Probabilistic Concepts in Engineering Design

Class 8

Jan. 18th 2010

Statistics of Extremes

1. Introduction

Extreme values of random variables: largest, smallest

Statistically, these pertain to the maximum and minimum values from a set of observations.

Conceivably, if the set of observations(samples of size n) were repeated, other maximum and minimum values will be obtained; thus, the possible largest and smallest values comprise populations of their own.

→ modeled as random variables

2

Extreme values from observational data are of special importance to many engineering applications.

In the structural safety

the high loads and low resistance

Gumbel(1954)

3

2. Probability Distribution of Extremes

2.1 Extract Distributions

- The largest and smallest values from samples of size n are also random variables and therefore have probability distributions.
- These distributions can be expected to be related to the distribution of the initial variate.

Let X be the initial random variable with known initial distribution function $F_X(x)$

Consider samples of size n taken from the population (sample space) of X;

Each sample will be a set of observations(x1, x2, …….xn), representing the first, second and nth observed values.

4

Since every observed value is unpredictable prior to actual observation, we may assume that each observation is the value of random variable.

The set of observations(x1, x2, ……xn) is the realization of the sample random variables(X1, X2,…….Xn).

The largest value and the smallest value from samples of size n taken from a population X are also random variables whose probability distributions may derived from that of initial variate X.

5

If Yn, the largest among(X1,X2, .. Xn), is less than some value y, then all other sample random variables must necessarily also be less than y.

Assume that X1, X2, ……are statistically independent,

$$F_{X1}(x) = F_{X2}(x) = \ldots = F_{Xn}(x) = F_X(x)$$

On these bases,
the distribution function of Yn is

$$\begin{aligned} F_{Yn}(y) &= P(Y_n \le y) \\ &= P(X_1 \le y, X_2 \le y, \ldots X_n \le y) \\ &= [F_X(y)]^n \end{aligned}$$

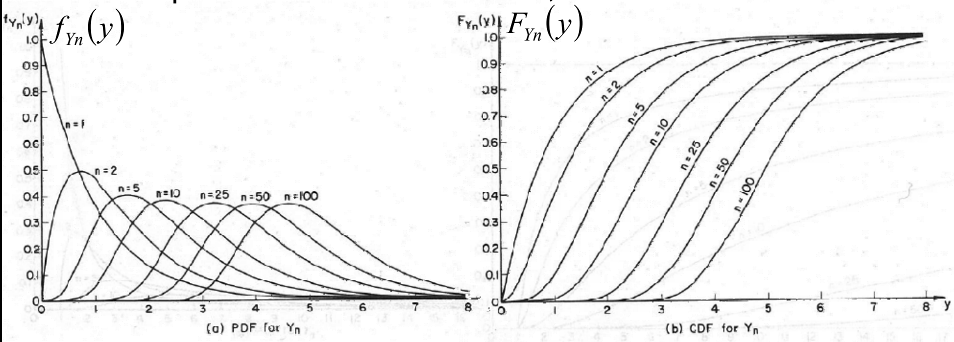
The corresponding density function for Yn is

$$f_{Yn}(y) = \frac{\partial F_{Yn}(y)}{\partial y} = n[F_X(y)]^{n-1} f_X(y)$$

6

From these equations, we see that for a given y the probability $[F_X(y)]^n$ decreases with n; this means that the functions $F_{Yn}(y)$ and $f_{Yn}(y)$ will sift to the right with increasing values of n.

