Lecture 2: Parameter Estimation

STA261 – Probability & Statistics II

Ofir Harari

Department of Statistical Sciences

University of Toronto

Outline

Radioactive Decay (Section 8.2 in the book)

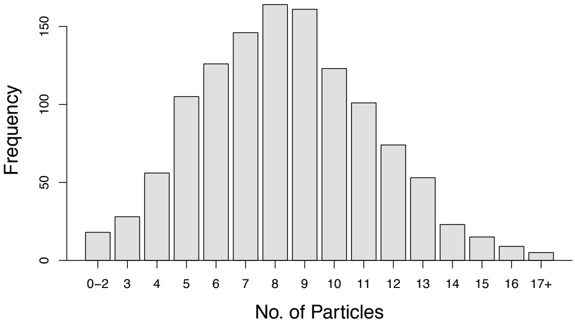
- Radioactive material, emits α-particles
- period of observation $\ll$ half-life ($\Longrightarrow$ emission rate $\approx$ constant)
- Rate, λ, classifies material
- We observe counts of α-particles
- What do the counts tell us about λ?
- Have X - count for a 10 sec interval
- Have 1207 non-overlapping intervals
- Have X_i, $i = 1, \ldots, 1207$

x	Observed Count
0,1,2	18
3	28
4	56
5	105
6	126
7	146
$\vdots$	$\vdots$
15	15
16	9
≥ 17	5

Table: Intervals were categorized by count

The data

```
> observed <- c(18, 28, 56, 105, 126, 146, 164, 161,
+               123, 101, 74, 53, 23, 15, 9, 5)
> names <- c("0-2", paste(3:16), "17+")
>
> bp <- barplot(observed, names=names, axes=TRUE,
+               xlab="No. of Particles", ylab="Frequency")
```



Randomness as a form of ignorance

- Clearly the number of particles emitted per time unit varies
- As science develops, the reasons for that may be uncovered
- For example, when a fair die is rolled, we say "$X \sim \mathrm{U}\{1, \ldots, 6\}$..."
    - Perhaps if we knew the initial position of the die, the release speed and angle etc., we could tell the result with certainty?
    - But that would be difficult...
- Turning to statistics/probability is oftentimes an admission of the limitations/inadequacy of a scientific theory
- "Essentially, all models are wrong, but some are useful." (George E.P. Box)
- With that in mind, we try to fit the best possible model to quantify the uncertainty with respect to all possible outcomes

Back to the α particles example

- Rate of emission is approximately constant
- No two particles can be emitted simultaneously
- This means that the Poisson distribution represents a reasonable model for the data:

$$X_1 \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Pois}(\lambda) \qquad (n = 1207)$$

- Note that the distribution of the number of particles emitted is completely determined by λ - the *parameter* of the distribution
- If we are to make any probabilistic statements about potential future outcomes (number of emitted particles), value must be provided for λ
  - or rather – learned from the data!

Parameter estimation

Reminder

Any function of the sample is called a *statistic*.

Definition

Let $X_1, \ldots, X_n$ be a random sample from some distribution, and let θ be a parameter of that distribution. Any statistic $U = U(X_1, \ldots, X_n)$ that is used to estimate θ, is called an *estimator* of θ.

- We usually denote an estimator of θ by $\widehat{\theta}$
  - In the particle emission example: $\widehat{\lambda} =?$
- In the next few weeks we will –
  - Introduce methods of estimation,
  - Discuss properties of estimators, and
  - Learn how to assess the performance of estimators

Method of Moments estimation

Definition

Let $X_1, \dots, X_n$ be a random sample from some distribution.

1. The kth *moment* of the distribution (if exists) is defined as

$$\mu_k = \mathbb{E}\left[X^k\right].$$

2. The kth *sample moment* is defined as

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

- The method of moments estimator of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ is the solution to the system

$$\begin{cases} \mu_1(\hat{\boldsymbol{\theta}}) = m_1 \\ \vdots \\ \mu_p(\hat{\boldsymbol{\theta}}) = m_p \end{cases}$$

Example: the normal distribution

Example

For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, find the method of moments estimators of μ and σ.

