Lecture 6: More MLES

Consider the Bernoulli but w/ a parameter

$$\eta = p/1-p = \text{odds}$$

What is the MLE for η?

$$\eta = p/1-p \iff p = \eta/(1+\eta)$$

Get the likelihood in terms of η

$$L(p) = p^{N\bar{x}}(1-p)^{N-N\bar{x}}$$

$$L(\eta) = \left(\frac{\eta}{1+\eta}\right)^{N\bar{x}}\left(1-\frac{\eta}{1+\eta}\right)^{N(1-\bar{x})}$$

$$\hat{\eta} = \underset{\eta}{\operatorname{argmin}}\, L(\eta)$$

1. Get $\frac{\partial l}{\partial \eta}$

$$l(\eta) = \log L(\eta) = N\bar{x}\log(\eta/1+\eta) + N(1-\bar{x})\log\left(\frac{1}{1+\eta}\right)$$

$$= N\bar{x}(\log(\eta) - \log(1+\eta)) - N(1-\bar{x})\log(1+\eta)$$

$$\frac{\partial \ell}{\partial \eta} = N\bar{x}/\eta - \frac{\cancel{N\bar{x}}}{\cancel{1}+\eta} - N\frac{1}{1+\eta} + \frac{\cancel{N\bar{x}}}{\cancel{1}+\eta}$$

$$= \frac{N\bar{x}}{\eta} - \frac{N}{1+\eta}$$

2) Set eq. to zero, solve for η.

$$\frac{\cancel{N}\bar{x}}{\eta} = \frac{\cancel{N}}{1+\eta} \Rightarrow \cancel{N}\bar{x} + \cancel{N}\eta\bar{x} = \eta\cancel{N}$$

$$\Rightarrow \bar{x} = \eta(1-\bar{x})$$

$$\Rightarrow \boxed{\hat{\eta} = \frac{\bar{x}}{1-\bar{x}}}$$

Recall: $\eta = p/1-p$ and $\hat{p} = \bar{x}$

curiously, $\hat{\eta} = \hat{p}/1-\hat{p}$

Theorem: Transformations of MLEs

If $\hat{\theta}$ is the MLE for θ then the MLE for $g(\theta)$ is $g(\hat{\theta})$.

Ex. $X_n \overset{iid}{\sim} \text{Pois}(\lambda)$, $\lambda > 0$.

Let's get MLE $\hat{\lambda}$.

① $L(\lambda) = f_\lambda(\underline{x}) = \prod_n f(x_n) = \prod_n \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$

$$= \lambda^{\sum_n x_n} e^{-N\lambda} \prod_n \left(\frac{1}{x_n!}\right)$$

$$\ell(\lambda) = \left(\sum_n x_n\right) \log(\lambda) - N\lambda + \log\left(\prod_n \frac{1}{x_n!}\right)$$

② Take deriv, set to zero.

$$\frac{\partial \ell}{\partial \lambda} = \left(\sum_n x_n\right)\frac{1}{\lambda} - N = 0$$

$$\Rightarrow \boxed{\hat{\lambda} = \frac{1}{N}\sum_n x_n = \bar{X}}$$

What's the MLE of $\theta = \log(\lambda)$?

$$\hat{\theta} = \log(\hat{\lambda}) = \log(\bar{X}).$$

Ex. $X_n \overset{iid}{\sim} \text{Exp}(\lambda)$

Let's get MLE $\hat{\lambda}$.

1. $$L(\lambda) = \prod_n f(x_n) = \prod_n \lambda e^{-\lambda x_n} \mathbb{1}(x_n > 0)$$
$$= \lambda^N e^{-\lambda \sum_n x_n}$$

$$\ell(\lambda) = \log L(\lambda)$$
$$= N \log(\lambda) - \lambda \sum_n x_n$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{N}{\lambda} - \sum_n x_n = 0$$

$$\Rightarrow \frac{N}{\lambda} = \sum_n x_n \Rightarrow \boxed{\hat{\lambda} = \frac{N}{\sum_n x_n} = \frac{1}{\bar{x}}}$$

$$\mathbb{E}[\text{Exp}(\lambda)] = 1/\lambda$$

Sometimes people parameterize the Exp in terms of
$$\beta = 1/\lambda$$

i.e. $f(x_n) = \frac{1}{\beta} e^{-x_n/\beta}$ so that $\mathbb{E}[\text{Exp}(\beta)] = \beta$

