 97.5% quantiles of $\xi(\theta|x)$, this $[\theta_0, \theta_1]$ is called the 95% *central credible interval* of θ.
- In general, for θ_0 and θ_1 being the $\alpha/2$ and $1-\alpha/2$ quantile of $\xi(\theta|x)$, $[\theta_0, \theta_1]$ forms a $(1-\alpha) \times 100\%$ central credible interval.
- (Draw a figure.)
- *Note that here the probability statement is made from an after-experiment view-point. This is only possible in the Bayesian perspective.*

Example: political poll

- Prior $\xi(\theta)$ is Beta(α, β).
- Observation $X = x$ out of n interviewees support governmer.
- Posterior $\xi(\theta|x)$ is Beta$(\alpha + x, \beta + n - x)$.
- For example if $n = 100$, $x = 40$, and $\xi(\theta)$ is Beta(12,12), then $\xi(\theta|x)$ is Beta(52,72), and the central 95% credible interval for θ is

  $$[0.33, 0.51].$$

  The quantiles can be computed using R's `qbeta` command.
- Question: What is a 95% credible interval for $\frac{p}{1-p}$, the odds for support?
- It is $\left[\frac{0.33}{1-0.33}, \frac{0.51}{1-0.51}\right] = [0.49, 1.04]$

In general,

- If $[\theta_0, \theta_1]$ is a $(1-\alpha)\%$ credible interval for a parameter θ, and
- if $g(\cdot)$ is a monotone increasing (or decreasing) function, then

$$[g(\theta_0), g(\theta_1)] \quad \text{(or } [g(\theta_1), g(\theta_0)] \text{ when } g \text{ is decreasing)}$$

  is a $(1-\alpha)\%$ credible interval for $g(\theta)$. Why?
- This is because

$$P(g(\theta_0) \leq g(\theta) \leq g(\theta_1)|x) = P(\theta_0 \leq \theta \leq \theta_1|x).$$

- In particular, if $[\theta_0, \theta_1]$ is a central $(1-\alpha)\%$ credible interval for θ, then so is $[g(\theta_0), g(\theta_1)]$ (or $[g(\theta_1), g(\theta_0)]$ when g is decreasing) for $g(\theta)$.
- We may call this the *invariance property* of credible intervals.

Interval estimation from a sampling point of view

- Now let us again assume a sampling perspective.
- θ is an unknown *fixed* quantity.
- A probabilistic statement about θ such as "given the data, θ is in $[\theta_0, \theta_1]$ with 95% chance" is now out of reach.
- The goal of interval estimation is still to construct a range of values that contain θ *with high probability*.
- But now *"with high probability"* must take on a different meaning.

- The only randomness is on the data X before the experiment is carried out.
- We form an interval guess $[\theta_l(X), \theta_u(X)]$about θ based on the data according to certain rules.
- Note that from a before-experiment perspective, *this interval itself* is random—its upper and lower bounds depend on the data.
- The parameter it is trying to cover is *not random*!

- If we repeat the experiment many times and the fraction of times for $[\theta_l(X), \theta_u(X)]$ to cover the unknown θ is 95%, then we say that we are 95% *confident* that the interval covers θ.
- Again, note that it is the interval that is random, not the parameter.
- After the experiment, the data $X = x$ is observed, and we end up with realization of this interval $[\theta_u(x), \theta_l(x)]$.
- It either covers θ or it doesn't. We *cannot* say that θ is in $[\theta_u(x), \theta_l(x)]$ with 95% chance.

Confidence interval

- A random interval $[\theta_l(\mathbf{X}), \theta_u(\mathbf{X})]$ that contains θ the true parameter 95% of the time, is called a 95% *confidence interval* of θ.
- A realization of this interval after observing the data, $[\theta_l(\mathbf{x}), \theta_u(\mathbf{x})]$ is also called a 95% *confidence interval* of θ.
- However always keep in mind that this 95% chance is not a statement about this particular realized interval, but about the corresponding random interval $[\theta_l(\mathbf{X}), \theta_u(\mathbf{X})]$.
- The shooting range analogy for confidence intervals and credible intervals.
- The general public typically mistakenly interpret confidence intervals as if they were credible intervals! After all, it is more natural for people to think in the after-experiment perspective.