For an exponential initial distribution,



7

The distribution function for Y1, the smallest value from samples of size n, can be similarly derived. In this case, we observe that if Y1, the smallest among(X1, X2, …Xn) is larger than y, then all the other values in the same sample must larger than y. Hence, the survival function is

$$
\begin{aligned}
1-F_{Y1}(y) &\equiv P(Y_1 > y) \\
&= P(X_1 > y, X_2 > y,,, X_n > y) \\
&= [1-F_X(y)]^n
\end{aligned}
$$

The distribution function of Y1 is

$$F_{Y1}(y) = 1-[1-F_X(y)]^n$$

And corresponding density function becomes

$$f_{Y1}(y) = n[1-F_X(y)]^{n-1} f_X(y)$$

8

Example - 1

Consider the initial variate X having the exponential density function as follows:

$$f_X(x) = \lambda e^{-\lambda x}; \quad x \geq 0$$

The corresponding distribution function is

$$F_X(x) = 1 - e^{-\lambda x}$$

Therefore, the largest value from samples of size n will have the distribution function

$$F_{Yn}(y) = \left(1 - e^{-\lambda y}\right)^n; \quad y \geq 0$$

The corresponding density function

$$f_{Yn}(y) = \lambda n \left(1 - e^{-\lambda y}\right)^{n-1} e^{-\lambda y}$$

9

The CDF of the smallest value from the above initial distribution is $F_{Y1}(y) = 1 - \left[1 - \left(1 - e^{-\lambda y}\right)\right]^n$

$$= 1 - e^{-n\lambda y}; \quad y \geq 0$$

The corresponding PDF is

$$f_{Y1}(y) = n\lambda e^{-n\lambda y}; \quad y \geq 0$$

10

Example - 2

Consider an initial variate with

$$f_X(x) = \frac{1}{x^2}; \quad x > 1$$
$$= 0; \quad x < 1$$

Then

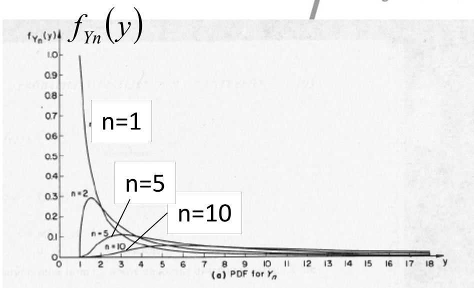
$$F_X(x) = 1 - \frac{1}{x}; \quad x \ge 1$$

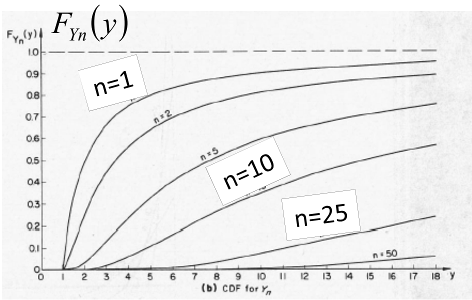
CDF

$$F_{Yn}(y) = \left(1 - \frac{1}{y}\right)^n$$

and

$$f_{Y1}(y) = \frac{n}{y^2}\left(1 - \frac{1}{y}\right)^{n-1}; \quad y \ge 1$$
$$= 0; \quad y < 1$$





11

Example - 3

Consider the initial variate X having the standard normal distribution with density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2}$$

The corresponding cumulative distribution function is

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(1/2)z^2} dz = \Phi(x)$$

The largest value from samples of size n will be a CDF given as follows:

$$F_{Y_n}(y) = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(1/2)z^2} dz\right)^n = [\Phi(x)]^n$$

and the corresponding PDF

$$f_{Y_n}(y) = \frac{n}{\sqrt{2\pi}} [\Phi(y)]^{n-1} e^{-(1/2)y^2}$$

In this case, the distribution function cannot derived analytically, will require numerical integration.

12

2.2 Asymptotic Distribution(漸近分布)

TOKYO TECH Pursuing Excellence

Characteristics of the distributions of the extremes as n become large.

Are there limiting or asymptotic forms of
as $n \to \infty$

13

3. The Three Asymptotic Forms

TOKYO TECH Pursuing Excellence

3.1 Gumbel's Classification

The asymptotic distributions of the extremes tend to converge on certain forms for large n; specifically to the double exponential form or two different single exponential forms.

