Recap

Method of moments to estimate θ

- d = # parameters in the model
- set up d equations:

$$\underbrace{E[X^k]}_{\text{involves } \theta} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^k}_{\text{involves data}}, \quad k = 1, 2, \ldots, d$$

and solve for θ. Solution $= \hat{\theta}_{MOME}$.

θ

Method of Maximum Likelihood

$X_1, \ldots, X_n$ ~ indep and identically distributed, follow $f_\theta(x)$.

Likelihood function of data: Joint pdf or pmf of sample data considered as a function of θ with data held fixed at the observed values $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

$$\underbrace{L(\theta)}_{\text{a fn. of } \theta} = L(\theta; \underbrace{x_1, x_2, , x_n}_{\text{fixed}}) \overset{\text{indep.}}{=} \prod_{i=1}^{n} f_\theta(x_i) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n)$$

- A function of θ — the data are held fixed.

Maximum likelihood estimator (MLE) of θ: The value $\hat{\theta}$ of θ that maximizes the likelihood function as a function of θ.

- Can think of MLE as the value of θ that is "most likely" to have led to the observed data.
- Essentially a calculus problem.

How to find MLE?

Direct approach: Directly maximize the likelihood function.

Ex: Let $X_1, X_2, \ldots, X_n$ represent a random sample from a Uniform $(0, \theta)$ distribution where $\theta > 0$. Find the MLE of θ.

Recall:

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & o/w. \end{cases}$$

MOME of θ?

$$\frac{\theta}{2} = E[X] = \bar{X} \Rightarrow \theta = 2\bar{X} = \hat{\theta}_{MOME}.$$

(uniform)

MLE of θ:

Intuitively:

$\theta =$ max. value in the population.

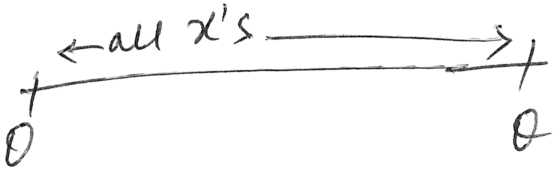
$\max\{X_1, \ldots, X_n\}$ is a natural estimator of θ.

will see that this is MLE.

$$L(\theta) = \underbrace{\frac{1}{\theta} I(0 \le x_1 \le \theta)}_{f_\theta(x_1)} \cdot \frac{1}{\theta} I(0 \le x_2 \le \theta) \cdots \underbrace{\frac{1}{\theta} I(0 \le x_n \le \theta)}_{f_\theta(x_n)}.$$

$I(a \le x \le b) = \begin{cases} 1, & a \le x \le b \\ 0, & \text{o/w} \end{cases}$. indicator function

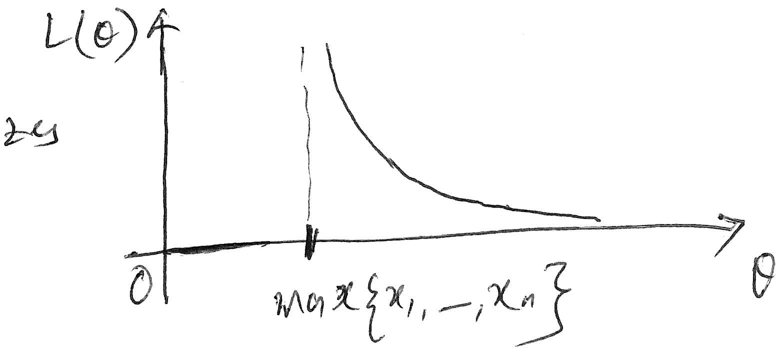
$= \frac{1}{\theta^n}$, provided



$= \frac{1}{\theta^n}$ provided all the x's are $\le \theta$

[No need to worry about x's being negative because by def. all x's ≥ 0].

$= \frac{1}{\theta^n}$, $\max\{x_1, \ldots, x_n\} \le \theta$.



Graph $\Rightarrow$ The value of θ that maximizes $L(\theta)$ is $\theta = \max\{x_1, \ldots, x_n\}$

$\overset{\text{def}}{=} \hat{\theta}_{MLE}$.

Q. Have two estimators for θ — $\hat{\theta}_1$ and $\hat{\theta}_2$. Which one will you choose?

A. The one that is better in the sense of having smaller mean-squared error (MSE) defined as

$$MSE[\hat{\theta}] = \underbrace{E[(\hat{\theta}-\theta)^2]}_{\text{Avg distance b/w est. } \hat{\theta} \text{ and the parameter } \theta.} \text{ — may depend on } \theta.$$

HW4 supplement: Use Monte Carlo simulation to compare $\hat{\theta}_{MOME}$ and $\hat{\theta}_{MLE}$ of θ to figure out which is better.

