Chapter 7 Probability Distributions Assessment

"*A reasonable probability is the only certainty*."

E.W. Howe

7.1 Introduction....2

7.2 Assessing Probabilities from Experts....3
7.2.1 Eliciting Probabilities from Expert using Reference Lotteries or Deals....3
7.2.2 Biases in Probability Encoding by Experts....5
7.2.3 Fitting an Expert's CDF to a Theoretical Distribution....6

7.3 Learning Probabilities from Data....10
7.3.1 Relative Frequencies....10
7.3.2 Fitting Theoretical Distributions to Data....11
7.3.3 Normal Distribution with Two Known Percentile Values....19

7.4 Discrete Approximation of Continuous Probability Distribution....22
7.4.1 Introduction....22
7.4.2 Desired Properties of the Discrete Approximation....22
7.4.3 Limit on the number of moments that can be preserved....23
7.4.4 Discrete Approximation for the Uniform Distributions....24
7.4.5 Discrete Approximation for Normal Distributions....25
7.4.6 3-Branch Discretization of the Asymmetric Triangular Distribution....27
7.4.7 Discrete Approximation for Continuous Distributions in DPL....28
7.4.8 Three-Point Quick Approximation Methods....28

References....29

Exercises....30

7.1 Introduction

- To perform Decision Analysis, we require the probability distributions of all the uncertain variables in the decision model.

- These probability distributions may be obtained by the following methods:

    1. Assessed directly via **Experts' Judgments** using **Probability Assessment Protocols**.

    2. **Learned from Data** using *Statistical or Machine Learning*.

    3. **Generated** from *Stochastic Models* such as queuing theory, Monte Carlo simulation, discrete event simulation, Markov chain, etc. But these models also require probability distribution assessments.

    4. Combinations of any of the above.

- We will cover methods 1 and 2 briefly in this chapter.

- Method 3 is covered in other IE modules.

- Note that although we have adopted the subjective view of the meaning of probabilities, we may still combine probabilities obtained objectively from data/models with those elicited directly from experts.

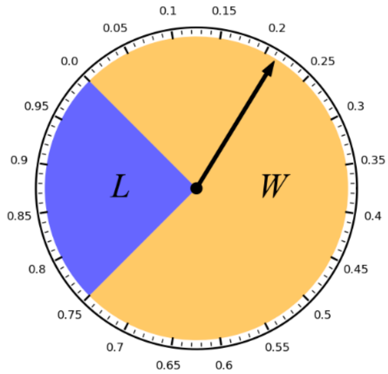
7.2 Assessing Probabilities from Experts

7.2.1 Eliciting Probabilities from Expert using Reference Lotteries or Deals

Assessing the Probability Distribution of a Discrete Variable

- For a discrete variable, the method will produce a probability mass function.
- Let X be a discrete uncertain variable with two outcomes: x_1 and x_2.
- The steps to assess the probability distribution of X are as follows:

1. Let W and L be two rewards such that $W \succ L$.
   For example, W could be a small cash amount and L is nothing.
2. Set the probability wheel with the orange sector at p.
3. **Repeat**



4. Ask the expert to choose between the two options:

   A: Spin the probability wheel.
   If the outcome is orange, he receives W.
   Otherwise, he receives L.

   B: Do not spin the wheel.
   If the outcome of $X = x_1$, he receives W.
   Otherwise, he receives L.

5. If the expert chose *A,* reduce the value of p set on the wheel.
   If the expert chose *B*, increase the value of p set on the wheel.

6. **Until** the expert is indifferent between options *A* and *B*.

7. Result: $\Pr\{X = x_1\} = p$
   $\Pr\{X = x_2\} = 1 - p.$

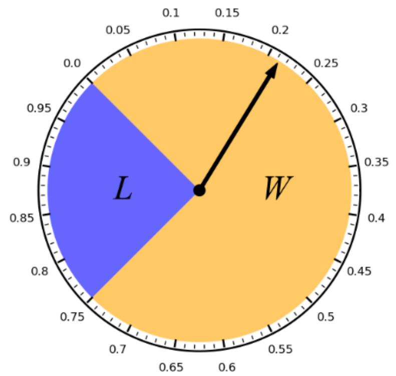
Discrete Variables with more than 2 outcomes:

- If X has more than 2 outcomes, the procedure above is repeated for outcomes x_2 to x_{n-1} to obtain $p_2, \ldots, p_{n-1}$, and let $p_n = 1 - \sum_{i=1}^{n-1} p_i$.
- The output is a probability mass function for X.

Assessing the Probability Distribution for a Continuous Variable

- For a continuous uncertain variable, the method will produce a CDF.
- Let X be a continuous uncertain variable.
- The steps to assess its CDF are as follows:

1. Let W and L be two rewards such that $W \succ L$.
   For example, W could be a small cash amount and L is nothing.
2. Select a value of x from the range of possible values of X based on a selected assessment protocol.
3. Set the probability wheel with the orange sector at p.
4. **Repeat**

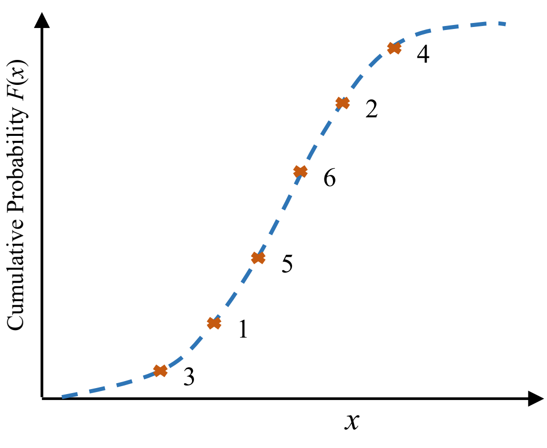


5. Ask the expert to choose between the two options:

   A: Spin the probability wheel.
   If the outcome is orange, he receives W.
   Otherwise, he receives L.

   B: Do not spin the wheel.
   If the actual outcome of $X \le x$, he receives W.
   Otherwise, he receives L.

6. If the expert chose A, reduce the value of p set on the wheel.
   If the expert chose B: increase the value of p set on the wheel.
7. **Until** the expert is indifferent between options A and B.
8. Encode $\Pr(X \le x) = p$ on the CDF for X.
9. Repeat Steps 1 to 8 for a number of other selected values of x based on the selected assessment protocols. Each of these values results in a point on CDF for X.
10. Plot the CDF for X.



- Note: The choice of x values and their sequence is important to avoid biases in the results. See Stanford/SRI protocol for guidelines on how to choose the sequence of values.

7.2.2 Biases in Probability Encoding by Experts

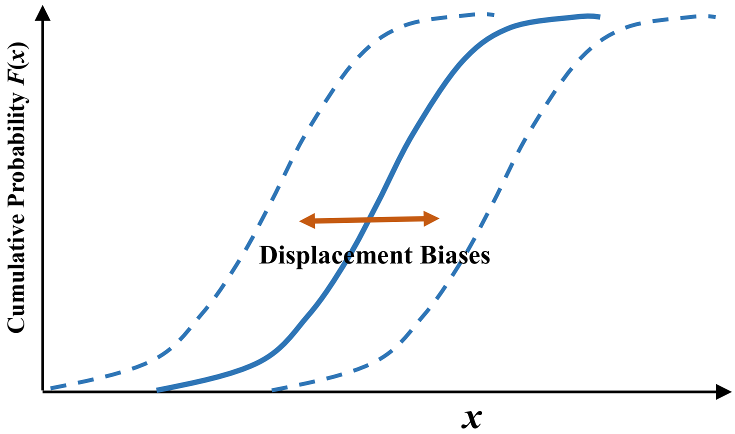
- The task of the analyst is to elicit from the expert a probability distribution that best describes the underlying knowledge of the expert.

- Conscious or subconscious discrepancies between the expert's responses and an accurate description of his underlying knowledge are called **biases**.

- There are two types of biases:
  1. Displacement bias
  2. Variability bias

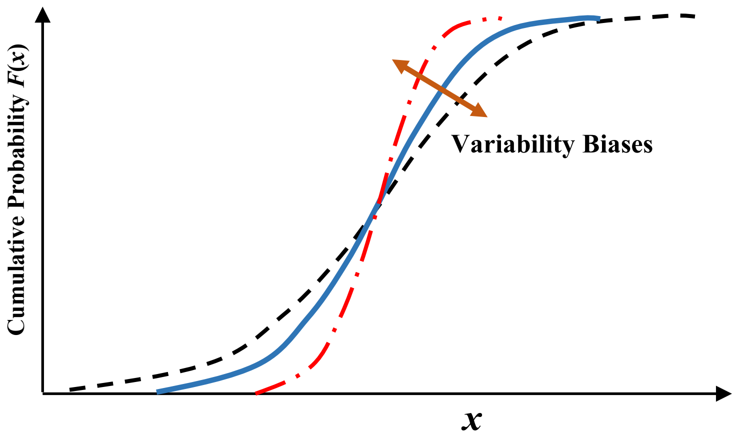
Displacement Bias

- Here there is a shift of the estimated probability distribution from the correct distribution in one direction.



Variability Bias

- Here there is an error in the shape of the distribution compared with the correct distribution.



- When the assessed distribution becomes tighter (has less spread) than the actual distribution (i.e., the red distribution), it is known as **Central Bias**.
- It shows overconfidence in the estimation of the variability of the variable by the expert.

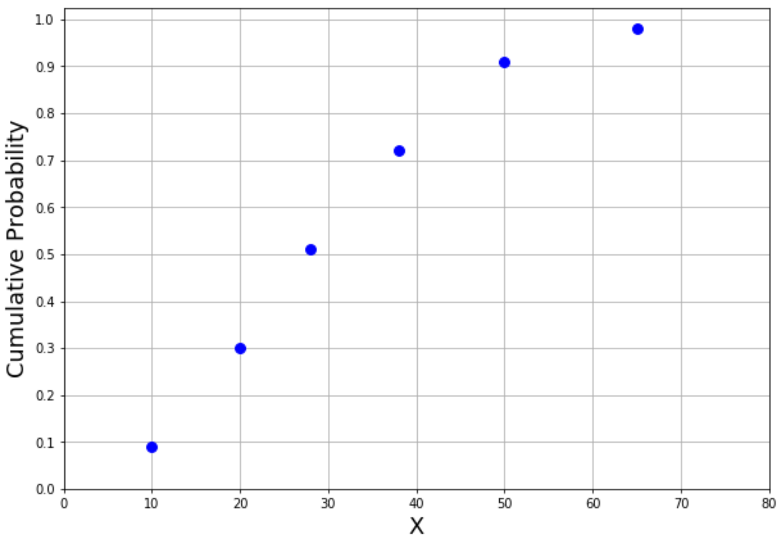
7.2.3 Fitting an Expert's CDF to a Theoretical Distribution

- The probability distribution for a continuous variable assessed by an expert is a CDF in graphical form.