Solution:

We need to solve the system

$$\begin{cases} \mu_1 & = \mathbb{E}[X] = \mu = m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}, \text{ and} \\ \mu_2 & = \mathbb{E}[X^2] = \{\mathbb{E}[X]\}^2 + \mathrm{Var}[X] = \mu^2 + \sigma^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \end{cases}$$

for which the solution is

$$\mu = m_1 = \overline{X}, \text{ and}$$

$$\sigma^2 = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

then subsitute estimator in place of true param

Example: the normal distribution (cont.)

Solution (cont.):

The Method of Moments estimators are then obtained by replacing each unknown parameter θ by $\widehat{\theta}$. In this case –

$$\begin{cases} \widehat{\mu} & = \overline{X}, \text{ and} \\ \widehat{\sigma}^2 & = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2. \end{cases}$$

⋆ Note that

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\overline{X}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2. \end{aligned}$$

Example: the Gamma distribution

Example

For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \lambda)$ (with pdf $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $x \geq 0$), find the method of moments estimators of α and λ.

Solution:

We need to solve the system

$$\begin{cases} \mu_1 & = \mathbb{E}[X] = \frac{\alpha}{\lambda} = m_1 = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}, \text{ and} \\ \mu_2 & = \mathbb{E}[X^2] = \{\mathbb{E}[X]\}^2 + \text{Var}[X] = \left(\frac{\alpha}{\lambda}\right)^2 + \frac{\alpha}{\lambda^2} = m_2 = \frac{1}{n}\sum_{i=1}^n X_i^2, \end{cases}$$

Or, more compactly

$$\begin{cases} \frac{\alpha}{\lambda} & = \overline{X} \\ \frac{\alpha}{\lambda^2} & = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$$

remember the simplification here proved eariler

Example: the Gamma distribution (cont.)

Solution (cont.):

$$\begin{cases} \dfrac{\alpha}{\lambda} & = \overline{X} \\[2ex] \dfrac{\alpha}{\lambda^2} & = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{cases}$$

Dividing the second equation by the first one, we obtain

$$\lambda = \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \alpha = \lambda\overline{X} = \frac{\overline{X}^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

and finally, the Method of Moments estimators are given by

$$\widehat{\lambda} = \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \left(= \frac{\widehat{\mu}}{\widehat{\sigma}^2}\right) \quad \text{and}$$

$$\widehat{\alpha} = \frac{\overline{X}^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \left(= \frac{\widehat{\mu}^2}{\widehat{\sigma}^2}\right).$$

can be computed from data, statistics

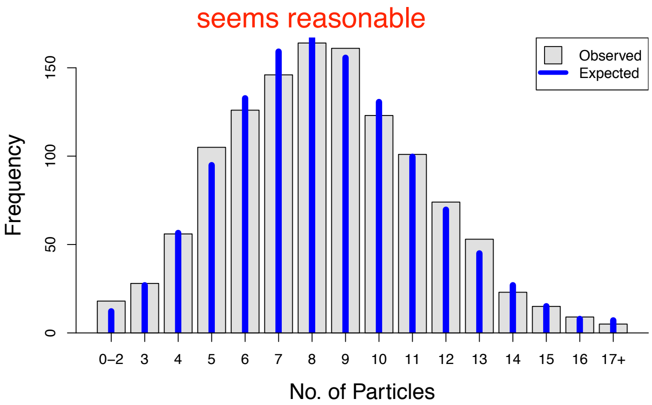
Back to the α particles example

- Here $X_1 \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Pois}(\lambda)$ $(n = 1207)$
- Method of moments estimator: solve –

$$\mu_1 = \mathbb{E}[X] = \lambda = m_1 = \overline{X} \Longrightarrow \widehat{\lambda} = \overline{X}$$

  - Our method of moments **estimator** is $\widehat{\lambda} = \overline{X}$
- We are told that for our dataset $\widehat{\lambda} = \overline{X} = 8.392$
  - Our method of moments **estimate** is $\widehat{\lambda} = 8.392$
- Estimators are statistics (random variables); estimates are their evaluations (numbers)
- Is the Poisson model with $\widehat{\lambda} = 8.392$ adequate for this kind of data?

