Extreme Value Distributions

- Weibull Distribution
- Gumbel Distribution

In the design of many engineering systems, we are concerned with the largest or smallest of a set of random variables:

- maximum demand on a production system
- maximum stress on a system component
- maximum flood level
- maximum floor loadings
- minimum strength of system components

... etc.

Let $X_1, X_2, \ldots X_n$ be a set of random variables,
and $Y = \max\{X_1, X_2, \ldots X_n\}$

Then the CDF (cumulative distribution function) of Y is $F_Y(y) = P\{Y \le y\} = P\{X_i \le y \text{ for } i=1,2,..n\}$

If the X_i's are ***independent***,

$$F_Y(y) = P\{X_1 \le y \ \&\ X_2 \le y \ \&\ \ldots \ \&\ X_n \le y\}$$

$$= P\{X_1 \le y\} \times P\{X_2 \le y\} \times \cdots \times P\{X_n \le y\}$$

$$= F_{X_1}(y) \times F_{X_2}(y) \times \cdots \times F_{X_n}(y)$$

and if the X_i's are also ***identically distributed***,

$$F_Y(y) = \left(F_X(y)\right)^n$$

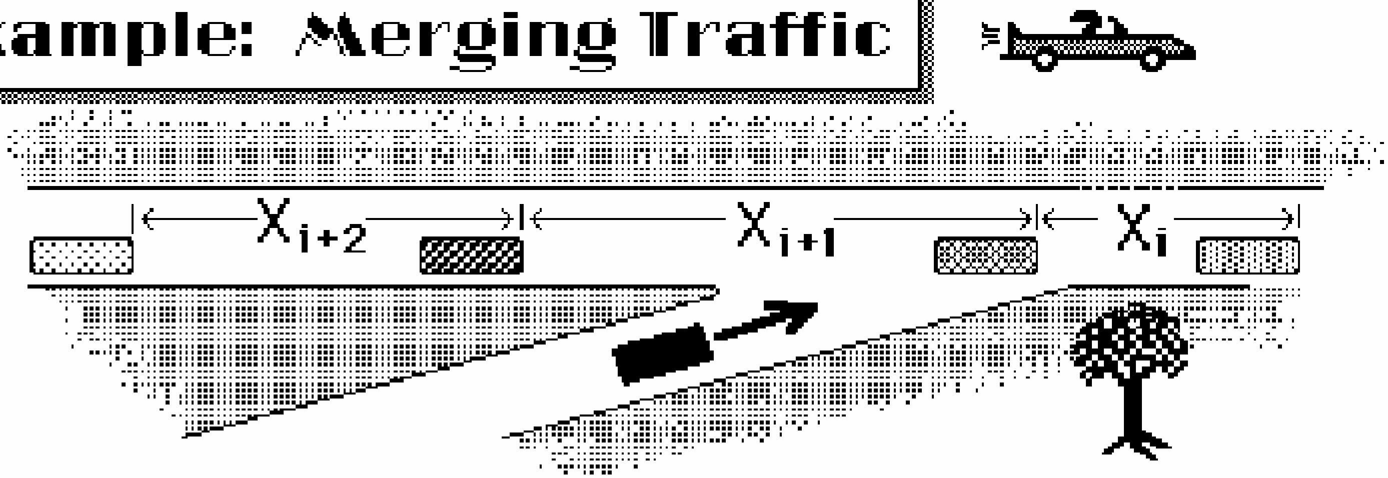
If X_i are continuous, independent, & identically-distributed random variables with density function $f_X(y)$, then
the ***density function*** for $Y = \max\{X_i\}$ is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [F_X(y)]^n$$

$$= n\,[F_X(y)]^{n-1} \frac{d}{dy} F_X(y)$$

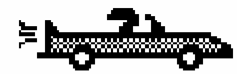
$$= n\,[F_X(y)]^{n-1} f_X(y)$$

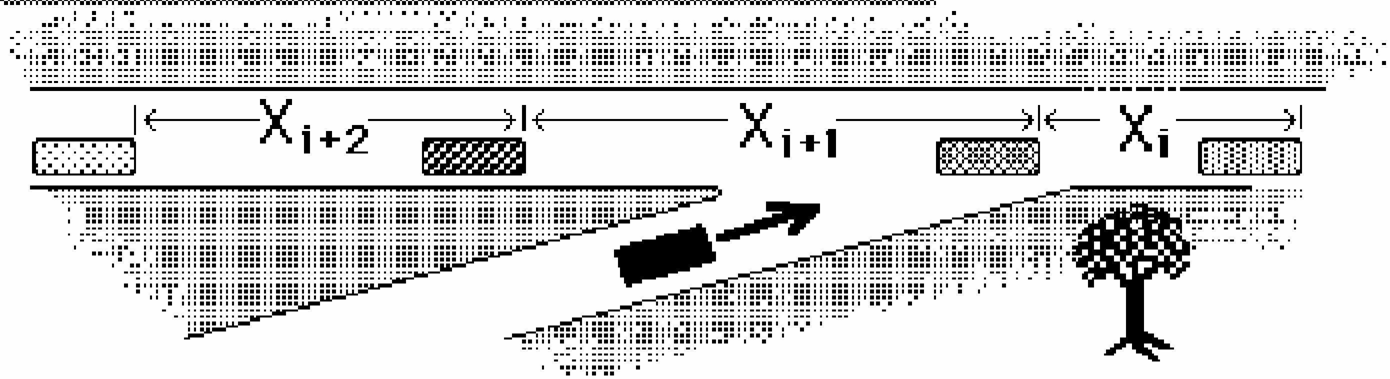
Example: Merging Traffic



A certain class of drivers will merge into freely flowing traffic ***only if*** the time X_i between oncoming cars is ***at least*** y seconds.

Example: Merging Traffic





Assume that the X_i's are i.i.d. random variables having an exponential distribution with parameter λ (the arrival rate).

What is the probability that a driver will NOT have merged after n cars have passed?

Example: Merging Traffic

Let $Y = \max\{X_1, X_2, \cdots, X_n\}$

Then $P\{\text{no merge after } n \text{ cars}\} = P\{Y \le y\}$

$$= F_Y(y) = [F_X(y)]^n$$

$$= (1 - e^{-\lambda y})^n$$

and the density function of Y is

$$f_Y(y) = n\lambda(1 - e^{-\lambda y})^{n-1} e^{-\lambda y} \quad (y \ge 0)$$

- Suppose that the rate of traffic is $\lambda = \dfrac{1 \text{ car}}{5 \text{ seconds}}$

and $y=10$ seconds between oncoming cars is required for merging.