Gumbel's classification

Type-1: the double exponential form, $\exp\left[-e^{-A(n)y}\right]$

Type-II: the exponential form, $\exp\left[-A(n)/y^k\right]$

Type III: the exponential form with upper bound ω, $\exp\left[-A(n)(\omega - y)^k\right]$

14

Because the extreme values of a random variable are invariably associated with the tails of its probability density function, the convergence of the distribution function of its extreme (largest or smallest) value to a particular limiting form, therefore, will **depend largely on the tail behavior of the initial distribution** in the direction of extreme.

The extreme value from an initial distribution with an exponentially decaying tail (in the direction of the extreme) will converge asymptotically to the type I limiting form.

For an initial variate that decays with a polynomial tail, its extreme value will converge to the type II.

If the extreme value has a finite lower bound, the corresponding extremal distribution will converge to the type III form.

15

3.1 The type I Asymptotic Form

The cumulative distribution function(CDF) of the type I asymptotic form for the distribution of the largest value is

$$F_{Xn}(x) = \exp\left[-e^{-\alpha_n(x-u_n)}\right]$$

where, α_n and u_n are the location and scale parameters defined as follows,

α_n : The characteristic largest value of the initial variate X
u_n : An inverse measure of dispersion of Xn

The corresponding probability density function (PDF) is

$$f_{Xn}(x) = \alpha_n e^{-\alpha_n(x-u_n)} \exp\left[-e^{-\alpha_n(x-u_n)}\right]$$

16

For the smallest value from an initial variate X with an exponential tail, the corresponding type I asymptotic form for the CDF is

$$F_{X1}(x) = 1 - \exp\left[-e^{-\alpha_1(x-u_1)}\right]$$

and the PDF is

$$f_{X1}(x) = \alpha_1 e^{-\alpha_1(x-u_1)} \exp\left[-e^{-\alpha_1(x-u_1)}\right]$$

Where the parameters are

α_1 : the characteristic smallest value of the initial variate X

u_1 : an inverse measure of dispersion of X1

17

The characteristic extremes

The characteristic largest value, u_n is a convenient measure of the central location of possible largest values.

In a sample of size n from an initial variate X, the expected number of sample values that are larger than x is

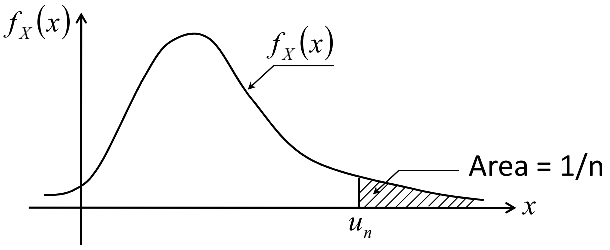
$$n[1 - F_X(x)]$$

The characteristic largest value, u_n , is defined as the particular value of X such that in a sample of size n form the initial population X, the expected number of sample values larger than is one; that is

$$n[1 - F_X(x)] = 1.0 \quad or \quad F_X(u_n) = 1 - \frac{1}{n}$$

18

In other words, u_n is the value of X with an exceedance probability of 1/n



The characteristic extremes, α_n and u_n are also the modal values(i.e., most probable values) of the respective extremal variates Xn and X1.

That is, u_n , is the most probable largest value from samples of X, and u_1 is the most probable smallest value from samples of X.

19

Standard Extremal Variates

We may introduce the standardized extremal variate

$$S = \alpha_n (X_n - u_n) \qquad \alpha_n = n f_X(u_n)$$

Its distribution(CDF) is

$$F_S(s) = \exp\left(-e^{-s}\right)$$

And its PDF is

$$F_S(s) = e^{-s} \exp\left(-e^{-s}\right)$$

The distribution of S remains of the type I asymptotic form, with parameters

$u_n = 0$ and $\alpha_n = 1.0$

The standard extremal variate serves the same purposed as the standard normal variate.

