S&DS 242/542: Theory of Statistics

Lecture 13: Asymptotic normality and the delta method

Recap of estimators

We've discussed two general methods for estimating a parameter $\theta \in \mathbb{R}^k$ given data $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$ from a parametric model:

- Method of moments, which chooses θ so that the first k *moments* of $f(x \mid \theta)$ match their sample estimates:

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \ldots, \quad \mathbb{E}[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

- Maximum likelihood, which chooses θ to maximize the likelihood (joint PDF or PMF of the observed data) or equivalently the *log-likelihood*

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(X_i \mid \theta)$$

We've computed these estimators analytically in some simple models, and also discussed numerical approaches for computation.

Large sample properties and confidence intervals

In this lecture, we will begin to discuss the statistical properties of these estimators in large samples, and aim to understand:

- In typical parametric models, why the sampling distributions of method of moments and maximum likelihood estimators are approximately normal for large n.
- General methods to compute the variances of these normal approximations.
- How to use these normal approximations to construct confidence intervals that quantify our uncertainty about the parameter value.

Example of the Poisson model

Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Poisson}(\lambda)$. Recall that the method of moments estimator and MLE are both given by the sample average $\hat{\lambda} = \bar{X}$.

Previously, we computed the bias and variance of $\hat{\lambda}$, showing that

$$\mathbb{E}_\lambda[\hat{\lambda}] = \lambda, \qquad \text{Var}_\lambda[\hat{\lambda}] = \lambda/n.$$

So $\hat{\lambda}$ is unbiased, with standard error $\sqrt{\lambda/n}$.

When n is large, asymptotic theory gives a more complete picture of the statistical behavior of $\hat{\lambda}$: By the LLN, $\hat{\lambda} \to \lambda$ in probability as $n \to \infty$. Furthermore, by the CLT

$$\sqrt{n}(\hat{\lambda} - \lambda) \to \mathcal{N}(0, \lambda)$$

in distribution as $n \to \infty$. Informally, the distribution of $\hat{\lambda}$ is approximately $\mathcal{N}(\lambda, \frac{\lambda}{n})$ for large n.

Confidence interval in the Poisson model

This normal approximation allows us to construct a confidence interval for λ: For a desired coverage level $1-\alpha \in (0,1)$, let $z^{(\alpha/2)}$ be the upper-$\alpha/2$ point of the standard normal distribution. Then an asymptotic $(1-\alpha)$-confidence interval is given by

$$\hat{\lambda} \pm z^{(\alpha/2)} \sqrt{\frac{\hat{\lambda}}{n}}$$

This satisfies, for large n,

$$\begin{aligned}\mathbb{P}_\lambda\left[\lambda \in \hat{\lambda} \pm z^{(\alpha/2)}\sqrt{\frac{\hat{\lambda}}{n}}\right] &= \mathbb{P}_\lambda\left[\hat{\lambda} - \lambda \in \pm z^{(\alpha/2)}\sqrt{\frac{\hat{\lambda}}{n}}\right] \\ &= \mathbb{P}_\lambda\left[\sqrt{\frac{n}{\hat{\lambda}}}(\hat{\lambda} - \lambda) \in \pm z^{(\alpha/2)}\right] \\ &\approx 1 - \alpha\end{aligned}$$

where $\mathbb{P}_\lambda$ denotes the probability over $X_1, \ldots, X_n \overset{IID}{\sim} \text{Poisson}(\lambda)$ with true parameter λ.

Consistency and asymptotic normality

An estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ in a parametric model is **consistent** if, for any true value of the parameter θ, given data $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$, we have

$$\hat{\theta} \to \theta$$

in probability as $n \to \infty$.

This estimator $\hat{\theta}$ is furthermore **asymptotically normal** if, for some asymptotic variance $v(\theta)$,

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}(0, v(\theta))$$

in distribution as $n \to \infty$.

In our previous example, the estimator $\hat{\lambda} = \bar{X}$ in the Poisson(λ) model is a consistent and asymptotically normal estimator of λ.

Implications of asymptotic normality

If an estimator $\hat{\theta}$ is asymptotically normal, with

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}(0, v(\theta))$$

then informally, for large n, this tells us:

- $\hat{\theta}$ is *asymptotically unbiased*. More precisely, the bias of $\hat{\theta}$ is of smaller order than $1/\sqrt{n}$. (Otherwise $\sqrt{n}(\hat{\theta} - \theta)$ would not converge to a distribution with mean 0.)
- The standard error of $\hat{\theta}$ is approximately $\sqrt{v(\theta)/n}$. In particular, this is on the order of $1/\sqrt{n}$, so the variance (rather than the squared bias) is the main contributing factor to the mean-squared-error.
- Under the true parameter θ, the sampling distribution of $\hat{\theta}$ is approximately $\mathcal{N}(\theta, \frac{v(\theta)}{n})$.

