Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

457.212 Statistics for Civil & Environmental Engineers
In-Class Material: Class 15
Extreme Value Distributions (A&T 4.2.3)

Given: The distribution model of a random quantity X, i.e. PDF or CDF

Question: From a sample of size n, the distribution of the minimum or maximum?

→ Deriving an ( ) distribution

e.g. maximum flood (or drought) in the next 100 years, maximum traffic load on bridge in the next 50 years

(More generally, the distribution of the *k*-th largest or smallest from a sample → "________ Statistics")

1. Deriving "Exact Distributions"

- Maximum: $Y_n = \max(X_1, X_2, ..., X_n)$

$$F_{Y_n}(y) = P(X_1 \leq y \quad X_2 \leq y \quad \cdots \quad X_n \leq y)$$

Under the assumption that $X_1, X_2, ..., X_n$ are statistically ( ) and ( ) distributed,

$$F_{Y_n}(y) = [\qquad]^n$$

Its corresponding PDF is,

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = n[\qquad]^{n-1} f_X(y)$$

- Minimum: $Y_1 = \min(X_1, X_2, ..., X_n)$

$$1 - F_{Y_1}(y) = P(X_1 > y \quad X_2 > y \quad \cdots \quad X_n > y) = [\qquad\qquad]^n$$

Therefore, the CDF of Y_1 is

$$F_{Y_1}(y) = 1 - [\qquad\qquad]^n$$

The corresponding pdf is

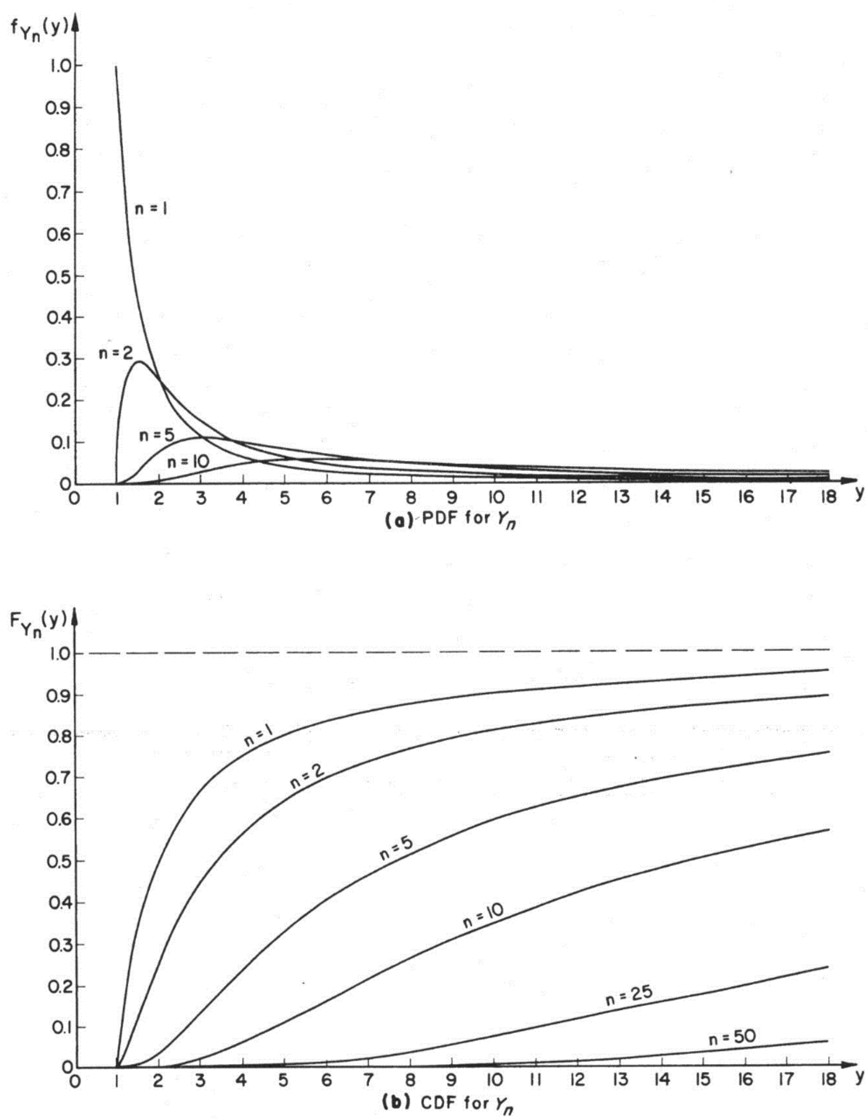
$$f_{Y_1}(y) = \frac{dF_{Y_1}(y)}{dy} = n[\qquad\qquad]^{n-1} f_X(y)$$

Instructor: Junho Song
junhosong@snu.ac.kr

Example 1: Suppose the PDF of a random variable X is given as below.

