Lecture 12
Parametric Inference

Data $D = \{X_1, X_2, \ldots, X_n\}$

$X_i \overset{iid}{\sim} F(x;\theta)$ parameter(s) unknown.

from a known family

Ex: (1) $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ both unknown

(2) $X_i \overset{iid}{\sim} N(\mu, 10^2)$ only μ unknown

(3) $X_i \overset{iid}{\sim} Exp(\lambda)$ unknown.

Estimation Methods

1. Method of Moments Estimation (MoM or MME)
2. Maximum Likelihood " (MLE)

1. Method of Moments Estimation & " " " Estimators

Suppose $D = \{X_1, X_2, \ldots, X_n\}$

$X_i \overset{iid}{\sim} F(x;\vec{\theta})$ where $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ is a vector of m unknown parameters

Then MME estimators can be found as follows: ②

Step ⓪: Identify the m unknown parameters

" 1: For each $k=1,\dots,m$, find k^{th} moment of $X \sim F$ as $\mu_k = E[X^k]$

" 2: For each $k=1,\dots,m$, find k^{th} sample moment (based on data) as

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

" 3: Set $\hat{\mu}_k = \mu_k$ for $k=1,2,\dots,m$ & solve for θ_i's, $i=1,2,\dots,m$

i.e. set

$$\hat{\mu}_1 = \mu_1$$

$$\hat{\mu}_2 = \mu_2$$

& μ_i's are functions of θ_i

$$\vdots$$

$$\hat{\mu}_m = \mu_m$$

So, we will have m equations in m unknowns.

- How to compute μ_k's? (i.e. k^{th} moments)

Recall $E[X^k] = \begin{cases} \sum_{x \in \mathcal{X}} x^k p_X(x) & \text{if } X \text{ is discrete} \\ \int x^k f_X(x)\,dx & \text{" " " cts} \end{cases}$

So, $\boxed{\text{MME method boils down to setting } E[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{ for } k=1,\dots,m \text{ \& solve for } \theta_i\text{'s.}}$

Ex 1) $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$. Find mme $\hat{p}$.

Here, $m=1$, p is the only parameter, so we set

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$p = \overline{X}_n \implies \hat{p}_{mme} = \overline{X}_n$$

Ex 2) $X_1, \dots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, mme $\hat{\lambda}$?

$m=1$, λ is the parameter

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\lambda = \overline{X}_n \implies \hat{\lambda}_{mme} = \overline{X}_n$$

e.g. if $D = \{0.7, 21.3, 39.4\}$ then $\hat{\lambda}_{mme} = \frac{0.7+21.3+39.4}{3} = \frac{61.4}{3} \approx 20.47$

Ex 3) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\hat{\mu}_{mme}$? $\hat{\sigma}^2_{mme}$?

$m=2$, μ & σ^2 are the parameters

$$\left\{ \begin{aligned} E[X] &= \overline{X}_n \\ E[X^2] &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 \end{aligned} \right\} \quad (**)$$

$$\Rightarrow \quad E[X] = \mu$$

$$\& \quad E[X^2] = Var(X) + (E[X])^2 \quad \text{(why?)}$$

$$= \sigma^2 + \mu^2$$

So, the two eq'ns in (*) become

$$\left.\begin{aligned} \mu &= \overline{X}_n \\ \sigma^2+\mu^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{aligned}\right\} \Rightarrow \begin{aligned} \mu &= \overline{X}_n \\ \sigma^2 &= \frac{1}{n}\sum X_i^2 - \mu^2 \\ &= \frac{1}{n}\sum X_i^2 - \overline{X}_n^2 \end{aligned}$$

$$\Rightarrow \quad \hat{\mu}_{mme} = \overline{X}_n$$

$$\hat{\sigma}^2_{mme} = \frac{1}{n}\sum X_i^2 - \overline{X}_n^2 = \frac{1}{n}\sum (X_i - \overline{X}_n)^2.$$