- One way to use this information in a decision tree or an influence diagram software is to fit a theoretical probability distribution to the data.

Example (Fitting an expert's CDF using Python scipy.optimize.curve_fit())

- An expert assessed the following 6-point CDF for an uncertain variable X.



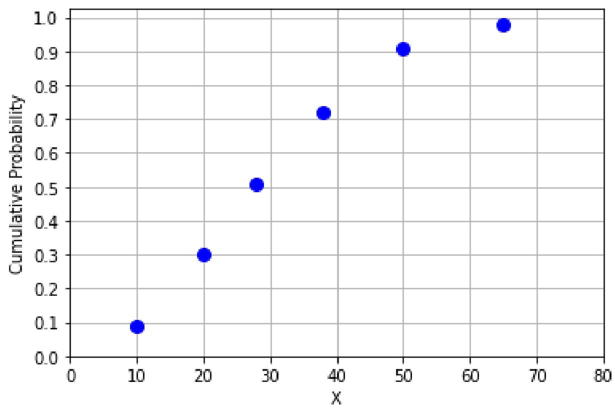
x	$F(x)$
10	0.09
20	0.30
28	0.51
38	0.72
50	0.91
65	0.98

Python:

In [1]:
```
""" Fitting Expert Assessed CDF """
import numpy as np
import scipy.stats
import matplotlib.pyplot as plt
```

In [2]:
```
# Data to fit CDF
xdata = [  10,   20,   28,   38,   50,   65 ]
ydata = [0.09, 0.30, 0.51, 0.72, 0.91, 0.98 ]
```

In [3]:
```
# Visualize the CDF data points
fig0, ax0 = plt.subplots()
ax0.plot(xdata, ydata, 'bo', ms=8)
ax0.set_xlim(0, 80)
ax0.set_yticks(np.linspace(0, 1, 11))
ax0.set_ylabel("Cumulative Probability")
ax0.set_xlabel("X")
ax0.grid()
plt.show()
```

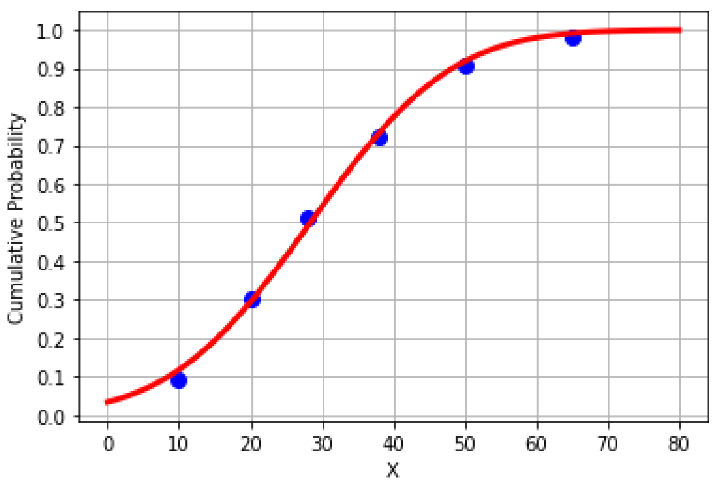


In [4]:
```
# First, try fitting a Normal Distribution
# The CDF to fit
fnorm = lambda x,mu,sigma: scipy.stats.norm(loc=mu, scale=sigma).cdf(x)

# Fit the data
popt, pcov = scipy.optimize.curve_fit(fnorm, xdata, ydata, p0=(50,10))
print(f"\nFitted Normal distribution with parameters ",
      ', '.join(f"{p:.4f}" for p in popt))
print(f"Mean = {scipy.stats.norm(*popt).mean():.4f}")
print(f"Std Dev = {scipy.stats.norm(*popt).std():.4f}")
```

```
Fitted Normal distribution with parameters  28.4258, 15.4810
Mean = 28.4258
Std Dev = 15.4810
```

In [5]:
```
# Visualize the fitted distribution against the data
xfit = np.linspace(0, 80, 81)
yfit1 = fnorm(xfit, *popt)
fig1, ax1 = plt.subplots()
ax1.plot(xdata, ydata, 'bo', ms=8)
ax1.plot(xfit,  yfit1, 'r-', lw=3 )
ax1.set_yticks(np.linspace(0,1,11))
ax1.set_ylabel("Cumulative Probability")
ax1.set_xlabel("X")
ax1.grid()
plt.show()
```



- The fitted normal distribution looks reasonable, but the left tail does not fit well and goes theoretically to minus infinity.

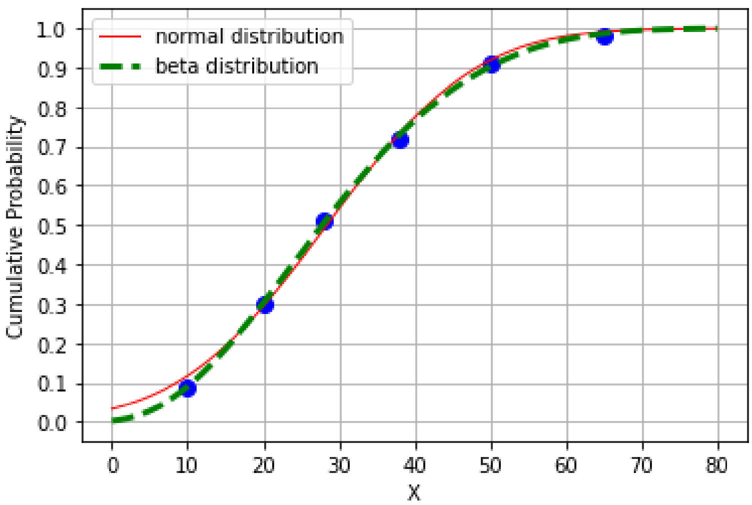
- Also, the expert's CDF is skewed a bit to the left. Let's try to fit an asymmetrical beta distribution to the data instead.

In [6]:
```
# Try fitting a beta distribution and compare it with the normal distribution
# The CDF to fit
fbeta = lambda x, a, b, L, S: scipy.stats.beta(a,b,loc=L,scale=S).cdf(x)

# Fit the data
popt, pcov = scipy.optimize.curve_fit(fbeta, xdata, ydata, p0=(3, 7, 0, 100))
print("\n\nFitted Beta distribution with parameters:",
      ', '.join(f"{p:.2f}" for p in popt))
print(f"Mean = {scipy.stats.beta(*popt).mean():.4f}")
print(f"Std Dev = {scipy.stats.beta(*popt).std():.4f}")
```

```
Fitted Beta distribution with parameters: 2.84, 5.93, -3.12, 99.79
Mean = 29.2084
Std Dev = 14.9350
```

In [7]:
```
# Visualize the fitted distribution against the data
yfit2 = fbeta(xfit, *popt)
fig2, ax2 = plt.subplots()
ax2.plot(xdata, ydata, 'bo', ms=8)
ax2.plot(xfit,  yfit1, 'r-',  lw=1, label="normal distribution")
ax2.plot(xfit,  yfit2, 'g--', lw=3, label="beta distribution")
ax2.set_yticks(np.linspace(0,1,11))
ax2.set_ylabel("Cumulative Probability")
ax2.set_xlabel("X")
ax2.legend()
ax2.grid()
plt.show()
```

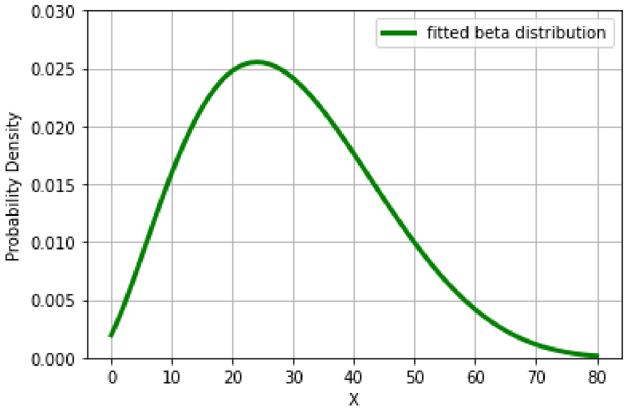


- The beta distribution is a better fit to the data.

In [8]:

```
# Finally, plot the PDF of the fitted beta distribution
print("\nPDF of fitted Beta distribution")
fbeta_pdf = lambda x, a, b, L, S: scipy.stats.beta(a,b,loc=L,scale=S).pdf(x)
yfit2pdf = fbeta_pdf(xfit, *popt)
fig3, ax3 = plt.subplots()
ax3.plot(xfit, yfit2pdf, 'g-', lw=3, label="fitted beta distribution")
ax3.set_ylim(0, 0.03)
ax3.set_ylabel("Probability Density")
ax3.set_xlabel("X")
ax3.legend()
ax3.grid()
plt.show()
```

```
PDF of fitted Beta distribution
```



- This distribution can be used directly in DPL.

7.3 Learning Probabilities from Data

- We may make use of data to access probabilities. Some of the methods of using data for estimating probabilities are discussed here.

7.3.1 Relative Frequencies

- The **relative frequency** of past events is often a good starting point to obtain an estimate of the probabilities of a discrete variable with few outcomes.

Example

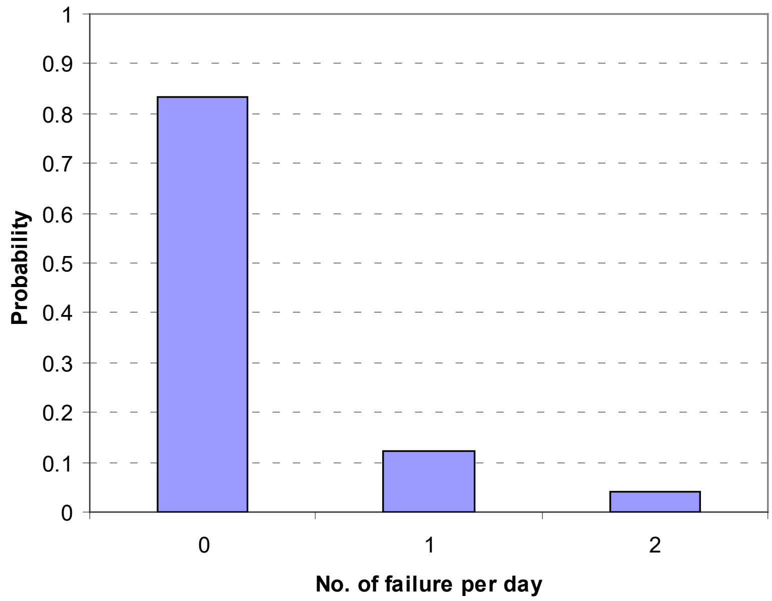
- Let X = number of machine failures per day.

