

Math 4261: Mathematical Statistics I

Lecture 12: Delta method

(References: DS 6.3; many examples credited to Zhou Fan)

Fall 2026

Outline

Recap

Asymptotic normality of estimators

Delta method

Plug-in estimators and standard errors

Outline

Recap

Asymptotic normality of estimators

Delta method

Plug-in estimators and standard errors

Recap: Parameter estimation and MLE

Suppose data $X_1, \dots, X_n$ came from joint distribution with PMF/PDF parametrized by $\theta \in \Omega \subseteq \mathbb{R}^k$:

$$f(x_1, \dots, x_n \mid \theta).$$

Goal: estimate θ . Ω is the parameter space.

An **estimator** $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ is a function of the data. $\hat{\theta}$ is a random variable.

The **likelihood function** and **log-likelihood** are

$$\text{lik}(\theta) = f(X_1, \dots, X_n \mid \theta), \quad \ell_n(\theta) = \log \text{lik}(\theta).$$

The **maximum likelihood estimator** is

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Omega} \ell_n(\theta).$$

MLE $\hat{\theta}$ is the value of θ that **maximizes the joint PMF/PDF at the observed data**.

Recap: Example of Bernoulli MLE

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$, $p \in [0, 1]$.

1. Write the **log-likelihood**: $\ell_n(p) = n\bar{X} \log p + n(1 - \bar{X}) \log(1 - p)$.
2. **Differentiate** ℓ_n and set to zero (for $0 < p < 1$):

$$\ell'_n(p) = \frac{n\bar{X}}{p} - \frac{n(1 - \bar{X})}{1 - p} = 0.$$

3. Solve for the **MLE**: $\hat{p} = \bar{X}$.

Recap: Example of multiple parameters

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Gamma}(\alpha, \lambda)$, $\theta = (\alpha, \lambda) \in \Omega = (0, \infty)^2$.

1. Find the **gradient** of ℓ_n : $\nabla \ell_n = (\partial_\alpha \ell_n, \partial_\lambda \ell_n)^\top$.
2. Set $\nabla \ell_n(\hat{\alpha}, \hat{\lambda}) = 0$. This gives $\hat{\lambda} = \hat{\alpha} / \overline{X}$. where $\hat{\alpha}$ satisfies

$$\log \hat{\alpha} - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \log \overline{X} - \frac{1}{n} \sum_{i=1}^n \log X_i.$$

- ▶ There is **no closed-form solution** for $\hat{\alpha}$.
- ▶ Nevertheless, we showed that a **unique** solution $\hat{\alpha}$ exists.

Recap: Example of non-i.i.d. model

We studied the Hardy–Weinberg equilibrium, which models the genotype counts in a population based on the recessive allele frequency $\theta \in [0, 1]$, as

$$(Y_{AA}, Y_{Aa}, Y_{aa}) \sim \text{Mult}(n, \{(1 - \theta)^2, 2\theta(1 - \theta), \theta^2\}).$$

The counts sum to n and are **dependent**.

The **log-likelihood** is,

$$\ell_n(\theta) = \text{Const.} + (2Y_{AA} + Y_{Aa}) \log(1 - \theta) + (Y_{Aa} + 2Y_{aa}) \log \theta.$$

Differentiating, setting to zero and solving gives the **MLE**

$$\hat{\theta} = \frac{Y_{Aa} + 2Y_{aa}}{2n}.$$

Outline

Recap

Asymptotic normality of estimators

Delta method

Plug-in estimators and standard errors

Large sample properties

In this lecture, we continue to investigate the statistical properties of estimators in large samples. In particular, we 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 estimate functions of parameters using **plug-in estimators**, and approximate their sampling distributions.

Recap: CLT

Let $X_1, \dots, X_n$ be i.i.d., with mean μ and variance $0 < \sigma^2 < \infty$.

Central limit theorem (CLT)

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Equivalently,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- ▶ For large n , $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$ distributed.
- ▶ The LLN says the sampling error vanishes. The CLT describes its **approximately Gaussian shape** and its **$1/\sqrt{n}$ scale**.

