Extreme Value Distributions

The Gumbel Distribution

© D. L. Bricker
Dept of Mechanical & Industrial Engineering
The University of Iowa

Asymptotic Distributions

From the *Central Limit Theorem* we know that, for "large" n,

$Y = \sum_{i=1}^{n} X_i$ has approximately a *Normal* distribution

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 \right\}$$

$Y = \prod_{i=1}^{n} X_i$ has approximately a *Lognormal* distribution

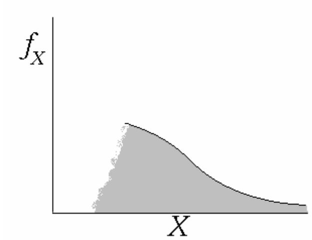
$$f_Y(y) = \frac{1}{y\sigma_X\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{\ln\left(y/\mu\right)}{\sigma}\right)^2 \right\}$$

Asymptotic Distributions

Is there such a limiting distribution for

$$Y = \max\{X_1, X_2, \ldots X_n\}$$

as n $\rightarrow\infty$?



If the right tail of each density function f_X falls off in an "exponential manner", then Y has approximately a **Gumbel** distribution.

Examples of such distributions:

exponential, Erlang, Gamma, Normal, ...

GUMBEL

Consider $Y = \max_i \{X_i\}$

where each X_i has a CDF of the form

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

and g(x) is an increasing function of x.

For example, if *exponential* distribution, $g(x) = \lambda x$.

The asymptotic distribution of Y as n→∞ is known as the *Gumbel* distribution, with

CDF	$F_Y(y) = \exp\{-\exp[-\alpha(y-u)]\}$
density	$f_Y(y) = \alpha \exp\{-\alpha(y-u) - \exp[-\alpha(y-u)]\}$
mean value	$\mu_Y = u + \frac{0.577}{\alpha}$
standard deviation	$\sigma_Y = \frac{1.282}{\alpha}$

The parameters α & u of the Gumbel distribution may be estimated by several methods, e.g.

- Method of Moments, i.e., matching the sample mean and standard deviation
- Linear regression

EXAMPLE: METHOD OF MOMENTS

An engineer has good estimates for the mean and standard deviation of the peak annual flow *Y* in a small stream.

Mean: $\mu_Y = 100 \text{ cfs } (\text{ft}^3/\text{sec})$

Std Deviation: $\sigma_Y = 50 \text{ cfs}$

Since

peak annual flow *Y* = maximum of 365 daily flows,

he expects that *Y* has approximately a *Gumbel* distribution.

What is the probability that, next year, the annual flow exceeds 200 cfs?

To estimate the parameters α & u of the Gumbel distribution, we use the relationships

$$\left.\begin{aligned} \sigma_Y &= \frac{1.282}{\alpha} \Rightarrow \alpha = \frac{1.282}{\sigma_Y} = 0.0256 \\ \mu_Y &= u + \frac{0.577}{\alpha} \Rightarrow u = \mu_Y - \frac{0.577}{\alpha} \end{aligned}\right\} \Rightarrow u = 77.5 \text{ cfs}$$

So the peak annual flow is assumed to have the distribution

$$F_Y(y) = \exp\{-\exp[-\alpha(y-u)]\}, \alpha = 0.0256, u = 77.5$$

We can now compute the probability that the peak annual flow next year will exceed 200 *cfs*:

$$\begin{aligned} P\{Y \geq 200\} &= 1 - F_Y(200) \\ &= 1 - \exp\{-\exp[-0.0256(200-77.5)]\} \\ &= 0.043 \end{aligned}$$

That is, a flow exceeding 200 *cfs* will occur every

$1/0.043$ years ≈ 23 years.

(Example, continued)

The so-called ***"hundred-year flood"*** is the value of y for which $P\{Y \geq y\} = 1 - F_Y(y) = 0.01$

i.e.,

$$\exp\{-\exp[-0.0256(y-77.5)]\} = 0.99$$

$$-\exp[-0.0256(y-77.5)] = \ln 0.99 = {}^{-}0.01005033585$$

$$-0.0256(y-77.5) = \ln\ 0.01005033585 = {}^{-}4.600149227$$

$$y - 77.5 = \frac{4.600149227}{0.0256} = 179.6933292$$

$$y = 77.5 + 179.6933292 = 257.193\ (cfs)$$

Example The annual maximum rate of flow of a particular river has a mean of 10K cfs with standard deviation of 3K cfs. Assume that this maximum rate of flow has a Gumbel distribution.