- The following data were collected over a period of 260 days:

No. of failures	No. of occurrences
0	217
1	32
2	11

- The probability mass function for X may be approximated by:

No. of Failures	Probability
0	217/260 = 0.835
1	32/260 = 0.123
2	11/260 = 0.042



7.3.2 Fitting Theoretical Distributions to Data

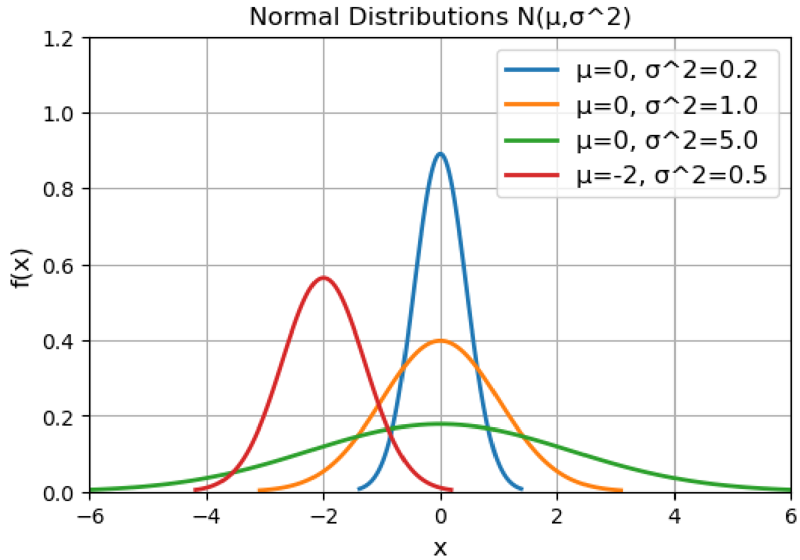
- If a process or an event is known to follow or may be assumed to follow a standard or theoretical probability distribution, then the problem is one of finding the best estimate of the parameters for the distribution given the data observed.

Some Commonly Used Probability Distributions

Normal distribution: $N(\mu, \sigma^2)$

PDF: $f(x;\mu,\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-1/2((x-\mu)^2/\sigma^2)}$ for $-\infty < x < \infty$

μ = mean, σ = standard deviation, *location* = μ, *scale* = σ

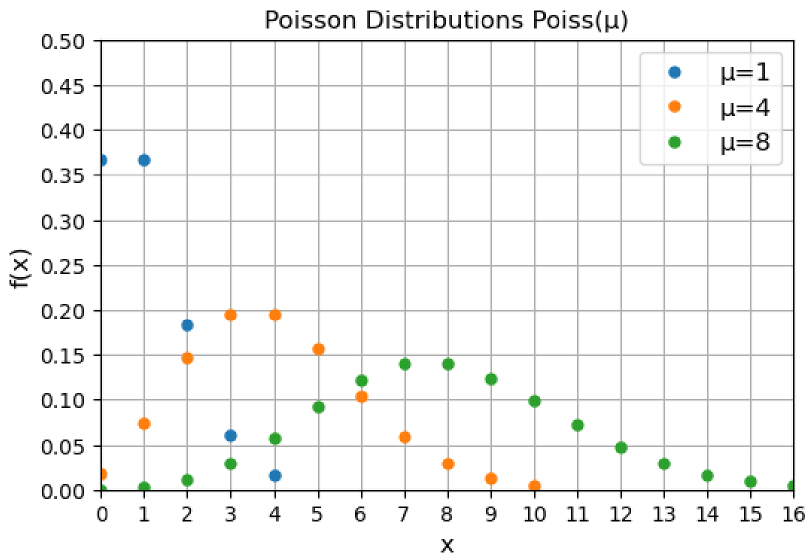


Poisson distribution: Pois(μ)

- The Poisson distribution is a discrete distribution for modeling the number of occurrences in a given time interval.

PMF: $$f(x;\mu) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!} & \text{if } x = 1, 2, 3, \ldots \\ 0 & \text{Otherwise} \end{cases}$$

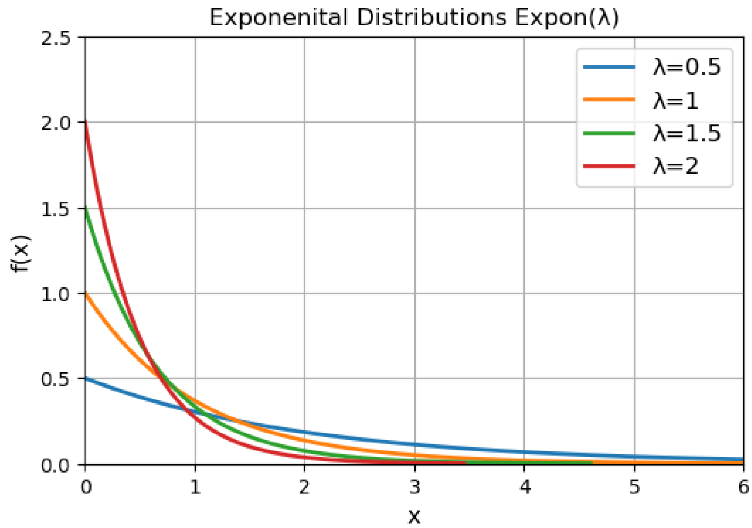
μ = average per unit time, variance = μ, location = 0, scale = μ.



One-Parameter Exponential distribution: $Expon(\lambda)$

PDF: $f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$

$\lambda > 0$ is the rate parameter, mean = $1/\lambda$, variance = $1/\lambda^2$, *location* = 0, *scale* = $1/\lambda$

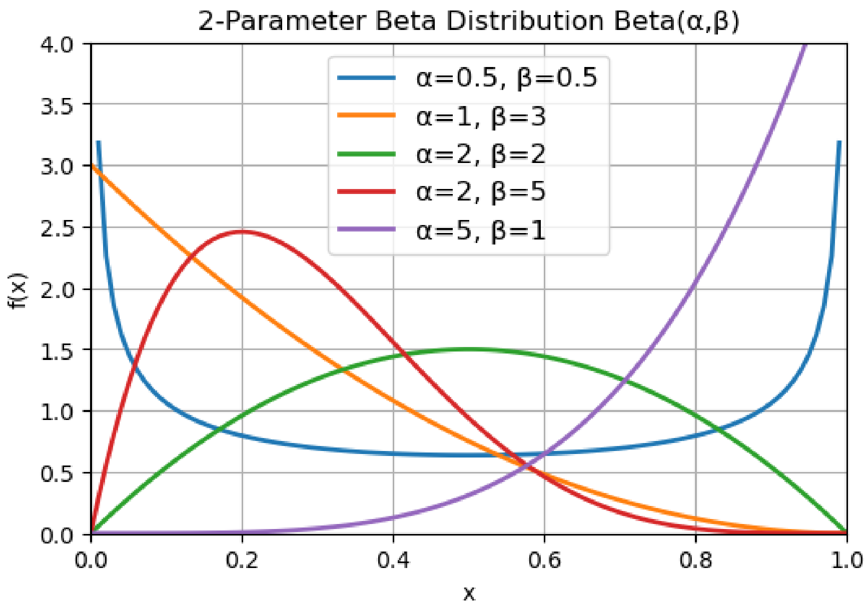


- The exponential distribution is popular for modeling the inter-arrival time of a Poisson process.

Standard two-parameter Beta distribution: Beta(α, β).

PDF: $f(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt} & 0 \le x \le 1 \\ 0 & \text{Otherwise} \end{cases}$

$\alpha > 0$ and $\beta > 0$ are shape parameters, mean $= \dfrac{\alpha}{\alpha+\beta}$, variance $= \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$



- The standard 2-parameter Beta distribution is useful for representing asymmetric distributions by specifying the shape parameters α and β when $0 \le x \le 1$.

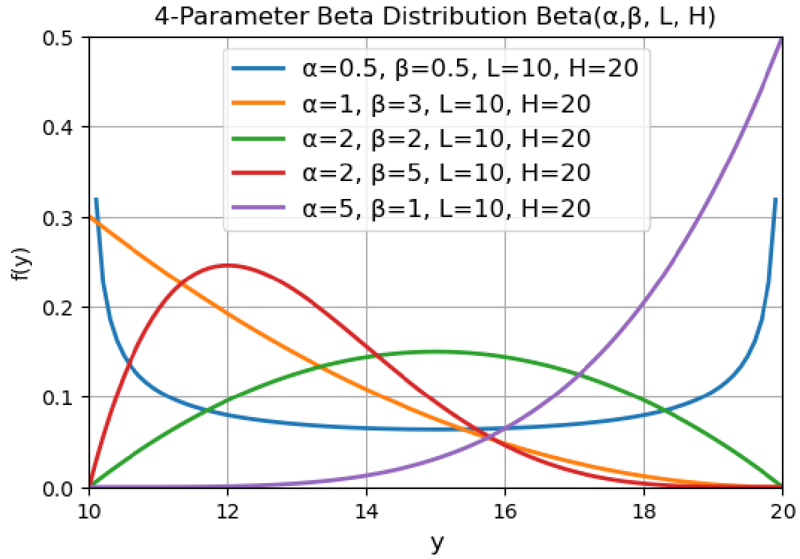
Four-parameter Beta distribution: Beta(α, β, L, H)

- We can use the 4-parameter Beta distribution when $L \le x \le H$, by applying the linear transformation $y = x\,(H - L) + L$. Y is Beta distributed with 4 parameters α, β, L, H if and only if $X = (Y - L)/(H - L)$ is Beta distributed with 2 parameters α and β.

PDF: $$f(y;\alpha,\beta,L,H) = \begin{cases} \dfrac{f(x;\alpha,\beta)}{(H-L)} & L \le y \le H \\ 0 & \text{Otherwise} \end{cases} \qquad \text{where } x = \frac{(y-L)}{(H-L)}$$

$\alpha > 0$ and $\beta > 0$ are shape parameters.

