Financial Models and @RISK

Yoo-Suk Hong <yhong@snu.ac.kr>
Dept of Industrial Engineering
Seoul National University

Seila *et. al.*, Chapter 3

@RISK

An add-in that can be used to perform simulations in Excel.

- Sampling from various distributions
  $\Leftarrow$ Additional functions for random variate generation
- Defining the simulation experiments
  $\Leftarrow$ Specifying cells to observe and the number of replications
- Running the simulation and collecting data
- Analyzing the output data and displaying the results graphically
- Fitting a distribution to data
  $\Leftarrow$ BestFit (part of @RISK) software

Geometric Random Walk

A model for the price of a stock at a future point in time.

The *log* of the *percentage increase* in the stock price between now and time t has a *normal* distribution.

$$\ln\left(\frac{P(t)}{P(0)}\right) = \underbrace{\left(\mu - \frac{\sigma^2}{2}\right) t}_{\text{mean}} + \underbrace{\sigma\sqrt{t}}_{\text{s.d.}}\ Z$$

where Z denotes a standard normal random variable.

$$P(t) = P(0)\, e^{\left(\mu-\frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}\, Z}$$

Note: $\mu =$ Drift parameter; $\sigma =$ Volatility parameter.

@RISK Functions for Sampling from Distributions

uniform, beta, binomial, chi-square, exponential, gamma, lognormal, normal, Poisson, and Weibull

extreme value, negative binomial, Pareto, Person types 5 and 6.

`RiskNormal()`, `RiskBeta()`, `RiskExpon()`, `RiskGamma()`, `RiskLognorm()`, `RiskPoisson()`, `RiskUniform()`, etc.

Remark: Refer to @RISK help menu for a complete listing.
@RISK + Help + Distribution Functions. . .

@RISK Toolbar Buttons

Simulation Settings: specifies the options for sampling from distributions for input parameters; the number of iterations (replications); etc. (Figs 3.3 and 3.4)

Add Output: specifies what cells to observe and record.

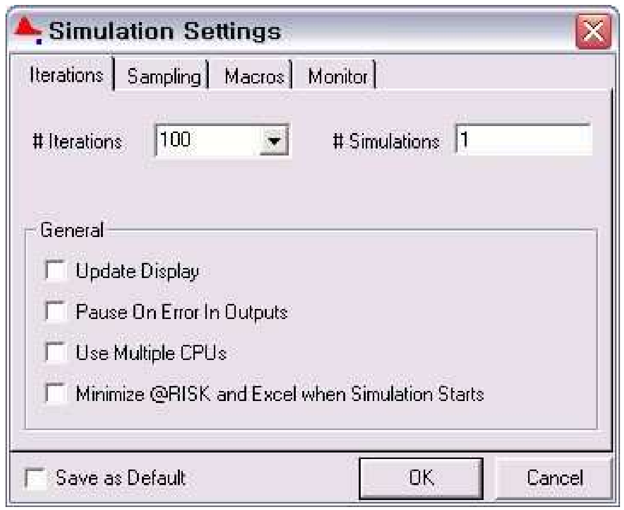
Run Simulation: actually performs the replications and produces *Summary Statistics* window. (Figs 3.5 and 3.6)

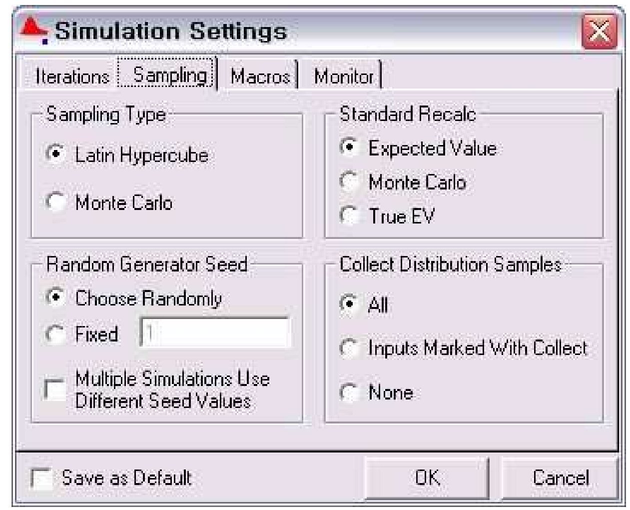
Detailed Statistics: produces a window with more statistics, i.e., *Detailed Statistics* window. (Fig 3.7)

Data Window: accesses the output data corresponding to the selected output.

Graph: produces a graph for the distribution of the selected output (Figs 3.11 and 3.12).

Simulation Settings





 # Iterations: Corresponds to what we call *replications*.

 # Simulations:

- Leave the value equal to 1.
- Used when a model is simulated multiple times with different parameters.

Simulation Settings

Sampling Type:

- Latin Hypercube: designed to more evenly distribute samples
- Monte Carlo: simple IID observations

Collect Distribution Samples: Whether to collect input data (All; Marked only; None)

Standard Recalc: What to display in each cell containing a formula that samples from a distribution

- Monte Carlo: displays the actual sampled values
- Expected Value: displays (the integer nearest to) the mean
- True EV: displays the mean even in the discrete case

Random Generator Seed: How the RNG is initialized

Calculation of a Confidence Interval

1. Click the **Run Simulation** button.
2. Click the **Detailed Statistics** button.
3. Select the menu item **Results + Report Settings...** (Fig 3.8)

   Tick the option "Generate Excel Reports Selected Below". Select "Detailed Statistics" and "Active Workbook".
4. Click **Generate Report Now**.

   Two new worksheets labeled *Input (Output) Statistics Report* will be generated. (Fig 3.9)
5. Copy the sample mean and standard deviation to another worksheet and compute the confidence interval. (Fig 3.10)

Options and Futures: Definitions

Derivatives are securities whose value is derived from the value of an underlying asset. Usually, the price or value of the underlying asset is currently known but subject to considerable uncertainty in the future.

Future: an agreement to buy a particular asset at a specified price at a specified time in the future. A future involves the *obligation* to purchase the asset.

Option: the *right*, but not the obligation, to buy (call option) or sell (put option) an asset at a specified price sometime in the future. $\longrightarrow$ More flexible than futures; Often used as a hedge against adverse events.

- European: must be exercised on the expiration date
- American: can be exercised on or before the expiration date

Return from an Option

Call option: Earns money only if the stock price goes *up*.

On the expiration date, if the stock price P_T is higher than the strike price S, the holder of the call will buy the stock for S and immediately sell it to realize a profit of $P_T - S$ per share.

$$R_T = \begin{cases} 0 & \text{if } P_T \leq S \\ P_T - S & \text{if } P_T > S \end{cases}$$

Put option: Earns money only if the stock price goes *down*.

On the expiration date, if the strike price S is higher than the stock price P_T, the holder will buy the stock at market and immediately sell it at S to realize a profit of $S - P_T$ per share.

Estimating the Price of an Option

The present value of the cash flow for an option is

$$R_0 = R_T\, e^{-rT}$$

if r is the risk-free rate used to discount the cash flow.

Financial theory says that the fair price of the option is $E(R_0)$.

To compute $E(R_0)$, we must know the distribution of P_T. *Either* probability calculus *or* simulation can be used to estimate the expected value.

Example: Option pricing with @RISK (Fig 3.13)

Hedging Using Put Options

Definition *"Risk"*

- the variance or standard deviation of the return on investment, i.e., a measure of the *uncertainty* in the return.
- the probability of an unfavorable outcome, e.g., downside risk.

If you buy a put option that allows you to sell the stock at a given price, then the loss due to dropping prices can be limited.

(Figs 3.14 and 3.15) $\Rightarrow$ Portfolio *vs.* Stock alone

- The mean returns on the two investments do not differ significantly.
- The use of a put option as a hedge is effective to improve the distribution of return without sacrificing the expected return.

Dynamic Financial Models

Definition *Dynamic models* are those that observe a variable of interest over time. (*e.g.*, cash position model of an insurance company; cash flow for an Asian option).

The stock price after a delay of δt

$$P(t+\delta t) = P(t)\, e^{\left(\mu - \frac{\sigma^2}{2}\right)\delta t \,+\, \sigma\sqrt{\delta t}\, Z}$$

Implementation of the model with @RISK (Fig 3.16)
Remark: Use the absolute referencing in spreadsheet (see p. 94).

