CE 513: Statistical Methods in Civil Engineering

Lectures- 10: Parameter Estimation

Dr. Budhaditya Hazra
Room: N-307
Department of Civil Engineering

Probability Paper Plot

Let X be a random variable with PDF $P_X(x)$.

Let $\{x_i\}_{i=1}^{n}$ be a sample of X.

Probability paper is a special plotting device in which y-axis is scaled in such a way that the PDF function appears as a straight line.

Example

$$P_X(x) = 1 - \exp(-\lambda x) \quad x \geq 0$$

$$1 - P_X(x) = G_X(x) = \exp(-\lambda x)$$

$$\log G_X(x) = -\lambda x$$

The complement of the cumulative PDF appears as a straight line.

Probability Paper Plot

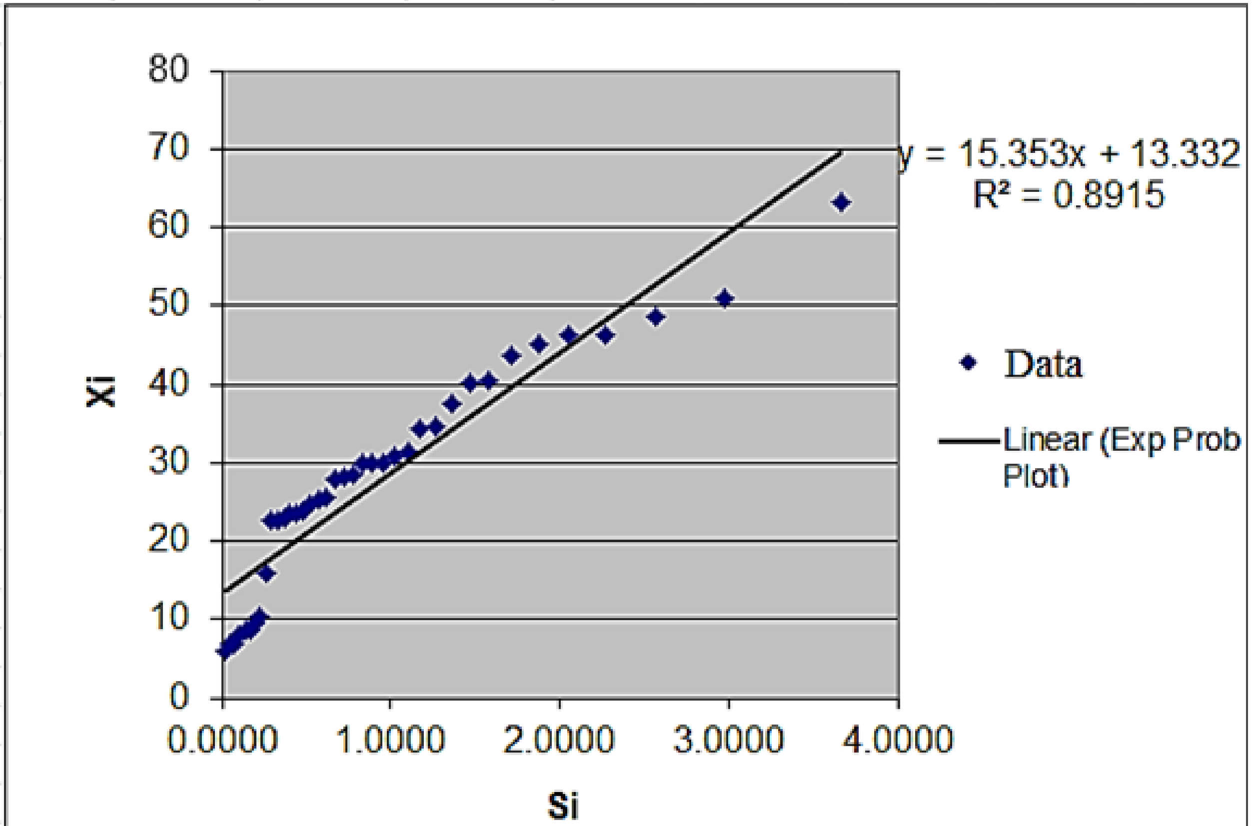
Data (in increasing order)	Rank	Pi
5.96	1	0.03
6.83	2	0.05
6.84	3	0.08
8.17	4	0.10
8.68	5	0.13
8.74	6	0.15
9.41	7	0.18
10.36	8	0.21
15.9	9	0.23
22.5	10	0.26
22.7	11	0.28
23	12	0.31
23.509	13	0.33
23.6	14	0.36
23.7	15	0.38
24.7	16	0.41
25.3	17	0.44
25.407	18	0.46
28	19	0.49
28.2	20	0.51
28.5	21	0.54
30	22	0.56
30	23	0.59
30	24	0.62

