Parameter estimation

Lecture 13a (STAT 24400 F24)

A parametric model

Suppose that we observe data generated from
a *known* distribution with *unknown* parameter(s).

For example,

- The data is $\mathrm{N}(\mu, \sigma^2)$, with μ unknown (and σ^2 known).
- The data is $\mathrm{N}(\mu, \sigma^2)$, with μ and σ^2 unknown.
- The data is Exponential(λ), with λ unknown.
- The data is Binomial(n, p), with n known and p unknown.

How can we estimate the unknown parameter(s)?

How can we perform inference on the unknown parameter(s)?

(Inference includes confidence intervals, credible interval, hypothesis testing, etc.)

Parametric model notations

General notation:

- $X_1, \dots, X_n$ = data drawn i.i.d. from the distribution
- θ = the unknown parameter(s)
- θ lies in Θ = subspace of $\mathbb{R}$ (or $\mathbb{R}^2$ if two parameters, etc)
- We will write $f(x \mid \theta)$ for the density or PMF of the distribution

  e.g., Exponential(λ) $\rightsquigarrow$ density $f(x \mid \lambda) = \lambda e^{-\lambda x}$

  Poisson(λ) $\rightsquigarrow$ PMF $f(x \mid \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

  - If frequentist, θ is non-random — not a conditional density/PMF
  - If Bayesian, θ is random — it's a conditional density/PMF

Parameter estimation

Given data $X_1, \dots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$, would like to estimate the unknown θ

An estimator $\widehat{\theta}$ is any map from $\mathbb{R}^n$ to Θ,
mapping the data $(X_1, \dots, X_n)$ to an estimate $\widehat{\theta}$ of θ

Some familiar examples:

- $\mathrm{N}(\mu, \sigma^2)$, with μ unknown (& σ^2 known) $\rightsquigarrow \widehat{\mu} = \bar{X}$
- $\mathrm{N}(\mu, \sigma^2)$, with μ & σ^2 unknown $\rightsquigarrow (\widehat{\mu}, \widehat{\sigma}^2) = (\bar{X}, S^2)$
- Binomial(n, p), with n known and p unknown

  $\rightsquigarrow \widehat{p} = \bar{X}$ = the observed fraction of Heads

Note: any estimator $\widehat{\theta}$ must be a function of $X_1, \dots, X_n$ only
(it cannot depend on the true value θ since that's unknown)

Unbiasedness of an estimator $\hat{\theta}$

An estimator $\hat{\theta}$ is **unbiased** if $\mathbb{E}(\hat{\theta}) = \theta$, for all values $\theta \in \Theta$.

treat θ as fixed, and draw $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$

For example, for $\mathsf{N}(\mu, \sigma^2)$ with μ unknown (& σ^2 known):

- $\widehat{\mu} = \bar{X}$ is unbiased, since $\mathbb{E}(\widehat{\mu}) = \mathbb{E}(\bar{X}) = \mu$ for any value of μ
- $\widehat{\mu} = 2\bar{X} - 1$ is not unbiased: $\mathbb{E}(\widehat{\mu}) = 2\mu - 1 \;\rightsquigarrow\; \mathbb{E}(\widehat{\mu}) = \mu$ only if $\mu = 1$

If θ is multidimensional, $\widehat{\theta}$ is unbiased if each of its components is unbiased.

For example, for $\mathsf{N}(\mu, \sigma^2)$ with μ & σ^2 unknown:

- $(\widehat{\mu}, \widehat{\sigma}^2) = (\bar{X}, S^2)$ is unbiased, since $\mathbb{E}(\widehat{\mu}) = \mathbb{E}(\bar{X}) = \mu$ and $\mathbb{E}(\widehat{\sigma}^2) = \mathbb{E}(S^2) = \sigma^2$, for any μ, σ^2

Sampling distribution and standard error of estimator $\widehat{\theta}$

The **sampling distribution** is the distribution of $\widehat{\theta}$ (since $\widehat{\theta}$ is a function of the sample)

treat θ as fixed, and draw $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$

(the sampling distribution generally depends on the unknown θ)

Standard error (SE) refers to any estimate of the standard deviation of $\widehat{\theta}$

(true SD may depend on θ, while SE depends on the data but not on θ)

Parameter estimation (normal mean)

Example: normal sampling — $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$

- If $\theta = \mu$ is unknown while σ^2 is known:
  - The family of densities is
  $$f(x \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$
  - Our usual estimator is $\widehat{\theta} = \bar{X}$
  - It is unbiased since $\mathbb{E}(\bar{X}) = \mu$
  - $\mathsf{SD}(\bar{X}) = \mathsf{SE}(\bar{X}) = \sigma/\sqrt{n}$.
  - Its sampling distribution is $\bar{X} \sim \mathsf{N}(\mu, \sigma^2/n)$

Parameter estimation (normal mean & variance)

Example: normal sampling — $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$

- If $\theta = (\mu, \sigma^2)$:
  - The family of densities is
  $$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$
  - Our usual estimators are $\widehat{\theta} = (\bar{X}, S^2)$.
  - It's unbiased since $\mathbb{E}(\bar{X}) = \mu$ and $\mathbb{E}(S^2) = \sigma^2$.
  - Joint sampling distribution:
  $\bar{X} \sim \mathsf{N}(\mu, \frac{\sigma^2}{n})$, $\quad \frac{n-1}{\sigma^2} \cdot S^2 \sim \chi^2_{n-1}$, $\quad \bar{X} \perp\!\!\!\perp S^2$
  - $\mathsf{SD}(\bar{X})=\sigma/\sqrt{n}$, estimated via $\mathsf{SE}(\bar{X}) = S/\sqrt{n}$
  - We have not computed $\mathsf{SE}(S^2)$ (relates to the χ^2 distrib.)