For example if $D = \{0.9, 1.9, 3.8, 0.2, 7.1\}$ is a random sample from $N(\mu, \sigma^2)$,

$$\hat{\mu}_{mme} = \frac{0.9+1.9+\cdots+7.1}{5} = \frac{13.9}{5} = 2.78$$

$$\sum x_i^2 = 69.31$$

So $\hat{\sigma}^2_{mme} = \frac{1}{n}\sum x_i^2 - \overline{x}_n^2 = \frac{1}{5}\cdot 69.31 - 7.7284 \cong 6.13$

<u>Ex 4)</u> $X_1, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(a,b)$ $\quad \hat{a}_{mme}$? $\hat{b}_{mme}$?

Also compute the MME estimates for the data in Ex 3). (Exercise).

Ex 5) $X = \begin{cases} 1 & \text{w.p. } \theta \\ 2 & \text{w.p. } 1-\theta \end{cases}$ $\hat{\theta}_{mme}$?

Sol'n: So, X has pmf $P_X(a) = \begin{cases} \theta & a=1 \\ 1-\theta & a=2 \\ 0 & \text{o.w.} \end{cases}$ and $\mathcal{X} = \{1,2\}$

$m=1$, the only parameter is θ

$$E[X] = \overline{X}_n$$

where $E[X] = \sum_{a \in \mathcal{X}} a\, P_X(a) = 1\,\theta + 2(1-\theta)$
$= \theta + 2 - 2\theta = 2 - \theta$

$\Rightarrow 2 - \theta = \overline{X}_n \Rightarrow \theta = 2 - \overline{X}_n$

i.e. $\underline{\hat{\theta}_{mme} = 2 - \overline{X}_n}$

For example if $D = \{2,1,1,1,2\}$

then, $\overline{X}_n = \dfrac{2+1+1+1+2}{5} = \dfrac{7}{5} = 1.4$

So, $\hat{\theta}_{mme} = 2 - 1.4 = \underline{0.6}$

• In Ex 5) $Var(X) = E[X^2] - (E[X])^2$

where $E[X^2] = \sum_{a \in \mathcal{X}} a^2 P_X(a) = 1^2\theta + 2^2(1-\theta)$
$= \theta + 4(1-\theta)$
$= 4 - 3\theta$

$\Rightarrow Var(X) = 4 - 3\theta - (2-\theta)^2$
$= \dots = \theta(1-\theta)$

Hence $Var(\hat{\theta}_{mme}) = Var(2 - \overline{X}_n) = Var(\overline{X}_n)$
$\overset{iid}{=} \dfrac{Var(X_1)}{n} = \dfrac{\theta(1-\theta)}{n}$

Therefore, $se(\hat{\theta}_{mme}) = \sqrt{Var(\hat{\theta}_{mme})}$

$$= \sqrt{\frac{\theta(1-\theta)}{n}}$$

- Note that (i) $se(\hat{\theta}_{mme}) = \sqrt{\frac{\theta(1-\theta)}{n}} \longrightarrow 0$ as $n \to \infty$

& (ii) $E[\hat{\theta}_{mme}] = E[2-\bar{X}_n] = 2 - E[\bar{X}_n]$

$$= 2 - E[X_1]$$

$$= 2 - (2-\theta) = \theta.$$

So, $\hat{\theta}_{mme}$ is unbiased.

Combining (i) & (ii) $MSE(\hat{\theta}_{mme}) = Var(\hat{\theta}_{mme}) \longrightarrow 0$ as $n \to \infty$.

Thus, <u>$\hat{\theta}_{mme}$ is a consistent estimator for θ.</u>

- Hence for the data D in Ex 5)

$Var(\hat{\theta}_{mme}) = \frac{\theta(1-\theta)}{n} \Rightarrow \widehat{Var}(\hat{\theta}_{mme}) = \frac{\hat{\theta}_{mme}(1-\hat{\theta}_{mme})}{n}$

$$= \frac{0.6\,(1-0.6)}{5} = \underline{\underline{0.048}}$$

& $\widehat{se}(\hat{\theta}_{mme}) \approx \sqrt{0.048} \approx \underline{\underline{0.22}}$

<u>Caveat:</u> Here $n=5$ may not be large enough!