Pricing of an Asian option: The cash flow depends on the *average* price over the term of the option. (Fig 3.21) Cell B7 has

```
=AVERAGE('Stock Price'!E4:E30)
```

Using BestFit to Fit a Distribution to Data

1. Select the range containing the data to be fitted.
2. (Optionally within the BestFit standalone program)
   Click the **Input Data Options** button.
3. Click the **Fit Distributions to Data** button. (Fig 3.31)
4. You will be presented with the **Fit Results** window. (Fig 3.32)

   Fitted Distributions: listed in order of best fit w.r.t. *Rank by*
   Comparison: graph for fitting the distribution
   Difference: difference between the density and the histogram
   P-P: probability-probability plot
   Q-Q: qualtile-quantile plot
   Stats: parameter values for fitted distribution and input data
   GOF: three goodness-of-fit tests (Ch-sq, A-D, K-S)

Selecting Input Probability Distributions

Input modeling deals with determining which probability distributions to use to model stochastic behavior in simulations.

No data available: Use distributions that "fit" the situation.
e.g.) exponential, normal, uniform, triangular, etc.

Data available: Use either one of the following approaches.

1. Use data values to define *empirical* distributions.
   (a) Case 1: Individual data values
   (b) Case 2: Grouped data
2. Use data values to *fit* a theoretical distribution.
   (a) Hypothesizing the family of distribution
   (b) Estimation of parameters
   (c) Validation of the distribution (Goodness-of-fit test, etc.)

Empirical Distributions (Individual)

Suppose we have data values $\{x_1, x_2, \ldots, x_n\}$.

(1) Sort data in increasing order. Let $x_{(i)}$ be the ith one in the sorted list so that $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.

(2) The "cumulative" probability associated with $x_{(i)}$ is defined as $F(x_{(i)}) = \left(\dfrac{i-1}{n-1}\right)$. The cdf $F(x)$ derived from interpolation is given by

$$F(x) = \begin{cases} 0 & \text{if } x < x_{(1)} \\ \frac{i-1}{n-1} + \frac{1}{n-1}\left(\frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}}\right) & \text{if } x_{(i)} \leq x < x_{(i+1)} \\ 1 & \text{if } x \geq x_{(n)} \end{cases}$$

Empirical Distributions (Individual): Example

The sample data set $\{2.4, 3.2, 4.1, 6.8, 8.2\}$. Calculate $F(6)$.

$$
\begin{aligned}
F(2.4) &= \frac{1-1}{5-1} = 0 \\
F(3.2) &= \frac{2-1}{5-1} = \frac{1}{4} \\
F(4.1) &= \frac{3-1}{5-1} = \frac{1}{2} \\
F(6.8) &= \frac{4-1}{5-1} = \frac{3}{4} \\
F(8.2) &= \frac{5-1}{5-1} = 1
\end{aligned}
$$

Since $4.1 \le 6 < 6.8$,

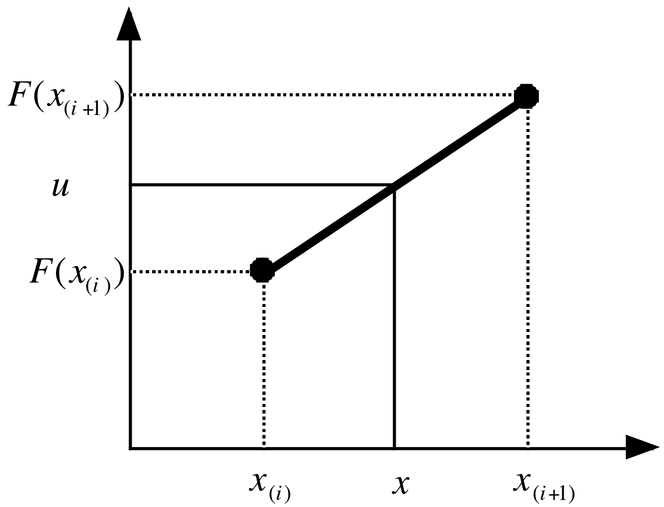
$$x_{(3)} \le 6 < x_{(4)} \rightarrow i = 3.$$

$$
\begin{aligned}
F(6) &= \frac{3-1}{5-1} + \frac{1}{5-1}\left(\frac{6-4.1}{6.8-4.1}\right) \\
&= 0.675
\end{aligned}
$$

Empirical Distributions (Individual): Algorithm

Using inverse transform, $F(x) = u$ yields

$$x = ((n-1)u - i + 1)\left(x_{(i+1)} - x_{(i)}\right) + x_{(i)}$$



We have to find i such that

$$x_{(i)} \le x < x_{(i+1)}$$

which is equivalent to

$$F(x_{(i)}) \le u < F(x_{(i+1)}).$$

Assignment: Show that $i = \lfloor (n-1)u \rfloor + 1$.

Empirical Distributions (Grouped Data)

Suppose we have grouped data consisting of k adjacent intervals $[a_0, a_1), [a_1, a_2), \cdots, [a_{k-1}, a_k)$ such that the jth contains n_j observations. Let $N = n_1 + n_2 + \cdots + n_k$.

A reasonable empirical distribution G can be

$$G(a_0) = 0,\ G(a_1) = \frac{n_1}{N}, \cdots, G(a_j) = \frac{n_1 + n_2 + \cdots + n_j}{N}.$$

The cdf $G(x)$ derived from interpolation is given by

$$G(x) = \begin{cases} 0 & \text{if } x < a_0 \\ G(a_{j-1}) + \left(\frac{x - a_{j-1}}{a_j - a_{j-1}}\right)[G(a_j) - G(a_{j-1})] & \text{if } a_{j-1} \le x < a_j \\ 1 & \text{if } x \ge a_k \end{cases}$$

Empirical Distributions (Group): Example

5 adjacent intervals $[0,2), [2,4), [4,6), [6,8), [8,10)$ which contains $10, 30, 50, 25, 6$ observations, respectively. Calculate $G(5.3)$.

$$\begin{aligned}
G(0) &= 0 \\
G(2) &= \frac{10}{121} = 0.08 \\
G(4) &= \frac{10+30}{121} = 0.33 \\
G(6) &= \frac{10+30+50}{121} = 0.74 \\
G(8) &= \frac{10+30+50+25}{121} = 0.95
\end{aligned}$$

Since $4 \le 5.3 < 6$,

$$\left[\begin{array}{ll} \left[\begin{array}{ll} 4 & G(4) = 0.33 \\ 5.3 & G(5.3) \end{array} \right] \\ \quad 6 \qquad G(6) = 0.74 \end{array} \right]$$

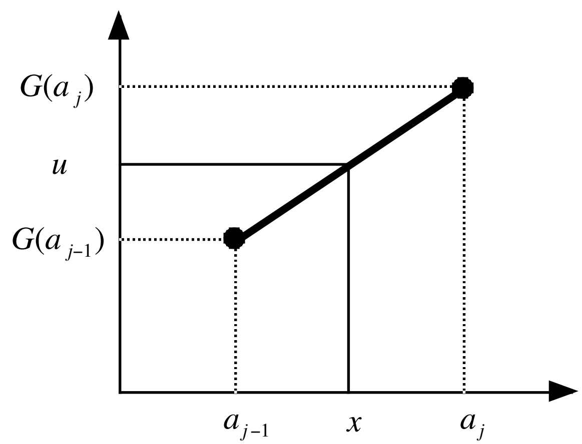
$$\frac{5.3-4}{6-4} = \frac{G(5.3)-0.33}{0.74-0.33}$$

$$G(5.3) = 0.33 + \frac{5.3-4}{6-4}(0.74-0.33) = 0.60$$

Empirical Distributions (Group): Algorithm

Using inverse transform, $G(x) = u$ yields

$$x = a_{j-1} + \left[\frac{a_j - a_{j-1}}{G(a_j) - G(a_{j-1})}\right](u - G(a_{j-1}))$$



We have to find j such that

$$a_{j-1} \le x < a_j$$

which is equivalent to

$$G(a_{j-1}) \le u < G(a_j).$$

Assignment: Show that j holds that $N\,u \in \left[\sum_{k=0}^{j-1} n_k, \sum_{k=1}^{j} n_k\right).$

Fitting a Theoretical Distribution to Data (I)

"What distribution is this data likely to have come from?"