What's the MLE for β? $1/\hat{\lambda} = 1/(1/\bar{x}) = \bar{x}$

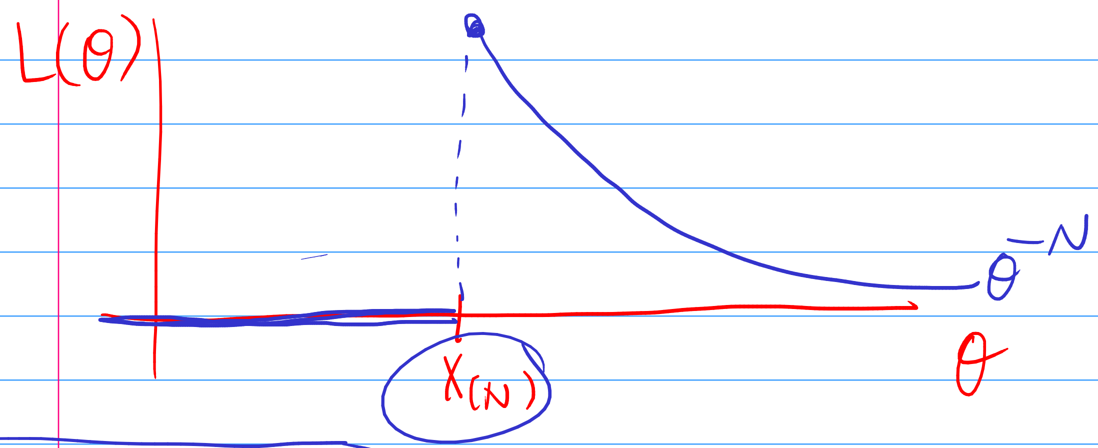
Ex. let $X_n \overset{iid}{\sim} U(0,\theta)$ for $\theta > 0$

What's the MLE for θ?

(1) $$L(\theta) = \prod_n \frac{1}{\theta} \mathbb{1}(0 \le x_n \le \theta)$$

$$= \theta^{-N} \prod_n \mathbb{1}(x_n \ge 0) \prod_n \mathbb{1}(x_n \le \theta)$$

$$= \theta^{-N} \mathbb{1}(x_{(1)} \ge 0) \mathbb{1}(x_{(N)} \le \theta)$$



$$\boxed{\hat{\theta} = X_{(N)}}$$

Ex. $X_n \overset{iid}{\sim} N(\theta, 1)$ where $\theta \ge 0$

Consider $$L(\theta) = \prod_n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_n - \theta)^2\right)$$

$$= (2\pi)^{-N/2} \exp\left(-\frac{1}{2}\sum_n (x_n - \theta)^2\right)$$

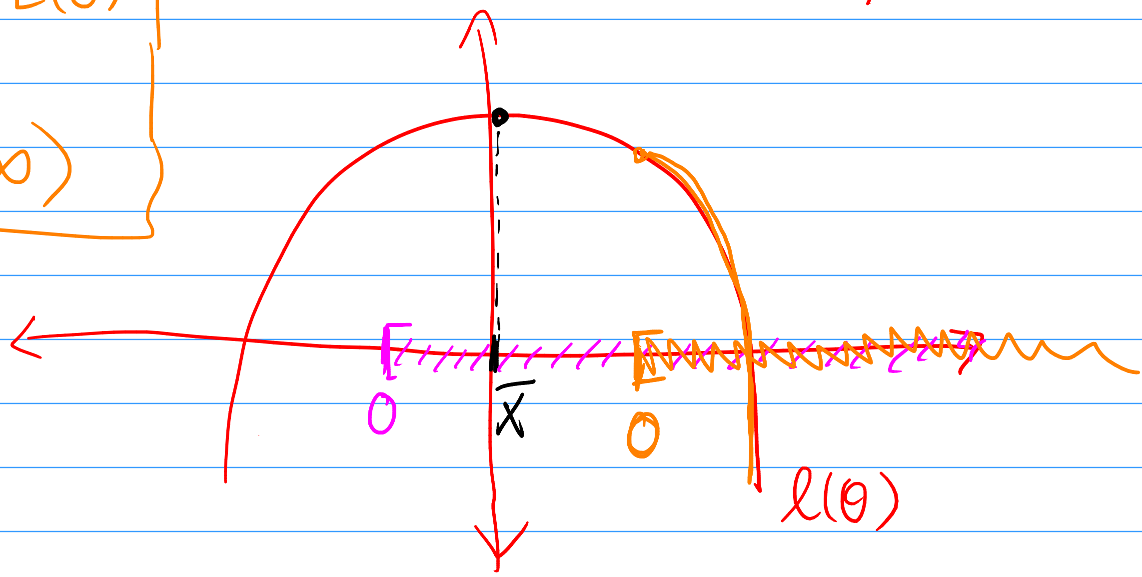
$$\ell(\theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_n (x_n-\theta)^2$$

a quadratic in θ

$\sim -\theta^2$

$\hat\theta = \underset{\theta\in\Theta}{\arg\max}\, L(\theta)$

$\Theta = [0,\infty)$



Case 1: $\bar{x} \geq 0$ so $\hat\theta = \bar{x}$

Case 2: $\bar{x} < 0$ so $\hat\theta = 0$

$$\hat\theta = \max(0, \bar{x})$$

Ex. $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both unknown.