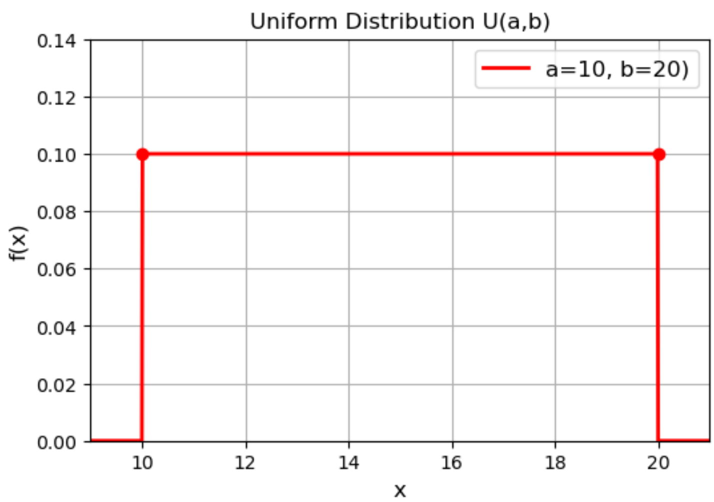
mean $= \dfrac{\alpha H + \beta L}{\alpha + \beta}$, variance $= \dfrac{\alpha\beta(H-L)^2}{(\alpha+\beta)^2(\alpha+\beta+1)}$, *location* $= L$, scale $= (H - L)$



Uniform distribution: $U(a, b)$

PDF: $$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

mean $= \dfrac{(a+b)}{2}$, variance $= \dfrac{(b-a)^2}{12}$, *location* $= a$, scale $= (b - a)$



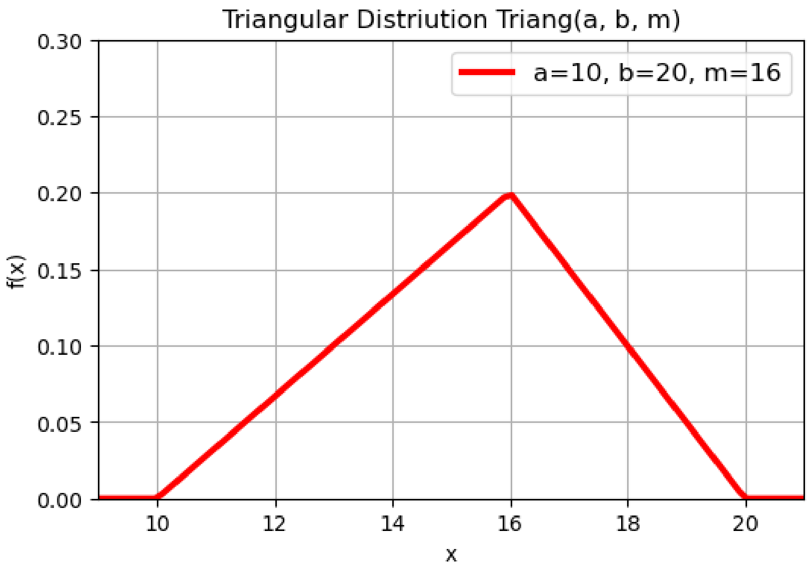
Triangular distribution: Triangular (a, b, m)

$$\text{PDF: } f(x) = \begin{cases} 0 & x < a \\ \dfrac{2(x-a)}{(b-a)(m-a)} & a \le x < m \\ \dfrac{2}{b-a} & x = m \\ \dfrac{2(b-x)}{(b-a)(b-m)} & s < x \le b \\ 0 & b < x \end{cases}$$

a = lower bound, b = upper bound, m = mode.

mean $= \dfrac{(a+b+m)}{3}$, variance $= \dfrac{a^2+b^2+m^2-ab=am-bm}{18}$,

$location = a,\ \ scale = \dfrac{m-a}{b-a}$



The Maximum Likelihood Estimation (MLE) Method

- This is a popular parameters estimation method and is easy to implement computationally.

- Let $\boldsymbol{D} = \{ x_1, x_2, \ldots, x_n \}$ be a set of observed data values. Suppose we want to fit data $\boldsymbol{D}$ to a probability distribution with PDF or PMF $f(x|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of unknown parameters.

- The **Likelihood Function** $L(\boldsymbol{\theta} \mid \boldsymbol{D})$ is the joint probability of observing all the x values in $\boldsymbol{D}$ given that the distribution to be fitted has the parameter $\boldsymbol{\theta}$. That is,

$$L(\boldsymbol{\theta} \mid \boldsymbol{D}) = f(x_1, x_2, \ldots, x_n \mid \boldsymbol{\theta})$$

- The MLE method finds the value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta} \mid \boldsymbol{D})$, that is

$$\hat{\boldsymbol{\theta}} = \arg \underset{\boldsymbol{\theta}}{Max}\, f(x_1, x_2, \ldots, x_n \mid \boldsymbol{\theta})$$

- If the data values are assumed to be mutually independent, then

$$\hat{\boldsymbol{\theta}} = \arg \underset{\boldsymbol{\theta}}{Max}\left(\prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta}) \right)$$

- It is usually more convenient to **Minimize the Negative Log-Likelihood Function** instead as log() is a monotonically increasing function. That is

$$\hat{\boldsymbol{\theta}} = \arg \underset{\boldsymbol{\theta}}{Min}\left(-\sum_{i=1}^{n} \log f(x_i \mid \boldsymbol{\theta}) \right)$$

- Some distributions (e.g. Normal, Poisson, Gamma) have known closed-form optimal solutions for $\hat{\boldsymbol{\theta}}$. Otherwise, numerical optimization algorithms are often used to find $\hat{\boldsymbol{\theta}}$.

Using Software Tools

1. **Excel Add-in @Risk.**

   - This is a Monte-Carlo simulation software that can fit both continuous and discrete probability distributions. It is available in the DecisionTools suite (http://www.palisade.com). Limited-time trial versions are available.

2. **Python scipy.stats**

   - You can directly fit discrete probability distributions supported by scipy.stats using the MLE or MM (method of moments) methods.

3. **Other Software**

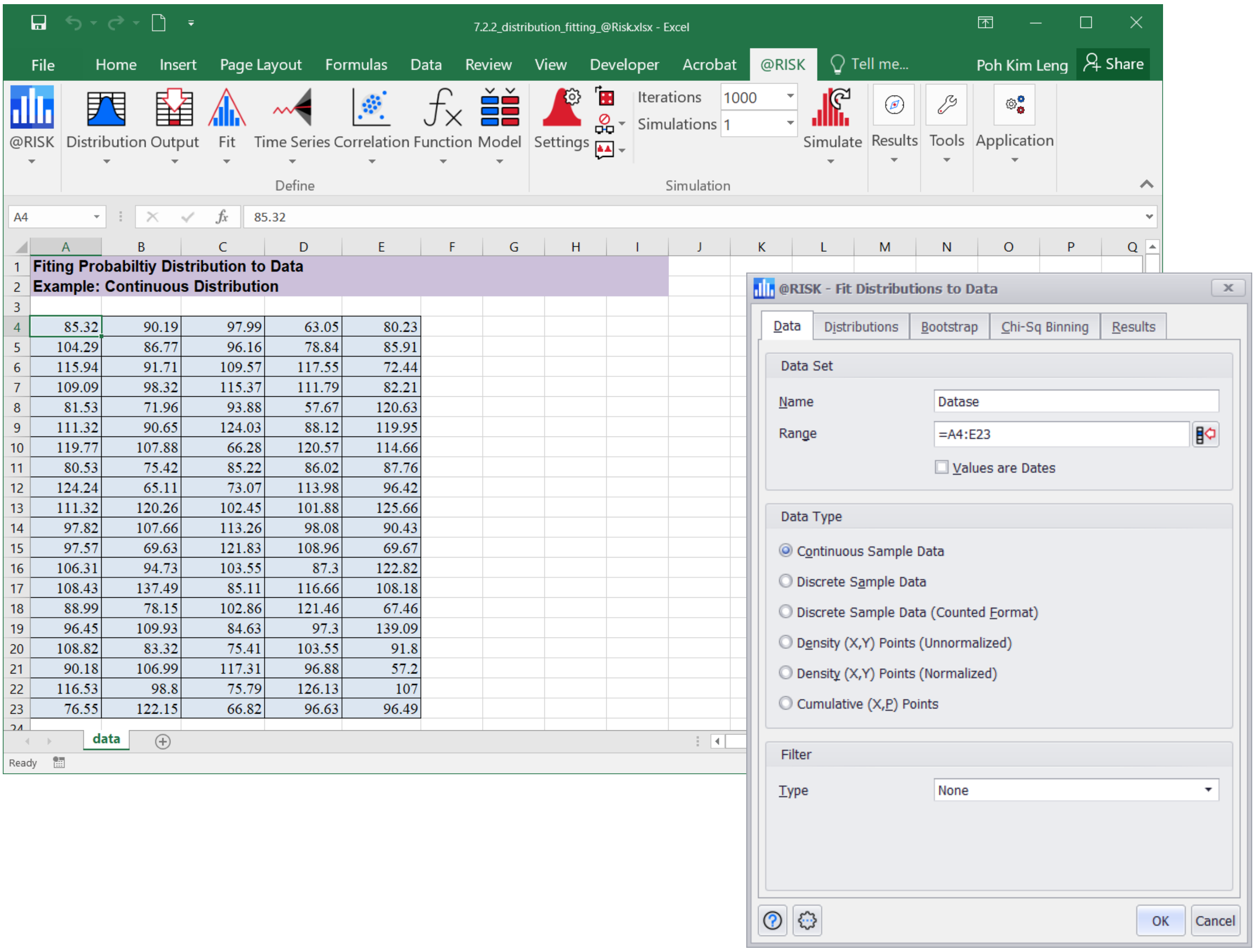
   - EasyFit
   - XLSTAT
   - ExpertFit from FlexSim

Example 1 (Continous Distributions)