Assessing the fit



$$\text{Expected}_k = 1207 \times e^{-\widehat{\lambda}} \frac{\widehat{\lambda}^k}{k!}$$

Consistency

Definition (Consistency)

Let $X_i, \dots, X_n \sim f_\theta$ (a sample from a distribution depending on parameter θ). We say that $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \dots, X_n)$ is a *consistent estimator* of θ if for any $\varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\left(|\widehat{\theta}_n - \theta| > \varepsilon\right) = 0.$$

- To be precise, the above property is called *consistency in probability*
- Sometimes we say that $\widehat{\theta}_n$ *converges in probability* to θ, or simply write $\widehat{\theta}_n \xrightarrow{\text{P}} \theta$
- In fact, we have all encountered one consistent estimator already... when we studied the Weak Law of Large Numbers (WLLN)!

i.e. P(l \bar{X_n} - \mu l > \epsilon) \stackrel{p}{\sim} 0

Consistency of method of moments estimators

Reminder (WLLN)

Let $X_i, \ldots, X_n$ be a random sample from some distribution with mean $\mathbb{E}[X]$. Then for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > \varepsilon\right) = 0.$$

- Thus m_1 $(= \overline{X})$ is a consistent estimator of the first moment μ_1 $(= \mathbb{E}[X])$

true regardless of distribution of X

- Replace X_i with X_i^k and $\mathbb{E}[X]$ with $\mathbb{E}[X^k]$ above, and you will learn that $m_k \xrightarrow{\text{P}} \mu_k$ for any k (as long as the latter exists)

Theorem

Let $\widehat{\theta}_n \xrightarrow{\text{P}} \theta$ and $\widehat{\eta}_n \xrightarrow{\text{P}} \eta$. Then

1. $\widehat{\theta}_n + \widehat{\eta}_n \xrightarrow{\text{P}} \theta + \eta$
2. $\widehat{\theta}_n \widehat{\eta}_n \xrightarrow{\text{P}} \theta\eta$
3. $g(\widehat{\theta}_n) \xrightarrow{\text{P}} g(\theta)$ for any continuous $g : \mathbb{R} \to \mathbb{R}$

UNIVERSITY OF TORONTO

Consistency (cont.)

- The last Theorem is the reason why method of moments estimators are usually consistent
- Consider, for example

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = m_2 - m_1^2$$

we obtained for the normal distribution earlier

- $m_1 \xrightarrow{\text{P}} \mathbb{E}[X]$, hence $m_1^2 \xrightarrow{\text{P}} \{\mathbb{E}[X]\}^2$ (why?)

by property 3 previous page

- $m_2 \xrightarrow{\text{P}} \mathbb{E}[X^2]$, therefore

$$\widehat{\sigma}^2 = m_2 - m_1^2 \xrightarrow{\text{P}} \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2 = \sigma^2$$

- Similarly,

because n—> infinity

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}\cdot\widehat{\sigma}^2 \xrightarrow{\text{P}} 1\cdot\sigma^2 = \sigma^2$$

sample variance is also consistent

UNIVERSITY OF TORONTO

Method of moments: concluding remarks

- Method of moments estimators are typically easy to derive
- Could lead to conflicting results (non-unique solutions) – in the Poisson example we obtained $\widehat{\lambda} = \overline{X}$, but note that

$$\mu_1 = \mathbb{E}[X] = \lambda = m_1,$$

$$\mu_2 = \mathbb{E}[X^2] = \mathrm{Var}[X] + \{\mathbb{E}[X]\}^2 = \lambda + \lambda^2 = m_2,$$

note here estimator derived from second order moments different from first

hence $\widehat{\lambda} = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \widehat{\sigma}^2$ is another method of moments estimator of λ