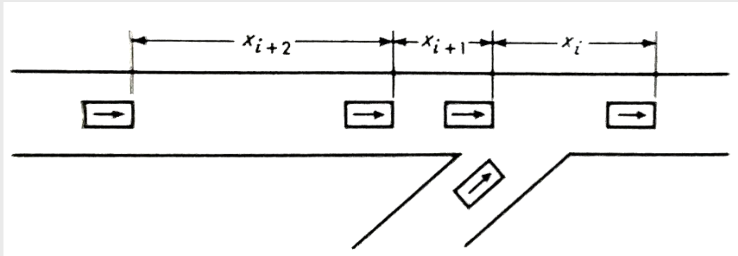
$$f_X(x) = \frac{1}{x^2}, \quad x \geq 1$$

When someone constructs a sample of size n, derive the CDF and PDF of the largest in the sample, i.e. $Y_n = \max(X_1, X_2, ..., X_n)$.



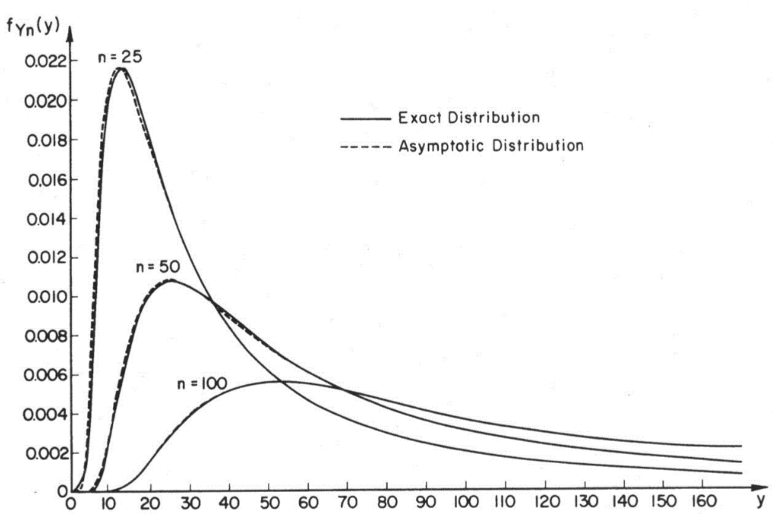
Instructor: Junho Song
junhosong@snu.ac.kr

Example 2: Suppose a certain class of drivers will merge into freely flowing traffic only if the time between passing cars, X_i is at least y sec. If the arrivals of cars are following a Poisson arrival process with the mean occurrence rate ν, what is the probability that after n cars have passed, a driver will not have been able to merge?



2. Asymptotic Distributions

An asymptotic distribution can be derived for large samples, i.e. $n \to \infty$, using Cramer's method (1946). For the extreme value distribution in **Example 1**, the exact (i.e. derived) and asymptotic distributions by Cramer's method are compared as follows.



(c) PDF for Y_n

Instructor: Junho Song
junhosong@snu.ac.kr

The asymptotic distributions of the extremes tend to converge on certain limiting forms (Gumbel 1958):

- **Type I**: The ( ) exponential form, $\exp\left[-e^{-A(n)y}\right]$

  Gumbel distribution (largest)

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

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

  Note: u_n ~ location parameter, $\beta_n = 1/\alpha_n$ ~ scale parameter

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

  Fréchet distribution (largest)

  $F_{Y_n}(y) = \exp\left[-\left(\frac{v_n}{y}\right)^k\right]$

  $f_{Y_n}(y) = \frac{k}{v_n}\left(\frac{v_n}{y}\right)^{k+1} \exp\left[-\left(\frac{v_n}{y}\right)^k\right]$

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

  Weibull distribution (smallest)

  $F_{Y_1}(y) = 1 - \exp\left[-\left(\frac{y-\varepsilon}{w_1-\varepsilon}\right)^k\right]$

The type is determined by the ( ) behavior of the original probability density function.