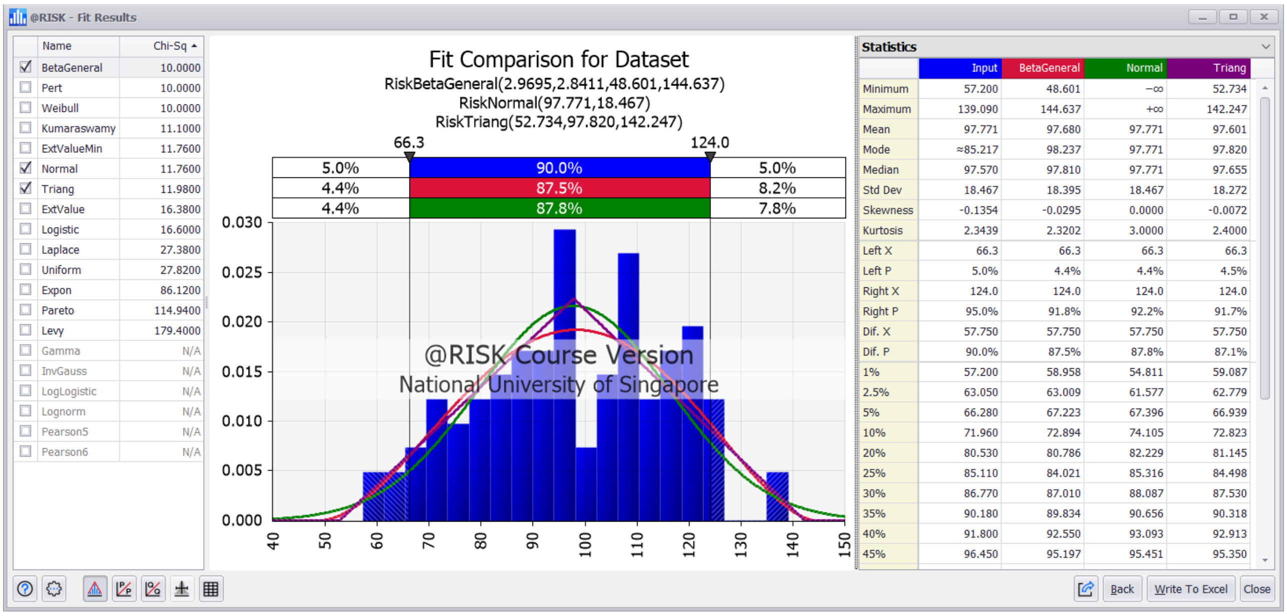
- Fit a continuous distribution to the following data set (100 observations):

85.32	90.19	97.99	63.05	80.23
104.29	86.77	96.16	78.84	85.91
115.94	91.71	109.57	117.55	72.44
109.09	98.32	115.37	111.79	82.21
81.53	71.96	93.88	57.67	120.63
111.32	90.65	124.03	88.12	119.95
119.77	107.88	66.28	120.57	114.66
80.53	75.42	85.22	86.02	87.76
124.24	65.11	73.07	113.98	96.42
111.32	120.26	102.45	101.88	125.66
97.82	107.66	113.26	98.08	90.43
97.57	69.63	121.83	108.96	69.67
106.31	94.73	103.55	87.3	122.82
108.43	137.49	85.11	116.66	108.18
88.99	78.15	102.86	121.46	67.46
96.45	109.93	84.63	97.3	139.09
108.82	83.32	75.41	103.55	91.8
90.18	106.99	117.31	96.88	57.2
116.53	98.8	75.79	126.13	107
76.55	122.15	66.82	96.63	96.49

Using @Risk:



- Fitted Results



Continuous Distributions Fitted:

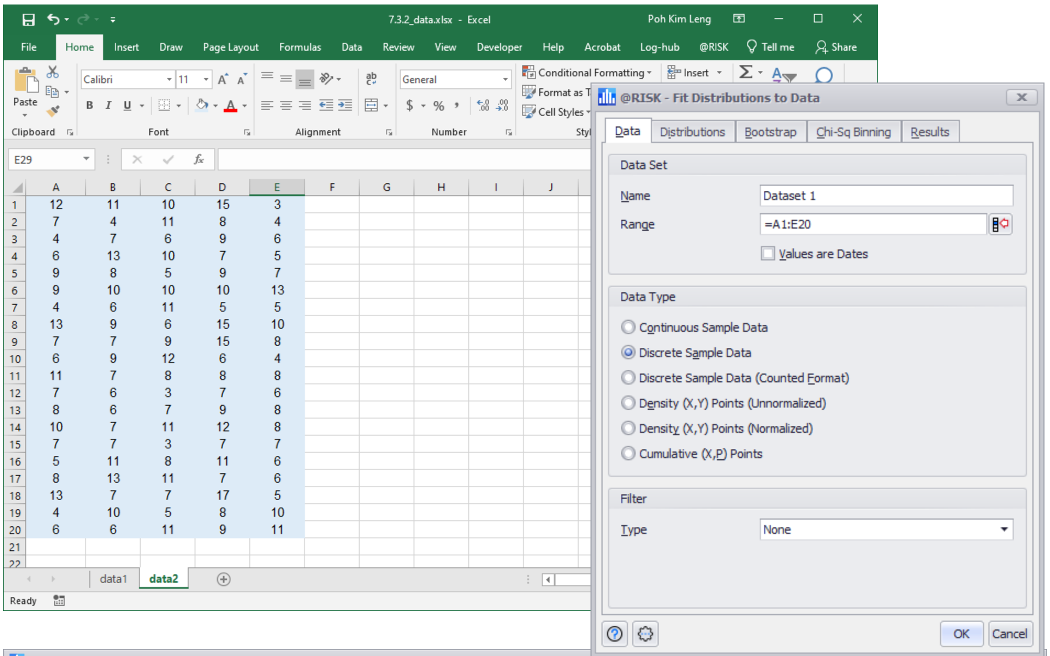
- Beta (2.9695, 2.8411, 48.601, 144.637)
- Triangular (52.734, 97.820, 142.247)
- Normal (97.771, 18.467)

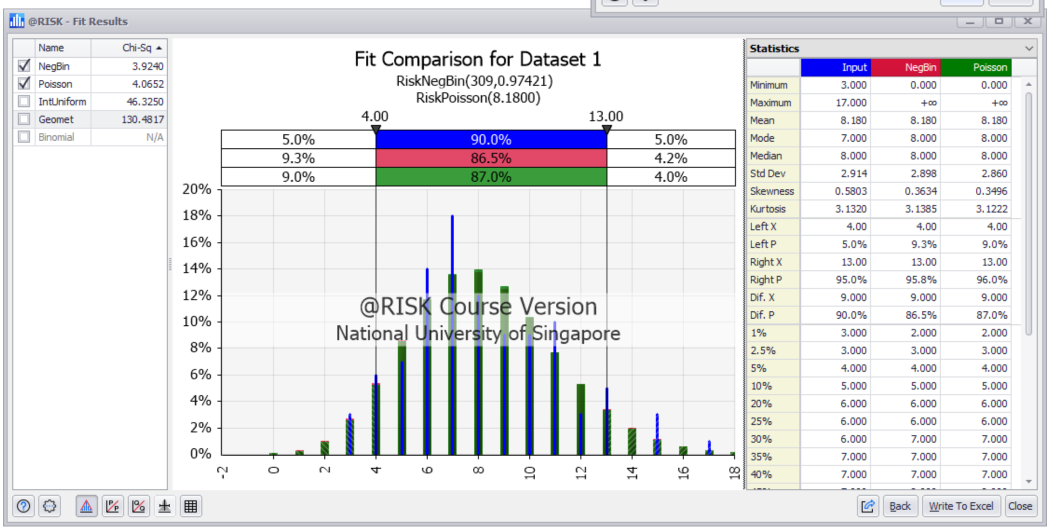
Example 2 (Discrete Distributions)

- Fit a distribution to the following 100 values observed from a discrete counting event.

12	11	10	15	3
7	4	11	8	4
4	7	6	9	6
6	13	10	7	5
9	8	5	9	7
9	10	10	10	13
4	6	11	5	5
13	9	6	15	10
7	7	9	15	8
6	9	12	6	4
11	7	8	8	8
7	6	3	7	6
8	6	7	9	8
10	7	11	12	8
7	7	3	7	7
5	11	8	11	6
8	13	11	7	6
13	7	7	17	5
4	10	5	8	10
6	6	11	9	11

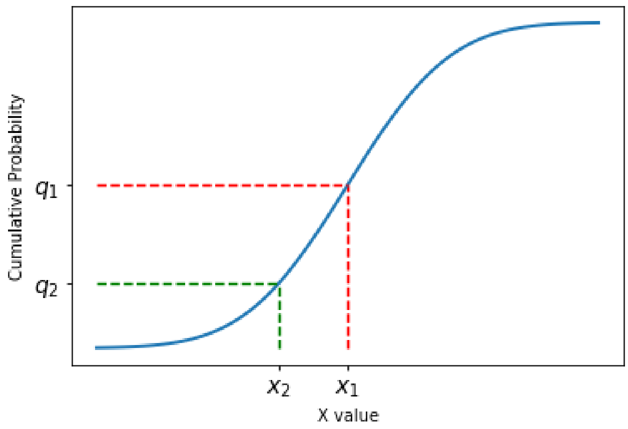
Using @Risk





7.3.3 Normal Distribution with Two Known Percentile Values

- Suppose an uncertain variable X follows the Normal distribution with two known percentile values as shown in the following CDF:



- That is, given $X \sim N(\mu, \sigma^2)$, and

$F(x_1; \mu, \sigma) = q_1$ (1)
$F(x_2; \mu, \sigma) = q_2$ (2)

where $F(x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$ and $erf(x) = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2}dt$

- How do we find the two parameters μ and σ?

Case 1: When the mean of the distribution is known.

- Consider the easy case when x_1 is the mean, i.e., $\mu = x_1$ and $q_1 = 0.5$
- We only need to find σ by solving equation (2):

$$F(x_2) = \frac{1}{2}\left[1 + erf\left(\frac{x_2-\mu}{\sigma\sqrt{2}}\right)\right] = q_2$$

- Therefore $\sigma = \dfrac{x_2 - \mu}{z_2}$

where $z_2 = \sqrt{2}erf^{-1}(2q_2 - 1)$ is the inverse CDF value of the standard normal distribution $N(0,1)$ at percentile q_2.

- The value of z_2 can be computed using:
  - Python: `z2 = scipy.stats.norm.ppf(q2)`
  - Excel: `=NORM.S.INV(q2)` or `=NORM.INV(q2, 0, 1 )`
  - R: `z2 <- qnorm(q2, mean=0, sd=1)`

Example

- It is observed that the mean of a variable X is 200, and 20% of the values of X is less than or equal to 100. If X follows the normal distribution, what is its standard deviation?