How do we find confidence intervals (CIs)?

- A common method is to construct CIs based on MLEs using their sampling distributions.
- Sometimes the exact sampling distribution of MLEs can be found and a CI can be constructed accordingly.
- More often, however, the exact sampling distribution of MLEs are difficult to find, we use approximate methods such as Fisher's approximation to build an approximate CI.
- We will learn each in turn through a sequence of examples.

Example: our first confidence interval

- Suppose our data are i.i.d. observations $X_1, X_2, \ldots, X_n$ from a $N(\mu, \sigma^2)$ distribution where σ^2 is known.
- The MLE for μ is $\hat{\mu} = \bar{X}$. It has a sampling distribution of $N(\mu, \sigma^2/n)$.
- So we know that

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

  is a $N(0,1)$ random variable.
- Therefore

$$P(\Phi^{-1}(0.025) < Z < \Phi^{-1}(0.975)) = 0.95.$$

  That is

$$P\left(\Phi^{-1}(0.025) < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \Phi^{-1}(0.975)\right) = 0.95.$$

From

$$P\left(\Phi^{-1}(0.025) < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \Phi^{-1}(0.975)\right) = 0.95$$

we have

$$P\left(\bar{X}-\Phi^{-1}(0.975)\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+\Phi^{-1}(0.975)\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Thus, a 95% CI for μ is

$$\left[\bar{X}-\Phi^{-1}(0.975)\frac{\sigma}{\sqrt{n}}, \bar{X}+\Phi^{-1}(0.975)\frac{\sigma}{\sqrt{n}}\right]$$

or

$$\left[\bar{X}-1.96\frac{\sigma}{\sqrt{n}}, \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right].$$

Exercise: Show that more generally, a $(1-\alpha)\times 100\%$ CI for μ is

$$\left[\bar{X}-\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, \bar{X}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right].$$

- $(1-\alpha)\times 100\%$ is called the confidence level.
- The textbook calls $\gamma = 1-\alpha$ the confidence coefficient.

- Now what if σ is also unknown? What will be a 95% CI for μ?
- In this case, the above interval involves σ, and so is not useful.
- We can estimate σ^2 by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- Now let us consider the statistic with σ replaced by the sample standard deviation s:
$$\frac{\sqrt{n}(\bar{X} - \mu)}{s}.$$
- What is its sampling distribution? If we know its sampling distribution, say its cdf is F, then

$$\begin{aligned}1 - \alpha &= P\left(F^{-1}(\alpha/2) \le \frac{\sqrt{n}(\bar{X} - \mu)}{s} \le F^{-1}(1 - \alpha/2)\right) \\ &= P\left(\bar{X} - F^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{X} + F^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}\right).\end{aligned}$$

- So a $(1 - \alpha) \times 100\%$ CI for μ is
$$\left[\bar{X} - F^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}, \bar{X} + F^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}\right].$$

- So let's try to find the sampling distribution of $\sqrt{n}(\bar{X}-\mu)/s$.
- Note
$$\frac{\sqrt{n}(\bar{X}-\mu)}{s} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{s^2/\sigma^2}}$$
- Let $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ and $W = \frac{(n-1)s^2}{\sigma^2}$. We know that $Z \sim N(0,1)$, $W \sim \chi^2_{n-1}$ and they are independent.
- Now
$$\frac{\sqrt{n}(\bar{X}-\mu)}{s} = \frac{Z}{\sqrt{W/(n-1)}} \sim t_{n-1}$$
its distribution is defined to be the *(student's) t distribution with $n-1$ degrees of freedom*.

t distribution with m degrees of freedom

- More generally, if Z is a $N(0,1)$ random variable, W is a χ^2_m random variable, and Z and W are independent, then we define the distribution of
$$\frac{Z}{\sqrt{W/m}}$$
to be the *t distribution with m degrees of freedom*, denoted by t_m.
- One can show that its p.d.f is
$$f_{t_m}(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-\frac{(m+1)}{2}} \quad \text{for } -\infty < x < \infty.$$
- You don't have to memorize the pdf and/or cdf.