1. Hypothesizing the family of distribution
   - Intrinsic property of distribution describing a certain situation
   - General characteristics: *e.g.* range, continuous/discrete
   - Summary statistics: *e.g.* mean (μ), median ($\tilde{\mu}$), skewness (ν)
   - Number of "shape" parameters
2. Estimation of parameters
   - Method of maximum likelihood
   - Method of moments
3. Validation of the distribution
   - Expert opinion (so-called face validity)
   - Graphing real data *vs.* fitted distribution
   - Goodness-of-fit (χ^2, Kolmogorov-Sminov, Anderson-Darling)

Summary Statistics

Population Parameter	*Estimate Statistics*	*Function*	*Distribution*
Min, Max	$x_{(1)}, x_{(n)}$	measure range	C, D
Mean (μ)	$\bar{x} = \frac{\sum x_i}{n}$	measure central tendency	C, D
Median ($\tilde{\mu}$)	$\tilde{x}(n) = \begin{cases} x_{(n+1/2)} & n \ odd \\ \frac{(x_{(n/2)} + x_{(n/2+1)})}{2} & n \ even \end{cases}$	measure central tendency	C, D
Variance (σ^2)	$S^2(n) = \frac{\sum (x_i - \bar{x})^2}{n}$	measure variability	C, D
Coefficient of Variation $CV = \sigma/\mu$	$\hat{CV}(n) = \frac{S(n)}{\bar{x}(n)}$	alternative measure of variability	C
Lexis ratio $\tau = \sigma^2/\mu$	$\hat{\tau}(n) = \frac{S^2(n)}{\bar{x}(n)}$	measure of variability	D
Skewness $\upsilon = \frac{E((x-\mu)^3)}{(\sigma^2)^{3/2}}$	$\hat{\upsilon} = \frac{\sum (x_i - \bar{x}(n))^3 / n}{(S^2(n))^{3/2}}$	measure of symmetry	C, D

Guidelines using Summary Statistics

Summary Statistics:
Mean $\mu \longrightarrow \bar{x}$, Median $\tilde{\mu} \longrightarrow \tilde{x}$, Variance $\sigma^2 \longrightarrow s^2$,

- Coefficient of variation (Continuous) $CV = \frac{\sigma}{\mu} \longrightarrow \widehat{CV}$
- Lexis ration (Discrete) $\tau = \frac{\sigma^2}{\mu} \longrightarrow \hat{\tau}$
- Skewness $\nu = \frac{E((X-\mu)^2)}{(\sigma^2)^{3/2}} \longrightarrow \hat{\nu}$

Symmetrical ($\hat{\nu} = 0$ or $\bar{x} = \tilde{x}$):
Normal, Uniform, Symmetric Triangular, Binomial $(p = 1/2)$

Skewed to right ($\hat{\nu} > 0$ or $\bar{x} > \tilde{x}$):
Exponential, Gamma, LogNormal, Weibull, Triangular

Skewed to left ($\hat{\nu} < 0$ or $\bar{x} < \tilde{x}$): Beta, Triangular

General Characteristics of Parameters

Distributions have parameters that determine

- Location (*e.g.* mean)
- Scaling (*e.g.* standard deviation)
- Shape

<u>Exponential Distribution</u>: $f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$

- $\mu = \beta$, $\sigma^2 = \beta^2 \quad \Rightarrow \quad CV = \frac{\sigma}{\mu} = 1$
- Skewed to right
- Single shape $\quad \Rightarrow \quad$ No shape parameter
- Good for (independent) arrival process
- Not good for processing times
- Memoryless property

Lifetime Distributions: Skewed to Right (I)

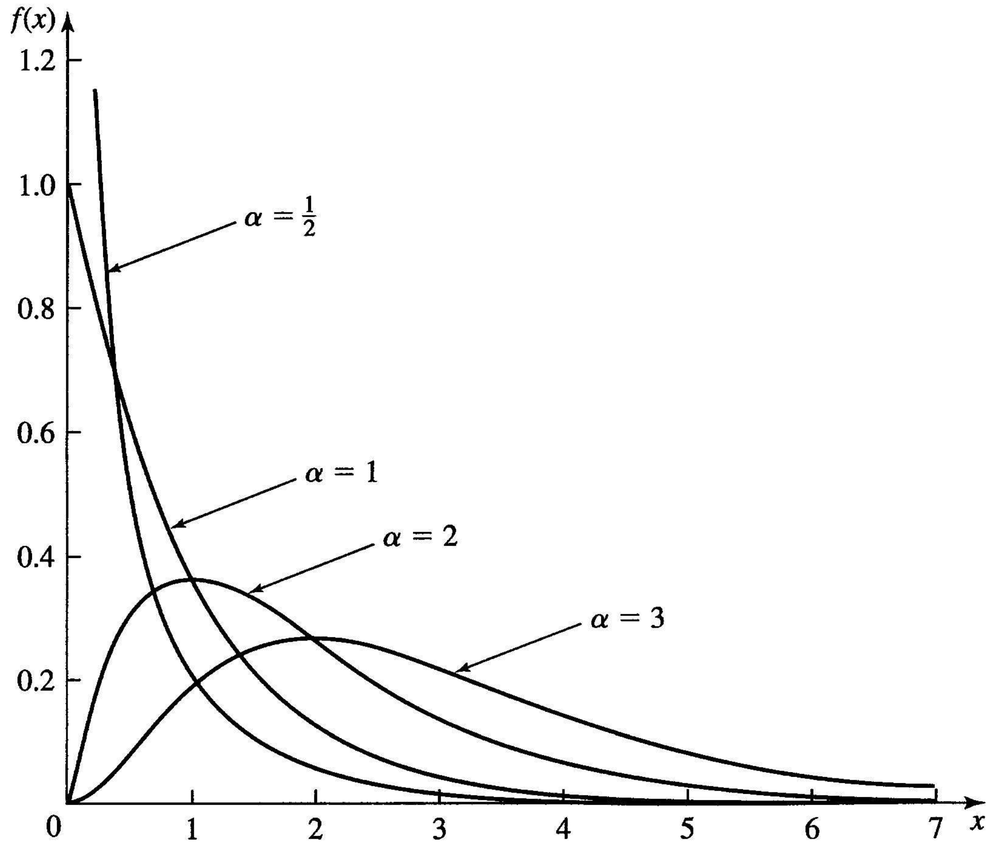
Gamma Distribution: $f(x) = \dfrac{x^{\alpha-1}\, e^{-x/\beta}}{\beta^{\alpha}\, \Gamma(\alpha)}, \quad x > 0$

Note: $\Gamma(z) = \int_0^{\infty} t^{z-1}\, e^{-t}\, dt$ yields

$$\Gamma(1) = 1, \quad \Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(z+1) = z\,\Gamma(z) = z!$$

- $\mu = \alpha\,\beta$, $\sigma^2 = \alpha\,\beta^2 \quad \Rightarrow \quad CV = \frac{\sigma}{\mu} = \frac{1}{\sqrt{\alpha}}$
  (Note that $CV < 1$, if $\alpha > 1$.)
- The parameter α is a shape parameter.
- If $\alpha = 1$, Gamma becomes the exponential distribution.
- As $\alpha \to \infty$, we get a normal (bell) shape.

Gamma Density Functions



$Gamma(\alpha, 1)$ Density Functions

Lifetime Distributions: Skewed to Right (II)

Weibull Distribution: $f(x) = \dfrac{\alpha\, x^{\alpha-1}\, e^{-(x/\beta)^{\alpha}}}{\beta^{\alpha}}, \quad x > 0$