Example: Poisson

(Revisited). Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$, where $\lambda > 0$. Recall that the MoM 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

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

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, $\hat{\lambda} \rightarrow \lambda$ in probability as $n \rightarrow \infty$, hence **consistent**. Furthermore, by the CLT

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

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

Consistency and asymptotic normality

Consistency

We say that an estimator $\hat{\theta}$ for θ is **consistent** if

$$\hat{\theta} \rightarrow \theta \quad \text{in probability, as } n \rightarrow \infty.$$

Asymptotic normality

The estimator is **asymptotically normal**, if, furthermore, for some asymptotic variance $v(\theta)$,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)), \quad \text{as } n \rightarrow \infty$$

Implications of asymptotic normality

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

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

then informally, for large n , this tells us the following:

- ▶ $\hat{\theta}$ 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 $\mathcal{N}(\theta, v(\theta)/n)$.

Outline

Recap

Asymptotic normality of estimators

Delta method

Plug-in estimators and standard errors

Motivation

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

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

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

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

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

Delta method

If S is an asymptotically normal statistic and g is a smooth function on an open interval containing its limit, then $g(S)$ will also be asymptotically normal. This is formalized as the **delta method**.

Theorem (Delta method)

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

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

in distribution as $n \rightarrow \infty$, when $X_1, \dots, X_n \stackrel{\text{IID}}{\sim} f(x \mid \theta)$.

If $g : I \rightarrow \mathbb{R}$ is continuously differentiable at $\mu(\theta)$, where I is an open interval containing $\mu(\theta)$, then

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

Proof sketch of the delta method

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

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

Taylor expand

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

Scale by $\sqrt{n}$:

$$\sqrt{n}(g(S(X)) - g(\mu(\theta))) \approx g'(\mu(\theta)) \underbrace{\sqrt{n}(S(X) - \mu(\theta))}_{\xrightarrow{d} N(0, v(\theta))}$$

$$\xrightarrow{d} N(0, g'(\mu(\theta))^2 \cdot v(\theta))$$

"□"

Example: Poisson

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$, where $\lambda > 0$. $\mathbb{E}[X_i] = \lambda$, $\text{Var}[X_i] = \lambda$.
CLT gives

$$\frac{\sqrt{n}}{\sqrt{\lambda}} (\bar{X}_n - \lambda) \xrightarrow{d} \mathcal{N}(0, 1). \quad g'(t) = \frac{1}{\sqrt{t}}$$

Apply the Delta method with $g(t) = 2\sqrt{t}$.

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda)$$

$$\sqrt{n}(g(\bar{X}_n) - g(\lambda)) = \sqrt{n}(2\sqrt{\bar{X}_n} - 2\sqrt{\lambda})$$

i.e.

$$2\sqrt{\bar{X}_n} \sim \mathcal{N}(2\sqrt{\lambda}, \frac{1}{n})$$

approximately

variance independent of λ !

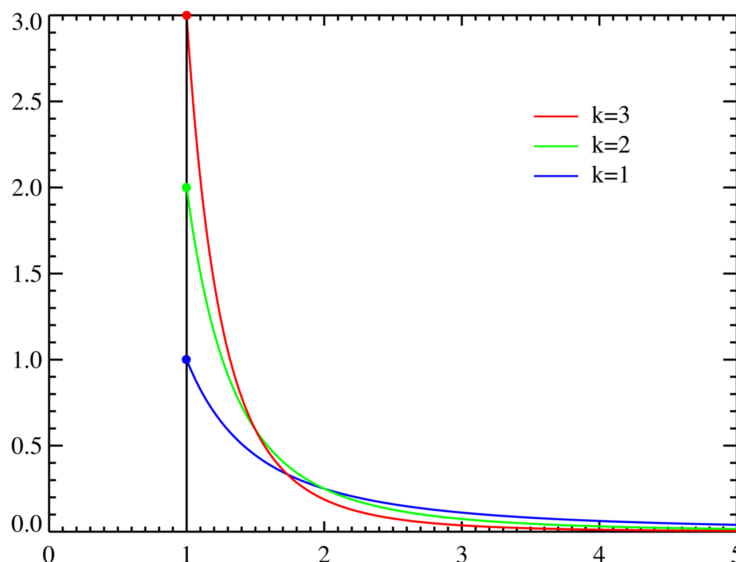
$$\xrightarrow{d} \mathcal{N}(0, \left(\frac{1}{\sqrt{\lambda}}\right)^2 \lambda) = \mathcal{N}(0, 1)$$

Delta method

Example: Pareto

Let $X \sim \text{Pareto}(\theta, 1)$, where $\theta > 0$ is the shape parameter and the lower cutoff is 1.

$$f(x \mid \theta) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & x \geq 1, \\ 0, & x < 1. \end{cases}$$



Density curves: $k = \theta$ in the plot.

