Point estimation

$$\theta_{true} \longrightarrow \boxed{P(x;\theta)} \longrightarrow \underbrace{x_1, x_2, \ldots x_n}_{\text{observed data}}$$

$$\hat{\theta} \longleftarrow \boxed{T(\cdot)} \longleftarrow x_1, x_2, \ldots x_n$$

$\hat{\theta} \leftarrow$ random variable

$$\hat{\theta} \xrightarrow{P} \theta_{true}$$

$$\hat{\theta} = T_n(X_1, \ldots, X_n)$$

Statistic
OR
estimator

Formulate an Estimation Problem

$(x_1, \ldots, x_n)$

1. Write down a parametric probabilistic model:

$$p(x;\theta)$$

2. Identify what is known and unknown

3. Understand the constraints

$\theta \in \Theta$ ← feasible set

4. Estimate!!

Point estimation route

$$\hat{\theta} = T_n(X_1, \ldots, X_n)$$

Bayesian estimation

$p(\theta)$: prior

$\downarrow$ Bayes theorem

$p(\theta \mid x_1, \ldots, x_n)$ posterior

How do we represent probabilistic models?

1. Obvious way

$$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} p(x; \theta)$$

Normal, Binomial, Exponential

(2.) Less obvious route

$$X_i = f(\theta, \epsilon_i) \quad , \quad \epsilon_i \overset{iid}{\sim} p(\epsilon)$$

In most cases, we will deal with:

$$X_i = \underbrace{g(\theta)}_{\text{signal}} + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$\Downarrow$$

$$p(x_i ; \theta) = \mathcal{N}(x_i ; g(\theta), \sigma^2)$$

Example: Find the joint distribution of an exponential random sample

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$$

$$\Downarrow$$

$$p(x_i ; \lambda) = \begin{cases} \lambda e^{-\lambda x} , & x > 0 \\ 0 , & \text{o.w.} \end{cases}$$

$$p(x_1, \dots, x_n; \lambda) = \prod_{i=1}^{n} p(x_i; \lambda)$$

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

<u>Definition</u> (Conditional independence)

We say a random variable X is conditionally independent of Y <u>given</u> Z if

$$p(x \mid y, z) = p(x \mid z)$$

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_{-i})$$

$-i$ denotes all the random variables besides i

<u>Example</u>: AR(1)

Consider a random sample $X_1, \dots, X_n$ s.t.

$$X_i = \alpha X_{i-1} + \varepsilon_i, \quad |\alpha| < 1, \quad \varepsilon_i \sim N(0, \sigma^2)$$

and we define $X_0 \sim p(X_0)$. What is the distribution $p(x_1, \dots, x_n ; \alpha)$.

$$p(x_0, x_1, \dots, x_n ; \alpha) = \boxed{p(x_0) \prod_{i=1}^{n} \underbrace{p(x_i \mid x_{i-1} ; \alpha)}_{\mathcal{N}(x_i ; \alpha X_{i-1}, \sigma^2)}}$$

$$P(x_1, \dots, x_n ; \alpha) = \int (\;) \, dx_0$$

Method of Moments

If I have an estimator $\hat{\Theta}$, I should expect that

$$\mathbb{E}_{p(x;\hat{\theta})}[X^k] \approx \underbrace{\mathbb{E}_{p(x;\theta)}[X^k]}_{\text{replace with sample moment}}$$

k is a positive integer

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow{a.s.} \mathbb{E}[X^k]$$

Theorem (Uniquess of Moments)

Let X be a random variable with pdf $p(x)$. Suppose that the MGF exists:

$$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

existence:

$$\mathbb{E}\left[\left|e^{tX}\right|\right] < \infty$$

and the moments $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, ... exist. Then the sequence of moments $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, ... uniquely determine the random variable X and its pdf.

