ESE 531 : Finishing up MVUE + Review

Today - project proposal → 11:59 PM TN!

How to evaluate estimators of θ ?

$$\boxed{MSE(\hat{\theta}) = bias^2(\hat{\theta}) + V[\hat{\theta}]}$$

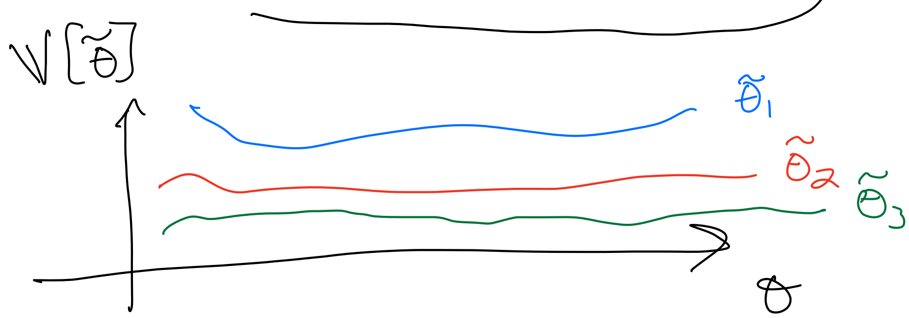
If $\hat{\theta}$ is unbiased → what is the best estimator?

$\hat{\theta}$ should have the smallest possible variance to be the best estimator in the MSE sense

MVUE - minimum variance unbiased estimator $\hat{\theta}$ that satisfied

$$\underline{V[\hat{\theta}] \leq V[\tilde{\theta}] \quad \forall \theta}$$

$\tilde{\theta}$ is any other unbiased estimator



$\tilde{\theta}_3$ is the MVUE

Cramer-Rao Lower Bound

$$\mathbb{V}[\tilde{\theta}] \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \log p(x_{1:n};\theta)}{\partial\theta^2}\right]}$$

Fisher information $I(\theta)$

An unbiased estimator is efficient if it achieves or attains the CRLB

$$\mathbb{V}[\hat{\theta}] = I^{-1}(\theta)$$

$\Rightarrow$ $\hat{\theta}$ is efficient.

- The MLE is asymptotically efficient

$\hat{\theta}$ is the MLE : $\lim_{n\to\infty} \mathbb{V}[\hat{\theta}] \to$ CRLB

not always tending to 0

Existence of an Efficient Estimator

NOT ON TEST

You can deduce if an efficient estimator exists if we can find an estimator s.t.

$$\mathbb{V}[\hat{\theta}] = \frac{1}{I(\theta)}$$

Cauchy-Schwartz inequality

For our choices of Y and Z in the proof of the CRLB, it can be shown that the bound is equal if and only if:

score function

$$\frac{\partial \log p(x_{1:n};\theta)}{\partial\theta} = a(\theta)\left(g(x_{1:n}) - \theta\right)$$

$a(\theta)$ and $g(x)$ are "any functions". It turns out

$\hat{\theta} = g(X_{1:n})$ is the MVUE and is efficient.

CRLB for Signals with White Gaussian Noise

· Suppose we have the following probabilistic model:

$$X_i = s_i(\theta) + \varepsilon_i\,,\quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

and s_i is an known (indexed) function (Example: $s_i(\theta) = \cos(2\pi i\theta i)$)

Then, the CRLB for θ is established as follows:

$$\boxed{\mathbb{V}[\tilde{\theta}] \geq \frac{\sigma^2}{\sum_{i=1}^{n}\left(\frac{\partial s_i}{\partial\theta}\right)^2}}$$

$\frac{1}{I(\theta)}$

directly compute the CRLB for any s_i

Proof.

$$p(x_{1:n}; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - s_i(\theta))^2}{2\sigma^2}\right)$$

$$\Downarrow$$ convert to log joint and get fisher information

Example: Sinusoidal Frequency Estimation (pg. 36 - kay)

Suppose we have:

$$X_i = A\cos(2\pi f_0 i + \phi) + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Find the CRLB of f_0 where $|f_0| \leq \frac{1}{2}$.