Let's get MLE for μ and σ^2. $[\mu\in\mathbb{R}, \sigma^2>0]$

① $$L(\mu,\sigma^2) = \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_n-\mu)^2\right)$$

$$= (2\pi)^{-N/2}(\sigma^2)^{-N/2}\exp\left(-\frac{1}{2\sigma^2}\sum_n (x_n-\mu)^2\right)$$

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2)$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_n (x_n-\mu)^2$$

2 $\frac{\partial \ell}{\partial \mu} = 0$ and $\frac{\partial \ell}{\partial \sigma^2} = 0$ $\left[\frac{\partial \ell}{\partial \tau} = 0\right]$

$$\ell(\mu, \tau) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\tau) - \frac{1}{2\tau}\sum_n (x_n-\mu)^2$$

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{2\tau}\sum_n 2(x_n-\mu)(-1) = \frac{1}{\tau}\sum_n (x_n-\mu)$$

$$= \frac{1}{\tau}\left(\sum_n x_n - N\mu\right) = 0$$

$$\Rightarrow \sum_n x_n - N\mu = 0$$

$$\Rightarrow \boxed{\hat{\mu} = \frac{1}{N}\sum_n x_n = \bar{x}}$$

$$\frac{\partial \ell}{\partial \tau} = -\frac{N}{\cancel{2}}\frac{1}{\tau} + \frac{1}{\cancel{2}\tau^2}\sum_n (x_n-\mu)^2 = 0$$

$$\Rightarrow \frac{-N}{\tau} + \frac{1}{\tau^2}\sum_n (x_n-\mu)^2 = 0$$

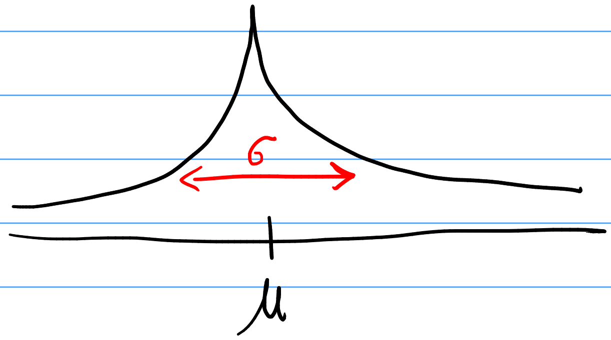
$$\Rightarrow \frac{1}{\tau}\sum_n (x_n-\mu)^2 = N$$

$$\Rightarrow \frac{1}{N}\sum_n (x_n-\mu)^2 = \tau$$

$$\boxed{\hat{\tau} = \hat{\sigma}^2 = \frac{1}{N}\sum_n (x_n-\hat{\mu})^2 = \frac{1}{N}\sum_n (x_n-\bar{x})^2}$$

Ex. $X_n \overset{iid}{\sim} \text{Laplace}(\mu, \sigma^2)$

$$f(x) = \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|x-\mu|\right)$$



what's the MLE?

1 $L(\mu, \sigma) = \prod_n \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|x_n-\mu|\right)$

$$= 2^{-N} \sigma^{-N} \exp\left(-\frac{1}{\sigma}\sum_n |x_n-\mu|\right)$$

$$\ell(\mu, \sigma) = -N\log(2) - N\log(\sigma) - \frac{1}{\sigma}\sum_n |x_n-\mu|$$

Problem, not diffable WRT μ.

Can look at $\frac{\partial \ell}{\partial \sigma}$.

$$\frac{\partial \ell}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^2}\sum_n |x_n-\mu| = 0$$

⋮

$$\boxed{\hat{\sigma} = \frac{1}{N}\sum_n |x_n - \hat{\mu}|}$$

$$e^{-a}$$

$$L(\mu, \sigma) \propto \exp\Big(-\frac{1}{\sigma} \underbrace{\sum_n |x_n - \mu|}_{a}\Big)$$

decreases as a increases.

So to maximize, should make a as small as possible.

$$\min \underbrace{\sum_n |x_n - \mu|}_{\text{total dist of } \mu \text{ to } x_n}$$

$$\boxed{\bar{x} = \operatorname*{argmin}_{\mu} \sum_n (x_n - \mu)^2}$$

Turns out μ should be median(x_n).