- Despite being consistent, they are often outperformed by other methods of estimation
- We may use them as initial values in iterative procedures when searching for Maximum Likelihood Estimators

UNIVERSITY OF TORONTO

Maximum Likelihood Estimation

Definition

Let $X_1 \ldots, X_n$ be continuous (discrete) random variables with joint pdf (pmf) $f(x_1, \ldots, x_n|\theta)$, where θ is a parameter. For a given vector of observations $(x_1, \ldots, x_n)$, the *likelihood* of θ is

$$\mathcal{L}(\theta) = f(x_1, \ldots, x_n|\theta).$$

- In words: "how *likely* are we to observe such a sample for this particular value of θ?"
- We are used to thinking of the joint pdf (pmf) as a function of randomly drawn X's with the parameters known and fixed
- Here the sample is observed and held fixed, while we let the (unknown) parameters vary
- Our goal is to choose among all possible values of θ, the one most likely to have produced the sample in hand

Example: Poisson distribution

- As before $X_1 \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Pois}(\lambda)$
- $P(X_i = x_i \big| \lambda) = \mathrm{e}^{-\lambda} \dfrac{\lambda^{x_i}}{x_i!}$, $x_i = 0, 1, 2, \ldots,\ i = 1, \ldots, n$
- The likelihood in this case is given by –

$$\begin{aligned}
\mathcal{L}(\lambda) = P(x_1, \ldots, x_n \big| \lambda) &\overset{\text{ind}}{=\!=} \prod_{i=1}^{n} P(X_i = x_i \big| \lambda) \\
&= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \Big/ \prod_{i=1}^{n} x_i! \\
&\propto \lambda^{n\bar{x}} e^{-n\lambda}
\end{aligned}$$

Definition

The *Maximum Likelihood Estimator* (MLE) of θ is

$$\widehat{\theta}_{\text{MLE}} = \arg\max_{\theta} \mathcal{L}(\theta)$$

(the value of θ that maximizes $\mathcal{L}(\theta)$).

Example: Poisson distribution (cont.)

$$\mathcal{L}(\lambda) \propto \lambda^{n\bar{x}} e^{-n\lambda}$$

- We wish to find $\widehat{\lambda}_{\mathrm{MLE}} = \arg\max_\lambda \mathcal{L}(\lambda)$
- It is often more convenient to work with the log-likelihood function,

$$\ell(\lambda) := \log \mathcal{L}(\lambda) = n\bar{x}\log\lambda - n\lambda + \text{const}$$

  - Turns products into sums
  - Prevents computer underflow errors
- Because the logarithm is a monotonically increasing function

$$\underline{\arg\max_\lambda \ell(\lambda) = \arg\max_\lambda \mathcal{L}(\lambda)}$$

justification why maximizing log likelihood still works

Example: Poisson distribution (cont.)

$$\ell(\lambda) = n\bar{x}\log\lambda - n\lambda + \text{const}$$

- Maximize $\ell(\lambda)$:

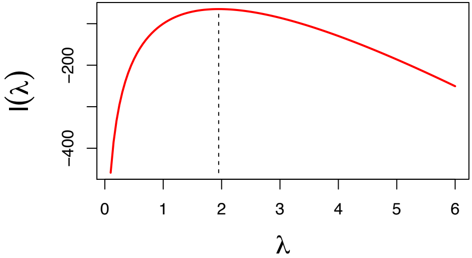
$$\frac{\partial \ell}{\partial \lambda} = \frac{n\bar{x}}{\lambda} - n = 0 \Longrightarrow \lambda = \bar{x}$$

