523301
COMPUTER STATISTICS

Paramate Horkaew, PhD

Introduction

- Definition of Estimation and Estimator
- Estimation Process
- Modeling an Estimator
- Comparing Possible Estimators
- Deriving Estimator: *Maximum Likelihood*
- Determining Error Bound
- Examples
- Conclusion

Estimation Theory

Concepts

Estimate the Values of Parameters based on measured empirical data that has a random component (RV).

Estimator

A means of approximating the unknown parameters using the measurements

Estimate the time used to commute between city A & B

Estimation Process

Find an **estimator** that takes the measured data as input and produces an estimate of the parameters with the corresponding accuracy.

The Definition of Optimal Estimator

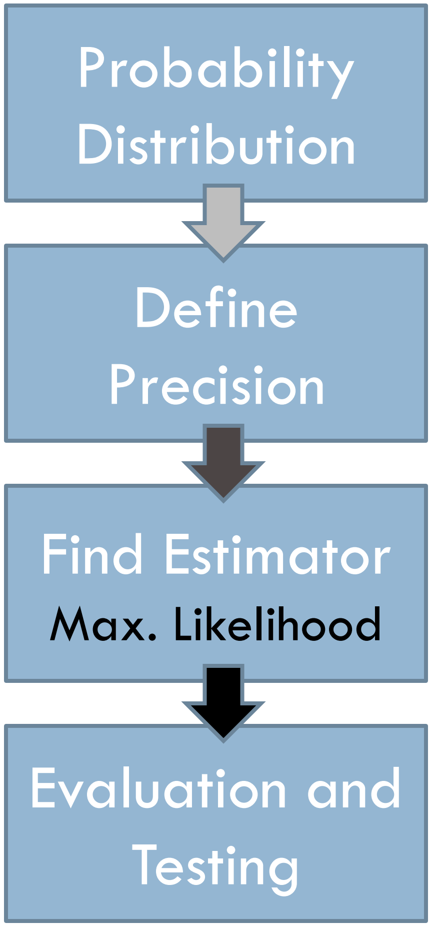
1) Minimum Average Error

over some class of estimator

2) Minimum Variance

(average squared error) …

computed between the estimated value and measured parameter



Modeling an Estimator

- Measure a set of N statistical samples taken from a random vector (RV) – $\mathbf{x}$
- Define a prob. distribution of M parameters – θ

  p $(x \mid \theta)$ maybe with their own distribution – π

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}. \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{bmatrix}, \quad \pi(\theta).$$

- Estimate θ' which minimizes (MMSE) $\mathbf{e} = \theta' - \theta$

Estimator Hypotheses

Let a discrete signal x [n] of *N* samples described by

$$x[n] = A + w[n],\ n = 0, 1, .., N-1$$

where A is unknown (to be estimated) and w [n] is a white noise defined by N (0, σ^2).

Suppose we want to compare 2 estimators:

1) $A_1 = x[0]$

2) $A_2 = (1/N)\ \Sigma_N\ x[n]$ or the sample mean

Comparing Estimators

1) Compute Expected Value

Discrete $\mathrm{E}[X] = \sum_{i=1}^{\infty} x_i\, p_i,$ Continuous $\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x.$

1) $E\,[A_1] = E\,[x\,[0]] = \mathbf{A}$

2) $E\,[A_2] = E\,[(1/N)\, \Sigma_N\, x\,[n]]$

$= (1/N)\, \Sigma_N\, E[x\,[n]] = (1/N)NA = \mathbf{A}$

2) Compute Variance

1) $\mathrm{var}\,[A_1] = \mathrm{var}\,[x\,[0]] = \boldsymbol{\sigma^2}$

2) $\mathrm{var}\,[A_2] = \mathrm{var}\,[(1/N)\, \Sigma_N\, x\,[n]]$

$= (1/N^2)\, \Sigma_N\, \mathrm{var}\,[x\,[n]] = (1/N^2)N\, \sigma^2 = \boldsymbol{\sigma^2/N}$

Maximum Likelihood (ML)

Since the pdf of the white noise is defined as

$$p(w[n]) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} w[n]^2\right)$$

By substituting w [n] with x [n] – A, we have x [n] = N (A, σ^2)

$$p(x[n]; A) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x[n] - A)^2\right)$$

If x [n] is independent, p (**x**; A) = Π p (x [n]; A)

$$p(\mathbf{x}; A) = \prod_{n=0}^{N-1} p(x[n]; A) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^N} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right)$$

GOAL Find A that maximizes p (**x**; A) (making it most probable)