Goodness of estimator, MSE

For a one-dimensional θ and an estimator $\widehat{\theta}$,

$$\text{MSE} = \mathbb{E}((\widehat{\theta} - \theta)^2)$$

treat θ as fixed, and draw $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$

- MSE generally depends on the unknown θ
- If $\widehat{\theta}$ is unbiased, then $\mathbb{E}(\widehat{\theta}) = \theta \Rightarrow \text{MSE} = \text{Var}(\widehat{\theta})$
- In general:

$$\text{MSE} = \mathbb{E}\Big(\big(\underbrace{(\mathbb{E}(\widehat{\theta}) - \theta)}_{=\text{ constant}} + (\widehat{\theta} - \mathbb{E}(\widehat{\theta})) \big)^2 \Big)$$

$$= (\mathbb{E}(\widehat{\theta}) - \theta)^2 + \underbrace{\mathbb{E}\Big((\widehat{\theta} - \mathbb{E}(\widehat{\theta}))^2 \Big)}_{=\text{Var}(\widehat{\theta})} + 2(\mathbb{E}(\widehat{\theta}) - \theta) \cdot \underbrace{\mathbb{E}(\widehat{\theta} - \mathbb{E}(\widehat{\theta}))}_{=0}$$

$$= \big(\underbrace{\mathbb{E}(\widehat{\theta}) - \theta}_{=\text{ bias}} \big)^2 + \text{Var}(\widehat{\theta})$$

The bias-variance tradeoff

$$\text{MSE} = \big(\underbrace{\mathbb{E}(\widehat{\theta}) - \theta}_{=\text{ bias}} \big)^2 + \text{Var}(\widehat{\theta})$$

The bias/variance tradeoff:
sometimes we can reduce one term at the cost of increasing the other.

For normal data:

- The sample mean $\widehat{\mu} = \bar{X}$ has zero bias, and variance $= \sigma^2/n$
- The estimator $\widehat{\mu} \equiv 0$ has high bias (for most μ's), and variance $= 0$

Method of moments - outline

The **method of moments** (MoM) is a strategy for constructing an estimator $\widehat{\theta}$.

If θ is one-dimensional:

- Compute the population mean $\mathbb{E}(X)$ as a function of θ
- Compute the sample mean $\bar{X} = \frac{1}{n}\sum_i X_i$
- Choose $\widehat{\theta}$ as the value of θ so that $\mathbb{E}(X) = \bar{X}$

If θ is k-dimensional:

- Compute population moments $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, ..., $\mathbb{E}(X^k)$ as functions of θ
- Compute the sample moments $\frac{1}{n}\sum_i X_i$, $\frac{1}{n}\sum_i X_i^2$, ..., $\frac{1}{n}\sum_i X_i^k$
- Choose $\widehat{\theta}$ as the value of θ so that $\mathbb{E}(X^j) = \frac{1}{n}\sum_i X_i^j$ for all j
(solving a system of k equations, for a k-dimensional unknown)

The method is better illustrated by examples.

Method of moments example: Normal mean (σ^2 is known)

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{N}(\mu, \sigma^2)$ for unknown $\mu \in \mathbb{R}$ (σ^2 is known.

Goal: Construct an estimator $\hat{\mu}$ for parameter μ using Method of moments.

- The density is

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

- For $X \sim f(\cdot \mid \mu)$, $\mathbb{E}(X) = \mu$
- Apply MoM to calculate $\widehat{\mu}$:

$$\bar{X} = \widehat{\mu}$$

Method of moments example: Normal μ, σ^2

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ for unknown $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Goal: Construct estimators $\hat{\mu}, \hat{\sigma}^2$ for parameters μ, σ^2 using MoM.

- The density is $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$
- For $X \sim f(\cdot \mid \mu, \sigma^2)$, $\mathbb{E}(X) = \mu$ and $\mathbb{E}(X^2) = \mu^2 + \sigma^2$
- Apply MoM to calculate $\widehat{\mu}, \widehat{\sigma}^2$:

$$\bar{X} = \widehat{\mu}$$

$$\frac{1}{n}\sum_i X_i^2 = \widehat{\mu}^2 + \widehat{\sigma}^2$$

compare to $\frac{1}{n-1}$ for S^2

$$\Rightarrow \quad \widehat{\mu} = \bar{X} \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_i X_i^2 - \widehat{\mu}^2 = \frac{1}{n}\sum_i X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$$

Method of moments example: Exponential

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$ for unknown $\lambda > 0$

- The density is

$$f(x \mid \lambda) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{x>0}$$

- For $X \sim f(\cdot \mid \lambda)$, $\mathbb{E}(X) = 1/\lambda$
- Apply MoM to calculate $\widehat{\lambda}$:

$$\bar{X} = \frac{1}{\widehat{\lambda}} \quad \Rightarrow \quad \widehat{\lambda} = \frac{1}{\bar{X}}.$$

Method of moments example: Uniform

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}[0, \theta]$ for unknown $\theta > 0$

- The density is

$$f(x \mid \theta) = \frac{1}{\theta} \cdot \mathbb{1}_{0 \le x \le \theta}$$

- For $X \sim f(\cdot \mid \theta)$,

$$\mathbb{E}(X) = \int_{x=0}^{\theta} x \cdot \frac{1}{\theta}\, \mathrm{d}\theta = \frac{\theta}{2}.$$

- Apply method of moments to calculate $\widehat{\theta}$:

$$\bar{X} = \frac{\widehat{\theta}}{2} \quad \Rightarrow \quad \widehat{\theta} = 2\bar{X}.$$