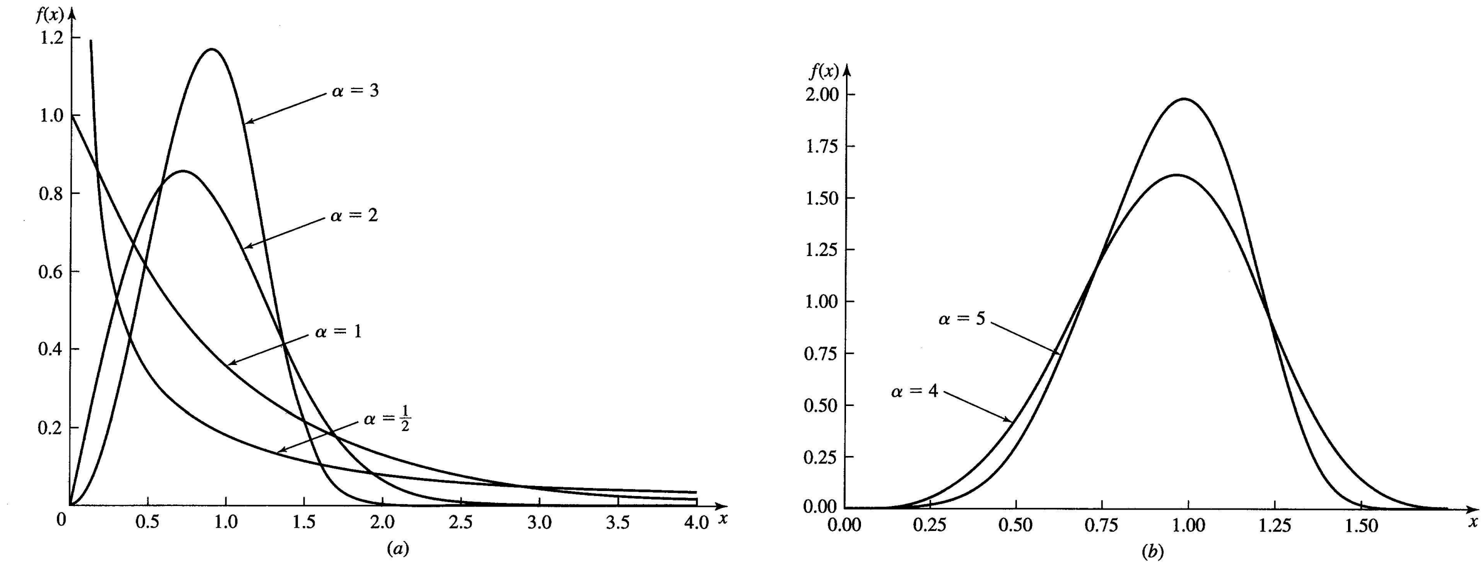
- $CV < 1$, if $\alpha > 1$.
- The parameter α is a shape parameter.
- If $\alpha = 1$, Weibull becomes the exponential distribution.
- As $\alpha \to \infty$, Weibull becomes degenerate at β. (Sharp peak)

LogNormal Distribution:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \; x > 0$$

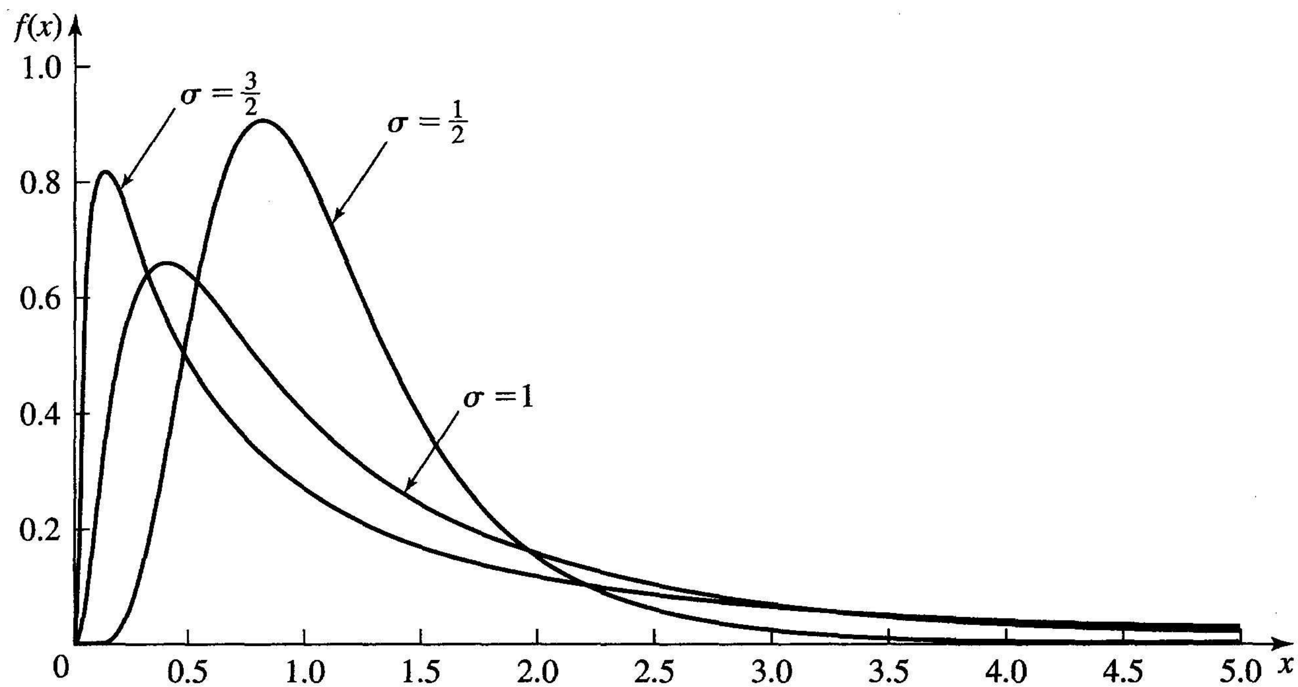
- $\mu = e^{\mu+\sigma^2/2}$, $\sigma^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \Rightarrow CV = \sqrt{e^{\sigma^2} - 1} > 1$
- The parameter σ is a shape parameter.

Weibull Density Functions



$Weibull(\alpha, 1)$ Density Functions

LogNormal Density Functions



$LN(0, \sigma^2)$ Density Functions

Comparison of Lifetime Distributions

	Gamma Distribution	Weibull Distribution	LogNormal Distribution
Parameters	Simple	Complex	Complex
Generation of Random Variates	No closed-form	Simple	Simple
Coefficient of Variation	$CV < 1$	$CV < 1$	$CV > 1$

Beta Distribution

Used as a rough model in the absence of data. The pdf is

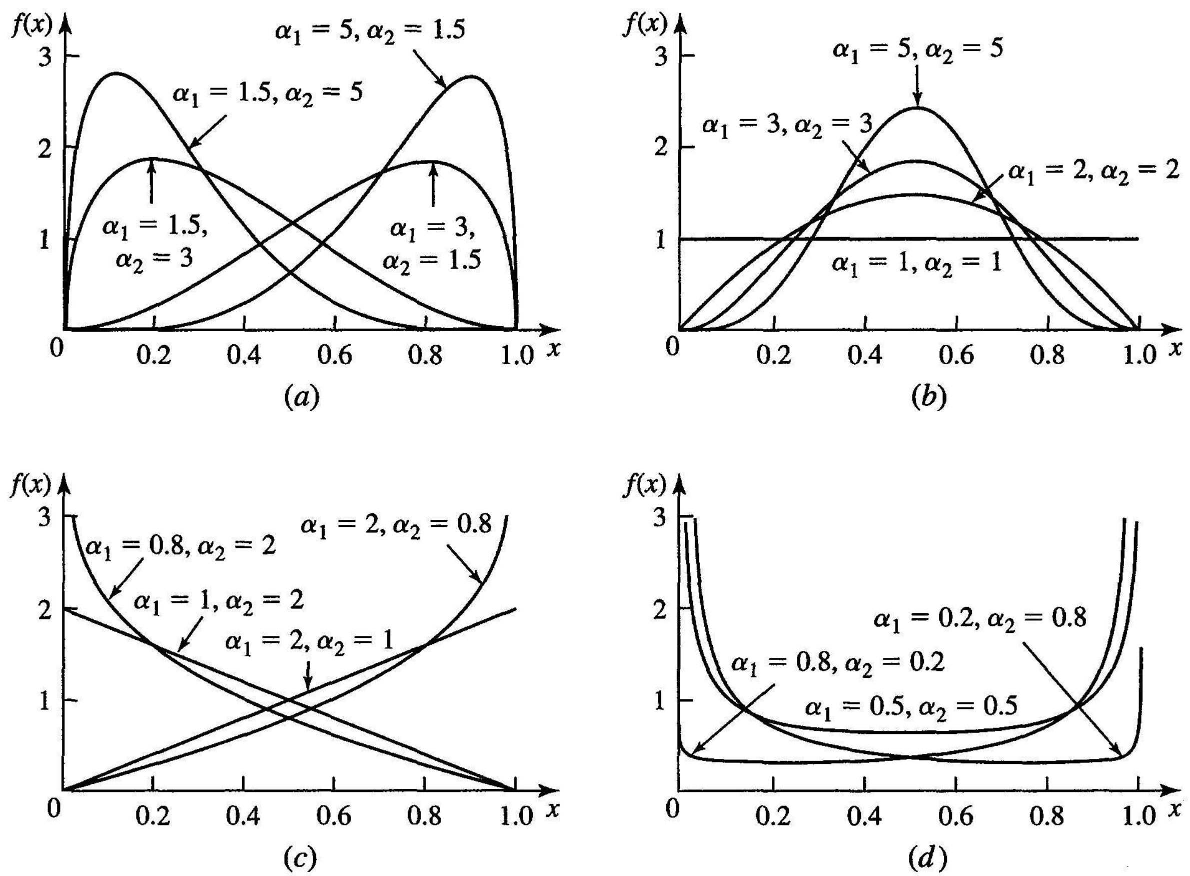
$$f(x) = \frac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)}, \quad 0 < x < 1,$$

where $B(z_1, z_2) = \int_0^1 t^{z_1 - 1}(1-t)^{z_2 - 1}\,dt$, whose properties are

$B(z_1, z_2) = B(z_2, z_1)$, $B(z_1, z_2) = \dfrac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}$, for any $z_1, z_2 > 0$.