In particular, the probabilities of S may be tabulated, from which the cumulative probabilities of a general Type I extremal variate Xn may be evaluated from such tables.

20

Chap.4
Empirical Determination of Distribution Models

1. Introduction

The functional form of the required probability distribution may not be easy to derive.
How to determine the distribution models from data?

Excel?

Empirical determination method---------------probability paper

21

2. Probability Paper

Graph papers for plotting observed experimental data and their corresponding cumulative frequencies (or probabilities) are called probability papers.

Probability papers are constructed such that a given probability paper is associated with a specific probability distribution: that is, different probability papers correspond to different probability distributions.

A probability paper is constructed using a transformed probability scale in such a manner as to obtain a linear graph between the cumulative probabilities of the underlying distribution and the corresponding values of the variate.

22

For example, in the case of the uniform distribution, the cumulative distribution function is linearly related to the values of the variate.

Experimental data may be plotted on any probability paper; the plotting position of each data point is determined as follows,

If there are N observations x1, x2, … x_N, the m_{th} value among the N observations(arranged in increasing order) is plotted at the cumulative probability

m/(N+1)

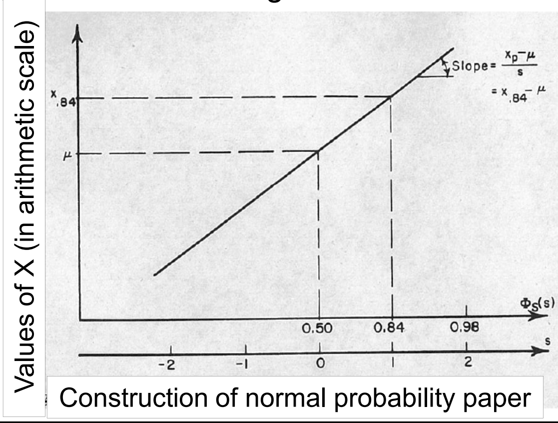
Gumbel(1954)

23

2.1 The Normal Probability Paper

The normal (or Gaussian) probability paper is constructed on the basis of the standard normal distribution function, as follows.

One axis (in arithmetic scale) represents the values of the variate X, as illustrated in Figure.

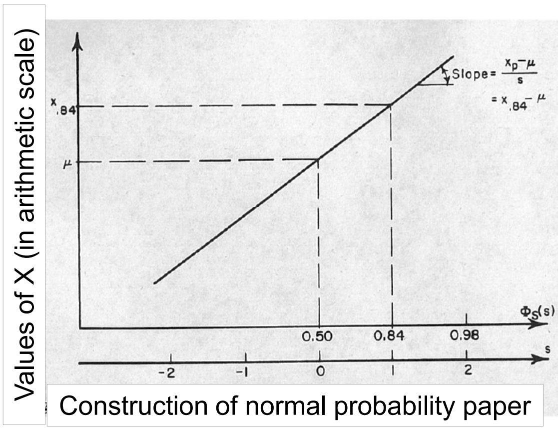


Construction of normal probability paper

24

TOKYO TECH
Pursuing Excellence

On the other axis are two parallel lines; one in arithmetic scale represents values of standard normal variate s,
whereas the other shows the cumulative probabilities Φs（s）corresponding to the indicated values of **s** as shown in figure.



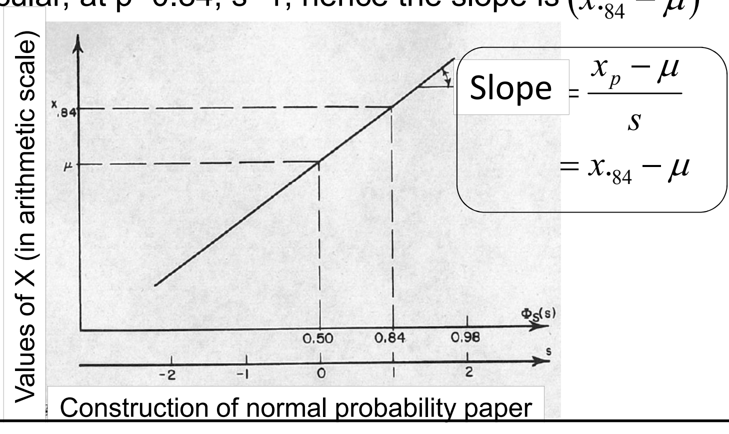
Construction of normal probability paper