- ❑ *What are the parameters of this distribution?*
- ❑ *Compute P[annual max flowrate ≥ 15K cfs]*

- Find an expression for the CDF of the river's maximum flow rate over the 20 year lifetime of an anticipated flood-control project. (Assume that the individual annual max flow rates are i.i.d. with Gumbel distribution, as before.)
- Compute P{20-year maximum flow rate ≥ 15K cfs}

Parameters of the Distribution

$$\sigma_Y \approx \frac{1.282}{\alpha} \Rightarrow \alpha \approx \frac{1.282}{\sigma_Y} = \frac{1.282}{3} = 0.4273333$$

$$\mu_Y \approx u + \frac{0.577}{\alpha} \Rightarrow u \approx \mu_Y - \frac{0.577}{\alpha} = 10 - \frac{0.577}{0.42733}$$

$$u = 8.649766$$

$$F_Y(y) = \exp\left[-e^{-\alpha(y-u)}\right], \ \alpha = 0.4273, \ u = 8.64977$$

P{annual max flowrate ≥ 15K cfs} = ?

t	F(t)	1-F(t)
10	0.57030	0.42969
11	0.69330	0.30669
12	0.78748	0.21251
13	0.85570	0.14429
14	0.90335	0.09664
15	0.93585	0.06414
16	0.95768	0.04231
17	0.97219	0.02780
18	0.98177	0.01822
19	0.98807	0.01192
20	0.99220	0.00779

The annual peak flow will exceed 15K cfs with probability approximately 6.4%

Let $Y = \text{Max}\{X_1, X_2, \ldots X_{20}\}$,
where X_i = peak flowrate in year i.
Each random variable X_i is assumed to have a Gumbel distribution:

$$F_X(t) = \exp\left[-e^{-0.4273(t-8.64977)}\right]$$

The 20-year maximum flowrate will therefore have the CDF:

$$F_Y(t) = [F_X(t)]^{20} = \left\{\exp\left[-e^{-0.4273(t-8.64977)}\right]\right\}^{20}$$

P{20-year maximum flow rate ≥ 15K cfs}

$$= 1 - F_Y(15)$$
$$= 1 - [F_X(15)]^{20}$$
$$= 1 - [0.93585]^{20}$$
$$= 1 - 0.26557$$
$$= 0.73443$$

Example

It has been verified experimentally that the velocity of an arbitrary wind gust has an exponential distribution, and hence a rapid convergence to a Gumbel distribution should be expected for the maximum gust velocity occuring during a thunder-storm.
This maximum gust velocity has an estimated mean of 15.6 ft/sec. with a standard deviation of 6.2 ft/sec.

What is the probability that the maximum gust velocity during a thunderstorm exceeds 30 mph?

What is the probability that the maximum gust velocity will be less than 10 mph?

Gumbel Distribution

Parameters: U= 12.809516, alpha= 0.20677419

t	F(t)	1-F(t)
5	0.006559	0.993440
10	0.167342	0.832657
15	0.529533	0.470466
20	0.797643	0.202356
25	0.922742	0.077257
30	0.971810	0.028189
35	0.989882	0.010117
40	0.996390	0.003609
45	0.998714	0.001285
50	0.999542	0.000457

P{Y ≤ 10} = 16.7%

P{Y ≥ 30} = 2.8%

Estimating the Gumbel parameters by Linear Regression

Suppose that we have sample observations $\{Y_1, Y_2, Y_3, \ldots Y_k\}$ of a random variable with Gumbel distribution.

Assume that they have been ordered so that

$$Y_1 \leq Y_2 \leq Y_3 \leq \ldots \leq Y_k$$

We would like to find the "best fit" of the function

$$F_Y(y) = \exp\{-\exp[-\alpha(y-u)]\}$$

to the data.