- Then

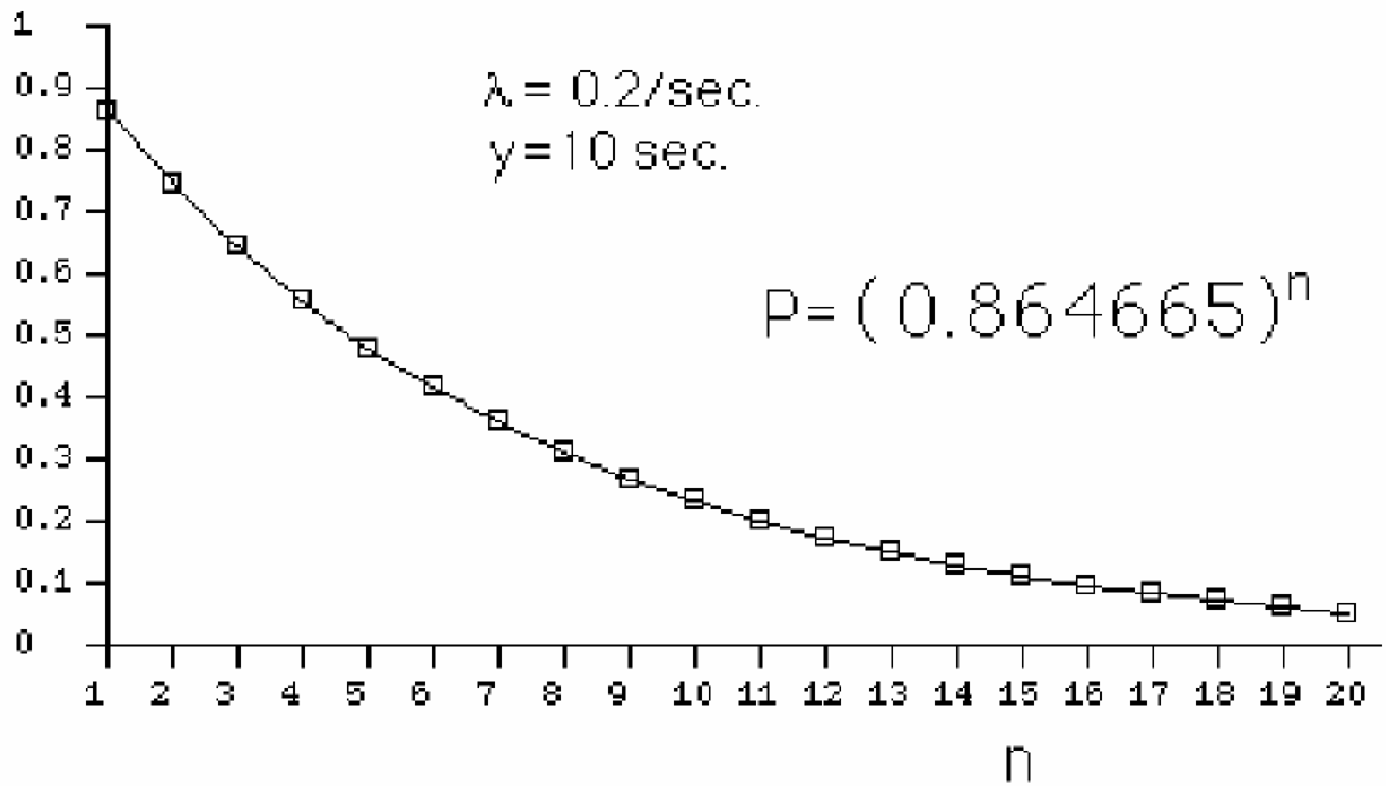
P{no merge after n cars have passed}

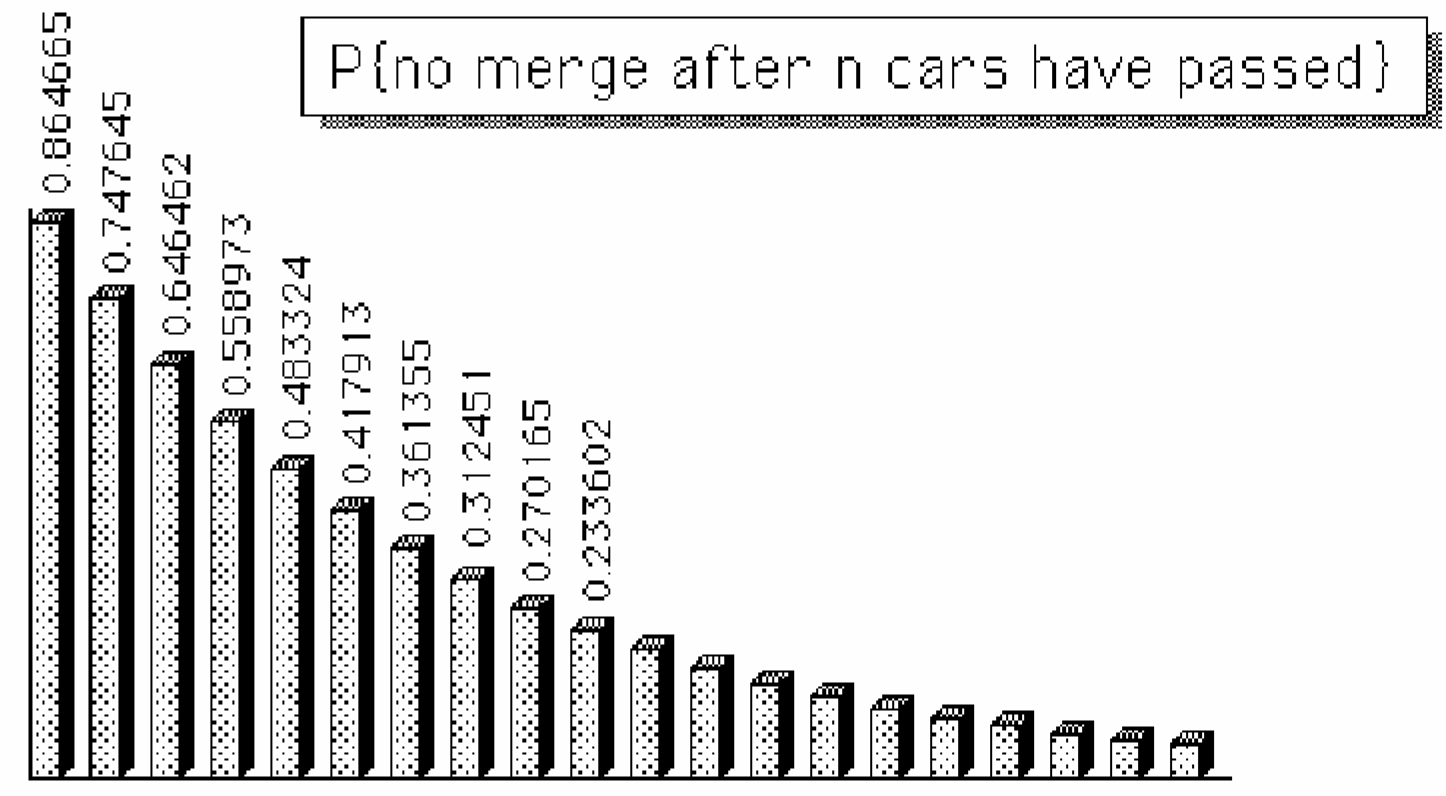
$$= \left(1 - e^{-\lambda y}\right)^n = \left(1 - e^{-2}\right)^n$$