Back to our CI example

- We know
$$\frac{\sqrt{n}(\bar{X}-\mu)}{s} = \frac{Z}{\sqrt{W/(n-1)}} \sim t_{n-1}.$$
Thus
$$P\left(F^{-1}_{t_{n-1}}(0.025) < \frac{\sqrt{n}(\bar{X}-\mu)}{s} < F^{-1}_{t_{n-1}}(0.975)\right) = 0.95,$$
and so
$$P\left(\bar{X} - F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}} < \mu < \bar{X} + F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}}\right) = 0.95.$$
Therefore a 95% CI for μ is
$$\left[\bar{X} - F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}}, \bar{X} + F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}}\right]$$

Example

If $n = 20$, $\bar{X} = 20.5$ and $s = 3.6$, then this 95% CI is

$$\left[20.5 - 2.093 \times \frac{3.6}{\sqrt{20}}, 20.5 + 2.093 \times \frac{3.6}{\sqrt{20}}\right].$$

That is

$$[18.82, 22.18].$$

It either contains μ or it doesn't. We can't say that μ is in [18.82,22.18] with 95% chance.

- *Question: What is a 95% CI for e^{μ}?*
- It is $[e^{18.82}, e^{22.18}]$. What does this mean?
- It is a realization of the CI

$$\left[\exp\left(\bar{X} - F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}}\right), \exp\left(\bar{X} + F^{-1}_{t_{n-1}}(0.975)\frac{s}{\sqrt{n}}\right)\right].$$

More generally, if $[\theta_l(\mathbf{X}), \theta_u(\mathbf{X})]$ is a $(1-\alpha)\%$ confidence interval for θ, then

- $[g(\theta_l(\mathbf{X})), g(\theta_u(\mathbf{X}))])$ (or $[g(\theta_l(\mathbf{X})), g(\theta_u(\mathbf{X}))])$ when g is decreasing) is a $(1-\alpha)\%$ CI for $g(\theta)$.
- Why?
- You may call this the "invariance property for CIs".

A general strategy for finding CI's

- First find a quantity called the *pivotal quantity* which is a transformation of the data (often the MLE) and the parameter of interest θ
$$S(\mathbf{X}, \theta)$$
whose sampling distribution F is known and does not depend on the unknown parameter.
- Then
$$P\left(S(\mathbf{X}, \theta) \in [F^{-1}(\alpha/2), F^{-1}(1-\alpha/2)]\right) = 1 - \alpha.$$
- Thus a $(1-\alpha) \times 100\%$ CI for θ is given by solving the range of θ, $[\theta_l(\mathbf{X}), \theta_u(\mathbf{X})]$, such that
$$F^{-1}(\alpha/2) \leq S(\mathbf{X}, \theta) \leq F^{-1}(1-\alpha/2).$$

- A common way to construst pivotal statistics is to use standardized MLE.
- The previous examples show a rather *uncommon* scenario: the exact sampling distribution of the MLE is known, and thus so is that of the pivotal statistic.
- Other times, the exact sampling distribution of the MLE is difficult to get, and we resort to CLT and Fisher's approximation to get a normal approximation to the sampling distribution when the amount of data, n, is large.