Definition (Method of Moments Estimator) (MOME)

The MOME estimator is the estimator $\hat{\theta}$ obtained by solving the following system of equations

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

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

OR

$$\mathbb{E}_{p(x;\hat{\theta})}\left[\left(X - g_1(\hat{\theta})\right)^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\vdots$$

$$g_k(\hat{\theta}) = \mathbb{E}_{p(x;\hat{\theta})}[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

If the system of equations is not soluable, we can use an optimization approach. Define a loss

$$l_j(X_1,\dots,X_n;\hat{\theta}) = \left(g_j(\hat{\theta}) - \frac{1}{n}\sum_{i=1}^{n} X_i^j \right)^2$$

and solve the following minimization problem:

$$\hat{\theta} = \arg\min_{\theta} \sum_{j=1}^{k} l_j(X_1,\dots,X_n;\theta)$$

Example: Binomial random sample

$$X_1,\dots,X_n \overset{iid}{\sim} \text{Binomial}(N,\pi)$$

we will assume N is known

$$p(x;\pi) = \binom{N}{x} \pi^{x} (1-\pi)^{N-x}$$

$x = 0,\dots,N$

$0 \leq \pi \leq 1$

Bernoulli $\quad Z \sim \text{Bernoulli}(\pi)$

$$p(z;\pi) = \pi^{\mathbb{1}(z=1)} (1-\pi)^{\mathbb{1}(z=0)} = \begin{cases} \pi, & z=1 \\ (1-\pi), & z=0 \\ 0, & \text{o.w.} \end{cases}$$

$$\mathbb{E}[Z] = \sum_{z=0}^{1} z \cdot p(z;\pi) = (0)(1-\pi) + (1)(\pi)$$
$$= \pi$$

$$\mathbb{E}[X] = \mathbb{E}[Z_1 + Z_2 + \cdots + Z_N] = \sum_{j=1}^{N} \mathbb{E}[Z_j] = N\pi$$

$$\mathbb{V}[Z] = \pi(1-\pi) \quad \Longrightarrow \quad \mathbb{V}[X] = N\pi(1-\pi)$$

Now find MoME:

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

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

Example: Gamma Random Sample

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$$

$$p(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad \begin{matrix}\alpha>0\\ \beta>0\\ x>0\end{matrix}$$

$$\Gamma(\alpha) = \int_0^{\infty} t^{z-1} e^{-t}\, dt$$

$$\Gamma(\alpha) = (\alpha-1)!$$ if α is an integer

$$\mathbb{E}[X] = \frac{\alpha}{\beta} \quad \text{and} \quad \mathbb{V}[X] = \frac{\alpha}{\beta^2}$$

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

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

For the MOME

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\frac{\hat{\alpha}(1+\hat{\alpha})}{\hat{\beta}^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2}$$

$$\boxed{\hat{\alpha} = \hat{\beta}\bar{X}}$$ (from the first equation)

$$\frac{\cancel{\hat{\beta}}\bar{X}(1+\hat{\beta}\bar{X})}{\hat{\beta}^{\cancel{2}}} = \overline{X^2}$$

$$\bar{X}(1+\hat{\beta}\bar{X}) = \hat{\beta}\,\overline{X^2}$$

$$1+\hat{\beta}\bar{X} = \hat{\beta}\left(\frac{\overline{X^2}}{\bar{X}}\right)$$

$$1 = \hat{\beta}\left(\frac{\overline{X^2}}{\bar{X}}\right) - \hat{\beta}(\bar{X})$$

$$1 = \hat{\beta}\left(\frac{\overline{X^2}}{\bar{X}} - \bar{X}\right) = \hat{\beta}\left(\frac{\overline{X^2} - \bar{X}^2}{\bar{X}}\right)$$

$$\boxed{\hat{\beta} = \frac{\overline{X}}{\overline{X^2} - \overline{X}^2}} \rightarrow \hat{\alpha} = \hat{\beta}\,\overline{X}$$

Properties of MOME

1. $\hat{\theta}$ is a consistent estimator for θ under weak conditions

$$\hat{\theta} \xrightarrow{P} \theta$$

Why do you think that is?

$$g_1(\hat{\theta}) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\overline{X}_n \xrightarrow{a.s.} \mu = \mathbb{E}[X] \quad \text{(SLLN)}$$

$$\xrightarrow{P} \mu = \mathbb{E}[X] \quad \text{(WLLN)}$$

$$g_1(\hat{\theta}) = \overline{X}_n \xrightarrow{P} \mu$$

If the inverse of g_1 exists h_1, then by the continous mapping Theorem

inverse function of g_1

$$\hat{\theta} \xrightarrow{P} h_1(\mu)$$

We can get the limiting distribution by applying the Delta method to $h_1(\overline{X}_n)$