- Two shape parameters $\alpha_1 > 0$ and $\alpha_2 > 0$.
- $\alpha_1 = \alpha_2$ implies symmetry (e.g., $U(0,1) = Beta(1,1)$).
- As α_1, $\alpha_2 \to \infty$, we get Gamma.
- As α_1, $\alpha_2 \to 0$, we get Bernoulli.

Beta Density Functions



$Beta(\alpha_1, \alpha_2)$ Density Functions

Use of Lexis Ratio (Discrete Distribution)

Poisson: $p(x) = \dfrac{e^{-\lambda}\,\lambda^x}{x!}, \quad x = 0, 1, 2, \cdots$

$$\mu = \lambda,\ \sigma^2 = \lambda \quad \Rightarrow \quad \tau = \frac{\sigma^2}{\mu} = \frac{\lambda}{\lambda} = 1$$

Binomial: $p(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x\,(1-p)^{n-x},\ x = 0, 1, \cdots$

$$\mu = n\,p,\ \sigma^2 = n\,p\,(1-p) \quad \Rightarrow \quad \tau = \frac{\sigma^2}{\mu} = (1-p) < 1$$

Negative Binomial: $p(x) = \begin{pmatrix} k+x-1 \\ x \end{pmatrix} p^k\,(1-p)^x,$
$x = 0, 1, \cdots$

$$\mu = \frac{k(1-p)}{p},\ \sigma^2 = \frac{k(1-p)}{p^2} \quad \Rightarrow \quad \tau = \frac{\sigma^2}{\mu} = \frac{1}{p} > 1$$

Review of Summary Statistics

Summary Statistic	Indicates
$\bar{x} \approx \tilde{x}$	Symmetric distributions (Normal, Uniform, Tri)
$\widehat{CV} \approx 1$	Exponential distribution
$\widehat{CV} > 1$ and $\hat{\nu} > 0$	LogNormal distribution
$\widehat{CV} < 1$ and $\hat{\nu} > 0$	Gamma or Weibull distribution with $\alpha > 1$
$\hat{\tau} \approx 1$	Poisson distribution
$\hat{\tau} < 1$	Binomial distribution
$\hat{\tau} > 1$	Negative Binomial distribution
$\hat{\nu} \approx 0$	Symmetric distributions (Normal, Uniform, Tri)
$\hat{\nu} < 0$	Skewed to left (Tri, Beta)
$\hat{\nu} > 0$	Skewed to right (Exponential, LogNormal, or Gamma and Weibull with $\alpha > 1$)

Fitting a Theoretical Distribution to Data (II)

"Which parameter value gives the highest probability of observing the data set?"

1. Hypothesizing the family of distribution
   - Intrinsic property of distribution describing a certain situation
   - General characteristics: *e.g.* range, continuous/discrete
   - Summary statistics: *e.g.* mean (μ), median ($\tilde{\mu}$), skewness (ν)
   - Number of "shape" parameters
2. Estimation of parameters
   - Method of maximum likelihood
   - Method of moments
3. Validation of the distribution
   - Expert opinion (so-called face validity)
   - Graphing real data *vs.* fitted distribution
   - Goodness-of-fit (χ^2, Kolmogorov-Sminov, Anderson-Darling)

Method of Maximum Likelihood

Example: Suppose we have a sample $\{2, 3, 2, 7, 4, 8, 10, 5, 3, 4\}$ and we hypothesize that it comes from Binomial($n = 10$).

$$b(n = 10, p, x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \cdots, 10$$

What value of the parameter p is most likely to give rise to the observed data set?

Suppose $p = 0.2$. What is the probability that we get the data set?

$$\left[\binom{10}{2} 0.2^2 0.8^8\right] \left[\binom{10}{3} 0.2^3 0.8^7\right] \cdots \left[\binom{10}{4} 0.2^4 0.8^6\right] = 4.5 \times 10^{-23}$$

$p = 0.3 \rightarrow 1.23 \times 10^{-15}$,
$p = 0.4 \rightarrow 4.03 \times 10^{-13}$, etc.

Maximum Likelihood Estimator

The likelihood function

$$L(\theta) = f(x_1;\theta)\, f(x_2;\theta) \cdots f(x_n;\theta)$$

measures how likely the n observations $x_1, x_2, \cdots, x_n$ are.

In case of a discrete distribution, $L(\theta)$ gives the probability of observing the data set for various values of θ.

We estimate the parameter θ with the value that maximizes the likelihood (probability) $L(\theta)$ of getting the observed data set.

$$\frac{d\,L(\theta)}{d\theta} = 0 \quad \text{or} \quad \frac{d\,\ln L(\theta)}{d\theta} = 0$$

gives the value of θ that maximizes $L(\theta)$ as a function of x_i's.

The corresponding statistic $\hat{\theta} = u_1(X_1, X_2, \cdots, X_n)$ is called the maximum likelihood estimator (MLE) of θ.

Example of MLEs

Exponential Distribution: $f(x) = \dfrac{1}{\beta} e^{-x/\beta}, \quad x > 0$

$$
\begin{aligned}
L(\beta) &= f(x_1;\beta)\, f(x_2;\beta) \cdots f(x_n;\beta) \\
&= \left(\frac{1}{\beta} e^{-x_1/\beta}\right) \left(\frac{1}{\beta} e^{-x_2/\beta}\right) \cdots \left(\frac{1}{\beta} e^{-x_n/\beta}\right) \\
&= \frac{1}{\beta^n} e^{-\sum x_i/\beta} \quad \leftarrow \quad \text{Take log on both side.} \\
l(\beta) &= -n \ln \beta - \frac{\sum x_i}{\beta} \quad \leftarrow \quad \text{Derivative w.r.t. } \beta. \\
\frac{d\,l}{d\beta} &= -\frac{n}{\beta} + \frac{\sum x_i}{\beta^2} = 0 \quad \rightarrow \quad \beta = \frac{\sum x_i}{n} = \overline{x}
\end{aligned}
$$

Summary of Some Easily Computed MLEs

Distribution	Density function	MLE
Uniform	$f(x)=\dfrac{1}{b-a} \quad a \le x \le b$	$\hat{a}=x_{\min}\,,\quad \hat{b}=x_{\max}$
Exponential	$f(x)=\dfrac{1}{\beta}e^{-x/\beta} \quad x>0$	$\hat{\beta}=\bar{x}$
Normal	$f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad x\in\Re$	$\hat{\mu}=\bar{x}\,,\quad \hat{\sigma}=\left[\left(\dfrac{n-1}{n}\right)\dfrac{\sum(x_i-\bar{x})^2}{n}\right]^{1/2}$
LogNormal	$f(x)=\dfrac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \quad x>0$	$\hat{\mu}=\dfrac{\Sigma\ln(x_i)}{n},\quad \hat{\sigma}=\left[\dfrac{\sum(\ln(x_i)-\hat{\mu})^2}{n}\right]^{1/2}$
Bernoulli	$p(x)=\begin{cases}1-p & \text{if } x=0\\ p & \text{if } x=1\end{cases}$	$\hat{p}=\bar{x}$
Discrete Uniform	$p(x)=\begin{cases}1/(b-a+1) & x\in\{a,a+1,\dots,b\}\\ 0 & \text{otherwise}\end{cases}$	$\hat{a}=x_{\min}\,,\quad \hat{b}=x_{\max}$
Binomial	$p(x)=\dbinom{n}{x}p^x(1-p)^{n-x} \quad x=0,1,\dots,n$	if n is known $\hat{p}=\dfrac{\bar{x}}{n}$
Geometric	$p(x)=p(1-p)^x \quad x=0,1,\dots$	$\hat{p}=\dfrac{1}{\bar{x}+1}$
Negative Binomial	$p(x)=\dbinom{k+x-1}{x}p^k(1-p)^x \quad x=0,1,\dots$	if k is known $\hat{p}=\dfrac{k}{\bar{x}+k}$
Poisson	$p(x)=\dfrac{e^{-\lambda}\lambda^x}{x!} \quad x=0,1,2,\dots$	$\hat{\lambda}=\bar{x}$

Method of Moments

MLE preferable, but difficult for Gamma, Weibull, Beta, etc.

k-th moment of distribution: $E(X^k) = \int_{-\infty}^{+\infty} x^k f(x)\, dx$[1]

k-th moment of data: $m_k = \dfrac{\sum x_i^k}{n}$

Suppose we need to estimate parameters $\theta_1, \theta_2, \cdots, \theta_r$.