<u>Properties of MMEs</u>

1) MME is consistent (i.e. $\hat{\theta}_{mme} \xrightarrow{P} \theta$ so $\theta \approx \hat{\theta}_{mme}$) for large n

2) MME is asymptotically normal, i.e. $\hat{\theta}_{mme} \overset{approx}{\sim} N(\theta, se^2)$

That is $\hat{\theta}_{mme} \overset{approx}{\sim} N(\theta, \hat{se}^2(\hat{\theta}_{mme}))$

$\Rightarrow$ $(1-\alpha)$ CI for θ would be

$$\hat{\theta}_{mme} \pm z_{\alpha/2}\, \hat{se}(\hat{\theta}_{mme})$$

(Normal-Based $(1-\alpha)100\%$ CI)

2) Maximum Likelihood Estimation & MLE Estimators

Data $D = \{X_1, X_2, \ldots, X_n\}$ $\quad X_i \overset{iid}{\sim} F(x; \vec{\theta})$

F ↑ known, $\vec{\theta}$ ↑ unknown.

The MLE tries to find $\vec{\theta}$ that best fits the data or explains the data "best", in the sense that

$$\boxed{\hat{\vec{\theta}}_{MLE} \text{ is the value that } \underline{\text{maximizes}} \text{ "Prob(observing the data set } D\text{)".}}$$

= joint pdf or pmf at the observed data

For $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_X$ with pmf P_X, (i.e. X_i is discrete)

joint pmf of $X_1, X_2, \ldots, X_n$ is

$$P_{\vec{X}}(x_1, x_2, \ldots, x_n; \vec{\theta}) = P_X(x_1;\vec{\theta})P_X(x_2;\vec{\theta}) \cdots P_X(x_n;\vec{\theta}) \text{ (by iid property)}$$

$$= \prod_{i=1}^{n} P_X(x_i;\vec{\theta}).$$

This joint pmf is called likelihood function (when viewed as a function of $\vec{\theta}$)

Similarly if $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_X$
with pdf $f_X(x;\vec{\theta})$ (i.e. X is cts)
then likelihood is

$$L(\vec{\theta}) = \prod_{i=1}^{n} f_X(x_i;\vec{\theta})$$

Notice that likelihood is not a probability or pmf or pdf as a function of $\vec{\theta}$ ($\vec{\theta}$ is not a RV.)

Hence $\hat{\theta}_{MLE}$ <u>maximizes the likelihood function</u> (over $\vec{\theta}$ values).

(Hence the name "<u>Maximum likelihood</u>", b/c likelihood function is maximized)

<u>Ex:</u> $D = \{X_1, X_2, \ldots, X_n\}$, $X_i \overset{iid}{\sim} Exp(\lambda)$ ← unknown.

Then likelihood fcn is

$$L(\lambda) = \prod_{i=1}^{n} f_X(x_i;\lambda)$$

$$= \prod_{i=1}^{n} \frac{1}{\lambda} e^{-x_i/\lambda} = \frac{1}{\lambda^n} e^{-\sum x_i/\lambda}$$

So, <u>$\hat{\lambda}_{MLE}$ maximizes the above function.</u>

- How to find λ that maximizes $L(\lambda)$?

Can use <u>calculus</u> if $L(\lambda)$ is <u>differentiable</u>.

Find $\frac{dL(\lambda)}{d\lambda}$ & set it equal to 0 & solve for λ

In general this is a difficult task b/c of product or reciprocals or exponentials.

Fact: If $f(x)$ is positive differentiable function, then the extrema (max & min) of $f(x)$ & $\log f(x)$ occur at the same x value(s).

Pf: Let $l(x) = \log f(x)$.