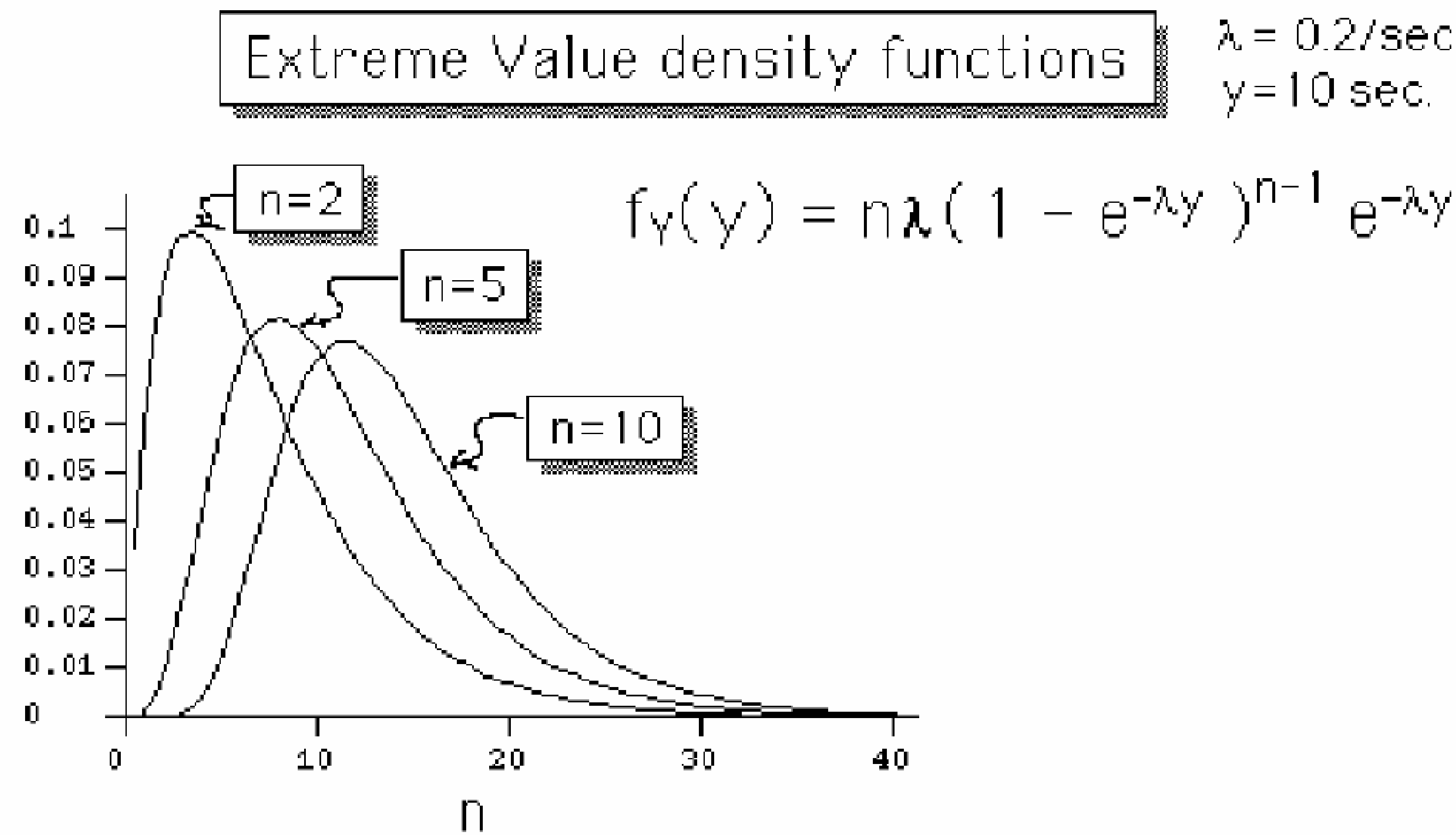
$$= (0.864665)^n$$

P{no merge after n cars have passed}

n	P
1	0.864665
2	0.747645
3	0.646462
4	0.558973
5	0.483324
6	0.417913
7	0.361355
8	0.312451
9	0.270165
10	0.233602




P{no merge after n cars have passed}
0.864665
0.747645
0.646462
0.558973
0.483324
0.417913
0.361355
0.312451
0.270165
0.233602



Extreme Value density functions
λ = 0.2/sec.
y = 10 sec.
fY(y) = nλ(1 - e^-λy)^n-1 e^-λy
n=2
n=5
n=10
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
n

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*** dist'n

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

Asymptotic Distributions

Are there such limiting distributions for

$$\max\{X_1, X_2, \cdots, X_n\}$$

and

$$\min\{X_1, X_2, \cdots, X_n\}$$

as $n \to \infty$?