$$s_i(f_0) = A\cos(2\pi f_0 i + \phi)$$

$$\frac{\partial s_i}{\partial f_0} = -A\sin(2\pi f_0 i + \phi)\, 2\pi i$$

$$\left(\frac{\partial s_i}{\partial f_0}\right)^2 = A^2 (2\pi i)^2 \sin^2(2\pi f_0 i + \phi)$$

and therefore,

$$\boxed{V\left[\tilde{f}_0\right] \geq \frac{\sigma^2}{\sum_{i=1}^{n} A^2 (2\pi i)^2 \sin^2(2\pi f_0 i + \phi)}}$$

Transformation of Parameters (CRLB)

- Invariance property of MLE: $\hat{\theta}$ is the MLE of $p(x_{1:n}; \theta)$

$$\Downarrow$$

then $\hat{\alpha} = \tau(\hat{\theta})$ is the MLE of $p(x_{1:n}; \tau(\theta))$

- Natural questions:

  1. Is efficiency preserved under a transformation of estimators

$\hat{\theta}$ we can show $\hat{\theta}$ is efficient

$$\Downarrow ?$$

$\hat{\alpha} = \tau(\hat{\theta})$ is efficient for α

No does not always preserve efficiency

But does if τ is linear

(2.) What is the CRLB for a transformed parameter?

$$\mathbb{V}[\tilde{\theta}] \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \log p(x_{1:n};\theta)}{\partial\theta^2}\right]}$$

$$\Downarrow$$

$$\mathbb{V}[\tilde{\alpha}] \geq \frac{\left(\frac{\partial \tau}{\partial\theta}\right)^2}{-\mathbb{E}\left[\frac{\partial^2 \log p(x_{1:n};\theta)}{\partial\theta^2}\right]}$$

this is a function of θ

If τ is an invertible function, we can write the bound in terms of α, but if not, we cannot.

Example: Unknown Constant Signal in Noise

$$X_i = \mu + \varepsilon_i \quad , \quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

Find the MLE of μ and show it is efficient.

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

This estimator is efficient

$$\mathbb{V}[\hat{\mu}] = \frac{\sigma^2}{n} = \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \log p(x_{1:n};\mu)}{\partial \mu^2}\right]}$$

What is the CRLB for $(\hat{\mu})^2$?

Here $\alpha = \tau(\mu) = \mu^2$. $\frac{\partial \tau}{\partial \mu} = 2\mu$

Here, τ is a nonlinear function $\rightarrow$ so not guaranteed that $\hat{\alpha}$ is efficient

$$\left(\frac{\partial \tau}{\partial \mu}\right)^2 = 4\mu^2$$

and so:

$$\mathbb{V}[\tilde{\alpha}] \geq 4\mu^2 \cdot \frac{\sigma^2}{n}$$

$$\hat{\alpha} = (\hat{\mu})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$$

$$\mathbb{V}[\hat{\alpha}] = \mathbb{V}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right]$$

$$= \mathbb{V}\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2\right]$$

$$= \mathbb{V}\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2 + \underbrace{2\sum_{j=1}^{n}\sum_{k=j+1}^{n} X_j X_j}\right)\right]$$

remainder should be 0

$\mathbb{V}[X_i; X_j]$

$$= \frac{1}{n^4}\,\mathbb{V}\left[\sum_{i=1}^{n} X_i^2\right]$$

$$= \frac{1}{n^4}\sum_{i=1}^{n} \underbrace{\mathbb{V}[X_i^2]}$$

$\sigma^2 + \mu^2$

$\mathbb{V}[X] + \mathbb{E}[X]^2$

$$\mathbb{V}[X_i^2] = \mathbb{E}[X_i^4] - \overbrace{\mathbb{E}[X_i^2]}^{2}$$

$$\overset{?}{=} 3\sigma^4 - (\sigma^2 + \mu^2)^2$$

This can shown in general to not attain the CRLB for α.