Disadvantages

1. MOME estimators in general can be biased

$$\mathbb{E}[\hat{\theta}] \neq \theta \quad \text{for MOME estimators in general}$$

2. $\hat{\theta}$ may not be in the feasible set Θ

Example: MOME for Uniform

$$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{U}(a, b) \iff p(x_i; a, b)$$

$$= \begin{cases} \dfrac{1}{b-a}, & a \le x_i \le b \\ 0, & \text{o.w.} \end{cases}$$

$$\hat{a} = \overline{X} - \sqrt{3\left(\overline{X^2} - \overline{X}^2\right)}$$

$$\hat{b} = \overline{X} + \sqrt{3\left(\overline{X^2} - \overline{X}^2\right)}$$

Suppose we observe data $(x_1, \dots, x_5) = (0, 0, 0, 0, 1)$

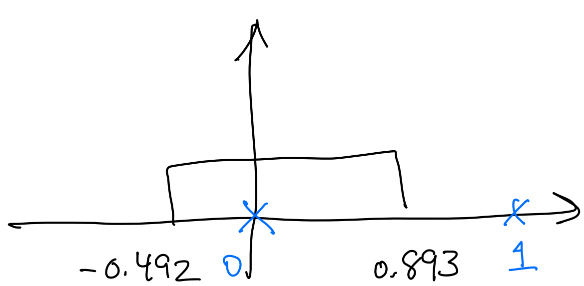
sample mean

sample second moment

$$\overline{X} \qquad\qquad \overline{X^2}$$

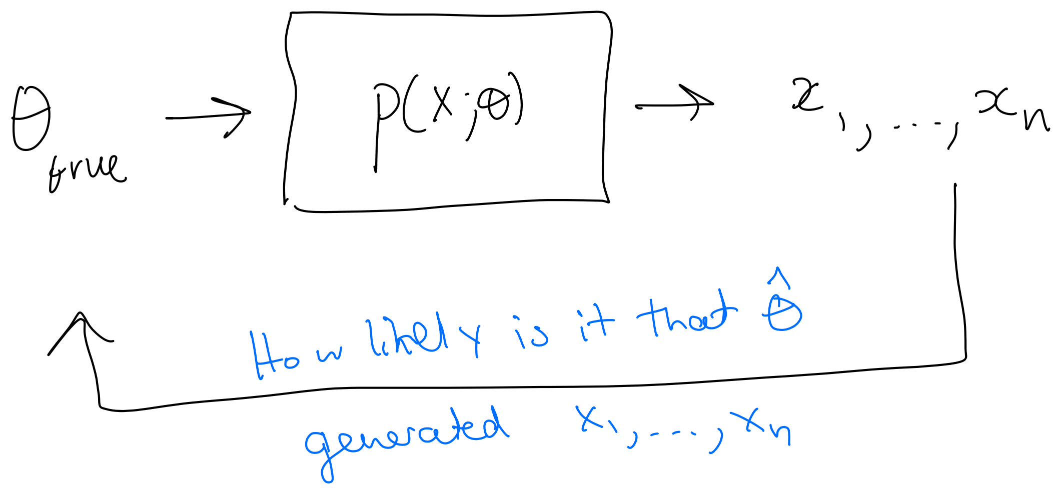
$$\hat{a} \approx -0.492$$

$$\hat{b} \approx 0.893$$



Maximum likelihood estimator (MLE)

Concept:



Definition (Likelihood Function)

Let $X_1, \dots, X_n$ be a random sample from a population $p(x;\theta)$ with $\theta \in \Theta$. Then, fixing the random sample $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ the likelihood function is a function of θ defined as:

$$L(\theta; x_1, \dots, x_n) \triangleq p(x_1, \dots, x_n; \theta)$$

Definition (Maximum Likelihood Estimator)

The MLE is defined as the estimator that solves the following optimization problem

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\ L(\theta; x_1, \dots, x_n)$$

Practical steps:

1. Find the likelihood function $L(\theta; x_1, \ldots, x_n)$
2. Determine the log-likelihood
$$\ell(\theta; x_1, \ldots, x_n) = \log L(\theta; x_1, \ldots, x_n)$$
3. Reduce $\ell(\theta; x_1, \ldots, x_n)$ to only depend on θ
4. Take derivative (w.r.t. θ) and set equal to 0:
   1. You can solve for $\frac{\partial \ell(\theta; x_1, \ldots, x_n)}{\partial \theta} = 0$ and obtain $\hat{\theta}$. If $\hat{\theta} \in \Theta$, then you're done
   2. Consider either a constrained optimization technique or resort to numerical optimization

Example: Binomial Random Sample

$X_1, \ldots, X_n \sim \text{Binomial}(N, \pi)$

$$p(x_i; \pi) = \binom{N}{x_i} \pi^{x_i} (1-\pi)^{N-x_i}$$

Step 1: Find $p(x_1, \dots, x_n; \pi)$ and treat as a likelihood function $L(\pi; x_1, \dots, x_n)$

$$p(x_1, \dots, x_n; \pi) = \prod_{i=1}^{n} p(x_i; \pi)$$

$$= \prod_{i=1}^{n} \binom{N}{x_i} \pi^{x_i} (1-\pi)^{N-x_i}$$

S = # of successes $\qquad$ F = # of failures

$$= \left(\prod_{i=1}^{n} \binom{N}{x_i}\right) \pi^{\sum_{i=1}^{n} x_i} (1-\pi)^{\sum_{i=1}^{n}(N-x_i)}$$

$$L(\pi; x_1, \dots, x_n) = \left(\prod_{i=1}^{n} \binom{N}{x_i}\right) \pi^{S} (1-\pi)^{F}$$

$$\propto \pi^{S}(1-\pi)^{F}$$

2. Step 2: Take the log

$$\tilde{\ell}(\pi; x_1, \dots, x_n) = \log\left(\pi^{S}(1-\pi)^{F}\right)$$

$$= S\log(\pi) + F\log(1-\pi)$$

(3.) Take derivative a set equal to 0

$$\left.\frac{\partial \tilde{\ell}}{\partial \pi}\right|_{\pi=\hat{\pi}} = \frac{S}{\hat{\pi}} + \frac{F}{1-\hat{\pi}}(-1) = 0$$

$$\Downarrow$$

$$\frac{S}{\hat{\pi}} = \frac{F}{1-\hat{\pi}}$$

$$\Downarrow$$

$$\frac{1-\hat{\pi}}{\hat{\pi}} = \frac{F}{S}$$

$$\frac{1}{\hat{\pi}} - 1 = \frac{F}{S}$$

$$\frac{1}{\hat{\pi}} = \frac{F+S}{S}$$

$$\Rightarrow \boxed{\hat{\pi} = \frac{S}{F+S}}$$

Example: Gaussian MLE

$X_1, \ldots, X_n \overset{iid}{\sim} N(0, \sigma^2)$. Find the MLE of σ^2.

$$p(x_i ; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x_i^2}{2\sigma^2}\right)$$

1. Find likelihood

$$p(x_1, \ldots, x_n ; \sigma^2) = \prod_{i=1}^{n} p(x_i ; \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x_i^2}{2\sigma^2}\right)$$

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

2. Take the log

$$\ell(\sigma^2 ; x_1, \ldots, x_n) = \frac{-n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2$$

③ Take derivative a set equal to 0.

$$\left.\frac{\partial \ell}{\partial \sigma^2}\right|_{\sigma^2 = \hat{\sigma}^2} = \frac{-n}{2} \frac{1}{\cancel{2\pi}\hat{\sigma}^2} \cancel{2\pi} + \frac{(\hat{\sigma}^2)^{-2}}{2}\left(\sum_{i=1}^{n} x_i^2\right)$$

$$\frac{\partial (\sigma^2)^{-1}}{\partial \sigma^2} = -(\sigma^2)^{-2}$$

$$\frac{n}{2} \frac{1}{\cancel{\hat{\sigma}^2}} = \frac{(\hat{\sigma}^2)^{-\cancel{2}\,1}}{2}\left(\sum_{i=1}^{n} x_i^2\right)$$

$$\hat{\sigma}^2 \cdot \frac{n}{\cancel{2}} = \sum_{i=1}^{n} x_i^2$$

$$\hat{\sigma}^2 = \frac{\cancel{2}}{n} \sum_{i=1}^{n} x_i^2 = \boxed{\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2}$$