Differentiation technique: Maximize the log-likelihood function $\ln L(\theta)$ with respect to θ instead of $L(\theta)$ as the former tends to be easier. The value of θ that maximizes $L(\theta)$ also maximizes $\ln L(\theta)$. (**Why?**)

Since log fn. is monotone, the value of θ that maximizes $L(\theta)$ is also the value that maximizes $\log L(\theta)$

Step 1: Set up the log-likelihood function.

$$\log L(\theta) = \log \prod_{i=1}^{n} f_\theta(x_i) = \sum_{i=1}^{n} \log f_\theta(x_i)$$

Step 2: Find the *likelihood equation* by partially differentiating $\ln L(\theta)$ with respect to θ and setting the derivative to equal to zero.

$$\boxed{\frac{\partial}{\partial\theta} \log L(\theta) = 0} \rightarrow \text{likelihood equation.}$$

Step 3: Solve the likelihood equation for θ. The solution is MLE if it is a point of maxima (no need to verify).

~~Solution of~~ $\hat{\theta}_{MLE}$ = solution of this equation.

Recall: Some useful properties of natural log:

[$\ln = \log$].

- $\ln(ab) = \ln(a) + \ln(b)$
- $\ln(a^b) = b\ln(a)$
- $\ln(e^a) = a$

Ex: Let $X_1, X_2, \ldots, X_n$ represent a random sample from an Exponential (λ) distribution where $\lambda > 0$. Find the MLE of λ.

Recall: $f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \quad \lambda > 0. \\ 0, & \text{otherwise.} \end{cases}$

Assume that all x's are positive.

Step1: $L(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n})$

$= \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$

$= \lambda^n e^{-n\bar{x}\lambda}$

[$\bar{x} = \frac{\sum_i x_i}{n}$]

$$\Rightarrow \log L(\lambda) = \log\left[\lambda^n e^{-n\bar{x}\lambda}\right]$$

$$= n\log(\lambda) - n\bar{x}\lambda.$$

<u>Step 2:</u>

$$0 = \frac{d}{d\lambda}\log L(\lambda) = \frac{n}{\lambda} - n\bar{x}$$

<u>Step 3:</u>

$$\Rightarrow \frac{\cancel{n}}{\lambda} = \cancel{n}\bar{x}$$

$$\Rightarrow \lambda = \boxed{\frac{1}{\bar{x}} = \hat{\lambda}_{MLE}}.$$

<u>MOME?</u>

$$\frac{1}{\lambda} \underset{\text{know}}{\overset{\uparrow}{=}} E[X] = \bar{X}$$

$$\Rightarrow \lambda = \boxed{\frac{1}{\bar{X}} = \hat{\lambda}_{MOME}}$$

Using R to get MLE

Ex: Recall the CPU data — CPU times for $n = 30$ randomly chosen jobs (in seconds): 70, 36, 43, 69, 82, 48, 34, 62, 35, 15, 59, 139, 46, 37, 42, 30, 55, 56, 36, 82, 38, 89, 54, 25, 35, 24, 22, 9, 56, 19. Graphics suggested that the distribution of these CPU times may be right-skewed. Suppose we ***assume*** that the parent distribution is Gamma (α, λ), with both parameters unknown. What are MLE's of these parameters?

Recall: $f(x) = \frac{\lambda^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\lambda x}$, $x > 0$, $\alpha > 0, \lambda > 0$

positive

See book: For MOME.

MLE: $\underset{(\alpha,\lambda)}{\log L(\theta)} = \sum_{i=1}^{n} \log \left[\frac{\lambda^\alpha}{\Gamma\alpha} x_i^{\alpha-1} e^{-\lambda x_i} \right]$.

— Difficult to maximize in closed form.

```
# We will continue working with the CPU data
# that we saw earlier

cpu <- scan(file="cputime.txt")

# Negative of log-likelihood function assuming gamma
# parent distribution

neg.loglik.fun <- function(par, dat)
{
result <- sum(dgamma(dat, shape=par[1], rate=par[2],
log=TRUE))
return(-result)
}

# Minimize -log (L), i.e., maximize log (L)

ml.est <- optim(par=c(3, 0.1), fn=neg.loglik.fun,
```

(α, λ) → data

α λ

? optim

initial values are close to MOME see book.

minimizes by default.

```
method = "L-BFGS-B", lower=rep(0,2), hessian=TRUE,
dat=cpu)
```

allows constraints. $\alpha > 0, \lambda > 0$.

```
# > ml.est
# $par
# [1] 3.63149628 0.07529459
```

— $\hat{\alpha}, \hat{\lambda}$

```
# $value
# [1] 136.561
```

= min value of $\log L(\alpha, \lambda)$.

```
# $counts
# function gradient
     # 20        20

# $convergence
# [1] 0
```

— good.