Then, let $E(X^k) = m_k$ beginning with $k = 1$, continuing until there are enough equations to provide unique solutions to $\theta_1, \theta_2, \cdots, \theta_r$.

[1] $E(X^k) = \sum x_i^k\, p(x_i)$ in case of discrete distribution.

Examples of MOM

Exponential Distribution: $f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$

$$\begin{aligned} E(X) &= \int_0^\infty x \, \frac{1}{\beta} e^{-x/\beta} \, dx = \beta \\ m_1 &= \frac{\sum x_i}{n} = \overline{x} \end{aligned}$$

Gamma Distribution: $f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad x > 0$

We can easily show that $E(X) = \alpha\beta$ and $E(X^2) = \alpha\beta^2(1+\alpha)$.

$\alpha\beta = \frac{\sum x_i}{n}$ and $\alpha\beta^2(1+\alpha) = \frac{\sum x_i^2}{n} \quad \rightarrow \quad \alpha = \frac{\overline{x}^2}{s^2}$ and $\beta = \frac{s^2}{\overline{x}}$.

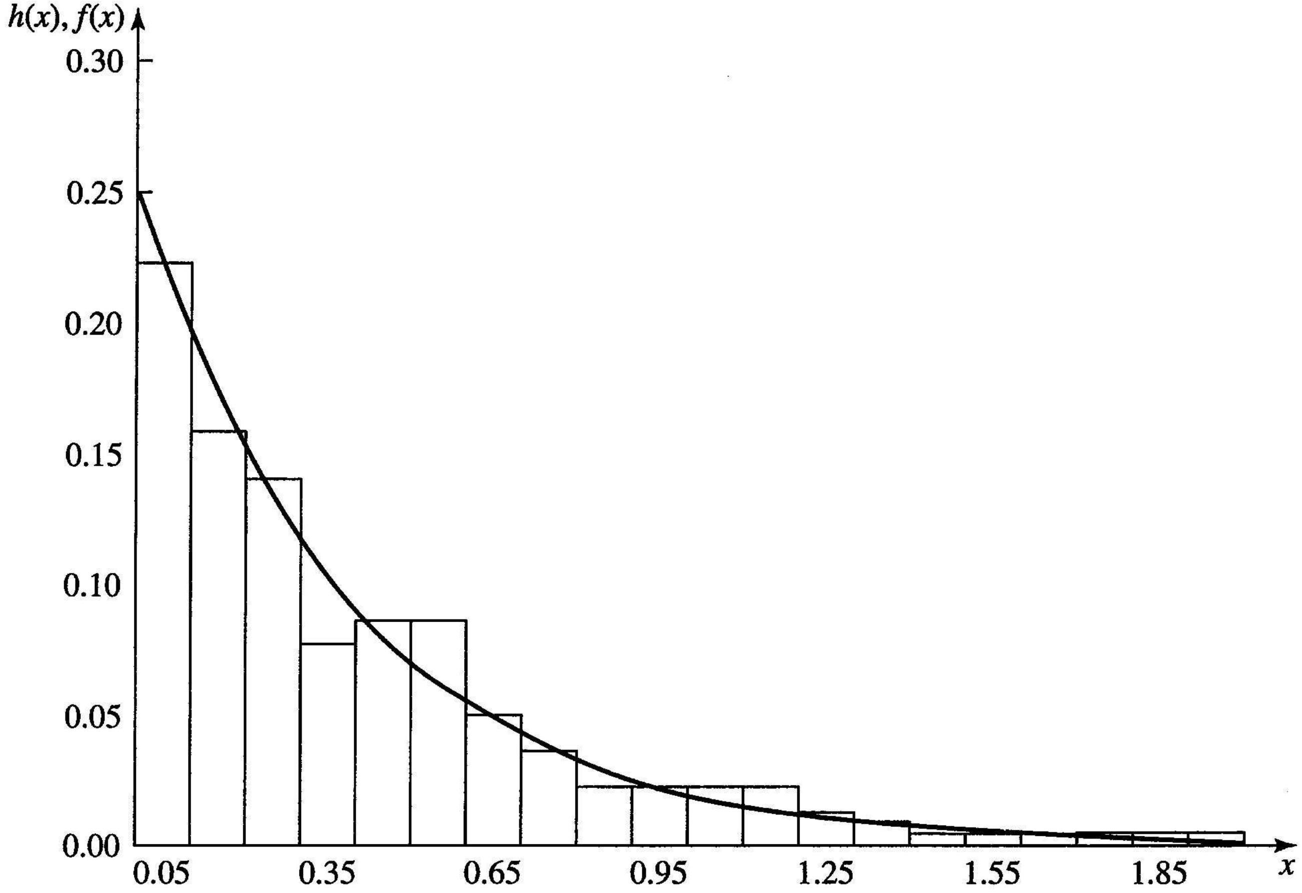
Assignment: Show above!

Fitting a Theoretical Distribution to Data (III)

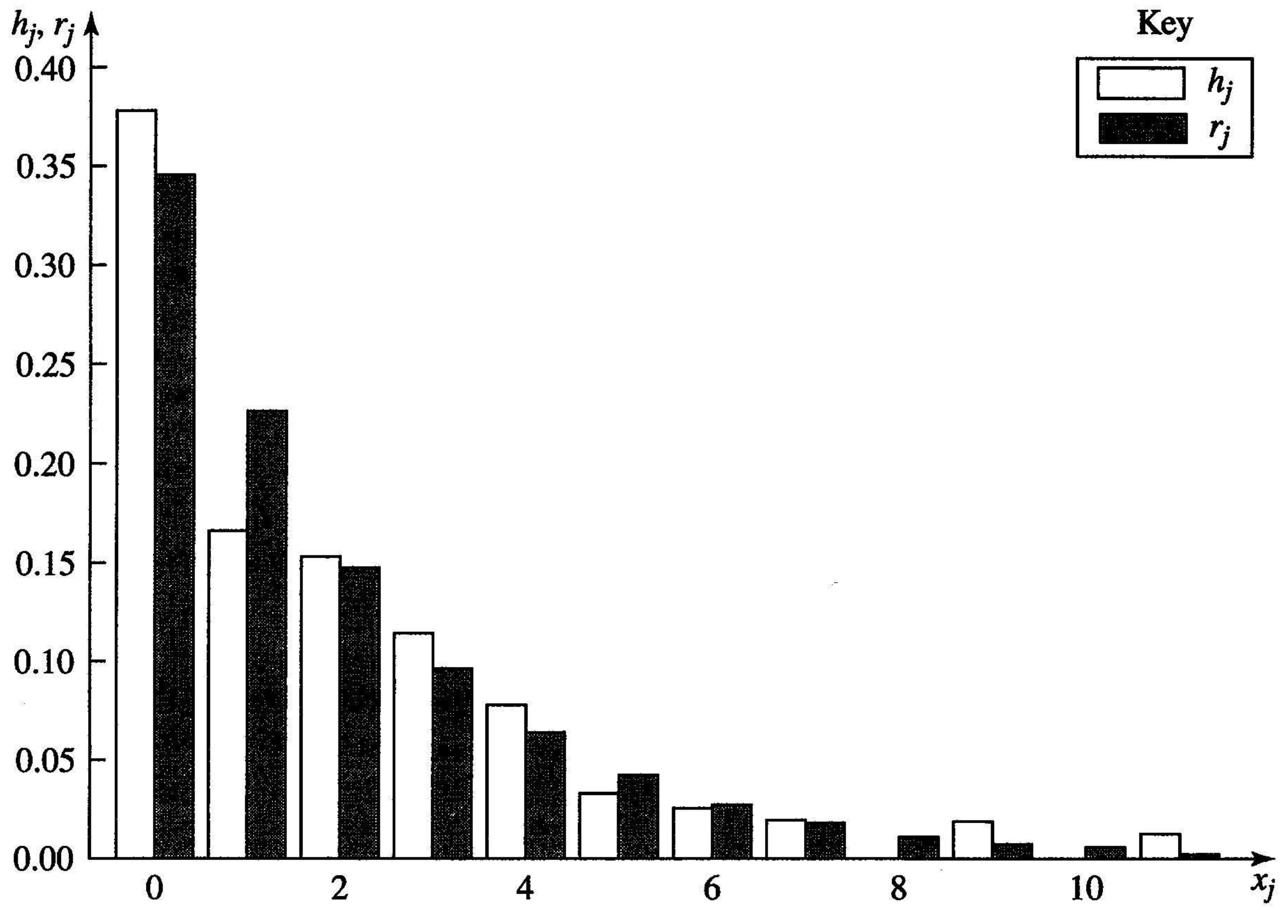
"How representative are the fitted distributions to the observed data set?"