The limiting distribution for Y depends upon the shape of the tail of the density functions f_X.

- ☞ Type 1: f_X falls off in "exponential" manner in the tail of interest
- ☞ Type 3: f_X falls to zero in the tail of interest

"tail of interest" is right for $Y=\textbf{\textit{max}}\{X_i\}$, and left for $Y=\textbf{\textit{min}}\{X_i\}$.

Type 1 Maximum of X_i's, which have a common CDF of the form

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

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

(For example, if exponential distribution, $g(x) = \lambda x$. The normal and gamma distributions are also of this type.)

The limiting distribution of $\max\{X_1, X_2, \cdots, X_n\}$ is the ***Gumbel*** distribution:

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

density $$f_Y(y) = \alpha \exp\left[-\alpha(y-u) - e^{-\alpha(y-u)}\right]$$

with

mean $$\mu_Y \approx u + \frac{0.577}{\alpha}$$

std. deviation $$\sigma_Y \approx \frac{1.282}{\alpha}$$

Example - Gumbel Distn.

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

(Peak annual flow = largest of 365 daily flows)

mean: $\mu_Y = 100$ cfs (ft^3/sec.)

st'd dev'n: $\sigma_Y = 50$ cfs.

What is the probability that, next year, the peak flow will exceed 200 cfs?

Since the peak annual flow is the maximum of a large number (365) daily flows, we will assume a ***Gumbel*** distribution.

First we use the mean and standard deviation to compute the parameters α and u:

$$\left.\begin{array}{l} \sigma_Y \approx \dfrac{1.282}{\alpha} \Rightarrow \alpha \approx \dfrac{1.282}{\sigma_Y} = 0.0256 \\ \mu_Y \approx u + \dfrac{0.577}{\alpha} \Rightarrow u \approx \mu_Y - \dfrac{0.577}{\alpha} \end{array}\right\} \quad u = 77.5 \text{ cfs}$$

$$F_Y(y) = \exp\left[-e^{-\alpha(y-u)}\right],\ \alpha = 0.256,\ u = 77.5$$

We can now compute the probability that the peak flow in a particular year will exceed 200 cfs:

$$P\{Y \geq 200\} = 1 - F_Y(200)$$

$$= 1 - \exp\{-e^{-0.256(200-77.5)}\}$$

$$= 0.043$$

That is, a flow exceeding 200 cfs will occur once every $\frac{1}{0.043} \approx 23$ years.

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 $Y = \mathrm{Max}\{X_1, X_2, \ldots X_{20}\}$,
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

$P\{Y \le 10\} = 16.7\%$

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

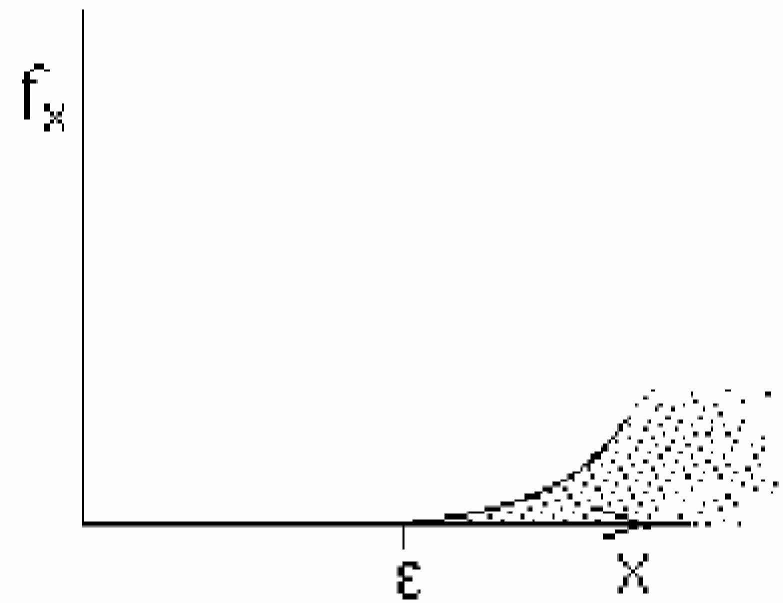
$P\{Y \ge 30\} = 2.8\%$

Type 3 density falls to zero in tail of interest, e.g., if $\min\{X_i\}$, $f_X(x) = 0$ for $x \le \varepsilon$

where ε is the *lower bound* on X_i

and $F_X(x) \approx c\,(x-\varepsilon)^k$ for x near ε and $x \ge \varepsilon$

Assume that X 's are i.i.d.



The limiting distribution for $Y=\min\{X_i\}$ as $n \to \infty$ is

CDF

$$F_Y(y) = 1 - \exp\left[-\left(\frac{y-\varepsilon}{u-\varepsilon}\right)^k\right]$$

density function

$$f_Y(y) = \frac{k}{u-\varepsilon}\left(\frac{y-\varepsilon}{u-\varepsilon}\right)^{k-1} \exp\left[-\left(\frac{y-\varepsilon}{u-\varepsilon}\right)^k\right]$$

mean

$$\mu_Y = \varepsilon + (u-\varepsilon)\,\Gamma\left(1 + \frac{1}{k}\right)$$

variance

$$\sigma_Y^2 = (u-\varepsilon)^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]$$

If $\varepsilon = 0$, this simplifies to

CDF $$F_Y(y) = 1 - e^{-(y/u)^k}$$

mean $$\mu_Y = u\,\Gamma\left(1+\frac{1}{k}\right)$$

variance $$\sigma_Y^2 = u^2\left\{\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right\}$$

which is the ***Weibull*** distribution.

The "Gamma" function is defined for all nonnegative x and satisfies

$$\Gamma(1+x) = x!$$

for integer values of x.