Confidence intervals

Given a coverage level $1 - \alpha \in (0, 1)$, a **$(1 - \alpha)$-confidence interval** for θ is an interval $\widehat{S} = \widehat{S}(X_1, \ldots, X_n)$ such that

$$\mathbb{P}_\theta[\theta \in \widehat{S}] = 1 - \alpha$$

where $\mathbb{P}_\theta$ denotes the probability when $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$.

Here, θ is a *fixed* parameter, and $\mathbb{P}_\theta$ is the probability over the randomness of the data $X_1, \ldots, X_n$ defining $\widehat{S}$.

The interval $\widehat{S}$ is an *asymptotic* $(1 - \alpha)$-confidence interval for θ if

$$\mathbb{P}_\theta[\theta \in \widehat{S}] \to 1 - \alpha$$

as $n \to \infty$. Informally, the probability that $\widehat{S}$ covers θ is approximately $1 - \alpha$ for large sample sizes n.

Confidence intervals from asymptotically normal estimators

Proposition

Suppose $\hat{\theta}$ is a consistent and asymptotically normal estimator of θ, with

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}(0, v(\theta))$$

in distribution as $n \to \infty$, and $v(\theta)$ is a continuous function that is non-zero at θ. Then for any $1 - \alpha \in (0, 1)$,

$$\hat{\theta} \pm z^{(\alpha/2)} \sqrt{\frac{v(\hat{\theta})}{n}}$$

is an asymptotic $(1 - \alpha)$-confidence interval for θ.

Note that estimators with smaller asymptotic variance $v(\theta)$ will yield narrower confidence intervals.

Confidence intervals from asymptotically normal estimators

Proof: $\mathbb{P}_\theta\left[\theta \in \hat{\theta} \pm z^{(\alpha/2)}\sqrt{\frac{v(\hat{\theta})}{n}}\right]$

$$= \mathbb{P}_\theta\left[\hat{\theta} - \theta \in \pm z^{(\alpha/2)}\sqrt{\frac{v(\hat{\theta})}{n}}\right]$$

$$= \mathbb{P}_\theta\left[\sqrt{\frac{n}{v(\hat{\theta})}}(\hat{\theta} - \theta) \in \pm z^{(\alpha/2)}\right]$$

- $\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}(0, v(\theta))$ by assumption of asymptotic normality
- $\hat{\theta} \to \theta$ in probability by consistency, so $v(\hat{\theta}) \to v(\theta)$ in probability by Cont. Mapping Theorem
- Then by Slutsky's Lemma: $\sqrt{\frac{n}{v(\hat{\theta})}}(\hat{\theta} - \theta) \to \mathcal{N}(0,1)$.

So $\mathbb{P}_\theta\left[\theta \in \hat{\theta} \pm z^{(\alpha/2)}\sqrt{\frac{v(\hat{\theta})}{n}}\right] \to 1-\alpha$ as $n \to \infty$.

Asymptotic normality of method of moments estimators

In the Poisson(λ) example, we showed asymptotic normality of the method of moments estimator $\hat{\lambda} = \bar{X}$ using the CLT.

More generally, for estimating a parameter $\theta \in \mathbb{R}$, the method of moments estimator equates

$$\mathbb{E}_\theta[X] = \bar{X}$$

Supposing that $\mathbb{E}_\theta[X] = \mu(\theta)$ and μ is a 1-to-1 function, the method of moments estimator is $\hat{\theta} = \mu^{-1}(\bar{X})$.

How can we deduce asymptotic normality of $\hat{\theta}$ from the asymptotic normality of $\bar{X}$?

The delta method

If S is an asymptotically normal statistic and $g : \mathbb{R} \to \mathbb{R}$ is a smooth function, then $g(S)$ will also be asymptotically normal. This is formalized as the **delta method**.

Theorem (Delta method)

Suppose $S = S(X_1, \ldots, X_n)$ satisfies

$$\sqrt{n}(S(X) - \mu(\theta)) \to \mathcal{N}(0, v(\theta))$$

in distribution as $n \to \infty$, when $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$.