The α & u which best fit the CDF

$$F_Y(y) = \exp\{-\exp[-\alpha(y-u)]\}$$

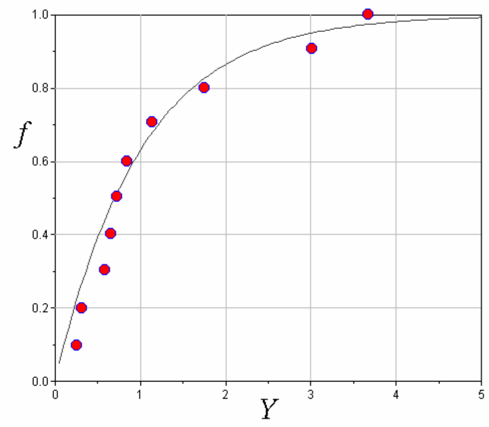
to the data should approximately satisfy the (*nonlinear*) equations:

$$\begin{cases} 1/k \approx \exp\{-\exp[-\alpha(Y_1-u)]\} \\ 2/k \approx \exp\{-\exp[-\alpha(Y_2-u)]\} \\ \vdots \\ k/k \approx \exp\{-\exp[-\alpha(Y_k-u)]\} \end{cases}$$

In order to use *linear* regression, we need to transform these equations so that they are *linear* in the unknown parameters α & u.

For example, let $f_i = i/k$ where k=10 is the number of observations. We plot the points (f_i, Y_i) and try to find a "best fit" of a curve of the required form

$$f_i = \exp\{-\exp[-\alpha(Y_i-u)]\}$$



through the data points, by choosing parameters α & u.

(a difficult task, but possible by using a nonlinear minimization algorithm!)

Linearization of the nonlinear curve:

$$f_i = \exp\{-\exp[-\alpha(Y_i-u)]\} \Rightarrow \ln f_i = -\exp[-\alpha(Y_i-u)]$$

$$\ln f_i = -\exp[-\alpha(Y_i-u)] \Rightarrow \ln(-\ln f_i) = -\alpha(Y_i-u)$$

$$\ln(-\ln f_i) = -\alpha(Y_i-u) \Rightarrow -\ln(-\ln f_i) = \alpha Y_i - \alpha u$$

Therefore, we plot the points $(-\ln[-\ln f_i],\ Y_i)$ and fit a straight line $\boxed{-\ln(-\ln f_i) = \alpha Y_i - \alpha u}$ through the points.

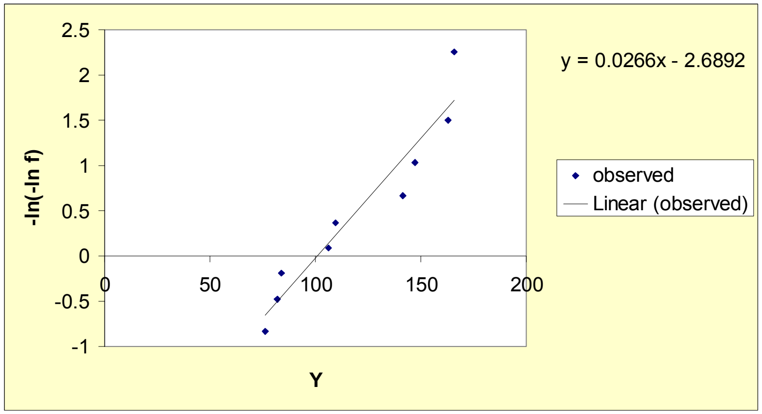
The *slope* of the line is the parameter α, and the *intercept* is –αu.

Example:

Y	F	ln F	–ln(-ln F)
125.1585	0.1	-2.30259	-0.834030
128.8174	0.2	-1.60944	-0.475880
136.6053	0.3	-1.20397	-0.185630
145.6062	0.4	-0.91629	0.087422
165.8850	0.5	-0.69315	0.366513
175.2260	0.6	-0.51083	0.671727
175.3826	0.7	-0.35667	1.030930
176.7497	0.8	-0.22314	1.499940
187.1741	0.9	-0.10536	2.250367
194.4379	1.0	0	*undefined*

Using the first 9 data points, we perform a linear regression, with *–ln(–ln F)* as the *dependent* variable and *Y* as the *independent* variable.

Results:



Note that the equation is really

$$-\ln(-\ln F) = 0.0266Y - 2.6892$$

Comparing

$$-\ln(-\ln F) = 0.0266Y - 2.6892$$

with

$$-\ln(-\ln f_i) = \alpha Y_i - \alpha u,$$

we conclude that

$\alpha = 0.0266$ and

$u = 2.6892/\alpha = 101.10$