Maximum (or Minimum) of $f(x)$ occurs at x for which $f'(x) = 0$ ... (1)

Maximum (or Minimum) of $l(x)$ occurs at x for which $l'(x) = 0$

But $l'(x) = (\log f(x))' = \dfrac{f'(x)}{f(x)} = 0 \Leftrightarrow f'(x) = 0$ (2) (since $f(x) > 0$)

From (1) & (2), $f(x)$ & $l(x)$ are maximized (or minimized) at the same x values.

- Thus, to find λ that maximizes $L(\lambda)$, we can find λ that maximizes $l(\lambda) = \log L(\lambda)$.

Hence, $l(\lambda) = \log L(\lambda) = \log\left(\lambda^{-n} e^{-\sum x_i/\lambda}\right)$

$$= -n \log \lambda - \frac{\sum x_i}{\lambda}$$

$$\Rightarrow l'(\lambda) = \frac{-n}{\lambda} + \frac{\sum x_i}{\lambda^2} = 0 \Rightarrow \frac{n}{\cancel{\lambda}} = \frac{\sum x_i}{\lambda^{\cancel{2}}} \Rightarrow \lambda = \frac{\sum x_i}{n}$$

$$\Rightarrow \hat{\lambda}_{MLE} = \overline{X}_n$$

(That is, $L(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda)$ is maximized when $\lambda = \overline{X}_n$)

- Taking the logarithm to find maximizers of $L(\vec{\theta})$ works in general as long as $L(\vec{\theta})$ is differentiable (with respect to $\vec{\theta}$).
- One also need to show $l''(\hat{\lambda}_{MLE}) < 0$ to verify $\hat{\lambda}_{MLE}$ is the maximizer.

Ex: $X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ Q: $\hat{p}_{MLE} = ?$

$$L(p) = \prod_{i=1}^{n} P_X(x_i)$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$
$$= p^{\sum x_i}(1-p)^{\sum(1-x_i)} = \underbrace{p^{\sum x_i}(1-p)^{n-\sum x_i}}_{\text{find } p \text{ that maximizes this function.}}$$

As before, take logarithm of $L(p)$,

$$\ell(p) = \log L(p) = \sum x_i \log p + (n - \sum x_i)\log(1-p)$$

So, $$\ell'(p) = \frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p} = 0$$

$$\Rightarrow \dots \Rightarrow \hat{p}_{MLE} = \frac{\sum x_i}{n} = \bar{X}_n.$$

Note: • Recall from Calculus that $f'(x)=0$ gives you extrema (i.e. where $f(x)$ is maximum (largest) or minimum (smallest)),

so, to determine the type we apply second derivative test.

If $f'(a)=0$ & $f''(a) > 0$ then f has minimum at $x=a$.
"   "   " < 0 "   "   " <u>maximum</u> "   "

• So, in the above example, we need to find

$$\ell''(p) = \frac{-\sum x_i}{p^2} - \frac{n-\sum x_i}{(1-p)^2} \Rightarrow \ell''(\hat{p}_{MLE}) = \ell''(\bar{X}_n) =$$

substituting $\sum X_i = n\bar{X}$

$$\Rightarrow l''(\bar{X}) = \frac{-n\bar{X}}{\bar{X}^2} - \frac{n-n\bar{X}}{(1-\bar{X})^2} = \dots = \frac{-n}{\bar{X}(1-\bar{X})} < 0 \quad \text{since } 0<\bar{X}<1.$$

Hence, $\hat{p}_{MLE} = \bar{X}$ is maximum (i.e. $L(p)$ is largest at $p=\bar{X}$).

What if $L(\theta)$ is not differentiable?