CRLB for Multiple Unknown Parameters (Vector Case)

- In reality, we usually have more than one unknown parameter:

$$X_i \sim N(\mu, \sigma^2)$$

$$X_i \sim N(\mu_0 + \mu_1 Z_1 + \mu_2 Z_2, \sigma^2)$$

So the question is, how do we establish a bound for

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_K \end{bmatrix}$$

The covariance matrix:

$$C_{\hat{\theta}} = \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}[\hat{\theta}]\right)\left(\hat{\theta} - \mathbb{E}[\hat{\theta}]\right)^T \right]$$

If $\hat{\theta}$ is unbiased $\mathbb{E}[\hat{\theta}] = \theta$

The CRLB for the vector parameter case establishes the following

$$\boxed{C_{\tilde{\theta}} \succsim \underbrace{I^{-1}(\theta)}_{\text{inverse of the Fisher Information matrix}}} \Rightarrow C_{\tilde{\theta}} - I^{-1}(\theta) \text{ is a PSD matrix}$$

for any θ, we have

$$\theta^T \left(C_{\tilde{\theta}} - I^{-1}(\theta) \right) \theta \geq 0$$

The Fisher information matrix $I(\theta) \in \mathbb{R}^{K \times K}$ is a matrix with the elements:

$$I_{jk}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log p(x_{1:n}; \theta)}{\partial \theta_j \, \partial \theta_k} \right]$$

Along the diagonal of I (if I is a diagonal matrix) we can arrive at CRLB for the individual parameters.

Example: Gaussian MLE with unknown μ and σ^2

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$$

If we solve the MLE, we have:

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

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

biased estimator

Does it make sense to find the CRLB for $\hat{\sigma}^2$?

$$MSE(\hat{\theta}) = \underline{\underline{bias^2(\hat{\theta})}} + V[\hat{\theta}]$$

It is possible for a biased estimator to achieve a MSE lower that CRLB

CRLB $\quad V[\hat{\theta}]$

$$= MSE[\hat{\theta}] \geq \frac{\sigma^2}{10}$$ if 10 samples

$\hat{\theta}_1$ unbiased $\quad MSE(\hat{\theta}_1) = MSE(\hat{\theta}_2)$

$\hat{\theta}_2$ biased

Which estimator has lower variance?

$V[\hat{\theta}_2] \leq V[\hat{\theta}_1]$

$$MSE(\hat{\theta}_1) = V[\hat{\theta}_1]$$

$$MSE(\hat{\theta}_2) = \underbrace{\text{bias}^2(\hat{\theta}_2)}_{>0} + \underbrace{V[\hat{\theta}_2]}$$

$\Rightarrow$

$$\text{bias} = \mathbb{E}[\hat{\theta} - \theta] = \mathbb{E}[\hat{\theta}] - \theta$$

Back to the question

$$p(x_{1:n}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$\log p(x_{1:n}; \mu, \sigma^2) = \frac{-n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

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

$$\boxed{\frac{\partial^2 \log p}{\partial \mu^2} = -\frac{n}{\sigma^2}}$$

$$\boxed{\frac{\partial^2 \log p}{\partial \mu \partial \sigma^2} = \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)}$$

$$= \frac{\partial^2 \log p}{\partial \sigma^2 \partial \mu}$$

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

$$\boxed{\frac{\partial^2 \log p}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{4\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$-\mathbb{E}\left[\frac{\partial^2 \log p}{\partial (\sigma^2)^2}\right] = \frac{n}{2\sigma^4}$$

$$-\mathbb{E}\left[\frac{\partial^2 \log p}{\partial \mu^2}\right] = \frac{n}{\sigma^2}$$

$$-\mathbb{E}\left[\frac{\partial^2 \log p}{\partial \sigma^2 \partial \mu}\right] = -\mathbb{E}\left[\frac{\partial^2 \log p}{\partial \mu \partial \sigma^2}\right] = 0$$

$\theta = (\mu, \sigma^2)$

$$I(\theta) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & \dfrac{n}{2\sigma^4} \end{bmatrix}$$

Does the sample variance attain the CRLB?