25

TOKYO TECH
Pursuing Excellence

A normal variate X with distribution N(μ、σ) would then be represented on this paper by a straight line passing throughΦs（s）＝0.50, and X=μ, with a slope $\left(x_p-\mu\right)/s=\sigma$

Where, xp is the value of the variate at probability p.
In particular, at p=0.84, s=1; hence the slope is $\left(x_{.84}-\mu\right)$



Construction of normal probability paper

26

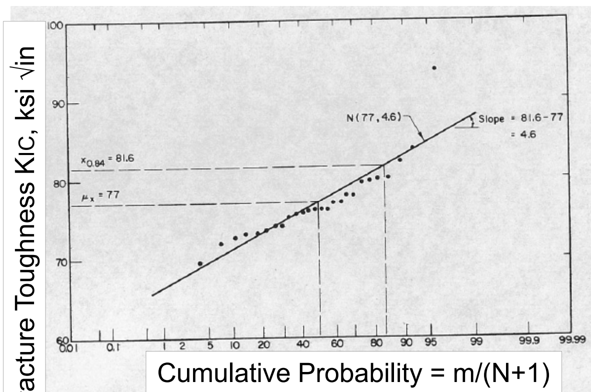
Such normal probability papers are available commercially.

Any set of data may be plotted on the normal probability paper; however, if the resulting graph of data plots shows a lack of linearity, this would suggest that the underlying population is not Gaussian.

27

Example - 1

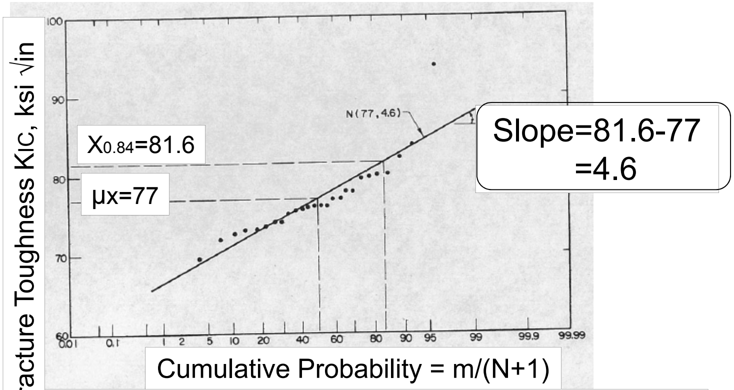
The data for fracture toughness of steel plate, given in the table is plott3d on the normal probability paper in Fig. E6.1, values of the fracture toughness Kc are plotted against the plotting positions m/(N+1), with N=26.



Fracture toughness plotted on normal probability paper

28

The straight line shown in fig.E6.1. is drawn by eye through the data points, from which we find μ＝77
also, the value of Kc at the 84% probability level is 81.6, thus σ=81.6-77=4.6 $ksi\sqrt{in}$



Fracture toughness plotted on normal probability paper

Table E6.1. Fracture Toughness of Base Plate of 18% Nickel Maraging Steel (Data From Kies et al., 1965)

m	K_{Ic} (ksi $\sqrt{\text{in.}}$)	$\frac{m}{N+1}$
1	69.5	0.0370
2	71.9	0.0741
3	72.6	0.1111
4	73.1	0.1418
5	73.3	0.1852
6	73.5	0.2222
7	74.1	0.2592
8	74.2	0.2963
9	75.3	0.3333
10	75.5	0.3704
11	75.7	0.4074
12	75.8	0.4444
13	76.1	0.4815
14	76.2	0.5185
15	76.2	0.5556
16	76.9	0.5926
17	77.0	0.6296
18	77.9	0.6667
19	78.1	0.7037
20	79.6	0.7407
21	79.7	0.7778
22	79.9	0.8148
23	80.1	0.8518
24	82.2	0.8889
25	83.7	0.9259
26	93.7	0.9630