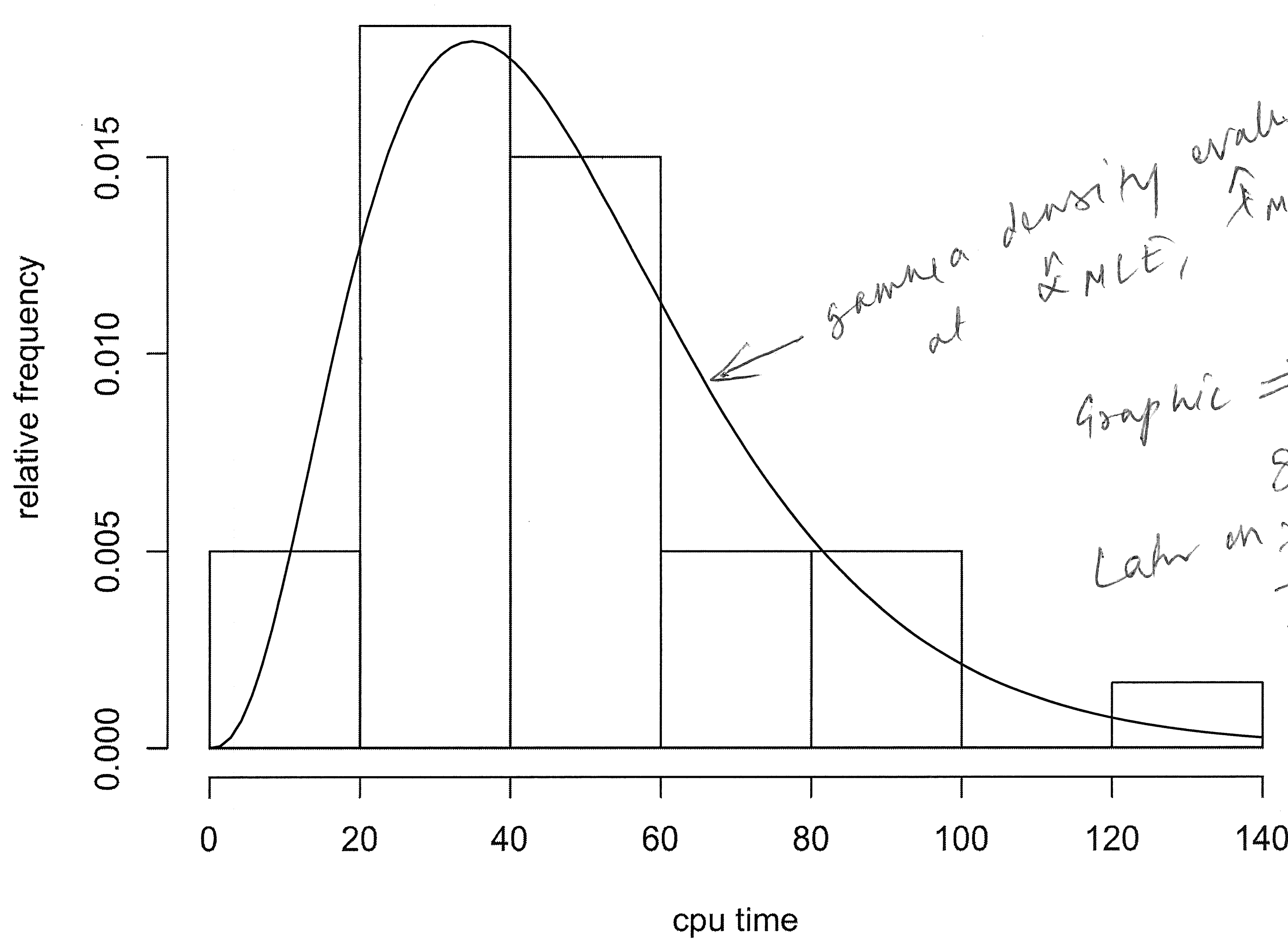
```
# $message
```

```
# [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

# $hessian
            # [,1]        [,2]
# [1,]    9.501374  -398.4584
# [2,] -398.458449 19223.5065
# >

# MLE

# > ml.est$par
# [1] 3.63149628 0.07529459
# >

# their standard errors

# > sqrt(diag(solve(ml.est$hessian)))
# [1] 0.89720941 0.01994668
```

```
# >

# How well the fitted model represents the data?

# relative frequency (density) histogram

hist(cpu, freq=FALSE, xlab="cpu time",
ylab="relative frequency",
main="histogram vs fitted gamma distribution")

# superimpose the fitted density

gamma.pdf <- function(x, shape=ml.est$par[1],
rate=ml.est$par[2])
{ dgamma(x, shape=shape, rate=rate) }

curve(gamma.pdf, from=0, to=140, add=TRUE)
```

add this plot to the existing plot

histogram vs fitted gamma distribution



gamma density evaluated at $\hat{\alpha}_{MLE}$, $\hat{\lambda}_{MLE}$.

Graphic $\Rightarrow$ fit is good.

Later on: more formal methods.