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Is

$$V[S^2] \stackrel{?}{=} \frac{2\sigma^4}{n} \quad ?$$

Five questions

→ you can bring a single sheet of paper for a cheat sheet

→ must be handwritten

→ can use front and back

① Random samples and their properties

True/False: If an estimator converges $\hat{\theta} \xrightarrow{P} \theta$, then it converges $\hat{\theta} \xrightarrow{a.s.} \theta$.

$$\hat{\theta} \xrightarrow{p} \theta \not\Longrightarrow \hat{\theta} \xrightarrow{a.s} \theta$$

Almost sure convergence implies convergence in probability but the converse is not true

TRUE

$$\hat{\theta} \xrightarrow{p} \theta \Longrightarrow \hat{\theta} \xrightarrow{d} \theta \ ?$$

TRUE/FALSE

Suppose $\hat{\theta} \xrightarrow{P} \theta$ and $\hat{\gamma} \xrightarrow{a.s.} a$, where a is a constant. Then,

$$\hat{\theta} + \hat{\gamma} \xrightarrow{p} \theta + a$$

By Slutsky's theorem, the statement is true

Continuous mapping Theorem

$\hat{\theta} \xrightarrow{P} \theta$ $\qquad\qquad$ $\theta \in (-\infty, \infty)$

$\gamma(\cdot) = \log(\cdot)$

$\gamma(\hat{\theta}) \xrightarrow{P} \gamma(\theta)$?

If γ is a continuous function, on the range of values θ can take then yes!

$\underline{\text{TRUE}}$ / False: If two random variables X_1 and X_2 have different MGFs, the their distributions are different.

True / $\underline{\text{False}}$: If two random variables have different MGFs, then their means cannot be the same.

For example

$X_1 \sim N(0, 1)$

$X_2 \sim N(0, 10)$

Same mean but different distributions

There will be 5 true and false questions.

- Method of Moments:

$\hat{\theta}$ is a MOME estimator, then

$$\hat{\theta} \xrightarrow{p} \theta$$

$$g_1(\theta) = \mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X]$$

$$g_2(\theta) = \mathbb{E}[X^2] = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathbb{E}[X^2]$$

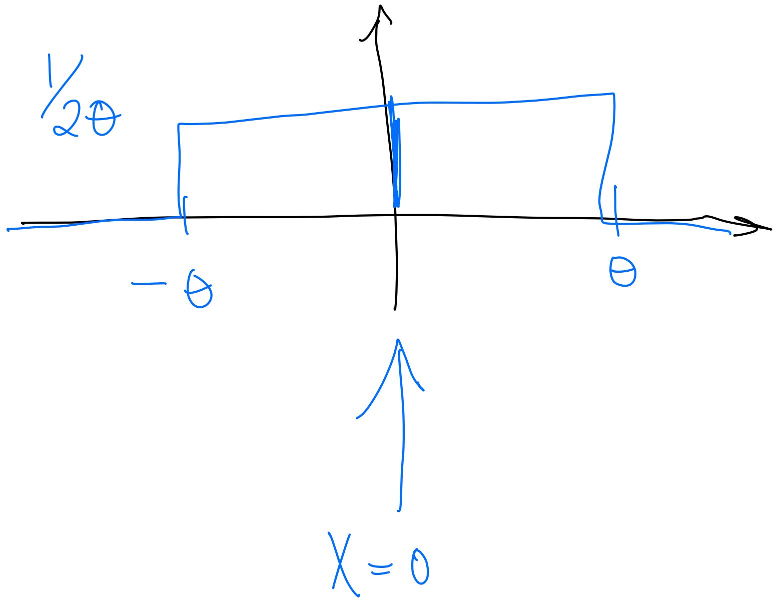
CMP

$$\hat{\theta} = g^{-1}\left(\underbrace{\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{bmatrix}}_{\text{converges by WLLN}}\right) \xrightarrow{p} \theta$$

Example:

$$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-\theta, \theta), \quad i = 1, \dots, n$$

$$\mathbb{E}[X_i] = 0$$



Theoretical moment does not depend on value of θ. In that case, we can use the second moment:

$$\mathbb{E}[X^2] = \int_{-\theta}^{\theta} x^2 \frac{1}{2\theta}\, dx$$

$$= \frac{1}{2\theta}\left(\frac{x^3}{3} \Big|_{-\theta}^{\theta} \right)$$

$$= \frac{1}{2\theta}\left(\frac{\theta^3}{3} - \frac{-\theta^3}{3} \right) = \frac{\cancel{2}\theta^{\cancel{3}\,2}}{3\,\cancel{6}\cancel{\theta}}$$

$$= \frac{\theta^2}{3} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$= \boxed{\hat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}}$$

