523301
COMPUTER STATISTICS

Paramate Horkaew, PhD

Introduction

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

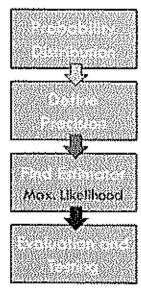
Find an **estimator** that takes the measured data as input and produces an estimate of the parameters with the corresponding accuracy.

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) – **x**
- Define a prob. distribution of *M* parameters – θ
  p (*x* | *θ*) 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) **e** = θ' – θ

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) Σ_N x [n] or the sample mean

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 $f(x = 4; 1.8) = 0.0723$

<u>PROBLEM</u> 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

523301
COMPUTER STATISTICS

Paramate Horkaew, PhD

Introduction

- Random Variable Review
- Definition of Expected Value
- Properties
- Examples
- Definition of Variance (by Expected Value)
- Expected Value of 2 Random Variables
- Chebyshev's Theorem
- Conclusion

Random Variables

Definition

A random variable **X** is a numerical measure of the outcomes of an experiment

Example

Let the experiment **X** = {0, 1, 2, 3, 4} each with equal prob. Find the prob. that P_1 (x >= 3) and P_2 (x < 3)

P_1= [N(3)+N(4)]/ N(**X**) = 2/5 = 0.4
P_2= [N(0)+N(1)+N(2)]/N(**X**) = 3/5 = 0.6

Mean of a Set of Observation

Suppose an experiment involves examination result of 3 subjects. The grade is either C (2), B (3) and A (4).

Suppose a student takes 15 subjects, and got C for 3 subj., B for 8 subj., and A for 4 subj.

Average number of grade (Grade Point Average)?

GPA = (2*3 + 3*8 + 4*4)/15 = 3.067

This can be rewritten as <u>weighted average</u>

GPA = 2*(3/15) + 3*(8/15) + 4*4(15)
= 2*P(*g* = 2) + 3*P(*g* = 3) + 4*P (*g*=*4*)

Definition of Expectation

A similar technique – "Taking the probability of an outcome times the value <u>associated</u> with the random variable for that outcome"

The sum of the these products is Expected Value of ...

Definition

An Expected Value is the value of an RV (or a function thereof) one would "expect" to find if one could repeat the random variable process an infinite number of times and take the average of the values obtained.

An Expected Value is a weighted average of all possible values

Formal Definition

If the probability distribution of **X** admits a probability density function (*pdf*) f (x), then the expected value can be computed as

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

An expected value of an arbitrary function g (x) of X is therefore

$$E[g(X)] = \mu_{g(X)} = \begin{cases} \sum_{all\ x} g(x) f(x)\ ; & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} g(x) f(x) dx\ ; & if\ X\ is\ continuous \end{cases}$$

Covariance

If X and Y are random variables with joint probability distribution f (x, y), the covariance, σ_{XY} , of X and Y is defined as

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$\sigma_{XY} = E[XY] - \mu_X \mu_Y$$

Note: It is used to look at linear relationship between two random variables. It can take negative value.

Their correlation coefficient is defined as

$$\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$$

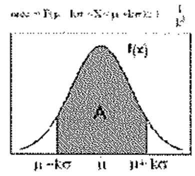
Which has the value between −1 and 1 If it is exactly −1 and 1, it means X and Y are exactly linearly related, i.e., Y = aX + b

Chebyshev's Theorem

Suppose that X is any random variable with mean E(X) = μ and variance Var (X) = σ^2 and standard deviation σ.