Increasing Function Technique

Since evaluating *exp* is difficult, applying *ln* (an increase function) to both side

$$\ln p(\mathbf{x};A) = -N\ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

Therefore, the goal is now to find $\hat{A} = \arg\max \ln p(\mathbf{x};A)$

Taking derivative with respect to A at both sizes

$$\frac{\partial}{\partial A}\ln p(\mathbf{x};A) = \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1}(x[n]-A)\right] = \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1}x[n] - NA\right]$$

The optimal A is at critical point where its derivative is zero

$$0 = \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1}x[n] - NA\right] = \sum_{n=0}^{N-1}x[n] - NA \Rightarrow \hat{A} = \frac{1}{N}\sum_{n=0}^{N-1}x[n]$$

Cramér–Rao Lower Bound (CRLB)

Compute Fisher Information Number of the Estimator

$$\mathcal{I}(A) = \mathrm{E}\left(\left[\frac{\partial}{\partial A}\ln p(\mathbf{x};A)\right]^2\right) = -\mathrm{E}\left[\frac{\partial^2}{\partial A^2}\ln p(\mathbf{x};A)\right]$$

From the above result, we have

$$\frac{\partial}{\partial A}\ln p(\mathbf{x};A) = \frac{1}{\sigma^2}\left[\sum_{n=0}^{N-1} x[n] - NA\right]$$

Taking the second derivative to find Fisher Information

$$\frac{\partial^2}{\partial A^2}\ln p(\mathbf{x};A) = \frac{1}{\sigma^2}(-N) = \frac{-N}{\sigma^2} \quad \Rightarrow \quad -\mathrm{E}\left[\frac{\partial^2}{\partial A^2}\ln p(\mathbf{x};A)\right] = \frac{N}{\sigma^2}$$

Then the BOUND is defined as $\mathrm{var}\left(\hat{A}\right) \geq \frac{1}{\mathcal{I}}$ or $\mathrm{var}\left(\hat{A}\right) \geq \frac{\sigma^2}{N}$

An Example of Normal Distribution

Let X1, …, Xn be a random sample from a normal distribution

$$
\begin{aligned}
f(x_1, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)} \\
&= (\frac{1}{2\pi\sigma^2})^{n/2} e^{-\sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma^2)}
\end{aligned}
$$

Similar to the previous example, we take *ln* on both side

$$
ln(f(x_1, \dots, x_n; \mu, \sigma^2)) = \frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
$$

Take derivative w.r.t. μ and σ^2 to find the parameter estimation

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

Hint: d ln (x)/d(x) = 1/x

An Example of Bernoulli Distribution

Bernoulli Distribution is derive from the experiment with two possible outcomes Success (/w probability p) or Failure (/w probability $1 - p$)

It is defined as

$$f(x\,;\,p) = p^x(1-p)^{1-x} \quad \text{for } x \in \{\text{Failure }(0), \text{Success }(1)\}$$

Given N sampled data $x_1, x_2, \ldots, x_n$, the independent joint probability $L(p\,;\,x)$ and corresponding likelihood ($\ln(L)$) become,

$$\mathbf{L}(p|\mathbf{x}) = p^{x_1}(1-p)^{(1-x_1)} \cdots p^{x_n}(1-p)^{(1-x_n)} = p^{(x_1+\cdots+x_n)}(1-p)^{n-(x_1+\cdots+x_n)}.$$

$$\ln \mathbf{L}(p|\mathbf{x}) = \ln p(\sum_{i=1}^{n} x_i) + \ln(1-p)(n - \sum_{i=1}^{n} x_i) = n(\bar{x}\ln p + (1-\bar{x})\ln(1-p)).$$

Diff. w.r.t. p and = zero, then $\frac{\partial}{\partial p}\ln \mathbf{L}(p|\mathbf{x}) = n\left(\frac{\bar{x}}{p} - \frac{1-\bar{x}}{1-p}\right) = n\frac{\bar{x}-p}{p(1-p)}$

$$p = \bar{x}.$$

Exercise: Poisson Distribution

Poisson Distribution probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event. Poisson probability distribution is defined as

$$f(x; \lambda) = (\lambda^x e^{-\lambda})/x!$$

where x is the number of events in the given interval

λ is the mean number of events per interval

e.g. Average birth rate is 1.8 (λ) births/hours, the probability of observing 4 birth in an hour is $f(x = 4; 1.8) = 0.0723$

PROBLEM Use Maximum Likelihood (ML) to prove that $\lambda = (1/N)\Sigma_N\, x[n]$

Conclusion

- Definition of Estimation and Estimator
- Estimation Process
- Modeling an Estimator
- Comparing Possible Estimators
- Deriving Estimator: *Maximum Likelihood*
- Determining Error Bound
- Examples
- Conclusion