$$\hat{\theta} = \sqrt{3\,\overline{X_n^2}}$$

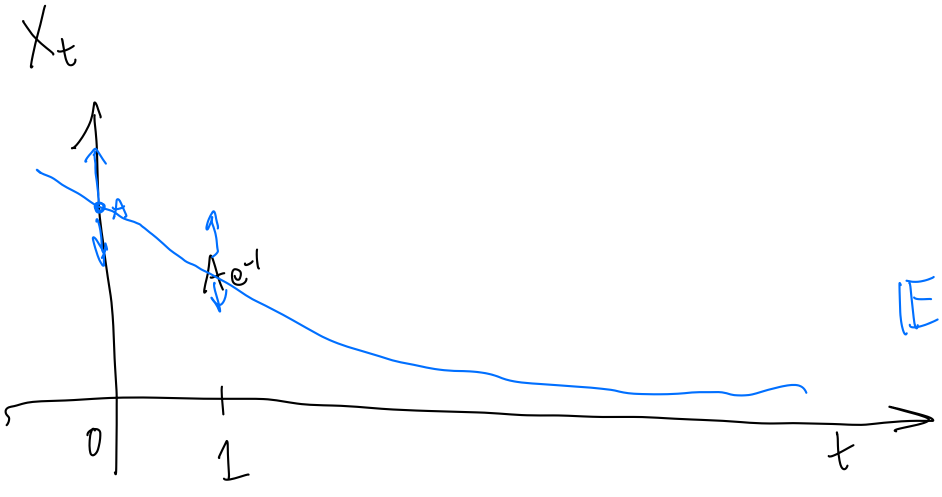
$$\overline{X_n^2} \xrightarrow{p} \mathbb{E}[X^2]$$

$$\overset{CMT}{\Longrightarrow} \hat{\theta} \xrightarrow{p} \theta$$

<u>MOME</u>: Suppose you have the following model:

$$X_t = Ae^{-t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

$$t = 1, \ldots, T$$



$t \to \infty$

$\mathbb{E}[X_t] \to 0$

Find the MoME of A.

$$\mathbb{E}[X_t] = Ae^{-t} \qquad N(Ae^{-t}, \sigma^2)$$

$$\mathbb{V}[X_t] = \sigma^2$$

In the case of an iid random sample

$$\mathbb{E}[X] = \mathbb{E}[X_t] \quad \forall \ i$$

MoME does not exist. Not an i.i.d. random sample. The X_t s do not have the same mean. Because the mean of $\mathbb{E}[X_t]$ changes with the time, we can't compute the MoME.

Example: Suppose we have

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{A} e^{-\hat{\gamma}} =$$

$$X_t = \underbrace{A e^{-\gamma}}_{\theta} + \varepsilon_t , \quad \varepsilon_t \sim N(0, \sigma^2)$$

Same as underdete

We will assume σ^2 is known.

$$\mathbb{E}[X] = \mathbb{E}[X_t] = Ae^{-\gamma}$$

$$\mathbb{E}[X^2] = \mathbb{E}[X_t^2] = \sigma^2 + A^2 e^{-2\gamma}$$

$$\underbrace{\mathbb{V}[X]}_{\sigma^2} = \mathbb{E}[X^2] - \underbrace{\mathbb{E}[X]}_{Ae^{-\gamma}}{}^2$$

$$A\overbrace{e^{-\gamma}}^{b} = \overline{X}_n \qquad \boxed{Ab = \overline{X}_n}$$

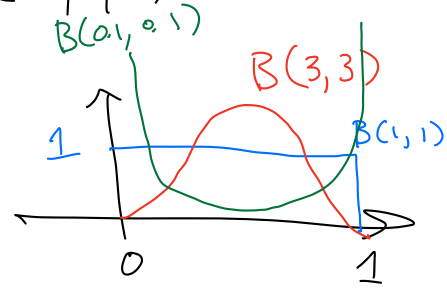
$$\sigma^2 + A^2 \overbrace{e^{-2\gamma}}^{b^2} = \overline{X_n^2} \qquad \sigma^2 + A^2 b^2 = \overline{X_n^2}$$

$$A = \frac{\overline{X}_n}{b}$$

not solvable?
will follow up

Suppose we have a Beta (α, β)

$$X_i \sim \text{Beta}(\alpha, \beta) \qquad i = 1, \dots, n$$



$$\mathbb{E}[X_i] = \frac{\alpha}{\alpha+\beta} \qquad \mathbb{V}[X_i] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Find the MoME of α and β.