- Hence $\widehat{\lambda}_{\text{MLE}} = \overline{X}$
- Verify maximum:

$$\left.\frac{\partial^2 \ell}{\partial \lambda^2}\right|_{\lambda=\bar{x}} = -\left.\frac{n\bar{x}}{\lambda^2}\right|_{\lambda=\bar{x}} = -\frac{n}{\bar{x}} < 0$$

Poisson example: R simulation

```
lambda <- 2
n <- 100
x <- rpois(n, lambda) #sampling 100 random Poisson RVs

lambdaValues <- seq(.1, 6, by=.05)
xBar <- mean(x) #the MLE
loglike <- n*xBar*log(lambdaValues) - n*lambdaValues

plot(lambdaValues, loglike, type='l')
lines(c(xBar, xBar), c(min(loglike), max(loglike)), lty=2)
```



Example: MLEs for the normal distribution

Example

For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, find the maximum likelihood estimators of μ and σ.

Solution:

The likelihood in this case is

$$\mathcal{L}(\mu, \sigma^2) = f(x_1, \ldots, x_n | \mu, \sigma^2) \overset{\text{ind}}{=\!=} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\},$$

and the log-likelihood is consequently

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log\sigma^2 - -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + \text{const.}$$

MLEs for the normal distribution (cont.)

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + \text{const.}$$

Now, solving

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \Longrightarrow \sum_{i=1}^{n} x_i = n\mu \Longrightarrow \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X},$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\Longrightarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2.$$

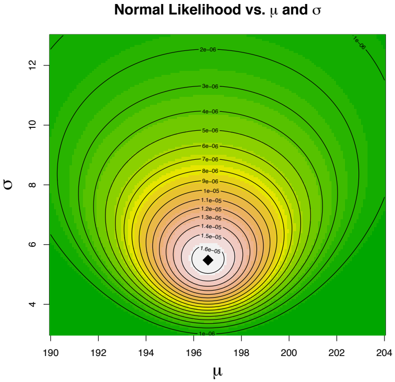
note we substituted \mu = mean in

⋆ Do not expect the MLEs to always coincide with the Method of Moments estimators...

note its each individual data - mean

Normal example: R simulation

```
y <- c(200.3, 195.0, 192.4, 205.2, 190.1)
n <- length(y)

mu.hat <- mean(y)
sigma.hat <- sqrt((sd(y))^2*(n-1)/n)

muVec <- seq(190, 204, length = 150)
sigmaVec <- seq(3, 13, length = 150)

LogLikeli <- function(mu, sigma)
{
        -n*log(sigma) - sum((y-mu)^2)/(2*sigma^2)
}

Z <- outer(muVec, sigmaVec,
     function(mu, sigma) mapply(LogLikeli, mu, sigma))

contour(muVec, sigmaVec, exp(Z), nlevel=20)
points(mu.hat, sigma.hat, cex=2.5, pch=18)
```



UNIVERSITY OF TORONTO

Parameter estimation for muon decay

Example

Denote by Θ the angle at which electrons are released in a muon decay, and let $X = \cos\Theta$. If we assume that X follows a distribution with pdf

$$f(x|\alpha) = \frac{1+\alpha x}{2}, \ -1 \leq x \leq 1, \ -1 \leq \alpha \leq 1,$$

find both method of moments and maximum likelihood estimators for α, based on a random sample $X_1, \ldots, X_n$.

Solution:

- First note that

note in this case pdf is continuous

$$\mathbb{E}[X] = \int x f(x|\alpha)\mathrm{d}x = \frac{1}{2}\int_{-1}^{1}(x+\alpha x^2)\mathrm{d}x = \frac{\alpha}{3},$$

thus the method of moments estimator of α can be derived by solving

$$\mu_1 = \mathbb{E}[X] = \frac{\alpha}{3} = m_1 = \overline{X} \Longrightarrow \widehat{\alpha} = 3\overline{X}$$

Muon decay example (cont.)