If $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable at $\mu(\theta)$, then

$$\sqrt{n}\Big(g(S(X)) - g(\mu(\theta))\Big) \to \mathcal{N}\Big(0, g'(\mu(\theta))^2 v(\theta)\Big).$$

Proof sketch of the delta method

$\sqrt{n}(S(X)-\mu(\theta)) \to \mathcal{N}(0, v(\theta))$ implies that

$S(X)-\mu(\theta)$ is typically small for large n.

Taylor expand:

$$g(S(X)) \approx g(\mu(\theta)) + g'(\mu(\theta)) \cdot (S(X)-\mu(\theta))$$

$$\Rightarrow \sqrt{n}\left[g(S(X)) - g(\mu(\theta))\right] \approx g'(\mu(\theta)) \cdot \underbrace{\sqrt{n}\left[S(X)-\mu(\theta)\right]}_{\to \mathcal{N}(0, v(\theta))}$$

$$\Rightarrow \sqrt{n}\left[g(S(X)) - g(\mu(\theta))\right] \approx \mathcal{N}\left(0, g'(\mu(\theta))^2 \cdot v(\theta)\right).$$

Method of moments in the Exponential model

Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Exponential(λ). Recall that $\mathbb{E}_\lambda[X] = \frac{1}{\lambda}$, so the method of moments estimator is $\hat{\lambda} = \frac{1}{\bar{X}}$.

By the CLT:

$$\sqrt{n}\left(\bar{X} - \frac{1}{\lambda}\right) \to \mathcal{N}(0, \underbrace{\mathrm{Var}_\lambda[X]}_{=1/\lambda^2}) = \mathcal{N}(0, 1/\lambda^2)$$

Apply delta method w/ $g(x) = \frac{1}{x}$, $g'(x) = -\frac{1}{x^2}$:

$$\sqrt{n}(\hat{\lambda} - \lambda) = \sqrt{n}(g(\bar{X}) - g(1/\lambda))$$

$$\to \mathcal{N}(0, \underbrace{g'(1/\lambda)^2 \cdot 1/\lambda^2}_{\left(-\frac{1}{(1/\lambda)^2}\right)^2 \cdot \frac{1}{\lambda^2} = \lambda^2})$$

Method of moments in the Exponential model

So to summarize:

$$\sqrt{n}(\hat{\lambda} - \lambda) \to \mathcal{N}(0, \lambda^2) \text{ in distribution as } n \to \infty.$$

An asymptotic $(1-\alpha)$-confidence interval for λ is:

$$\hat{\lambda} \pm z^{(\alpha/2)} \cdot \sqrt{\frac{\hat{\lambda}^2}{n}}$$

Method of moments in the Pareto model

Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Pareto}(\theta, 1)$. Recall that $\mathbb{E}_\theta[X] = \frac{\theta}{\theta-1}$ (for $\theta > 1$), so the method of moments estimator is $\hat{\theta} = \frac{\bar{X}}{\bar{X}-1}$.

$$\sqrt{n}\left(\bar{X} - \frac{\theta}{\theta-1}\right) \to \mathcal{N}(0, \underbrace{\text{Var}_\theta[X]}_{= \frac{\theta}{(\theta-1)^2(\theta-2)}})$$

(for $\theta > 2$)

Apply delta method w/ $g(x) = \frac{x}{x-1} = 1 + \frac{1}{x-1}$, $g'(x) = -\frac{1}{(x-1)^2}$

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\left(g(\bar{X}) - g\left(\frac{\theta}{\theta-1}\right)\right)$$

$$\longrightarrow \mathcal{N}\left(0, g'\left(\frac{\theta}{\theta-1}\right)^2 \cdot \frac{\theta}{(\theta-1)^2(\theta-2)}\right)$$

Method of moments in the Pareto model

$$g'\left(\frac{\theta}{\theta-1}\right)^2 \cdot \frac{\theta}{(\theta-1)^2(\theta-2)}$$

$$= \left[-\frac{1}{\left(\frac{\theta}{\theta-1}-1\right)^2}\right]^2 \cdot \frac{\theta}{(\theta-1)^2(\theta-2)}$$

$$= (\theta-1)^4 \cdot \frac{\theta}{(\theta-1)^2(\theta-2)} = \frac{\theta(\theta-1)^2}{\theta-2}$$

So to summarize:

$$\sqrt{n}(\hat{\theta}-\theta) \to \mathcal{N}\left(0, \frac{\theta(\theta-1)^2}{\theta-2}\right) \text{ in distribution as } n \to \infty.$$

An asymptotic $(1-\alpha)$-confidence interval for θ is

$$\hat{\theta} \pm z^{(\alpha/2)} \cdot \sqrt{\frac{\hat{\theta}(\hat{\theta}-1)^2}{(\hat{\theta}-2)n}}$$

Asymptotic normality of method of moments estimators

Proposition

Let $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$ *for a single parameter* $\theta \in \mathbb{R}$. *Suppose* $\mathbb{E}_\theta[X] = \mu(\theta)$ *and* $\mathrm{Var}_\theta[X] = v(\theta)$, *and consider the method of moments estimator* $\hat{\theta} = \mu^{-1}(\bar{X})$.