Recall Fisher's approximation

- When data $X_1, X_2, \ldots, X_n$ are i.i.d. from a distribution $f(x|\theta)$, and MLE for θ is a solution to the equation

$$\frac{d}{d\theta} \log L(\theta) = 0,$$

then for large n, the sampling distribution of the MLE $\hat{\theta}$ is approximately

$$N\left(\theta, \frac{\tau^2(\theta)}{n}\right)$$

where

$$\frac{1}{\tau^2(\theta)} = I(\theta) \quad \text{Fisher's information.}$$

- Thus for large n,

$$P\left(\Phi^{-1}(0.025) < \frac{\hat{\theta}-\theta}{\tau(\theta)/\sqrt{n}} < \Phi^{-1}(0.975)\right) \approx 0.95.$$

So

$$P\left(\hat{\theta}-\Phi^{-1}(0.975)\frac{\tau(\theta)}{\sqrt{n}} < \theta < \hat{\theta}+\Phi^{-1}(0.975)\frac{\tau(\theta)}{\sqrt{n}}\right) \approx 0.95.$$

- The lower and upper limit involves θ, so this does not give a CI.
- The bounds cannot depend on the unknown parameter for otherwise the interval is not useful.
- Fortunately, we know that for large n, the MLE $\hat{\theta}$ is typically pretty close to the true θ. (Which property of the MLE is this?) So hopefully $\tau(\hat{\theta})$ is not very different from $\tau(\theta)$ for the true θ. Therefore, an approximate 95% CI for θ is

$$\left[\hat{\theta}-\Phi^{-1}(0.975)\frac{\tau(\hat{\theta})}{\sqrt{n}} < \theta < \hat{\theta}+\Phi^{-1}(0.975)\frac{\tau(\hat{\theta})}{\sqrt{n}}\right]$$

More generally, for large n, an approximate $(1-\alpha)\times 100\%$ CI based on Fisher's approximation is given by

$$\left[\hat{\theta}-\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{\tau(\hat{\theta})}{\sqrt{n}}<\theta<\hat{\theta}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{\tau(\hat{\theta})}{\sqrt{n}}\right]$$

where $\tau(\hat{\theta})$ satisfies

$$\tau^2(\hat{\theta})=\frac{1}{I(\hat{\theta})}.$$

Example: Exponential data

- Suppose the data $X_1, X_2, \ldots, X_n$ are i.i.d. observations from a Exponential(λ) distribution with unknown mean λ.
- We know that the MLE for λ is $\hat{\lambda} = \frac{1}{\bar{X}}$.
- Let us construct an approximate 95% CI for λ based on Fisher's approximation.

As we did last time, the Fisher's information is

$$I(\lambda) = E\left[\left(\frac{d}{d\lambda}\log f(X_1|\lambda)\right)^2\right] = -E\left[\frac{d^2}{d\lambda^2}\log f(X_1|\lambda)\right] = \frac{1}{\lambda^2}.$$

So

$$\tau^2(\lambda) = \frac{1}{I(\lambda)} = \lambda^2.$$

Thus

$$\tau^2(\hat{\lambda}) = \tau^2\left(\frac{1}{\bar{X}}\right) = \frac{1}{\bar{X}^2},$$

and accordingly (note that all X_i's are positive)

$$\tau(\hat{\lambda}) = \frac{1}{\bar{X}}.$$

Therefore, an approximate $(1-\alpha)\times 100\%$ CI for λ is

$$\left[\hat{\theta}-\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{\tau(\hat{\theta})}{\sqrt{n}}<\theta<\hat{\theta}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{\tau(\hat{\theta})}{\sqrt{n}}\right]$$

That is,

$$\left[\frac{1}{\bar{X}}-\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{1}{\sqrt{n}\bar{X}},\frac{1}{\bar{X}}+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{1}{\sqrt{n}\bar{X}}\right].$$

For example, if $n=50$ and $\bar{X}=25.32$, then an approximate 95% CI is

$$[0.0285, 0.0504].$$

Question: What is an (approximate) 95% CI for $\theta=\frac{1}{\lambda}$?

- Exercise: Can you find it in four ways?