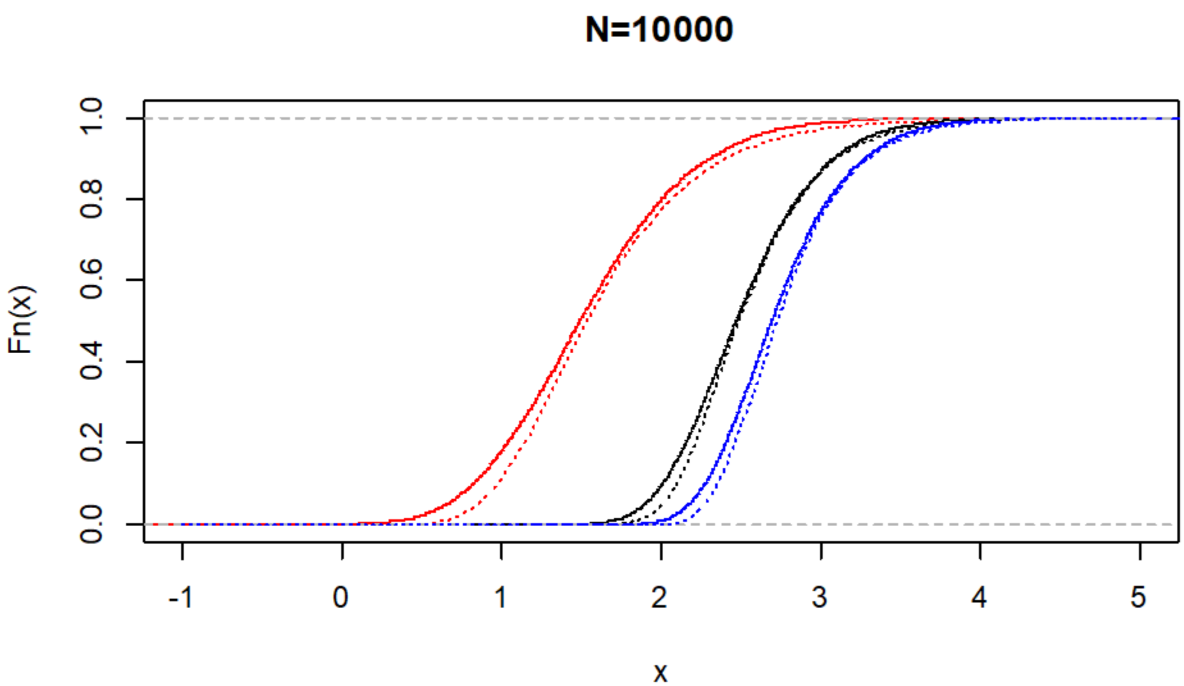
- Exponentially decaying tail (e.g. Normal) → **Type I**
- Polynomial tail (e.g. Example 3) → **Type II**
- Polynomial tail with the limited extreme value → **Type III**

Let us check whether the distribution of the maximum value among $n = 10, 100$ and 200 standard normal random variables converge to the Gumbel distribution:

```
install.packages('VGAM')
library('VGAM')

NormalMaxGen = function(n,N){
  # n: number of samples in each set
  # N: number of sets
  Sample = rnorm(n*N,0,1)
  Sample_matrix = matrix(Sample,n,N)
  SampleMax = apply(Sample_matrix,2,max) # Get maximum of each column
  return(SampleMax)
}
```