1. Hypothesizing the family of distribution
   - Intrinsic property of distribution describing a certain situation
   - General characteristics: *e.g.* range, continuous/discrete
   - Summary statistics: *e.g.* mean (μ), median ($\tilde{\mu}$), skewness (ν)
   - Number of "shape" parameters
2. Estimation of parameters
   - Method of maximum likelihood
   - Method of moments
3. Validation of the distribution
   - Expert opinion (so-called face validity)
   - Graphing real data *vs.* fitted distribution
   - Goodness-of-fit (χ^2, Kolmogorov-Sminov, Anderson-Darling)

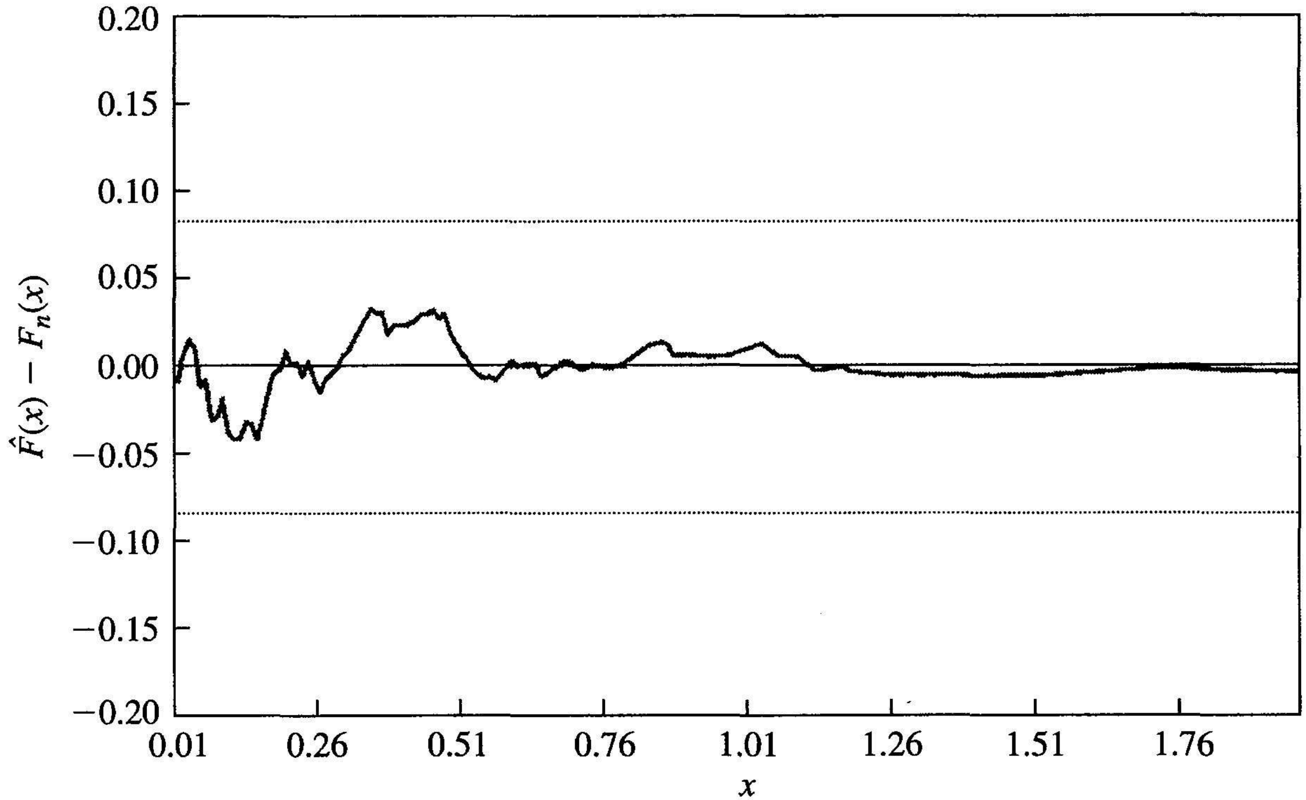
Density/Histogram Overplots



Frequency Comparison



Distribution Function Differences Plots



Probability Plots: Background

$\hat{F}(x)$: Distribution function of the fitted (model) distribution
$F_n(x)$: Empirical distribution function of the sample distribution

$$F_n(x) = \frac{\text{Number of } x_i\text{'s} \le x}{n}$$

Let $x_{(i)}$ be the ith smallest of the x_j's. Note that $F_n(x_{(i)}) = \frac{i}{n}$.

$\tilde{F}_n(x)$: Adjusted empirical cdf so as to avoid $F_n(x_{(n)}) = 1$:

$$\tilde{F}_n(x_{(i)}) = F_n(x_{(i)}) - \frac{0.5}{n} = \frac{i - 0.5}{n}$$

A straightforward procedure would be to plot $\tilde{F}_n(x_{(i)})$ vs $\hat{F}(x)$
$\longrightarrow$ Difficult to identify similarities/differences.

Probability Plots: Definitions (I)

Quantile-quantile (Q-Q) Plot

Let $q_i = \dfrac{i - 0.5}{n}$ for $i = 1, 2, \cdots, n$, so that $0 < q_i < 1$.

$$\begin{cases} q_i\text{-quantile of a fitted (model) distribution, } & x^M_{q_i} = \hat{F}^{-1}(q_i) \\ q_i\text{-quantile of the sample distribution, } & x^S_{q_i} = \tilde{F}^{-1}(q_i) = x_{(i)} \end{cases}$$

- Each ordinate value $q \;\longrightarrow\;$ Two quantiles x^M_q and x^S_q
- The Q-Q plot will amplify the differences between the *tails* of the distribution functions.

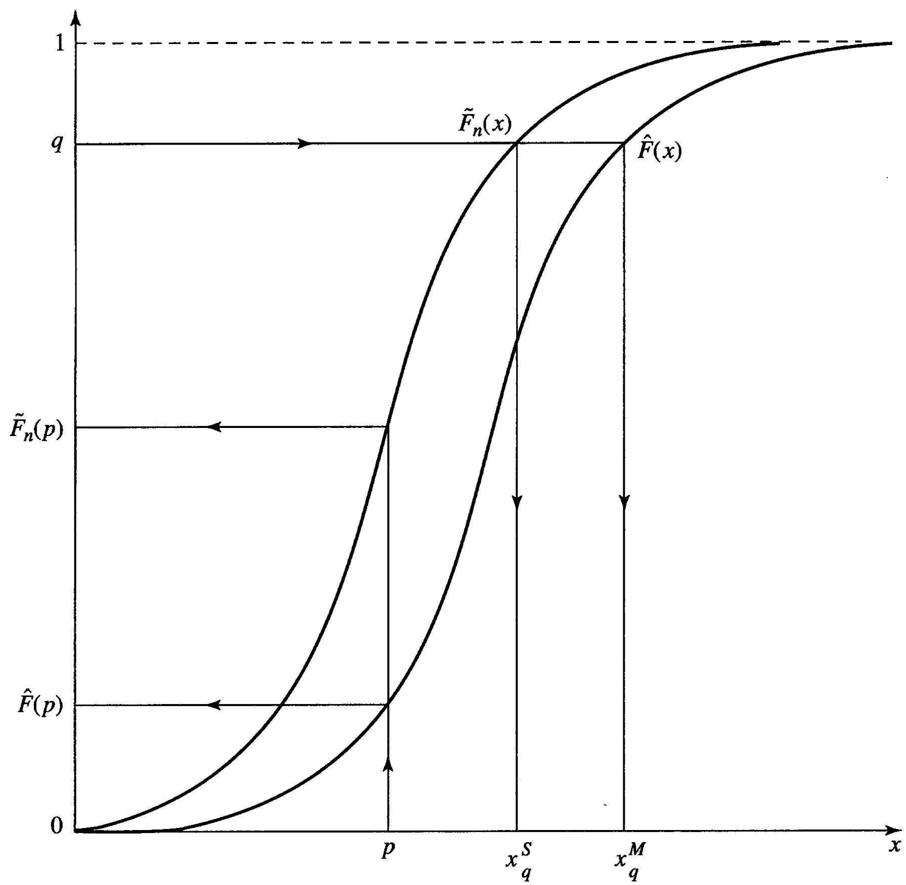
Probability Plots: Definitions (II)