$i/(N+1)$

PP plot

Si
0.0260
0.0526
0.0800
0.1082
0.1372
0.1671
0.1978
0.2296
0.2624
0.2963
0.3314
0.3677
0.4055
0.4447
0.4855
0.5281
0.5725
0.6190
0.6678
0.7191
0.7732
0.8303
0.8910
0.9555

mu=	46954
lambda=	54398



PP plot-for practice

5.96	28
6.83	28.2
6.84	28.5
8.17	30
8.68	30
8.74	30
9.41	30.88
10.36	31.38
15.9	34.28
22.5	34.5
22.7	37.407
23	40.03
23.509	40.48
23.6	43.53
23.7	45
24.7	46.31
25.3	46.397
25.407	48.74
	50.888
	63.319

Maximum Likelihood Estimation

General Mathematical Statement of Estimation Problem:

For… Measured Data $\mathbf{x} = [\, x[0]\; x[1]\; \ldots\; x[N\text{-}1] \,]$

Unknown Parameter $\theta = [\, \theta_1 \;\; \theta_2 \;\; \ldots \;\; \theta_p \,]$

θ is Not Random

$\mathbf{x}$ is an N-dimensional random data vector

Q: What captures all the statistical information needed for an estimation problem ?

A: Need the N-dimensional PDF of the data, parameterized by θ

$$p(\mathbf{x};\theta)$$

We'll use $p(\mathbf{x};\theta)$ to find $\hat{\theta} = g(\mathbf{x})$

Maximum Likelihood Estimation

Let $f(x;\theta)$ be the density function of population X

θ is the only parameter to be estimated

from a set of sample values $x_1, x_2, \ldots, x_n$

Joint density function of the sample

$f(x_1, x_2, \ldots \ldots \ldots \ldots, x_n; \theta)$

This is in general difficult to work with

- Simplify it by making independence assumption
- Each sample is sampled independently of the others
- Each sample belongs to the same parent distribution

Joint density simplifies to $f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$

Maximum Likelihood Estimation

A better and somewhat well behaved function: *Likelihood*

We define the *likelihood function* L of a set of n sample values from the population by

$$L(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta).$$

In the case when X is discrete, we write

$$L(x_1, x_2, \ldots, x_n; \theta) = p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta).$$

- Likelihood function L is a function of a single variable θ
- Method of maximum likelihood: Comprises of choosing, as an estimate of θ, the particular value of that maximizes L

Maximum Likelihood Estimation

The maximum of $L(\theta)$ occurs at the value of θ where $\mathrm{d}L(\theta)/\mathrm{d}\theta$ is zero. Hence, in a large number of cases, the *maximum likelihood estimate* (MLE) $\hat{\theta}$ of θ based on sample values $x_1, x_2, \ldots,$ and x_n can be determined from

$$\frac{\mathrm{d}L(x_1, x_2, \ldots, x_n; \hat{\theta})}{\mathrm{d}\hat{\theta}} = 0.$$

Gaussian with known sigma

- the log-likelihood is:

$$\sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} -\frac{1}{2}(\mathbf{x}_j - \mu)^t \Sigma^{-1}(\mathbf{x}_j - \mu) - \frac{1}{2}\ln (2\pi)^d |\Sigma|$$

- The gradient wrt to the mean is:

$$\nabla_{\mu} \sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} \Sigma^{-1}(\mathbf{x}_j - \mu)$$

- Setting the gradient to zero gives:

$$\sum_{j=1}^{n} \Sigma^{-1}(\mathbf{x}_j - \mu^*) = \mathbf{0} \qquad \Rightarrow \qquad \mu^* = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$$

Gaussian with unknown mean & sigma

- the log-likelihood is:

$$\mathcal{L} = \sum_{j=1}^{n} -\frac{1}{2\sigma^2}(x_j - \mu)^2 - \frac{1}{2}\ln 2\pi\sigma^2$$

- The gradient is:

$$\nabla_{\mu,\sigma^2}\mathcal{L} = \begin{bmatrix} \sum_{j=1}^{n} \frac{1}{\sigma^2}(x_j - \mu) \\ \sum_{j=1}^{n} -\frac{1}{2\sigma^2} + \frac{(x_j-\mu)^2}{2\sigma^4} \end{bmatrix} = 0$$

$$\mu^* = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu^*)^2$$

Question: Work out the case where sigma is known and varies at each point