Instructor: Junho Song
junhosong@snu.ac.kr

```
u = function(n){
  # location parameter u_n of Gumbel (Example 4.18 in A&T)
  sqrt(2*log(n))-(log(log(n))+log(4*pi))/2/sqrt(2*log(n))
}

b = function(n){
  # scale parameter b_n = 1/alpha_n (See Example 4.18 in A&T)
  1/sqrt(2*log(n))
}

X10 = NormalMaxGen(10,10000) # Generate 10 normal random var. 10,000 times
X100 = NormalMaxGen(100,10000)
X200 = NormalMaxGen(200,10000)

xv = seq(-1,5,length=100)

plot(ecdf(X10),main="N=10,000",xlim=c(-1,5),col='red')
lines(xv,pgumbel(xv,u(10),b(10)),col='red',lty='dotted')

lines(ecdf(X100),col='black')
lines(xv,pgumbel(xv,u(100),b(100)),col='black',lty='dotted')

lines(ecdf(X200),col='blue')
lines(xv,pgumbel(xv,u(200),b(200)),col='blue',lty='dotted')
```



Instructor: Junho Song
junhosong@snu.ac.kr

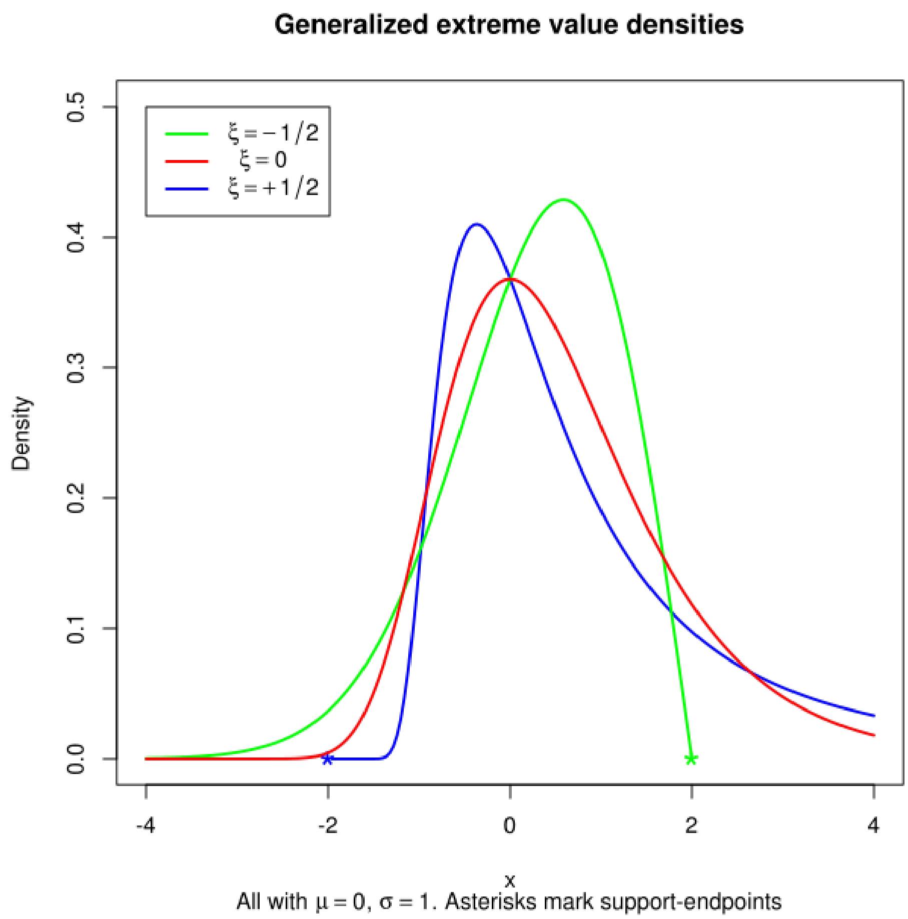
3. Generalized Extreme Value (GEV) Distribution

A general extreme value distribution model that can represent Type I (Gumbel), Type II (Fréchet) and Type III (Weibull) distributions.

The CDF of GEV distribution is given as

$$F_Y(y) = \exp\left[-\left(1+\xi\left(\frac{y-u}{\beta}\right)\right)^{-1/\xi}\right]$$

where u, $\beta(>0)$, and ξ are the location, scale and shape parameters. As $\xi \to 0$, GEV becomes Gumbel. For $\xi > 0$, GEV is Fréchet distribution. When $\xi < 0$, GEV is Weibull distribution. The "VGAM" package of R provides dgev(), pgev(), qgev() and rgev() functions. One can estimate the parameters from a data set using the package as well.



* Image source and more details:
https://en.wikipedia.org/wiki/Generalized_extreme_value_distribution