Chebyshev's Theorem gives a **conservative estimate** of the probability that the random variable X assumes a value within k standard deviations ($k\sigma$) of its mean μ, which is

$$P(\mu - k\sigma < X < \mu + k\sigma) \approx 1 - 1/k^2$$



If $\mu = 8$ and $\sigma = 3$, P ($-4 < X < 20$), $-4 = 8 - 3k$ and $20 = 8 + 3k$, then k = 4 therefore P (A) = 1 − 1/16
= 15/16

Conclusion

- Random Variable Review
- Definition of Expected Value
- Properties
- Examples
- Definition of Variance (by Expected Value)
- Expected Value of 2 Random Variables
- Chebyshev's Theorem
- Conclusion

523301
COMPUTER STATISTICS

Paramate Horkaew, PhD

Introduction

- Definition of Correlation
- Correlation Coefficient – *r*
- Properties of Correlation Coefficient
- Examples
- Definition of Regression
- Linear Regression and Least Squares Method
- Coefficient of Determination
- Conclusion

Correlation

Definition

Correlation is about finding (computing):

1) The Direction (positive or negative) "and"

2) The Strength (low- high) of

a relationship between 2 variables

Note: The correlation measures how **both variables** vary **jointly.** It makes **no distinction** between them.

Correlation Coefficient (1)

Let X and Y be 2 random variables of *n* samples each

The **direction** and **strength** of the relationship can be expressed by means of a correlation coefficient "**r**", which is mathematically defined as:

$$r = \frac{S_{XY}}{S_X S_Y} = \frac{SCP}{\sqrt{(SSX)(SSY)}}$$

SCP: The Sum of Cross Products of deviation

$$SCP = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}$$

Correlation Coefficient (2)

SSX: The Sum of Squares deviation of X

$$SSX = \sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

SSY: The Sum of Squares deviation of Y

$$SSY = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

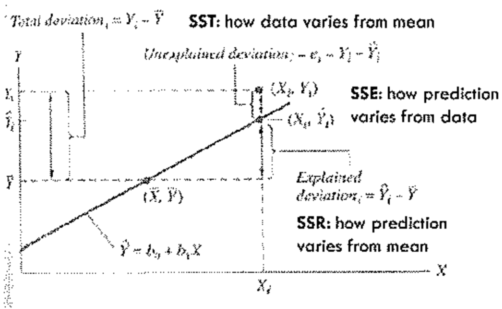
Pearson's r assumptions

The relationship is linear	Interval Measurement
P(X, Y) are normal distributed	Homogeneity of σ^2

Properties

- A correlation coefficient varies from -1 to +1
- −1 indicating a perfect **negative** relationship (one increases while the other decreases),
- 0 indicating no relationship (not correlate)
- +1 indicating a perfect **positive** relationship. (one increases with the other)
- The size of the correlation indicates the strength of the relationship:
  for example, the correlation coefficient −0.89 indicates a stronger relationship than a coefficient of +0.60.

Interpretation of Determination



Conclusion

- Definition of Correlation
- Correlation Coefficient – r
- Properties of Correlation Coefficient
- Examples
- Definition of Regression
- Linear Regression and Least Squares Method
- Coefficient of Determination
- Conclusion

523301
COMPUTER STATISTICS

Paramate Horkaew, PhD

Introduction

- Definition of Stochastic Process
- Parameters and Characterizations
- Classifications
- Second Order Process
- Examples of Stochastic Processes
- Exponential Distribution and Poisson Process
- Examples of Poisson Process
- Conclusion

Stochastic Process

Definition

Stochastic Process is a set of random variables or a set of observations, that changes over time.

Examples

Continuous-Parameter $X = \{x_t, t \geq 0\}$

Discrete-Parameter $X = \{x_n, n = 0, 1, 2, \dots\}$

Counting-Process

A discrete-valued, continuous-parameter stochastic process that increases by one each time some event occurs. The value of the process at time *t* is the number of events that have occurred up to (and including) time *t*.

Parameters and Characterizations

States are values assumed by X (*t*) or X (*n*)

State Space is a set of ALL possible **State**, called *I*.