TOKYO TECH
Pursuing Excellence

2.2 The Log-Normal Probability Paper

The log-normal probability paper can be obtained from the normal probability paper by simply changing the arithmetic scale for values of the variate X(on the normal probability paper) to a logarithmic scale.

In this case, the standard normal variate becomes $S = \dfrac{\ln(X / x_m)}{\zeta}$
where x_m is the median of X

31

Example - 2

Data for the fracture toughness of MIG welds are tabulated in table E6-2.

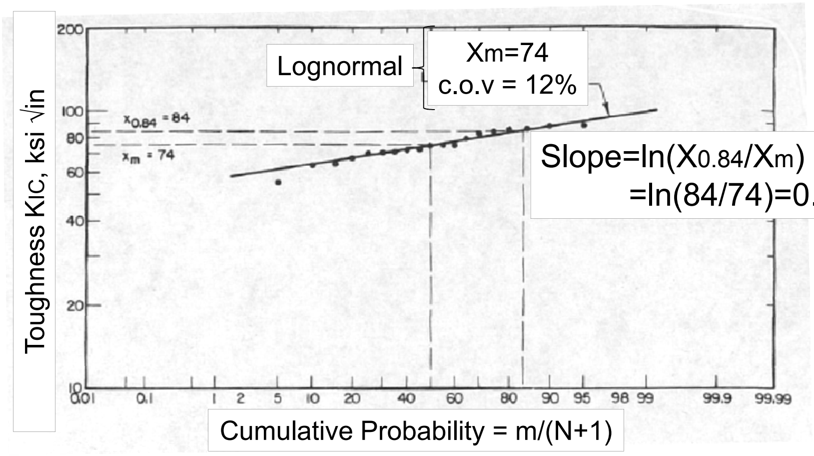
Table E6.2. Fracture Toughness of MIG Welds (Data from Kies et al., 1965)

m	K_{Ic} (ksi$\sqrt{\text{in.}}$)	$\dfrac{m}{N+1}$
1	54.4	0.05
2	62.6	0.10
3	63.2	0.15
4	67.0	0.20
5	70.2	0.25
6	70.5	0.30
7	70.6	0.35
8	71.4	0.40
9	71.8	0.45
10	74.1	0.50
11	74.1	0.55
12	74.3	0.60
13	78.8	0.65
14	81.8	0.70
15	83.0	0.75
16	84.4	0.80
17	85.3	0.85
18	86.9	0.90
19	87.3	0.95

32

TOKYO TECH
Pursuing Excellence

On the basis of linear graph of the plotted data shown in fig.E6.2, we may say that the fracture toughness of such welds has a log-normal distribution with a median of 74 and cov of 12%.



33

TOKYO TECH
Pursuing Excellence

Table E6.2. Fracture Toughness of MIG Welds (Data from Kies et al., 1965)

m	K_{Ic} (ksi$\sqrt{\text{in.}}$)	$\frac{m}{N+1}$
1	54.4	0.05
2	62.6	0.10
3	63.2	0.15
4	67.0	0.20
5	70.2	0.25
6	70.5	0.30
7	70.6	0.35
8	71.4	0.40
9	71.8	0.45
10	74.1	0.50
11	74.1	0.55
12	74.3	0.60
13	78.8	0.65
14	81.8	0.70
15	83.0	0.75
16	84.4	0.80
17	85.3	0.85
18	86.9	0.90
19	87.3	0.95

34

2.3 Construction of General Probability Paper

Probability papers are constructed in such a way that the values of the variate and the associated cumulative probabilities yield a straight line.