  We have $x_1 = 200,\ q_1 = 0.5$
  $x_2 = 100,\ q_2 = 0.2$

  $z_2 = F^{-1}(0.2) = -0.84162$

  $\sigma = (100 - 200)/\ (-0.84162) = 118.818$

- Hence $X \sim N(200, 118.818^2)$

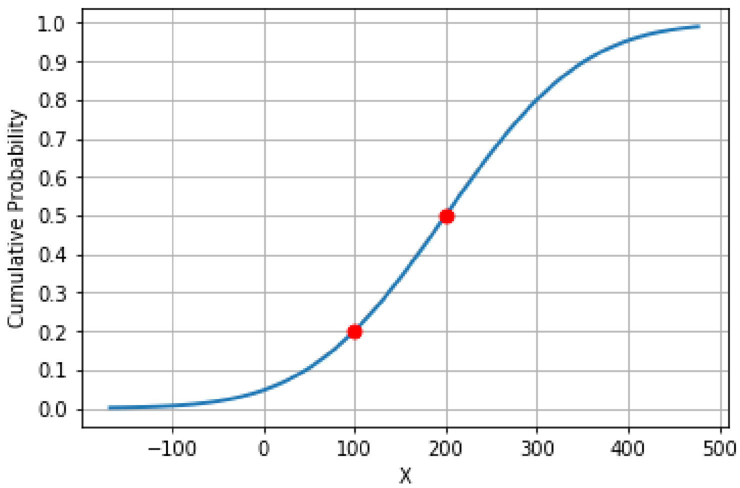
Python:

In [1]:
```
""" 7.3.3 Fit Normal Distribution with known mean and 1 percentile value """
from scipy.stats import norm
```

In [2]:
```
x1, q1 = 200, 0.5
x2, q2 = 100, 0.2
mu = x1
z2 = norm.ppf(q2)
sigma = (x2 - mu)/z2
print(f"mu = {mu}, sigma = {sigma}")
```

```
mu = 200, sigma = 118.81829498938903
```

In [3]:
```
import matplotlib.pyplot as plt
import numpy as np
xmin = norm.ppf(0.001, loc=mu, scale=sigma)
xmax = norm.ppf(0.990, loc=mu, scale=sigma)
x = np.linspace(xmin, xmax, 100)
y = norm.cdf(x, mu, sigma)
fig, ax = plt.subplots()
ax.plot(x, y, lw=2)
ax.plot([x1, x2],[q1,q2],'ro', ms=7)
ax.set_yticks(np.linspace(0, 1, 11))
ax.set_ylabel("Cumulative Probability")
ax.set_xlabel("X")
ax.grid()
plt.show()
```



Case 2 (When the mean is unknown)

- Consider the case when neither x_1 nor x_2 is the mean. In this case, we need to solve two equations with two unknowns μ and σ:

$$\frac{1}{2}\left[1+erf\left(\frac{x_1-\mu}{\sigma\sqrt{2}}\right)\right]=q_1 \quad \text{and} \quad \frac{1}{2}\left[1+erf\left(\frac{x_2-\mu}{\sigma\sqrt{2}}\right)\right]=q_2$$

Example:

- An uncertain variable follows the normal distribution. It is observed that 10% of its values are greater than or equal to 352.27 and 20% of its values are less than or equal to 100. What is the mean and standard deviation of X?

- Given $x_1 = 352.27$ $q_1 = 1 - 0.1 = 0.9$
  $x_2 = 100$ $q_2 = 0.2$

- Solving the 2 equations, we obtain mean =200.00, standard deviation =118.82.

Python:

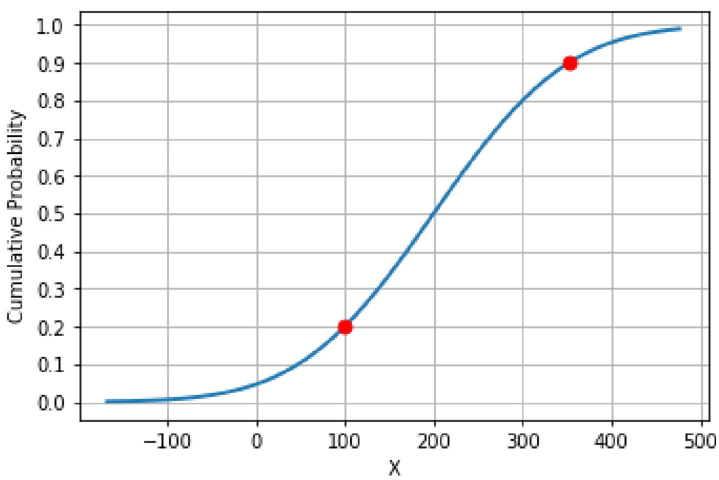
In [1]:
```
""" Fit Normal Distribution with 2 known percentile values """
from scipy.stats import norm
from scipy.optimize import root
```

In [2]:
```
# Equations to solve
eq = lambda param, x, q: (norm.cdf(x[0], *param) - q[0],
                          norm.cdf(x[1], *param) - q[1] )
```

In [3]:
```
x1, q1 = 352.27, 0.9
x2, q2 = 100, 0.2
guess = ((x1+x2)/2, abs(x1-x2)/2)
sol = root(eq, guess, args=((x1, x2),(q1, q2)))
print(sol.message)
mu, sigma = sol.x
print(f"mu = {mu}, sigma = {sigma}")
```

```
The solution converged.
mu = 199.99929759918126, sigma = 118.81746040871225
```

In [4]:
```
# Use the same code from previous example 1 to plot:
```

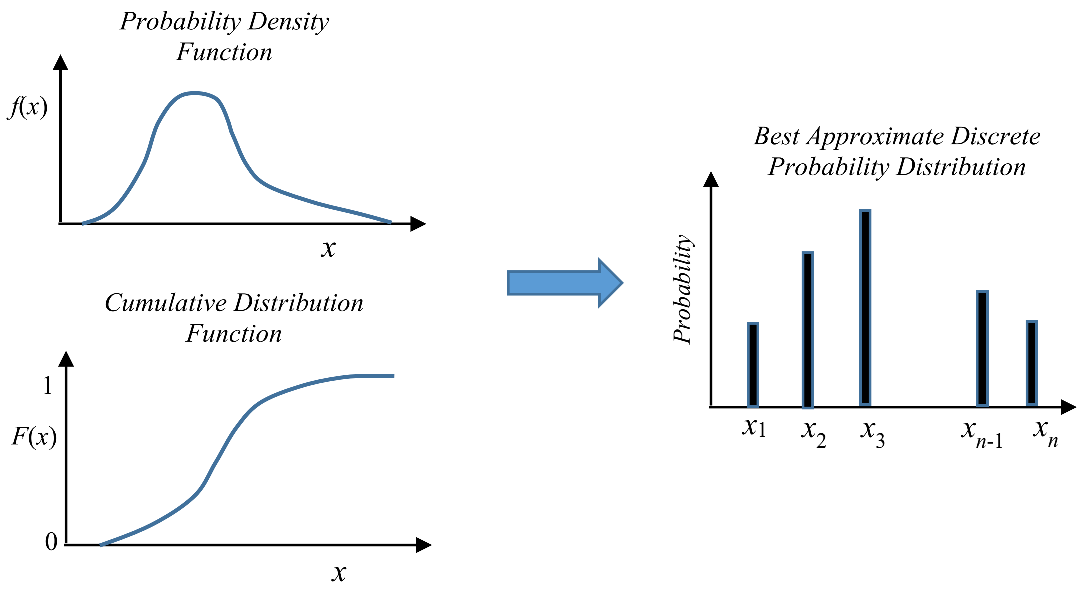


7.4 Discrete Approximation of Continuous Probability Distribution

7.4.1 Introduction

- Suppose we have a continuous uncertain variable with known PDF and we wish to represent it in a decision tree or influence diagram with a discrete chance node.

- We need to find the *best approximate discrete probability distribution* with *n* discrete outcomes to represent it in the decision tree or influence diagram.



- There are $2n$ unknowns to be determined: $p_1, x_1, p_2, x_2, \ldots, p_n, x_n$.

7.4.2 Desired Properties of the Discrete Approximation

- Given a continuous variable X with PDF $f(x)$, we seek to approximate it by a probability mass function (PMF) $p_1, x_1;\ p_2, x_2;\ \ldots, p_n, x_n$.

- An obvious requirement for the probabilities is $\sum_{i=1}^{n} p_i = 1$. But what else should also be satisfied?

- Another desired property is that the Expected Value on *any function* $g(x)$ should be preserved across the discretization. That is

$$\int_{-\infty}^{+\infty} g(x) f(x) dx = \sum_{i=1}^{n} p_i g(x_i) \qquad \text{for all functions } g(x)$$

- If $g(x)$ is the utility function, the expected utility will be preserved and there will be no error in optimal decision policy due to the discretization.

- Assuming that $g(x)$ can be approximated by the polynomial $g(x) = a_o + a_1 x + a_2 x^2 + \cdots$, we therefore, require that

$$\int_{-\infty}^{+\infty} (a_o + a_1 x + a_2 x^2 + \cdots) f(x) dx = \sum_{i=1}^{n} p_i (a_o + a_i x + a_2 x_i^{\,2} + \cdots).$$

- This may be rewritten in terms of the moments of the continuous and the discrete distributions:

$$a_o + a_1 E[x] + a_2 E[x^2] + \cdots = a_o + a_1 \sum_{i=1}^{n} p_i x_i + a_2 \sum_{i=1}^{n} p_i x_i^2 + \cdots$$

- Equating term by term:

$$E[x^k] = \sum_{i=1}^{n} p_i x_i^k \qquad \text{for } k = 0, 1, 2, \ldots$$

- Hence all moments must be preserved between the continuous and the discrete distributions.

k=0	$\Rightarrow$	$\sum_{i=1}^{n} p_i = 1$	discrete distribution must be valid.
k=1	$\Rightarrow$	$E[x] = \sum_{i=1}^{n} p_i x_i$	mean must be preserved.
k=2	$\Rightarrow$	$E[x^2] = \sum_{i=1}^{n} p_i x_i^{\,2}$	variance must be preserved (if the mean is already preserved)
k=3	$\Rightarrow$	$E[x^3] = \sum_{i=1}^{n} p_i x_i^{\,3}$	skewness must be preserved. (if the mean is already preserved)

7.4.3 Limit on the number of moments that can be preserved

- In principle, we need to achieve

$$E[x^k] = \int x^k f(x) dx = \sum_{i=1}^{n} p_i x_i^k \qquad \text{for all } k = 0, 1, 2, \ldots.$$

- For any desired number of discrete branches n, we need to determine the $2n$ values
  $p_1, x_1;\ \ p_2, x_2;\ \ p_3, x_3;\ \ldots,\ \ p_n, x_n$

- These $2n$ values can be determined by preserving first $(2n - 1)$ moments by solving the following $2n$ simultaneous equations:

$$\sum_{i=1}^{n} p_i = 1 \qquad (1)$$

$$\int x f(x) dx = \sum_{i=1}^{n} p_i x_i \qquad (2)$$

$$\int x^2 f(x) dx = \sum_{i=1}^{n} p_i x_i^{\,2} \qquad (3)$$

$$\ldots$$

$$\int x^{2n-1} f(x) dx = \sum_{i=1}^{n} p_i x_i^{\,2n-1} \qquad (2n)$$

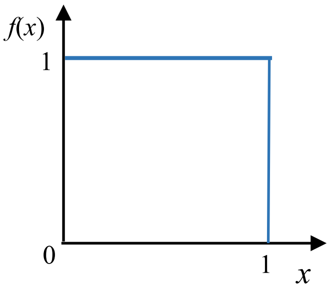
- Hence in practice, only the first $(2n\text{-}1)$ moments can be preserved.