$X(t_1)$ can be characterized by its distribution function,

$$F_{X(t_1)}(x_1) = P(X(t_1) \leq x_1)$$

Its Joint Distribution can also be determined

$$P[X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_k) \leq x_k]$$

Classification of Stochastic Process

There are 4 Classes of Stochastic Process

		time, Index set T	
		Discrete	*Continuous*
State/Sample Space I	*Discrete*	Discrete-time stochastic chain	Continuous-time stochastic chain
	Continuous	Discrete-time continuous-state process	Continuous-time continuous-state process

Discrete State Process → Chain

Discrete Time Process → Stochastic Sequence $\{X_n \mid n \in T\}$

e.g. Probing measures at every 10 ms.

Second Order Process

Let $\{X_t, t \in T\}$ be a real valued stochastic process, then (if exists, according to Komogorov)

A real-valued process $\{X_t, t \geq 0\}$ is called a second order process provided $E(X_t^2) < \infty$ for all $t \in T$. The *mean* and the *covariance function* of a second order process $\{X_t, t \geq 0\}$ are defined by

$$\begin{aligned} m_X(t) &= E(X_t) \\ \Gamma_X(s,t) &= \mathrm{Cov}(X_s, X_t) \\ &= E((X_s - m_X(s))(X_t - m_X(t))). \end{aligned}$$

The *variance* of the process $\{X_t, t \geq 0\}$ is defined by

$$\sigma_X^2(t) = \Gamma_X(t,t) = \mathrm{Var}(X_t).$$

Example 4

Customers arrive in a certain store according to a Poisson process with rate $\lambda = 4$/hour. Given that the store opens at 9:00am, Find

Let X(t) be the number of customers arrived between 9 to 9+t (t is hours)

1. The probability that exactly 1 customer has arrived by 9:30

$$P\{X(1/2) - X(0) = 1\} = \frac{(0.5\lambda)^1 e^{-0.5\lambda}}{1!} = 0.2707$$

2. (From 1), The probability that total of 5 has arrived by 11:30

$$\begin{aligned} P(X(0.5) = 1, X(2.5) = 5) &= P(X(0.5) = 1, X(2.5) - X(0.5) = 4) \\ &= P(X(0.5) = 1)P(X(2.5) - X(0.5) = 4) \\ &= \frac{(0.5\lambda)^1 e^{-0.5\lambda}}{1!} \frac{(2\lambda)^4 e^{-2\lambda}}{4!} \\ &= 0.0155 \end{aligned}$$

Example 5

The number of failures N(t), which occur in a computer network over the time interval [0, t), can be described by a Poisson process $\{ N(t), t \geq 0\}$. On an average, there is a failure after every 4 hours, i.e. $\lambda = 0.25$/hours.

What is the probability of at most 1 failure in [0, 8), at least 2 failures in [8, 16), and at most 1 failure in [16, 24) (time unit: hour)?

$$\begin{aligned} p &= P[N(8) - N(0) \leq 1, N(16) - N(8) \geq 2, N(24) - N(16) \leq 1] \\ &= P[N(8) - N(0) \leq 1]P[N(16) - N(8) \geq 2]P[N(24) - N(16) \leq 1] \\ &= P[N(8) \leq 1]P[N(8) \geq 2]P[N(8) \leq 1]. \end{aligned}$$

Since,

$$\begin{aligned} P[N(8) \leq 1] &= P[N(8) = 0] + P[N(8) = 1] \\ &= e^{-0.25 \cdot 8} + 0.25 \cdot 8 \cdot e^{-0.25 \cdot 8} = 0.406 \end{aligned}$$

$$P[N(8) \geq 2] = 1 - P[N(8) \leq 1] = 0.594.$$

⇨ Therefore,

$$\begin{aligned} p &= 0.406 \times 0.594 \times 0.406 \\ &= 0.098. \end{aligned}$$

Conclusion

- Definition of Stochastic Process
- Parameters and Characterizations
- Classifications
- Second Order Process
- Examples of Stochastic Processes
- Exponential Distribution and Poisson Process
- Examples of Poisson Process
- Conclusion