Values of $\Gamma\left(1+\frac{1}{k}\right)$ for k=0.1 through 9.9

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	∞	362880	120.0000	9.2605	3.3234	2.0000	1.5046	1.2658	1.1330	1.0522
1	1.0000	0.9649	0.9407	0.9236	0.9114	0.9027	0.8966	0.8922	0.8893	0.8874
2	0.8862	0.8857	0.8856	0.8859	0.8865	0.8873	0.8882	0.8893	0.8905	0.8917
3	0.8930	0.8943	0.8957	0.8970	0.8984	0.8997	0.9011	0.9025	0.9038	0.9051
4	0.9064	0.9077	0.9089	0.9102	0.9114	0.9126	0.9137	0.9149	0.9160	0.9171
5	0.9182	0.9192	0.9202	0.9213	0.9222	0.9232	0.9241	0.9251	0.9260	0.9269
6	0.9277	0.9286	0.9294	0.9302	0.9310	0.9318	0.9325	0.9333	0.9340	0.9347
7	0.9354	0.9361	0.9368	0.9375	0.9381	0.9387	0.9394	0.9400	0.9406	0.9412
8	0.9417	0.9423	0.9429	0.9434	0.9439	0.9445	0.9450	0.9455	0.9460	0.9465
9	0.9470	0.9474	0.9479	0.9484	0.9488	0.9492	0.9497	0.9501	0.9505	0.9509

E.g., $\Gamma\left(1+\frac{1}{2.5}\right) = \Gamma(1.4) = 0.8873$

Values of $\Gamma\left(1+\frac{2}{k}\right)$ for k=0.1 through 9.9

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	∞	2.433E18	3628800	2593.6	120.0000	24.0000	9.2605	5.0291	3.3234	2.4786
1	2.0000	1.7024	1.5046	1.3663	1.2658	1.1906	1.1330	1.0880	1.0522	1.0234
2	1.0000	0.9808	0.9649	0.9517	0.9407	0.9314	0.9236	0.9170	0.9114	0.9067
3	0.9027	0.8994	0.8966	0.8942	0.8922	0.8906	0.8893	0.8882	0.8874	0.8867
4	0.8862	0.8859	0.8857	0.8856	0.8856	0.8857	0.8859	0.8862	0.8865	0.8868
5	0.8873	0.8877	0.8882	0.8887	0.8893	0.8899	0.8905	0.8911	0.8917	0.8923
6	0.8930	0.8936	0.8943	0.8950	0.8957	0.8963	0.8970	0.8977	0.8984	0.8991
7	0.8997	0.9004	0.9011	0.9018	0.9025	0.9031	0.9038	0.9044	0.9051	0.9058
8	0.9064	0.9070	0.9077	0.9083	0.9089	0.9096	0.9102	0.9108	0.9114	0.9120
9	0.9126	0.9132	0.9137	0.9143	0.9149	0.9154	0.9160	0.9166	0.9171	0.9176

E.g., $\Gamma\left(1+\frac{2}{2.5}\right) = \Gamma\left(1.8\right) = 0.9314$

Computing Weibull Parameters

Given μ_Y & σ_Y^2 , we wish to solve for u & k:

$$\begin{cases} \mu_Y = u\,\Gamma\left(1+\dfrac{1}{k}\right) \\ \sigma_Y^2 = u^2\left\{\Gamma\left(1+\dfrac{2}{k}\right) - \Gamma^2\left(1+\dfrac{1}{k}\right)\right\} \end{cases}$$

This is a system of nonlinear equations with 2 unknowns, which might be solved by the Newton-Raphson method.

Solve for u in terms of the mean:

$$\mu_Y = u\,\Gamma\left(1+\frac{1}{k}\right) \Rightarrow u = \frac{\mu_Y}{\Gamma\left(1+\frac{1}{k}\right)}$$

Eliminate the variable u from the second equation:

$$\sigma_Y^2 = u^2\left\{\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right\}$$

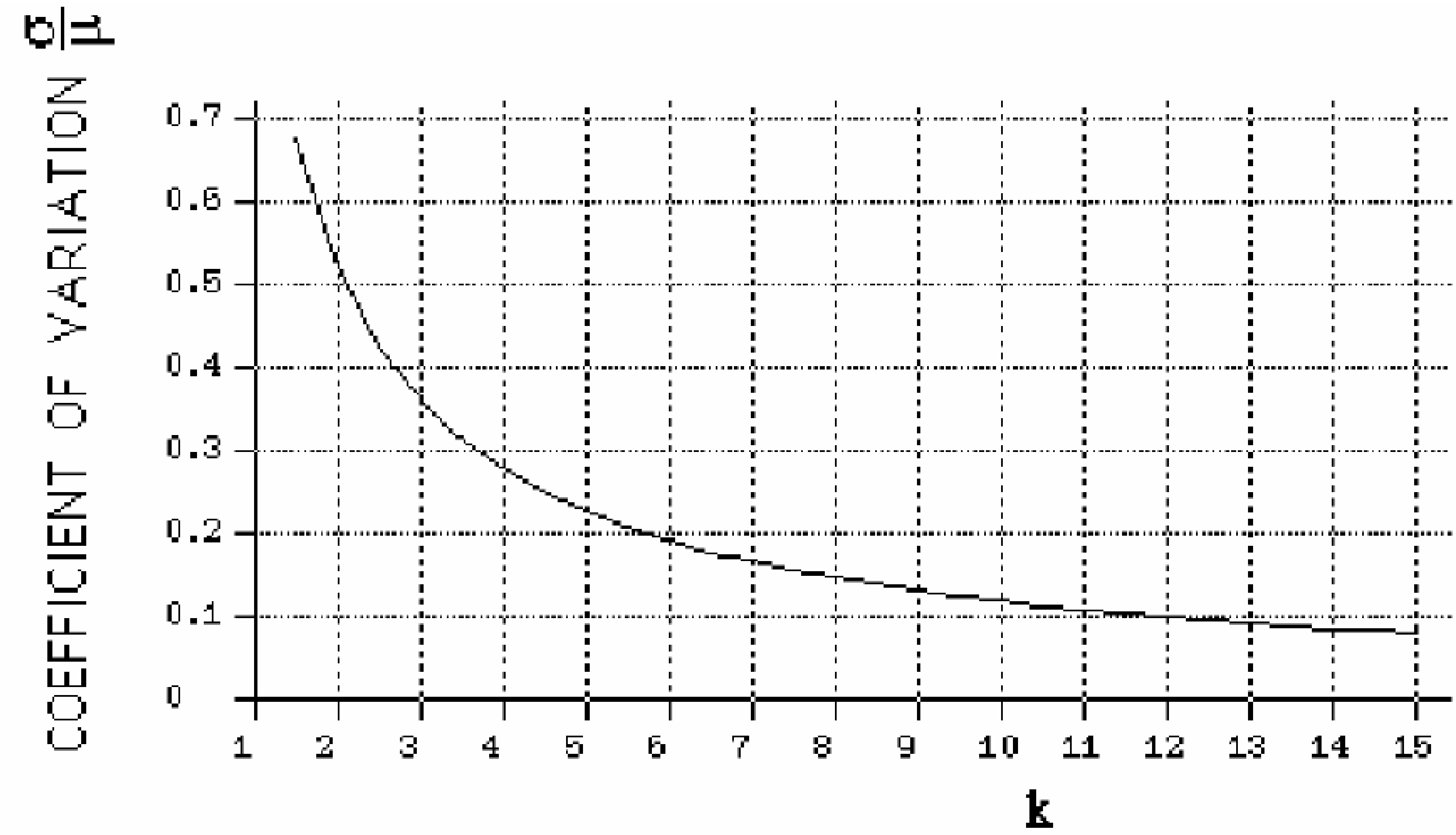
$$\Rightarrow \sigma_Y^2 = \frac{\mu_Y^2}{\Gamma^2\left(1+\frac{1}{k}\right)}\left\{\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right\}$$