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta} = \overline{X}_n$$

$$\mathbb{E}[X^2] = \mathbb{V}[X] + \mathbb{E}[X]^2$$

$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha^2}{(\alpha+\beta)^2} = \overline{X_n^2}$$

$$\overline{X}_n = \frac{\alpha}{\alpha+\beta}$$

$$\overline{X_n^2} = \frac{\overbrace{\alpha}^{\overline{X}_n}\ \overbrace{\beta}^{(1-\overline{X}_n)}}{(\alpha+\beta)(\alpha+\beta)(\alpha+\beta+1)} + (\overline{X}_n)^2 = \overline{X_n^2}$$

$$(\overline{X}_n)(1-\overline{X}_n)\frac{1}{(\alpha+\beta+1)} = \overline{X_n^2} - (\overline{X}_n)^2$$

$$\frac{1}{\alpha+\beta+1} = \frac{\overline{X_n^2} - (\overline{X}_n)^2}{\overline{X}_n(1-\overline{X}_n)}$$

$$\alpha + \beta + 1 = \frac{\bar{X}_n (1 - \bar{X}_n)}{\overline{X_n^2} - (\bar{X}_n)^2}$$

$$\alpha + \beta = \frac{\bar{X}_n (1 - \bar{X}_n)}{\overline{X_n^2} - (\bar{X}_n)^2} - 1$$

$$\hat{\alpha} = \bar{X}_n \left(\qquad \qquad \right)$$

$$\hat{\beta} = (1 - \bar{X}_n) \left(\qquad \qquad \right)$$

Optimization and EM → Known how to tell it a unique maximum for a likelihood function

- GMM ← Lecture
- Missing data in AR(1) ← HW

MLE

Example: Suppose I have i.i.d. random sample:

$$X_i \sim P(x_i;\theta)$$

where $P(x_i;\theta)$ is a discrete distribution of the form:

$$\left.\begin{aligned} p(X_i=0) &= p \\ p(X_i=1) &= 1-p \end{aligned}\right\}$$

what distribution?

Bernoulli RV

Find the MLE of p.

$$p(x_{1:n};\theta) = \prod_{i=1}^{n} p^{\mathbb{1}(x_i=0)} (1-p)^{\mathbb{1}(x_i=1)}$$

$$\ell(p;x_{1:n}) = \log p(x_{1:n};p)$$

$$= \sum_{i=1}^{n} \mathbb{1}(x_i=0)\log p + \mathbb{1}(x_i=1)\log(1-p)$$

$$= \log p \underbrace{\left(\sum_{i=1}^{n} \underbrace{\mathbb{1}(x_i=0)}_{(1-x_i)}\right)}_{n_0} + \log(1-p) \underbrace{\left(\sum_{i=1}^{n} \underbrace{\mathbb{1}(x_i=1)}_{x_i}\right)}_{n_1}$$

$$= n_0 \log p + n_1 \log(1-p)$$

$$\frac{\partial \ell}{\partial p} = \frac{n_0}{p} - \frac{n_1}{1-p} = 0$$

$$\frac{1-p}{p} = \frac{n_1}{n_0}$$

$$\frac{1}{p} - 1 = \frac{n_1}{n_0} \Rightarrow \frac{1}{p} = \frac{n_1 + n_0}{n_0}$$

$$\boxed{\hat{p} = \frac{n_0}{n_1 + n_0}}$$

Categorical distribution

$$p(x_i;\theta) = \begin{cases} \theta_1, & x_i = 0 \\ \theta_2, & x_i = 1 \\ \vdots & \\ \theta_K, & x_i = K \end{cases}$$

$$n_0 = \sum_{i=1}^{n} \mathbb{1}(x_i = 0)$$

$$\vdots$$

$$n_K = \sum_{i=1}^{n} \mathbb{1}(x_i = K)$$

$$\theta_1 = \frac{n_0}{\sum_{i=1}^{K} n_i} = \boxed{\frac{n_0}{n}}$$