Chapter 25

Properties of MLEs; Confidence Intervals

Lecture 25 | Devore 9e: 6.2, 7.1

25.1 More maximum likelihood

25.1.1 Several parameters at once

When there are m unknown parameters, set all m partial derivatives of the log-likelihood to zero and solve the system.

Example 25.1 (Normal, both parameters unknown)**.** $X_1, \ldots, X_n$ i.i.d. Normal(μ, σ^2). With $v = \sigma^2$,

$$\ell(\mu, v) = -\frac{n}{2}\log(2\pi v) - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Then

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{v}\sum_i (x_i - \mu) = 0 \implies \hat{\mu} = \bar{x},$$

$$\frac{\partial \ell}{\partial v} = -\frac{n}{2v} + \frac{\sum_i (x_i - \mu)^2}{2v^2} = 0 \implies \hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \hat{\mu})^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2.$$

So the MLE of σ^2 is the *biased* sample variance, $\frac{n-1}{n}S^2$.

25.1.2 Three properties worth knowing

Theorem 25.2 (Invariance)**.** *If $\hat{\theta}$ is the MLE of θ and g is any function, then $g(\hat{\theta})$ is the MLE of $g(\theta)$.*

This is a genuine convenience: having found $\hat{\sigma}^2$, the MLE of σ is $\sqrt{\hat{\sigma}^2}$ — no new optimization. (Note that unbiasedness has no such property: $\mathbb{E}[\hat{\theta}] = \theta$ does *not* give $\mathbb{E}[g(\hat{\theta})] = g(\theta)$.)

Theorem 25.3 (Large-sample behavior)**.** *Under mild regularity conditions, as $n \to \infty$ the MLE is*

- ***consistent**: $\hat{\theta} \to \theta$;*
- ***asymptotically unbiased**: $\mathbb{E}[\hat{\theta}] \to \theta$;*
- ***asymptotically normal and efficient**: no other estimator has smaller variance in the limit.*

Remark 25.4 (MLE versus method of moments)**.**

	Method of moments	**Maximum likelihood**
effort	solve algebraic equations	maximize a function
uses	the first m moments only	the whole density
guarantees	few	consistent, asymptotically efficient
risk	can leave the parameter space	usually well behaved

Method of moments is a good way to get a starting value; maximum likelihood is what you report.

25.2 From point estimates to intervals

A point estimate is a single number and is therefore almost surely wrong. Reporting $\hat{\mu} = 8.3$ says nothing about whether the truth is plausibly 8.2 or plausibly 40. What we want is an interval, together with a statement of how often that interval works.

Definition 25.5 (Confidence interval)**.** A $100(1-\alpha)\%$ **confidence interval** for θ is a pair of statistics $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ with

$$\mathbb{P}\{L \le \theta \le U\} = 1 - \alpha.$$

25.2.1 The normal mean, σ known

Let $X_1, \ldots, X_n$ be i.i.d. Normal(μ, σ^2) with σ known. From Lecture 21, $\bar{X} \sim \text{Normal}(\mu, \sigma^2/n)$ exactly, so

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1).$$

Let $z_{\alpha/2}$ be the value with $\mathbb{P}\{Z > z_{\alpha/2}\} = \alpha/2$. By symmetry,

$$\mathbb{P}\left\{-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} = 1 - \alpha.$$

Now solve the inequality for μ — this is the only real step:

$$\boxed{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;\le\; \mu \;\le\; \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

confidence	α	$z_{\alpha/2}$	interval
90%	0.10	1.645	$\bar{x} \pm 1.645\,\sigma/\sqrt{n}$
95%	0.05	1.96	$\bar{x} \pm 1.96\,\sigma/\sqrt{n}$
99%	0.01	2.576	$\bar{x} \pm 2.576\,\sigma/\sqrt{n}$

Example 25.6. A machine fills bottles with $\sigma = 0.5$ ml. A sample of $n = 25$ bottles has $\bar{x} = 500.2$ ml. A 95% confidence interval for μ is

$$500.2 \pm 1.96 \cdot \frac{0.5}{\sqrt{25}} = 500.2 \pm 0.196 = [500.00,\ 500.40].$$

Remark 25.7 (What "95% confident" means — and does not)**.** μ is a fixed unknown number; it is not random. The *interval* is random, because $\bar{X}$ is. The correct reading is:

if we repeated the whole experiment many times, about 95% *of the intervals so constructed would contain* μ*.*

It is **not** correct to say "μ has a 95% chance of lying in $[500.00, 500.40]$" — once the data is in, either it does or it does not.

Remark 25.8 (Width)**.** The width is $2z_{\alpha/2}\sigma/\sqrt{n}$. So:

- more confidence $\Rightarrow$ wider interval (you cannot get both);
- to halve the width, take 4 times as many observations — the $\sqrt{n}$ law again.

25.2.2 Large samples, σ unknown

In practice σ is unknown too. For large n, replace it by S and appeal to the CLT: for *any* underlying distribution with finite variance,

$$\mu \ \approx\ \bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (n \text{ large}).$$

For small n from a normal population, $z_{\alpha/2}$ must be replaced by a t-distribution quantile — which is the starting point of STAT401.

Remark 25.9. Everything in this course has been building to this sentence: *data* $\rightarrow$ *estimator* $\rightarrow$ *sampling distribution* $\rightarrow$ *interval with a guarantee*. The probability half of the course exists precisely to make the third arrow possible.