Conversely, therefore, a straight line on a specific probability paper represents a particular distribution (consistent with that of the probability paper) with given values of the parameters.

For this purpose, a probability paper should be constructed so that it is independent of the values of the parameters of the distribution.

This is accomplished by defining a standard variate (if one exists) appropriate for the given distributions.

35

Example - 3

The density function of the shifted exponential distribution is $f_X(x) = \lambda e^{-\lambda(x-a)}\ ; x \geq a$

$$= 0\ ; \qquad x < a$$

Where λ is the parameter, and a is the minimum value of X. In this case, the standard variate is $S = \lambda(X - a)$

The density function of *S* is

$$f_S(s) = f_X\left(\frac{s}{\lambda} + a\right)\left|\frac{1}{\lambda}\right| = e^{-s};\ s \geq 0$$

$$= 0; \qquad s < 0$$

Corresponding CDF is

$$F_S(s) = 1 - e^{-s}; \qquad s \geq 0$$

36

TOKYO TECH
Pursuing Excellence

On this basis, we construct the exponential probability paper as follows,
On one axis, scale values of the standard variates (in arithmetic scale);
On the same or parallel axis, mark the corresponding cumulative probabilities

$$F_S(s) = 1 - e^{-s}$$

The other perpendicular axis will represent values of the variate X (in arithmetic scale)

For illustration, specific values of s and FS(s) have been calculated as summarized in Table E6.4a.

37

TOKYO TECH
Pursuing Excellence

Drawing grid lines for given FS(s) at the indicated values of **s** shown in table E6.4a, we obtain the resulting paper as shown in fig.E6.4a.

Table E6.4.a Specific Values of s and $F_S(s)$

s	$F_S(s)$	s	$F_S(s)$
0.11	0.10	2.53	0.92
0.22	0.20	2.66	0.93
0.36	0.30	2.81	0.94
0.51	0.40	3.00	0.95
0.69	0.50	3.10	0.955
0.80	0.55	3.22	0.96
0.92	0.60	3.35	0.965
1.05	0.65	3.51	0.97
1.20	0.70	3.69	0.975
1.39	0.75	3.91	0.98
1.61	0.80	4.20	0.985
1.90	0.85	4.61	0.99
2.30	0.90		

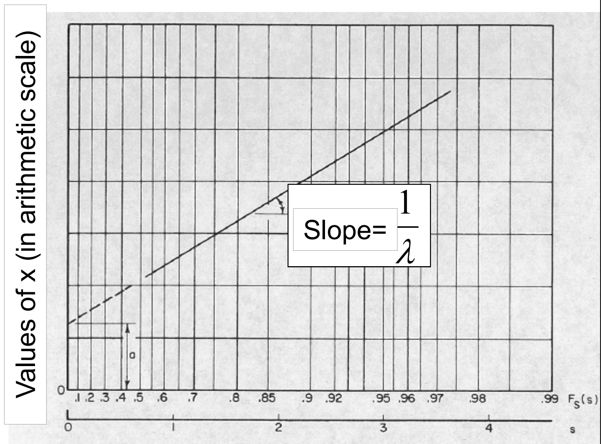


Figure E6.4a Construction of the exponential probability paper

38

Gumbel probability paper: type I asymptotic distribution of extremes

Its CDF for the largest value is given by the double exponential function.

$$F_X(x) = \exp\left[-e^{-\alpha(x-u)}\right] \quad -\infty < x < \infty$$

in which **u** is the characteristic largest value, and 1/α is a measure of dispersion.