If μ *is continuously differentiable at* θ *and* $\mu'(\theta) \neq 0$, *then*

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}\left(0, \frac{v(\theta)}{\mu'(\theta)^2}\right)$$

in distribution as $n \to \infty$.

Asymptotic normality of method of moments estimators

Proof: By CLT: $\sqrt{n}(\bar{X} - \mu(\theta)) \to \mathcal{N}(0, v(\theta))$

in distribution as $n \to \infty$.

By delta method:

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\mu^{-1}(\bar{X}) - \mu^{-1}(\mu(\theta)))$$

$$\to \mathcal{N}(0, \underbrace{(\mu^{-1})'(\mu(\theta))^2}_{= \frac{1}{\mu'(\mu^{-1}(\mu(\theta)))^2} = \frac{1}{\mu'(\theta)^2}} \cdot v(\theta))$$

$$\Rightarrow \sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}\left(0, \frac{v(\theta)}{\mu'(\theta)^2}\right)$$

Plug-in estimators

Sometimes we are interested in a function $g(\theta)$ of the parameter, rather than θ itself. A natural estimate of $g(\theta)$ is $g(\hat{\theta})$, where $\hat{\theta}$ is our estimate of θ. This is called the **plug-in estimate** of $g(\theta)$.

Example: You play a game where you flip a biased coin. If the coin lands heads, you give your friend \$1. If the coin lands tails, your friend gives you \$$x$. What is the value of x that makes this a fair game?

If the coin lands heads with probability p, then your expected winnings is $p \cdot (-1) + (1 - p) \cdot x$. The game is fair when

$$p \cdot (-1) + (1 - p) \cdot x = 0$$

i.e. when $x = \frac{p}{1-p}$ where $\frac{p}{1-p}$ is the *odds* of getting heads.

Plug-in estimators

The odds function $\frac{p}{1-p}$ is not symmetric about $p = \frac{1}{2}$. It is sometimes easier to interpret the log-odds or *logit*, which is $\log \frac{p}{1-p}$. The log-odds for p is the negative of that for $1 - p$.

To estimate the log-odds from n coin flips

$$X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(p)$$

we may first estimate p by $\hat{p} = \bar{X}$. (This is both the method of moments estimator and the MLE.)

Then the plugin estimate of $\log \frac{p}{1-p}$ is simply $\log \frac{\bar{X}}{1-\bar{X}}$.

Plug-in estimator for the log-odds

Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli(p), and consider the plugin estimate of $\log \frac{p}{1-p}$ given by $\log \frac{\bar{X}}{1-\bar{X}}$.

By CLT: $\sqrt{n}(\bar{X} - p) \to \mathcal{N}(0, p(1-p))$

Apply delta method w/ $g(x) = \log \frac{x}{1-x} = \log x - \log(1-x)$

$$\Rightarrow g'(x) = \frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)}$$

$$\Rightarrow \sqrt{n}\left(\log \frac{\bar{X}}{1-\bar{X}} - \log \frac{p}{1-p}\right) \to \mathcal{N}\Big(0, \underbrace{g'(p)^2}_{\left(\frac{1}{p(1-p)}\right)^2} \cdot p(1-p)\Big)$$

$$= \mathcal{N}\left(0, \frac{1}{p(1-p)}\right)$$

Confidence interval for the log-odds

Suppose we toss this coin $n = 100$ times and observe 60 heads, i.e. $\bar{X} = 0.6$. We would estimate the log-odds by $\log \frac{\bar{X}}{1-\bar{X}} \approx 0.41$.

We may estimate our standard error by $\sqrt{\frac{1}{n\bar{X}(1-\bar{X})}} \approx 0.20$.

An asymptotic $(1-\alpha)$-confidence interval for the log-odds $\log \frac{p}{1-p}$ is then given by $0.41 \pm 0.20 \cdot z^{(\alpha/2)}$.