- Useful for modeling wealth and quantities with occasional **very large observations**, such as insurance losses.
- A **heavy (power-law) tail**: $\mathbb{P}[X > x] = x^{-\theta}$ for $x \geq 1$. Smaller θ gives a heavier tail.

Example: Pareto

Let $X \sim \text{Pareto}(\theta, 1)$, where $\theta > 1$. Show that its mean is

$$\mathbb{E}_\theta[X] = \frac{\theta}{\theta - 1}.$$

$$\begin{aligned} \mathbb{E}X &= \int_1^\infty x \cdot \frac{\theta}{x^{\theta+1}} dx = \int_1^\infty \frac{\theta}{x^\theta} dx \\ &= \left. \frac{\theta x^{-\theta+1}}{-\theta+1} \right|_1^\infty = 0 - \frac{\theta}{-\theta+1} = \boxed{\frac{\theta}{\theta-1}} \text{ for } \theta > 1 \end{aligned}$$

$$(\mathbb{E}X = \infty \text{ if } \theta \leq 1)$$

Example: Pareto

$$X \sim \text{Pareto}(\theta, 1)$$

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Pareto}(\theta, 1)$, where $\theta > 1$. Derive the method of moments estimator of θ .

$$(1) \quad \mu_1 = E_{\theta} X = \frac{\theta}{\theta - 1}$$

$$(2) \quad \text{Invert} \quad \underbrace{\mu_1(\theta - 1)}_{\mu_1 \theta - \mu_1} = \theta \Rightarrow \mu_1 = \theta(\mu_1 - 1) \\ \Rightarrow \theta = \frac{\mu_1}{\mu_1 - 1}$$

$$(3) \quad \text{MOM} \quad \hat{\theta} = \frac{\bar{X}}{\bar{X} - 1}.$$

Example: Pareto

Show that the method of moments estimator is **consistent** for $\theta > 1$.

For $\theta > 2$, show that it is **asymptotically normal**, and give an approximate sampling distribution. You may use

$$\text{Var}_\theta[X] = \frac{\theta}{(\theta - 1)^2(\theta - 2)}.$$

$$\hat{\theta}_{\text{mom}} = \frac{\bar{X}}{\bar{X} - 1} = g(\bar{X}), \text{ where } g(t) = \frac{t}{t-1}$$

$g(t)$ is continuous for $t > 1$

(it suffices in Cont. mapping theorem that g cont. at the prob. limit of $\bar{X}$, which is $\frac{\theta}{\theta-1}$)

$$\cdot) \quad \bar{X} \xrightarrow{P} \frac{\theta}{\theta-1} \text{ by LLN}$$

$$\cdot) \quad \hat{\theta}_{\text{mom}} = g(\bar{X}) \xrightarrow{P} g\left(\frac{\theta}{\theta-1}\right) = \frac{\frac{\theta}{\theta-1}}{\frac{\theta}{\theta-1} - 1} = \frac{\frac{\theta}{\theta-1}}{\frac{\theta - \theta + 1}{\theta-1}} = \theta$$

Cont. mapping theorem

Example: Pareto Assume $\theta > 2$, $E_\theta X = \frac{\theta}{\theta-1}$, $Var_\theta X = \frac{\theta}{(\theta-1)^2(\theta-2)}$

$$CLT \quad \sqrt{n} \left(\bar{X} - \frac{\theta}{\theta-1} \right) \xrightarrow{d} N(0, \frac{\theta}{(\theta-1)^2(\theta-2)})$$

$$\text{Let } g(t) = \frac{t}{t-1}, \text{ so } g'(t) = -\frac{1}{(t-1)^2}$$

$$\sqrt{n} \left(\hat{\theta}_{MM} - \theta \right) \xrightarrow{d} N(0, \underbrace{g'(\frac{\theta}{\theta-1})^2}_{(*)} Var_\theta X)$$

$$g'(\frac{\theta}{\theta-1}) = -\frac{1}{(\frac{\theta}{\theta-1}-1)^2} = -(\theta-1)^2$$