This leaves us with a single nonlinear equation in k.

$$\sigma_Y^2 = \frac{\mu_Y^2}{\Gamma^2\left(1+\frac{1}{k}\right)} \left\{ \Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right) \right\} \Rightarrow \frac{\sigma_Y^2}{\mu_Y^2} = \frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1$$

$$\Rightarrow \quad \frac{\sigma_Y}{\mu_Y} = \sqrt{\frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1}$$

Coefficient of Variation



COEFFICIENT OF VARIATION $\frac{\sigma}{\mu}$
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
k

Coefficient of Variation $\frac{\sigma}{\mu}$

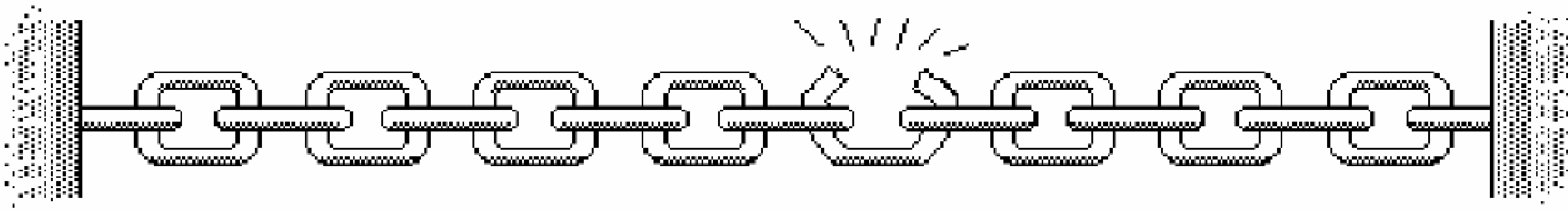
k	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	—	—	15.8	5.41	3.14	2.24	1.76	1.46	1.26	1.11
1	1	0.91	0.837	0.776	0.724	0.679	0.64	0.605	0.575	0.547
2	0.523	0.5	0.48	0.461	0.444	0.428	0.413	0.399	0.387	0.375
3	0.363	0.353	0.343	0.334	0.325	0.316	0.309	0.301	0.294	0.287
4	0.281	0.274	0.268	0.263	0.257	0.252	0.247	0.242	0.238	0.233
5	0.229	0.225	0.221	0.217	0.213	0.21	0.206	0.203	0.2	0.197
6	0.194	0.191	0.188	0.185	0.183	0.18	0.177	0.175	0.173	0.17
7	0.168	0.166	0.164	0.162	0.16	0.158	0.156	0.154	0.152	0.15
8	0.148	0.147	0.145	0.143	0.142	0.14	0.139	0.137	0.136	0.134
9	0.133	0.131	0.13	0.129	0.128	0.126	0.125	0.124	0.123	0.121

for example, if k=2.5, coefficient of variation is 0.428

Given the coefficient of variation ($\frac{\sigma}{\mu}$), we can either approximate k through use of the table or graph, or we can use a numerical method (e.g., the secant method) to solve the nonlinear equation

$$\frac{\sigma_Y}{\mu_Y} = \sqrt{\frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1}$$

Example - Weibull Dist'n



Let X_i = lifetime of link i of chain ($X_i \geq 0$)

$Y = \min\{X_1, X_2, \ldots X_n\}$ = lifetime of chain

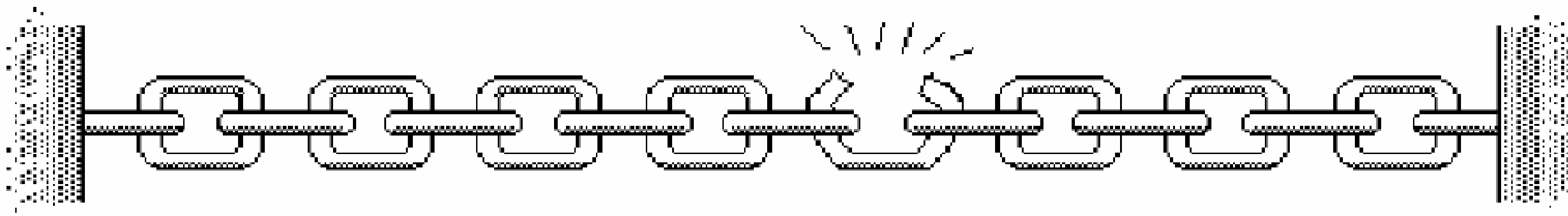
For large n (long chains), the distribution of Y should be approximately Weibull.

Suppose that we have estimates for the mean lifetime and standard deviation:

μ_Y = 150 hours σ_Y = 50 hours

What is the probability that the chain...

- fails before 100 hours of use?
- has not yet failed after 200 hours of use?



Since the lifetime of the chain is the minimum of the lifetimes (times until failure) of the individual links of the chain, and these lifetimes are each bounded below by zero, we will assume the ***Weibull*** distribution.