Probability-probability (P-P) Plot

$$\left.\begin{cases} \text{The fitted (model) probability, } \hat{F}(x_{(i)}) \\ \text{The sample probability, } \tilde{F}(x_{(i)}) = q_i \end{cases}\right\} \text{for } i = 1, 2, \cdots, n.$$

- Each abscissa value $p \ \longrightarrow\ $ Two probabilities $\hat{F}(p)$ and $\tilde{F}_n(p)$
- The P-P plot will amplify the differences between the *middles* of the distribution functions.

Probability Plots: Intuitive Appeal



Chi-square Goodness-of-Fit Tests

(1) $\begin{cases} H_0 : \text{Data comes from the distribution with cdf } \hat{F}(x) \\ H_1 : \text{Not } H_0 \end{cases}$

(2) Develop a histogram of data cells $[b_0, b_1), \cdots, [b_{k-1}, b_k)$.
Let $o_j =$ (observed) frequency of cell j.

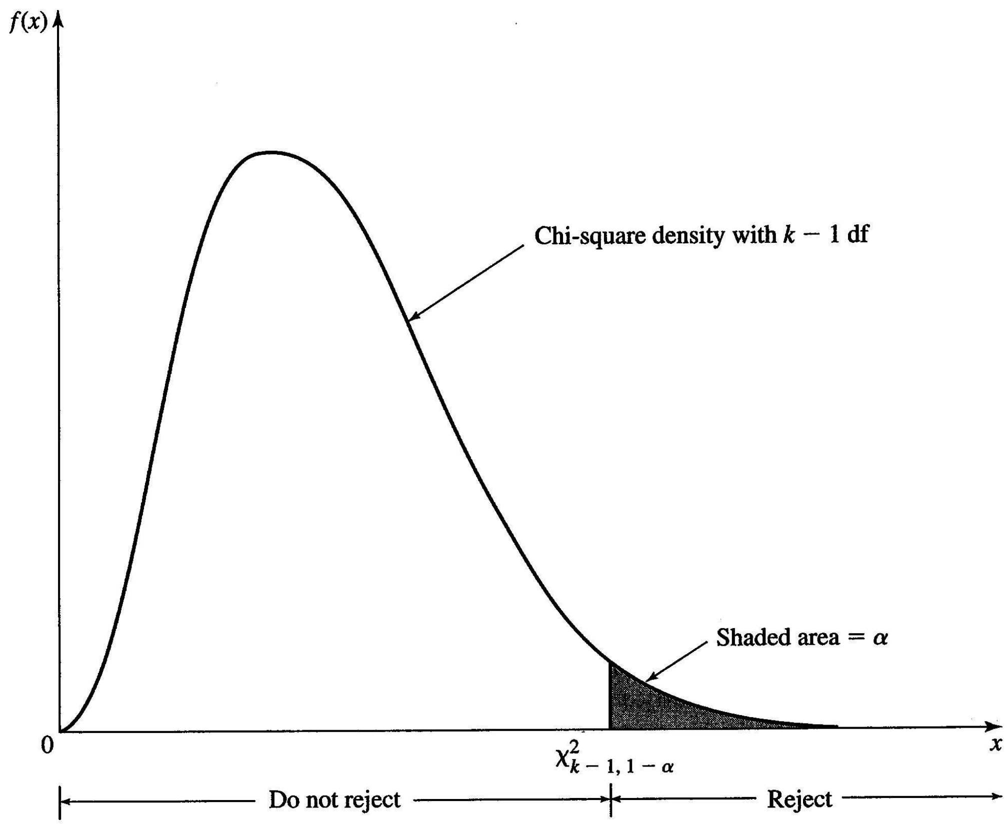
(3) For each cell, compute the expected frequency e_j assuming H_0 is correct: $e_j = n\left(\hat{F}(b_j) - \hat{F}(b_{j-1})\right)$

(4) For each cell, compute $\dfrac{(e_j - o_j)^2}{e_j}$.

(5) Compute $\chi^2 = \displaystyle\sum_{j=1}^{k} \frac{(e_j - o_j)^2}{e_j}$, which is $\chi^2(k-1)$ under H_0.

(6) Reject H_0 if $\chi^2 > \chi^2_{k-1,\alpha-1}$

Chi-square: Testing the Hypothesis



Validation of the Fitted Distribution: Summary

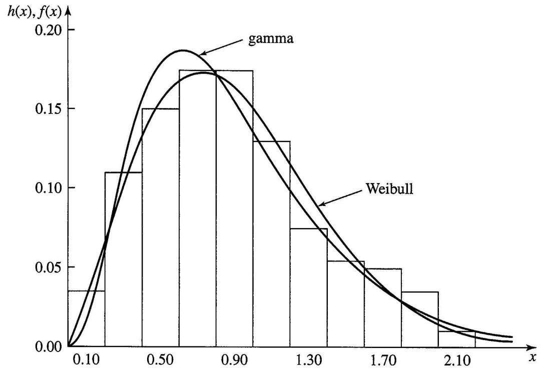


FIGURE 6.52
Density/histogram overplots for the service-time data and the Weibull and gamma distributions.

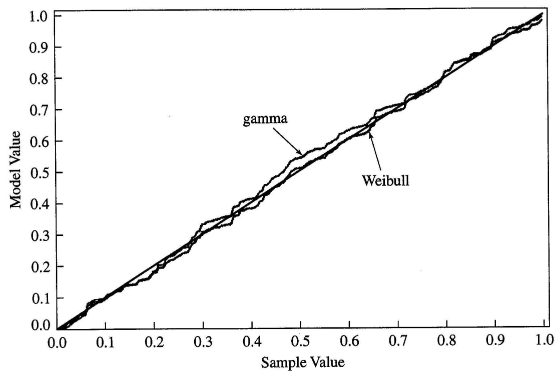


FIGURE 6.54
P–P plots for the service-time data and the Weibull and gamma distributions.

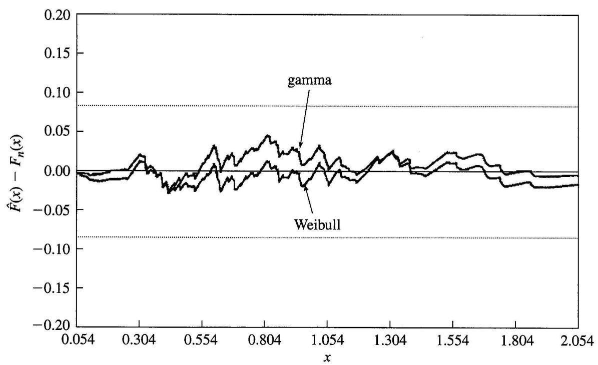


FIGURE 6.53
Distribution function differences plots for the service-time data and the Weibull and gamma distributions.

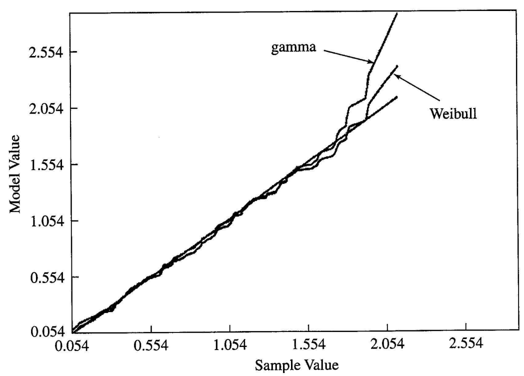


FIGURE 6.55
Q–Q plots for the service-time data and the Weibull and gamma distributions.