Asymptotic Variance of $\hat{\theta}_{MM}$:

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

Thus

$$\hat{\theta}_{MM} \sim N(\theta, \frac{\theta(\theta-1)^2}{n(\theta-2)})$$

approximately

Asymptotic normality of MoM

Proposition

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x \mid \theta)$ for a single parameter $\theta \in \Omega \subseteq \mathbb{R}$. Let $\mathbb{E}_\theta[X] = \mu(\theta)$ and $\text{Var}_\theta[X] = v(\theta) < \infty$. Suppose

- ▶ μ is **one-to-one**
- ▶ μ is continuously differentiable in a neighborhood of θ , with $\mu'(\theta) \neq 0$.

Then the MoM estimator $\hat{\theta} = \mu^{-1}(\bar{X})$ is **asymptotically normal** and satisfies

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

Asymptotic normality of MoM

Proof: Write $h = \mu^{-1}$, so $\hat{\theta} = h(\bar{X})$

By CLT, $\sqrt{n}(\bar{X} - \mu(0)) \xrightarrow{d} N(0, v(0))$

Delta method

$$\sqrt{n}(h(\bar{X}) - h(\mu(0))) \xrightarrow{d} N(0, h'(\mu(0))^2 v(0))$$

$\parallel$ $\parallel$
 $\hat{\theta}$ 0

$$h(\mu(0)) = 0 \xRightarrow{\text{chain rule}} h'(\mu(0)) \mu'(0) = 1$$
$$\Rightarrow h'(\mu(0)) = \frac{1}{\mu'(0)}$$

Outline

Recap

Asymptotic normality of estimators

Delta method

Plug-in estimators and standard errors

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: Coin log odds

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 $x = p/(1 - p)$ where $p/(1 - p)$ is the **odds** of getting heads.

Odds of heads to tails 10 : 1

means in 11 games, expect 10 heads, 1 tails

Example: Coin log odds

The odds function $p/(1-p)$ is not symmetric about $p = 1/2$. It is sometimes easier to interpret the **log-odds** or *logit*, which is $\log(p/(1-p))$.

To estimate the log-odds from n coin flips

$$X_1, \dots, X_n \stackrel{\text{IID}}{\sim} \text{Bernoulli}(p), \quad 0 < p < 1,$$

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

Then the **plug-in estimate** of $\log(p/(1-p))$ is simply

$$\log\left(\frac{\bar{X}}{1 - \bar{X}}\right), \quad 0 < \bar{X} < 1.$$

Example: Coin log odds

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$, where $0 < p < 1$, and consider the plug-in estimate of $\log(p/(1-p))$ given by $\log(\bar{X}/(1-\bar{X}))$, for a sample with $0 < \bar{X} < 1$. *How accurate is this?*

$$\text{By CLT, } \sqrt{n}(\bar{X} - p) \xrightarrow{d} N(0, p(1-p))$$

$$\text{Let } g(t) = \log \frac{t}{1-t} = \log t - \log(1-t)$$

$$g'(t) = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)}$$

$$\text{Delta method } \sqrt{n} \left(\log \frac{\bar{X}}{1-\bar{X}} - \log \frac{p}{1-p} \right) \xrightarrow{d} N(0, \underbrace{g'(p)^2 \cdot p(1-p)}_{(*)})$$

$$(*) = \frac{1}{p(1-p)^2 p(1-p)} = \frac{1}{p(1-p)} = N(0, \frac{1}{p(1-p)})$$

$$\text{Std error of } \log \frac{\bar{X}}{1-\bar{X}} \approx \frac{1}{\sqrt{n p(1-p)}}, \text{ estimate by } \frac{1}{\sqrt{n \bar{X}(1-\bar{X})}}.$$

Example: Coin 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

$$\log\left(\frac{\bar{X}}{1 - \bar{X}}\right) \approx 0.41.$$

We may estimate our **standard error** by

$$\sqrt{\frac{1}{n\bar{X}(1 - \bar{X})}} \approx 0.20.$$