In this case, the standard variate can be defined as

$$S = \alpha(X - u)$$

Then,

$$F_S(s) = \exp\left(-e^{-s}\right)$$

39

Table E6.5. Specific Values of s and $F_S(s)$

s	$F_S(s)$	s	$F_S(s)$	s	$F_S(s)$
−1.53	0.01	0.37	0.50	2.48	0.92
−1.10	0.05	0.51	0.55	2.62	0.93
−0.83	0.10	0.67	0.60	2.78	0.94
−0.64	0.15	0.84	0.65	2.97	0.95
−0.48	0.20	1.03	0.70	3.08	0.955
−0.33	0.25	1.25	0.75	3.20	0.96
−0.19	0.30	1.50	0.80	3.33	0.965
−0.05	0.35	1.82	0.85	3.49	0.97
0.09	0.40	2.25	0.90	3.68	0.975
0.23	0.45	2.36	0.91	3.90	0.98

Using the specific values of **s** and corresponding probabilities FS(s) calculated as summarized in Table E6.5., we constructed the Gumbel probability paper as follows,

40

Scale the values of **s** on one axis as shown in fig. E6.5. The other axis in Fig.E6.5. represents values of X, in arithmetic scale.

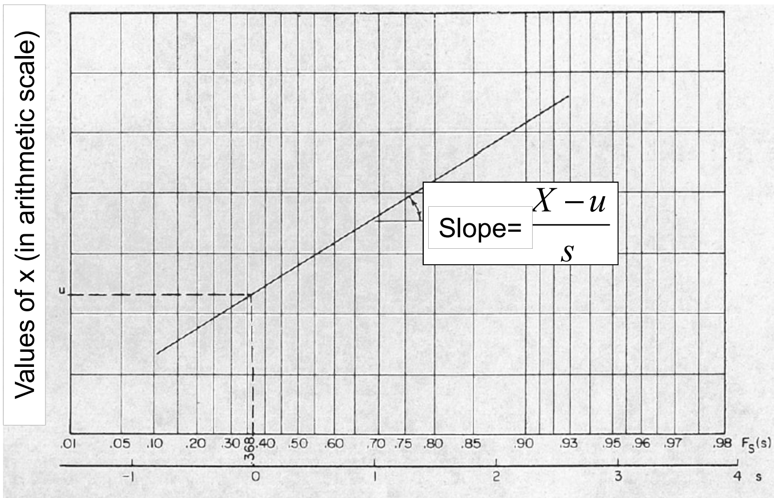


Figure E6.5 Construction of Gumbel probability paper

The straight line on this paper represents a particular type I extremal distribution—
The value of X on this line is the characteristic largest value **u**, whereas the slope of this line is α.

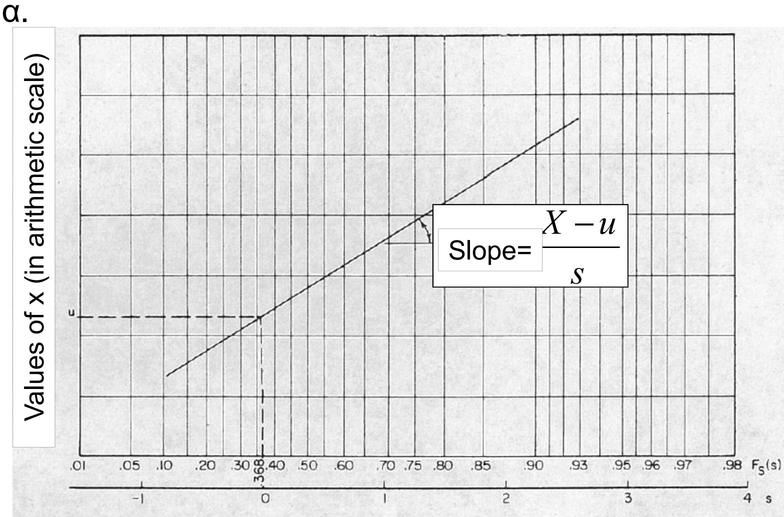


Figure E6.5 Construction of Gumbel probability paper

END

43