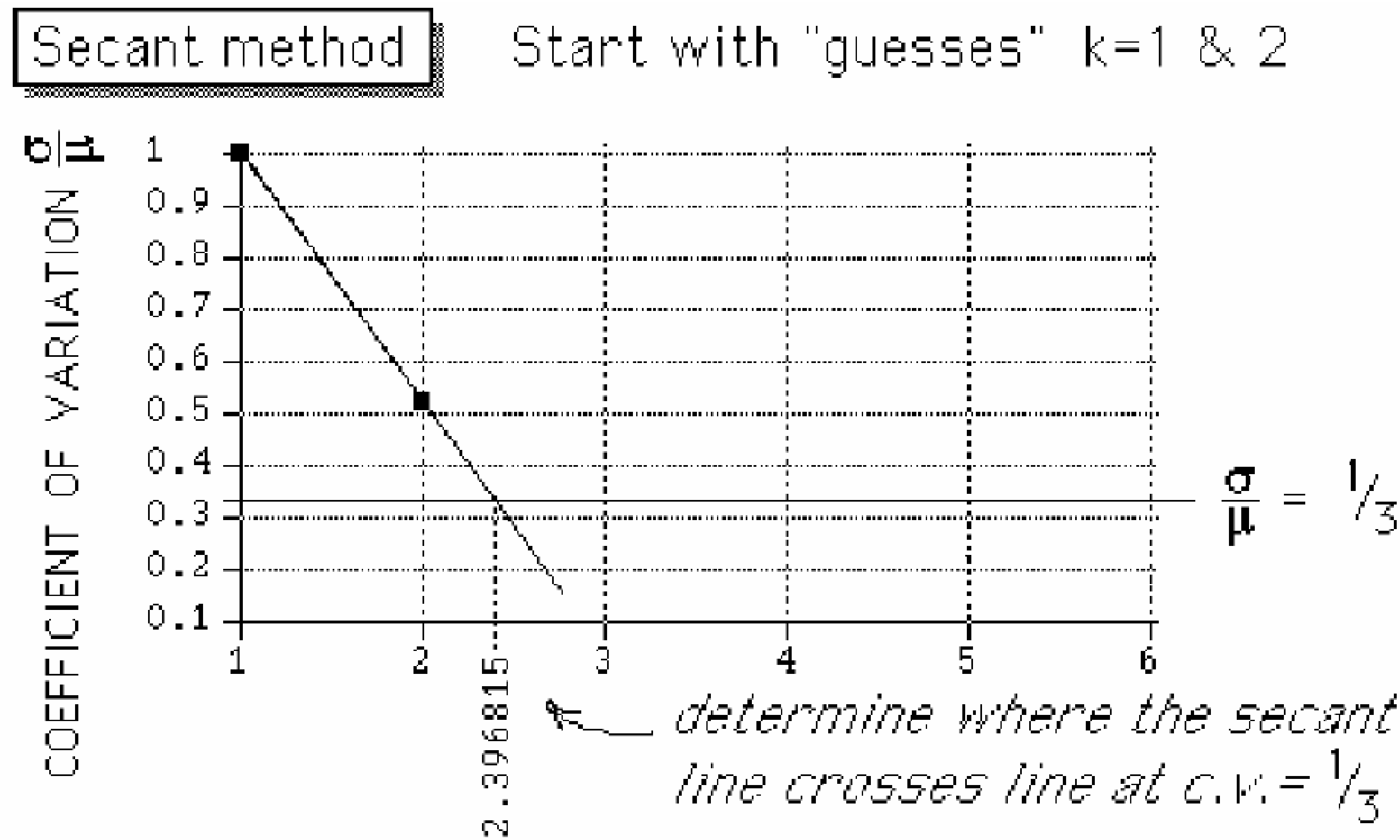
Computation of Parameter k

The coefficient of variation is $\frac{\sigma}{\mu} = \frac{1}{3}$

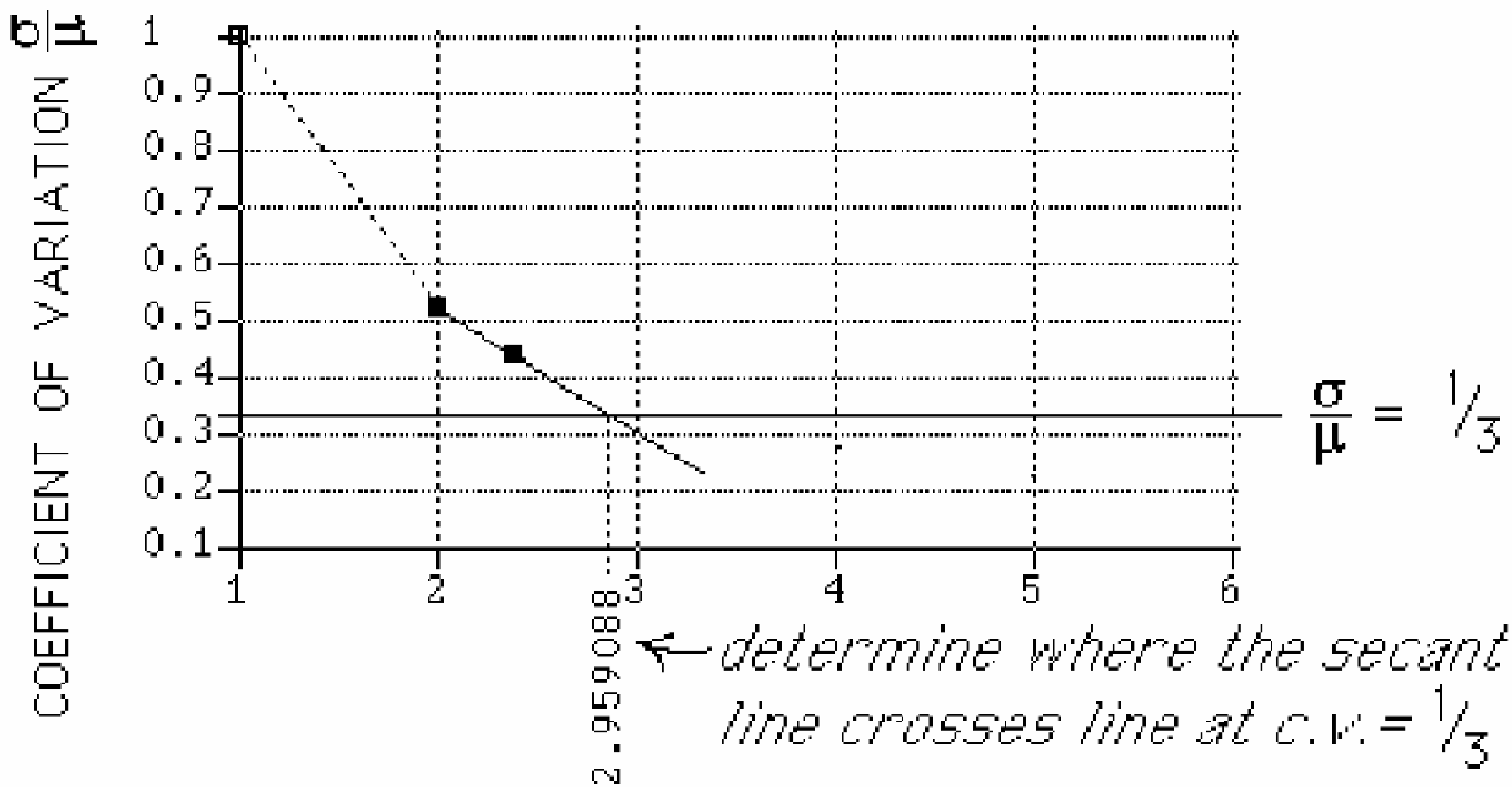
We will use the ***Secant Method*** to solve the nonlinear equation:

$$\frac{\sigma_Y}{\mu_Y} = \sqrt{\frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1}$$

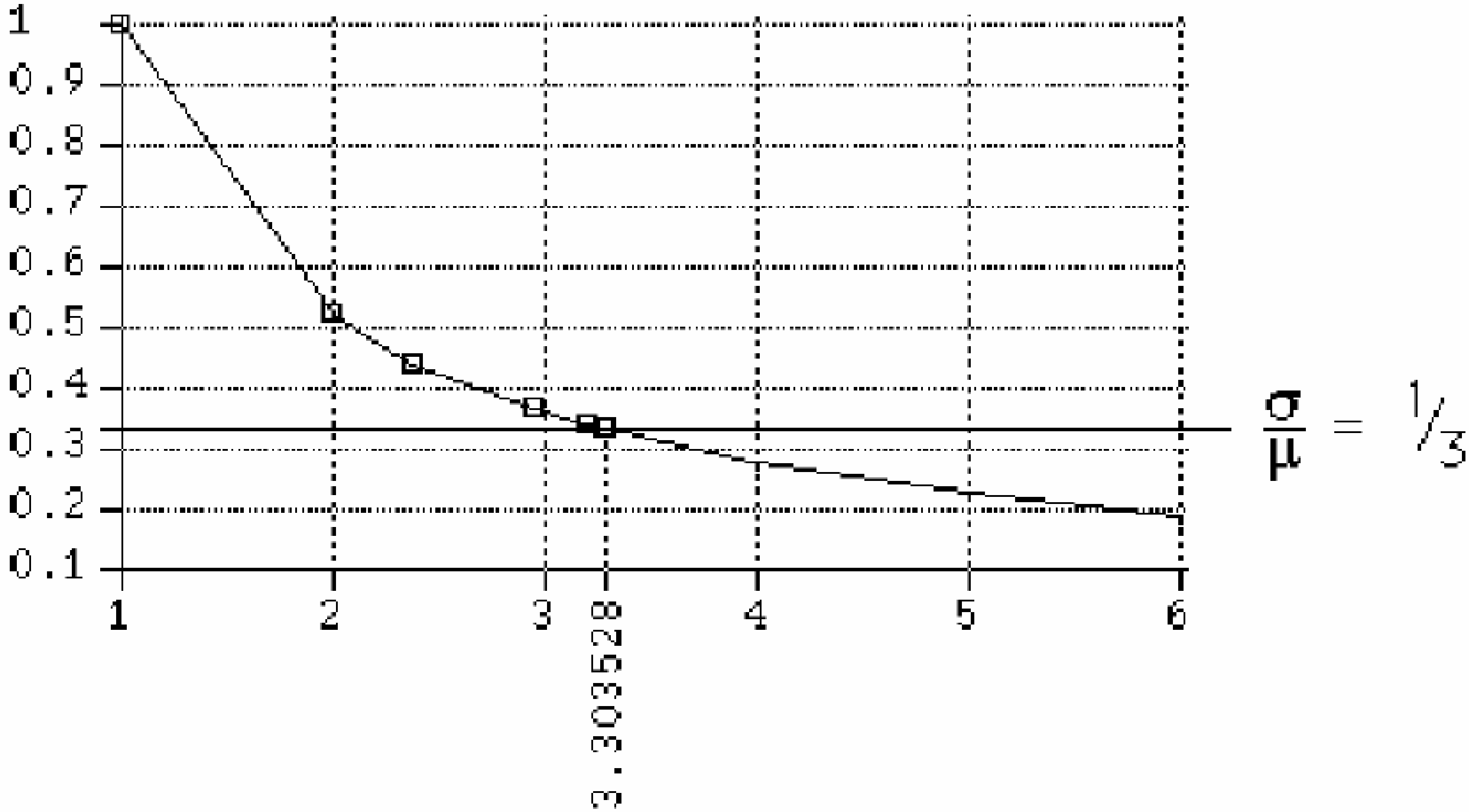
for the parameter **k**


Secant method
Start with "guesses" k=1 & 2
COEFFICIENT OF VARIATION σ/μ
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1
2
3
4
5
6
2.396815
σ/μ = 1/3
determine where the secant
line crosses line at c.v.= 1/3

Using last 2 values of k, draw a new secant line



Continue, until the procedure converges to $k = 3.303528$



Using the secant method, we get the following approximations to the value of k:

k	error
1.000000	0.666667
2.000000	0.189390
2.396815	0.111032
2.959088	0.034614
3.213779	0.008349
3.294735	0.000799
3.303305	0.000020
3.303528	0.000000

(6 iterations were performed!)

Given $\mathbf{k} = 3.303528$ and mean $\mu_Y = 150$, we can now solve for the parameter $\mathbf{u}$:

$$\mathbf{u} = \frac{\mu_Y}{\Gamma\left(1+\frac{1}{\mathbf{k}}\right)} = \frac{150}{\Gamma(1.3027)} = \frac{150}{0.897065} = 167.212$$

What is the probability that the chain...

- fails before 100 hours of use ?
- has not yet failed after 200 hours of use?