7.4.4 Discrete Approximation for the Uniform Distributions

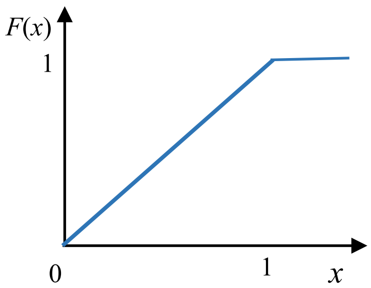
Standard Uniform Distribution $U(0,1)$

- The PDF and CDF are shown below:

PDF

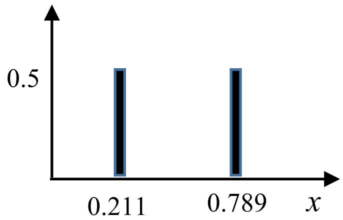


CDF



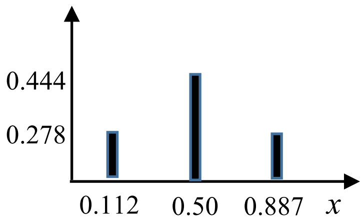
- If we wish to approximate it with a 2-branch probability distribution, we can only preserve the first 3 moments by solving the following 4 equations:

$$
\begin{aligned}
p_1 + p_2 &= 1 \\
x_1 p_1 + x_2 p_2 &= \int_0^1 x\,dx = \tfrac{1}{2} \\
x_1^2 p_1 + x_2^2 p_2 &= \int_0^1 x^2\,dx = \tfrac{1}{3} \\
x_1^3 p_1 + x_2^3 p_2 &= \int_0^1 x^3\,dx = \tfrac{1}{4}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
x_1 &= 0.211325, & p_1 &= 0.5 \\
x_2 &= 0.788675, & p_2 &= 0.5
\end{aligned}
$$



- If we wish to approximate it with a 3-branch probability distribution, we can only preserve the first 5 moments by solving the following 6 equations:

$$
\begin{aligned}
p_1 + p_2 + p_3 &= 1 \\
x_1 p_1 + x_2 p_2 + x_3 p_3 &= \int_0^1 x\,dx = \tfrac{1}{2} \\
x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 &= \int_0^1 x^2\,dx = \tfrac{1}{3} \\
x_1^3 p_1 + x_2^3 p_2 + x_3^3 p_3 &= \int_0^1 x^3\,dx = \tfrac{1}{4} \\
x_1^4 p_1 + x_2^4 p_2 + x_3^4 p_3 &= \int_0^1 x^4\,dx = \tfrac{1}{5} \\
x_1^5 p_1 + x_2^5 p_2 + x_3^5 p_3 &= \int_0^1 x^5\,dx = \tfrac{1}{6}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
x_1 &= 0.112702, & p_1 &= 0.277778 \\
x_2 &= 0.500000, & p_2 &= 0.444444 \\
x_3 &= 0.887298, & p_3 &= 0.277778
\end{aligned}
$$



- If we wish to approximate it with a 4-branch probability distribution, we can only preserve the first 7 moments by solving the following 8 equations:

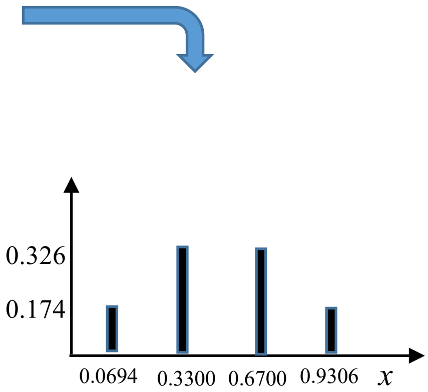
$$p_1 + p_2 + p_3 + p_4 = 1$$

$$x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 = \int_0^1 x\,dx = \tfrac{1}{2}$$

$$\vdots$$

$$x_1^6 p_1 + x_2^6 p_2 + x_3^6 p_3 + x_4^6 p_4 = \int_0^1 x^6\,dx = \tfrac{1}{7}$$

$$x_1^7 p_1 + x_2^7 p_2 + x_3^7 p_3 + x_4^7 p_4 = \int_0^1 x^7\,dx = \tfrac{1}{8}$$



<u>General Uniform Distribution *U*(*a*, *b*)</u>

- The n-branch discrete approximation for $U(a, b)$ is $(x_1, x_2, \ldots, x_n; p_1, p_2, \ldots p_n)$

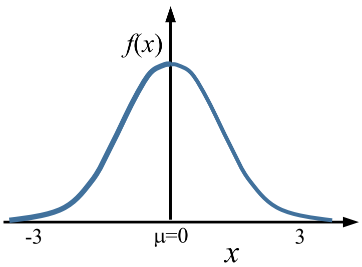
  where $x_i = a + (b - a)\, z_i$ for $i = 1$ to n

  and $(z_1, z_2, \ldots, z_n; p_1, p_2, \ldots p_n)$ is the n-branch discrete approximation for $U(0,1)$

7.4.5 Discrete Approximation for Normal Distributions

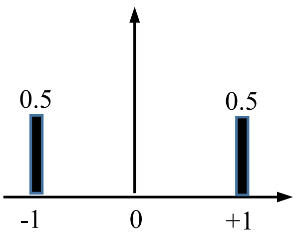
Standard Normal Distribution *N*(0, 1)

- The PDF is:



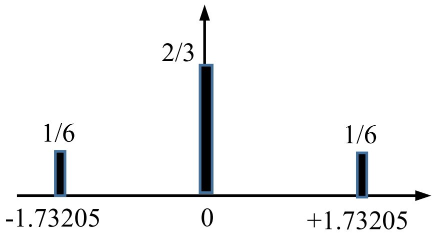
- 2-branch discrete approximation preserving the first 3 moments:

$x_1 = -1, \quad p_1 = 0.5$
$x_2 = +1, \quad p_2 = 0.5$



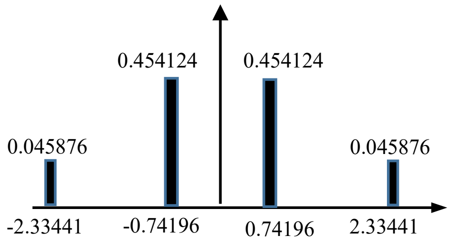
- 3-branch discrete approximation preserving the first 5 moments:

$x_1 = -1.73205, \quad p_1 = 1/6$
$x_2 = 0, \quad p_2 = 2/3$
$x_3 = +1.73205, \quad p_3 = 1/6$



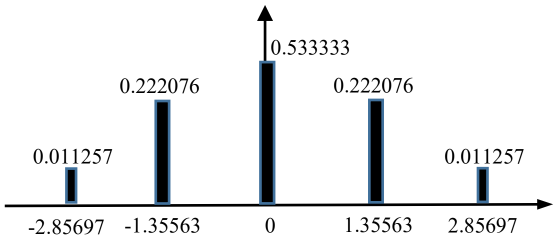
- 4-branch discrete approximation preserving the first 7 moments:

$x_1 = -2.33441, \quad p_1 = 0.045876$
$x_2 = -0.741964, \quad p_2 = 0.454124$
$x_3 = +0.741964, \quad p_3 = 0.454124$
$x_4 = +2.33441, \quad p_4 = 0.045876$



- 5-branch discrete approximation preserving the first 9 moments:

$x_1 = -2.85697, \quad p_1 = 0.011257$
$x_2 = -1.35563, \quad p_2 = 0.222076$
$x_3 = 0, \quad p_3 = 0.533333$
$x_4 = 1.35563, \quad p_4 = 0.222076$
$x_5 = 2.85697, \quad p_5 = 0.011257$



General Normal Distribution $N(\mu, \sigma^2)$

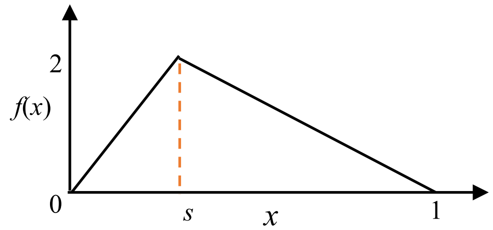
- The n-branch discrete approximation for $N(\mu, \sigma^2)$ is $(x_1, x_2, \ldots, x_n; p_1, p_2, \ldots p_n)$

  where $\quad x_i = \mu + \sigma z_i \quad$ for $i = 1$ to n

  and $\quad (z_1, z_2, \ldots, z_n; p_1, p_2, \ldots p_n)$ is the n-branch discrete approximation for $N(0,1)$

7.4.6 3-Branch Discretization of the Asymmetric Triangular Distribution

- The asymmetric distribution Triangular(0, 1, s) bounded between 0 and 1 has the PDF below:



- The 3-branch discretization for Triangular (0, 1, s) for some values of s are given below:

s	***x* value**	**Probability**
0.1	0.12793794 0.43298739 0.79603109	0.38778583 0.46905473 0.14315944
0.2	0.15441205 0.44738893 0.80108143	0.34663662 0.50021458 0.15314880
0.25	0.16315709 0.45364411 0.80321892	0.31965103 0.52050614 0.15984284
0.3	0.16952605 0.46060361 0.80546846	0.29192121 0.54076271 0.16731608
1/3	0.17277872 0.46591734 0.80708707	0.27426476 0.55302152 0.17271373
0.4	0.17779237 0.47831381 0.81059898	0.24308553 0.57211756 0.18479691
0.5	0.18377223 0.5 0.81622777	0.20833333 0.58333333 0.20833333

s	**Value**	**Probability**
0.6	0.189401 0.5216862 0.8222076	0.18479691 0.57211756 0.24308553
2/3	0.1929129 0.5340827 0.8272213	0.17271373 0.55302152 0.27426476
0.7	0.1945315 0.5393964 0.830474	0.16731608 0.54076271 0.29192121
0.75	0.1967811 0.5463559 0.8368429	0.15984284 0.52050614 0.31965103
0.8	0.1989186 0.5526111 0.845588	0.1531488 0.50021458 0.34663662
0.9	0.2039689 0.5670126 0.8720621	0.14315944 0.46905473 0.38778583

General Triangular (a, b, m) $\quad a \le m \le b$

- The n-branch discrete approximation for *Triangular*(a,b,m) is $(x_1, x_2, \ldots, x_n; p_1, p_2, \ldots p_n)$

  where $x_i = a + (b - a) z_i$ for $i = 1$ to n

  and $(z_1, z_2, \ldots, z_n; p_1, p_2, \ldots p_n)$ is the n-branch discrete approximation for *Triangular* (0, 1, s)
  $s = (m - a)/(b - a)$.

