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

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

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

Is there such a limiting distribution for $Y = \max \{X_1, X_2, \ldots, X_n\}$ as $n \to \infty$?

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, …
Consider 

\[ Y = \max \{ X_i \} \]

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 \( n \to \infty \) is known as the Gumbel distribution, with

<table>
<thead>
<tr>
<th>CDF</th>
<th>( F_Y(y) = \exp \left( -\exp \left( -\alpha (y - u) \right) \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>density</td>
<td>( f_Y(y) = \alpha \exp \left[ -\alpha (y - u) - \exp \left( -\alpha (y - u) \right) \right] )</td>
</tr>
<tr>
<td>mean value</td>
<td>( \mu_Y = u + \frac{0.577}{\alpha} )</td>
</tr>
<tr>
<td>standard deviation</td>
<td>( \sigma_Y = \frac{1.282}{\alpha} )</td>
</tr>
</tbody>
</table>

The parameters \( \alpha \) & \( u \) 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: \( \mu_Y = 100 \text{ cfs (ft}^3 / \text{sec)} \)

Std Deviation: \( \sigma_Y = 50 \text{ cfs} \)

Since

\[ Y = \text{maximum of 365 daily flows,} \]

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 $\alpha$ & $u$ of the Gumbel distribution, we use the relationships

\[
\begin{align*}
\frac{\sigma}{\alpha} &= \frac{1.282}{\sigma} = 0.0256 \\
\Rightarrow \quad \alpha &= \frac{1.282}{0.0256} \\
\mu &= u + \frac{0.577}{\alpha} \Rightarrow u = \mu - \frac{0.577}{\alpha}
\end{align*}
\]

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:

\[
P\{Y \geq 200\} = 1 - F_Y(200) = 1 - \exp\{-\exp[-0.0256(200-77.5)]\} = 0.043
\]

That is, a flow exceeding 200 cfs will occur every \(1/0.043 \approx 23\) years.

(Example, continued)

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

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

\[
-\exp[-0.0256(y-77.5)] = \ln 0.99 = 0.01005033585
\]

\[
y - 77.5 = \frac{4.600149227}{0.0256} = 179.6933292
\]

\[
y = 77.5 + 179.6933292 = 257.193 \text{ (cfs)}
\]

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

\[
\sigma_Y = \frac{1.282}{\alpha} \Rightarrow \alpha = \frac{1.282 \sigma_Y}{3} = 0.4273333
\]

\[
\mu_Y = u + \frac{0.577}{\alpha} \Rightarrow u = \mu_Y - \frac{0.577}{\alpha} = 10 - \frac{0.577}{0.4273333} = 8.649766
\]

---

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

\[P(\text{annual max flowrate } \geq 15\text{K cfs}) = ?\]

<table>
<thead>
<tr>
<th>t</th>
<th>F(t)</th>
<th>1-F(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.57030</td>
<td>0.42969</td>
</tr>
<tr>
<td>11</td>
<td>0.69330</td>
<td>0.30669</td>
</tr>
<tr>
<td>12</td>
<td>0.78748</td>
<td>0.21251</td>
</tr>
<tr>
<td>13</td>
<td>0.85570</td>
<td>0.14429</td>
</tr>
<tr>
<td>14</td>
<td>0.90335</td>
<td>0.09664</td>
</tr>
<tr>
<td>15</td>
<td>0.93565</td>
<td><strong>0.06414</strong></td>
</tr>
<tr>
<td>16</td>
<td>0.95768</td>
<td>0.04232</td>
</tr>
<tr>
<td>17</td>
<td>0.97219</td>
<td>0.02780</td>
</tr>
<tr>
<td>18</td>
<td>0.98177</td>
<td>0.01822</td>
</tr>
<tr>
<td>19</td>
<td>0.98807</td>
<td>0.01192</td>
</tr>
<tr>
<td>20</td>
<td>0.99220</td>
<td>0.00779</td>
</tr>
</tbody>
</table>

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 \( X_i = \text{peak flow rate in year } i \).

Each random variable \( X_i \) is assumed to have a Gumbel distribution:

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

The 20-year maximum flow rate will therefore have the CDF:

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

**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 occurring during a thunderstorm.  

This maximum gust velocity has an estimated mean of 15.5 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?*
The $\alpha$ & $u$ which best fit the CDF

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

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

$$\frac{1}{k} = \exp\left\{-\exp\left[-\alpha(Y_i-u)\right]\right\}$$

$$\frac{2}{k} = \exp\left\{-\exp\left[-\alpha(Y_2-u)\right]\right\}$$

$$\vdots$$

$$\frac{k}{k} = \exp\left\{-\exp\left[-\alpha(Y_k-u)\right]\right\}$$

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

**Linearization of the nonlinear curve:**

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

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

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

Therefore, we plot the points $(-\ln(-\ln f_i), Y_i)$ and fit a straight line $\ln(-\ln f_i) = \alpha Y_i - \alpha u$ through the points. The slope of the line is the parameter $\alpha$, and the intercept is $-\alpha u$.

For example, let $f_i = \frac{i}{k}$

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

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

through the data points, by choosing parameters $\alpha$ & $u$.

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

**Example:**

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$F$</th>
<th>$\ln F$</th>
<th>$-\ln(-\ln F)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>125.1585</td>
<td>0.1</td>
<td>-2.30259</td>
<td>-0.834030</td>
</tr>
<tr>
<td>128.8174</td>
<td>0.2</td>
<td>-1.60944</td>
<td>-0.475880</td>
</tr>
<tr>
<td>136.6053</td>
<td>0.3</td>
<td>-1.20397</td>
<td>-0.185630</td>
</tr>
<tr>
<td>145.6062</td>
<td>0.4</td>
<td>-0.91629</td>
<td>0.087422</td>
</tr>
<tr>
<td>165.8850</td>
<td>0.5</td>
<td>-0.69315</td>
<td>0.366513</td>
</tr>
<tr>
<td>175.2260</td>
<td>0.6</td>
<td>-0.51083</td>
<td>0.671727</td>
</tr>
<tr>
<td>175.3826</td>
<td>0.7</td>
<td>-0.35667</td>
<td>1.030930</td>
</tr>
<tr>
<td>176.7497</td>
<td>0.8</td>
<td>-0.22314</td>
<td>1.499940</td>
</tr>
<tr>
<td>187.1741</td>
<td>0.9</td>
<td>-0.10536</td>
<td>2.250367</td>
</tr>
<tr>
<td>194.4379</td>
<td>1.0</td>
<td>0</td>
<td>undefined</td>
</tr>
</tbody>
</table>

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:**

\[ y = 0.0266x - 2.6892 \]

Note that the equation is really

\[ -\ln(-\ln F) = 0.0266Y - 2.6892 \]

Comparing

\[ -\ln(-\ln F) = 0.0266Y - 2.6892 \]

with

\[ -\ln(-\ln f_i) = \alpha Y_i - \alpha u_i \]

we conclude that

\[ \alpha = 0.0266 \text{ and } u = \frac{2.6892}{\alpha} = 101.10 \]