- Onto finding the MLE of α
- Write the likelihood:

$$\mathcal{L}(\alpha) = f(x_1, \ldots, x_n \big| \alpha) = \prod_{i=1}^{n} f(x_i \big| \alpha) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + \alpha x_i)$$

- On the logarithmic scale:

$$\ell(\alpha) = -n \log 2 + \sum_{i=1}^{n} \log(1 + \alpha x_i)$$

- To find the $\widehat{\alpha}_{\mathrm{MLE}}$ we need to solve

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \frac{x_i}{1 + \alpha x_i} = 0$$

no closed form so iterative method

UNIVERSITY OF TORONTO

The Newton-Raphson method

- To find the MLE of θ we need to solve $\ell'(\theta)\Big|_{\theta=\widehat{\theta}} = 0$
- For θ_0 in the proximity of $\widehat{\theta}$,

$$\ell'(\hat{\theta}) \approx \ell'(\theta_0) + \ell''(\theta_0)(\hat{\theta} - \theta_0) \text{ (first order Taylor expansion about } \theta_0\text{)}$$

condition which derives the mle estimator

- However, $\ell'(\hat{\theta}) = 0$ (why?), hence

$$\ell'(\theta_0) + \ell''(\theta_0)(\hat{\theta} - \theta_0) \approx 0,$$

or, alternatively

$$\widehat{\theta} \approx \theta_0 - \frac{\ell'(\theta_0)}{\ell''(\theta_0)}$$

- This is the Newton-Raphson method –

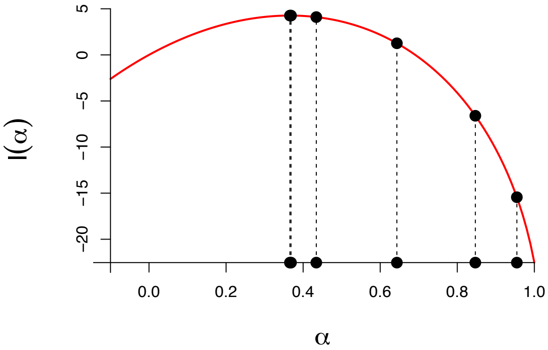
$$\boxed{\widehat{\theta}_{\text{new}} = \widehat{\theta}_{\text{old}} - \frac{\ell'(\widehat{\theta}_{\text{old}})}{\ell''(\widehat{\theta}_{\text{old}})} \text{ (iterate until convergence)}}$$

UNIVERSITY OF TORONTO

Back to the muon decay example

- $\ell'(\alpha) = \sum_{i=1}^{n} \frac{x_i}{1 + \alpha x_i}$
- $\ell''(\alpha) = -\sum_{i=1}^{n} \frac{x_i^2}{(1 + \alpha x_i)^2}$
- $\widehat{\alpha}_{\text{new}} = \widehat{\alpha}_{\text{old}} - \frac{\ell'(\widehat{\alpha}_{\text{old}})}{\ell''(\widehat{\alpha}_{\text{old}})}$
-
```
alpha <- runif(1) #random initial value
tolerance <- 5e-16 #largest difference between successive alphas
delta <- 1
LogLikeOld <- sum(log(1 + alpha*x)) #log-likelihood at initial alpha

while(delta > tolerance){
        lPrime <- sum(x/(1 + alpha*x)) #first derivative
        l2Prime <- -sum(x^2/(1 + alpha*x)^2) #second derivative
        alpha <- alpha - lPrime/l2Prime #Newton-Raphson update
        LogLike <- sum(log(1 + alpha*x)) #log-likelihood at new alpha
        delta <- abs(LogLike - LogLikeOld) #difference
        LogLikeOld <- LogLike
}
```

Muon decay example (cont.)




```
> alpha #maximum likelihood estimate
[1] 0.3657926
> (alphaHatMME <- 3*mean(x)) #method of moments estimate
[1] 0.3444427
```