- s, a and $(b - a)$ are also called the shape, location, and scale parameters of the distribution, respectively.

7.4.7 Discrete Approximation for Continuous Distributions in DPL

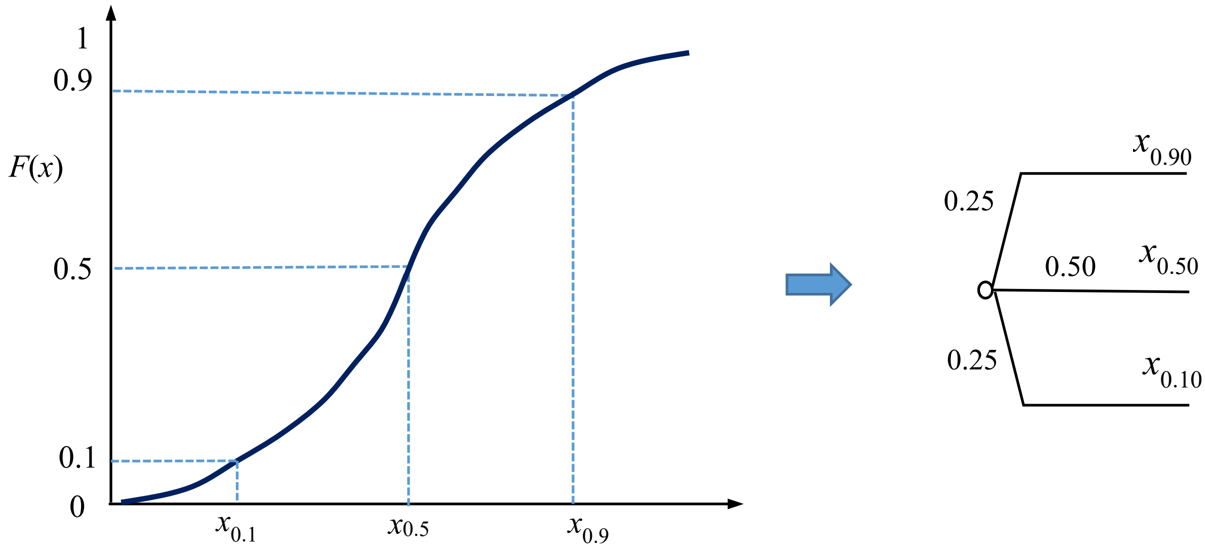
- In DPL, if a *discrete chance node* with *n* branches is created and a known standard continuous distribution is specified, DPL will automatically perform the $(2n - 1)$ moments matching and find the approximate branch values and probabilities for the decision tree. These values will be automatically used in creating the decision tree and generating the optimal policy tree.

- In DPL, if a *continuous chance node* is created and a known standard continuous distribution is specified, the resulting decision model can only be solved using Monte Carlo Simulation. There is no discrete optimal decision policy tree.

7.4.8 Three-Point Quick Approximation Methods

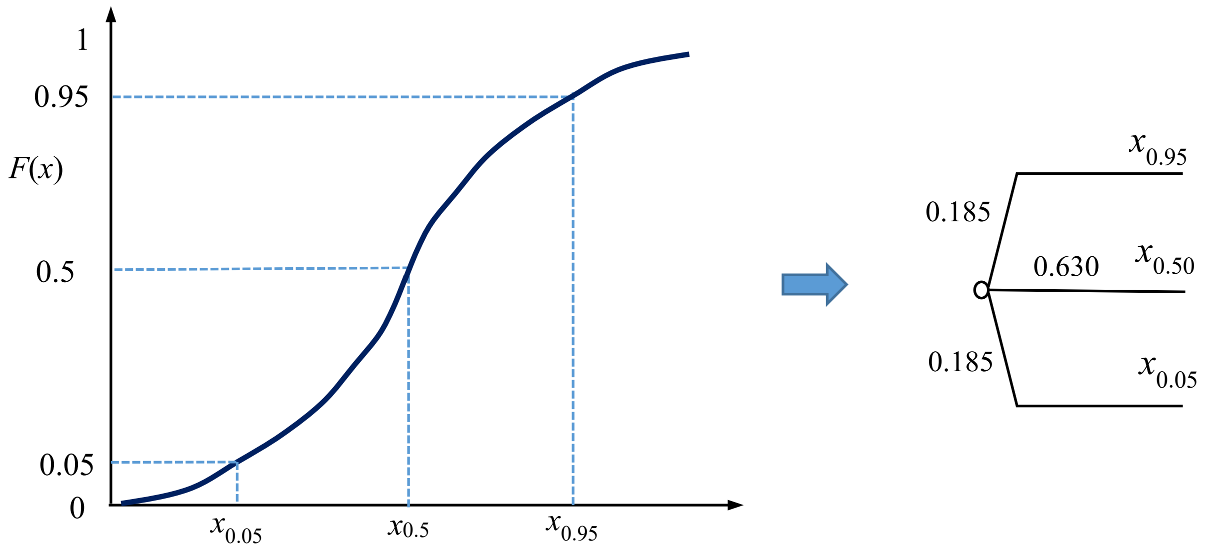
Stanford / SDG Approximation

- When the probability distribution is symmetrical, the 10, 50, and 90 percentiles provide a good approximation for a PMF with probabilities [0.25, 0.50, 0.25].



Pearson-Tukey Approximation

- When the probability distribution is symmetrical, the 5, 50 and 95 percentiles provide a good approximation for a PMF with probabilities [0.185, 0.630, 0.185].



References

1. A. Tversky and D. Kahneman (1974). Judgment under uncertainty: Heuristic and Biases. *Science* **185**:1124-1131. (Article 47 of Howard & Matheson, 1983).

2. A.C. Miller III and T.R. Rice (1983). Discrete Approximations of Probability Distributions, *Management Science* **29**(3):352-362.

3. D.L. Keefer and S.E. Bodily (1983). Three-Point Approximation for Continuous Random Variables, *Management Science* **29**(5):595-609.

Exercises

P7.1 Given the following data:

5.3299	4.2537	3.1502	3.7032	1.6070	6.3923
3.1181	6.5941	3.5281	4.7433	0.1077	1.5977
5.4920	1.7220	4.1547	2.2799		

(*a*) Plot a histogram representing the data.
(*b*) Fit a normal distribution to the data using any suitable method.
(*c*) Plot the PDF of the fitted normal distribution and compare it with the plot in (*a*).

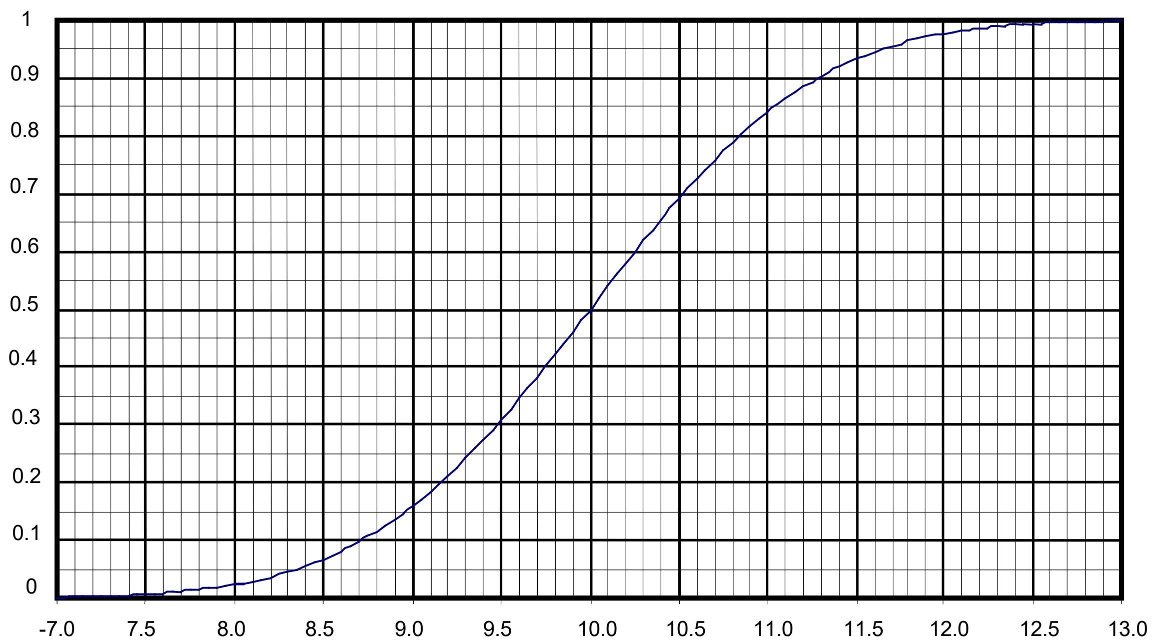
P7.2 Use DPL to find the discrete approximation of the following distributions with 3 branches:

(*a*) A triangular distribution with *min*=0*, max* =10*, mode*=5.
(*b*) An exponential distribution with *mean* =10.
(*c*) A uniform distribution with *min* =10 and *max* = 20.

P7.3 William faces an investment alternative whose net present value has a normal distribution with a mean of \$10 million and a standard deviation of \$1 million. William follows the Delta property and has a risk tolerance of \$5 million. Use DPL to determine and compare the certainty equivalent of this investment proposal using:

(*a*) Discrete 3-branch approximation (moments matching).
(*b*) Standard/SDG 3-branch quick approximation.
(*c*) Pearson-Tukey approximation.

The CDF for the distribution Normal (10, 1) is given below:



P7.4 A scientist collected the following weights (in grams) of laboratory animals:

9.79	9.23	9.11	9.62
8.73	11.93	10.39	8.68
9.76	9.59	11.49	9.86
11.41	9.60	7.24	

(***a***) Fit a probability distribution to the data using any suitable method.

(***b***) Plot the PDF of the fitted distribution and compare it with the observed data.

(c) Use the fitted distribution to estimate the probability that an animal's weight will be less than 9.5 grams.

P7.5 (Clement and Reilly 2001, Problem 10.11, modified)
A retail manager in a discount store wants to establish a policy regarding the number of cashiers to have on hand and also when to open a new cash register. The first step in this process is to determine the rate at which customers arrive at the cash register. One day, the manager observes the following times (in minutes) between arrivals of customers at the cash registers:

0.1	2.6	2.9	0.5
1.2	1.8	4.8	3.3
1.7	0.2	1.5	2.0
4.2	0.6	1.0	2.6
0.9	3.4	1.7	0.4

Assuming that customers' arrival is a poison process, fit an exponential distribution with one parameter λ to the data.

This page is blank.