Ex: $D = \{X_1, X_2, \dots, X_n\}$ $\quad X_i \overset{iid}{\sim} \text{Uniform}(0,\theta)$ $\quad \hat{\theta}_{MLE} = ?$

pdf: $f_X(x_i;\theta) = \begin{cases} \frac{1}{\theta} & 0<x_i<\theta \\ 0 & \text{o.w.} \end{cases} = \frac{1}{\theta} I(0<x_i<\theta)$

so $L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) =$

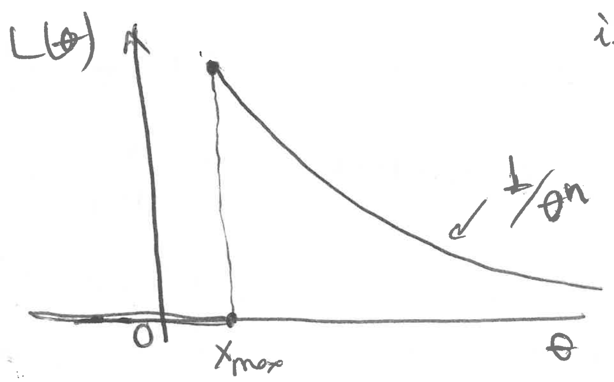
$$= \prod_{i=1}^{n} \frac{1}{\theta} I(0<X_i<\theta)$$

$$= \frac{1}{\theta^n} \underbrace{I(0<X_1<\theta)\, I(0<X_2<\theta) \dots I(0<X_n<\theta)}_{\text{product} = 1 \text{ iff all indicators} = 1}$$

$\Leftrightarrow 0 < \text{all } X_i < \theta \Leftrightarrow X_{max} < \theta$ and $0 < X_{min}$

$$= \frac{1}{\theta^n} I(0<X_{min})\, I(X_{max}<\theta) = \frac{1}{\theta^n} I(\theta > X_{max}) \quad \text{since } X_{min} > 0.$$

- Here derivative approach does not work! why?



i.e. $L(\theta)$ is decreasing as a function of θ for $\theta \geq X_{max}$ & 0 o.w.

so it is maximized at $\theta = X_{max}$

$$\Rightarrow \boxed{\hat{\theta}_{MLE} = X_{max}}$$

<u>Exer</u>: $D = \{X_1, X_2, \ldots, X_n\}$, $X_i \overset{iid}{\sim}$ Uniform(a, b)
find MLEs for a & b.

<u>Properties of MLEs</u>

1) $\hat{\theta}_{MLE}$ is consistent $\Rightarrow$ $\hat{\theta}_{MLE} \simeq \theta$ (for large n)
(ie. $\hat{\theta}_{MLE} \xrightarrow{P} \theta$)

2) $\hat{\theta}_{MLE}$ is asymptotically normal
i.e. $\hat{\theta}_{MLE} \overset{approx}{\sim} N(\theta, se^2(\hat{\theta}_{MLE}))$ for large n

$\Rightarrow (1-\alpha)$ CI for θ is
$$\hat{\theta}_{MLE} \pm z_{\alpha/2} \hat{se}(\hat{\theta}_{MLE})$$

3) MLE is <u>invariant</u> (AoS says "equivariant")
If $\hat{\theta}$ is an MLE for θ, then
$g(\hat{\theta})$ " " " " $g(\theta)$ for any g.

• (3) is a very useful property b/c it also implies
$g(\hat{\theta})$ is consistent for $g(\theta)$
& $g(\hat{\theta}) \overset{approx}{\sim} N(g(\theta), se^2(g(\hat{\theta})))$ for large n.
can get this by Δ-method

<u>Ex</u>: $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p), Find an MLE of $\frac{p}{1-p}$.
$\hat{p}_{MLE} = \overline{X}_n$, so $\frac{\overline{X}_n}{1-\overline{X}_n}$ is the MLE of $\frac{p}{1-p}$

<u>Ex</u>: $X_1, \ldots, X_n \overset{iid}{\sim}$ Uniform$(0, \theta)$. Find a <u>consistent</u> estimator for $Var(X_i) = \frac{\theta^2}{12}$. X_{max} is MLE (consistent) for θ
(consistent) " $\theta^2/12$.
$\Rightarrow X_{max}^2/12$ " "