Using a Computer Algebra System in Data Analysis

Larry Leemis • The College of William & Mary

February 12, 2014

Civilization advances by extending the number of important operations which we can perform without thinking about them.
—Alfred North Whitehead (1861–1947)

(joint work with John Drew, Matt Duggan, Diane Evans, Andy Glen, Billy Kaczynski, Daniel Luckett, Jeff Mallozzi, Raghu Pasupathy, Bruce Schmeiser, Jackie Taber, Erik Vargo, Keith Webb, Jeff Yang)

Outline

Example 2: Product of two independent random variables

Let X and Y be independent random variables:

$$X \sim \mathit{Triangular}(1, 2, 4)$$

$$Y \sim \mathit{Triangular}(1, 2, 3)$$

Find the distribution of $V = XY$.

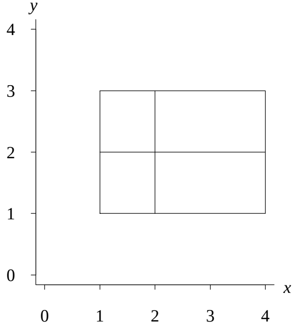
Example 2: Product of two independent random variables

Let X and Y be independent random variables:

$$X \sim Triangular(1, 2, 4)$$

$$Y \sim Triangular(1, 2, 3)$$

Find the distribution of $V = XY$.



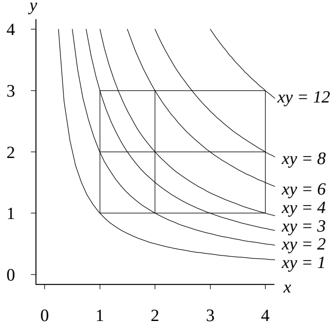
Example 2: Product of two independent random variables

Let X and Y be independent random variables:

$$X \sim \mathit{Triangular}(1, 2, 4)$$

$$Y \sim \mathit{Triangular}(1, 2, 3)$$

Find the distribution of $V = XY$.



Example 2: Product of two independent random variables (continued)

The APPL code is

```
> X := TriangularRV(1, 2, 4);
> Y := TriangularRV(1, 2, 3);
> V := Product(X, Y);
```

$$f_V(v) = \begin{cases} -\frac{4}{3}v + \frac{2}{3}\ln v + \frac{2v}{3}\ln v + \frac{4}{3} & 1 < v \le 2 \\ -8 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + \frac{10}{3}v - 4\ln v - \frac{5v}{3}\ln v & 2 < v \le 3 \\ -4 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + 2v - 2\ln v - \\ \quad v\ln v - 2\ln 3 - \frac{2v}{3}\ln 3 & 3 < v \le 4 \\ \frac{44}{3} - 14\ln 2 - \frac{7v}{3}\ln 2 - \frac{8}{3}v - 2\ln 3 + \\ \quad \frac{22}{3}\ln v - \frac{2v}{3}\ln 3 + \frac{4v}{3}\ln v & 4 < v \le 6 \\ \frac{8}{3} - 8\ln 2 - \frac{4v}{3}\ln 2 - \frac{2}{3}v + \frac{4}{3}\ln v + \\ \quad \frac{v}{3}\ln v + 4\ln 3 + \frac{v}{3}\ln 3 & 6 < v \le 8 \\ -8 + 8\ln 2 + \frac{2v}{3}\ln 2 + \frac{2}{3}v + 4\ln 3 - \\ \quad 4\ln v + \frac{v}{3}\ln 3 - \frac{v}{3}\ln v & 8 < v < 12 \end{cases}$$

Example 2: Product of two independent random variables (continued)

The APPL code is

```
> X := TriangularRV(1, 2, 4);
> Y := TriangularRV(1, 2, 3);
> V := Product(X, Y);
```

$$f_V(v) = \begin{cases} -\frac{4}{3}v + \frac{2}{3}\ln v + \frac{2v}{3}\ln v + \frac{4}{3} & 1 < v \le 2 \\ -8 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + \frac{10}{3}v - 4\ln v - \frac{5v}{3}\ln v & 2 < v \le 3 \\ -4 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + 2v - 2\ln v - \\ \quad v\ln v - 2\ln 3 - \frac{2v}{3}\ln 3 & 3 < v \le 4 \\ \frac{44}{3} - 14\ln 2 - \frac{7v}{3}\ln 2 - \frac{8}{3}v - 2\ln 3 + \\ \quad \frac{22}{3}\ln v - \frac{2v}{3}\ln 3 + \frac{4v}{3}\ln v & 4 < v \le 6 \\ \frac{8}{3} - 8\ln 2 - \frac{4v}{3}\ln 2 - \frac{2}{3}v + \frac{4}{3}\ln v + \\ \quad \frac{v}{3}\ln v + 4\ln 3 + \frac{v}{3}\ln 3 & 6 < v \le 8 \\ -8 + 8\ln 2 + \frac{2v}{3}\ln 2 + \frac{2}{3}v + 4\ln 3 - \\ \quad 4\ln v + \frac{v}{3}\ln 3 - \frac{v}{3}\ln v & 8 < v < 12 \end{cases}$$

Example 2: Product of two independent random variables (continued)

The APPL code is

```
> X := TriangularRV(1, 2, 4);
> Y := TriangularRV(1, 2, 3);
> V := Product(X, Y);
```

$$f_V(v) = \begin{cases} -\frac{4}{3}v + \frac{2}{3}\ln v + \frac{2v}{3}\ln v + \frac{4}{3} & 1 < v \le 2 \\ -8 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + \frac{10}{3}v - 4\ln v - \frac{5v}{3}\ln v & 2 < v \le 3 \\ -4 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + 2v - 2\ln v - v\ln v - 2\ln 3 - \frac{2v}{3}\ln 3 & 3 < v \le 4 \\ \frac{44}{3} - 14\ln 2 - \frac{7v}{3}\ln 2 - \frac{8}{3}v - 2\ln 3 + \frac{22}{3}\ln v - \frac{2v}{3}\ln 3 + \frac{4v}{3}\ln v & 4 < v \le 6 \\ \frac{8}{3} - 8\ln 2 - \frac{4v}{3}\ln 2 - \frac{2}{3}v + \frac{4}{3}\ln v + \frac{v}{3}\ln v + 4\ln 3 + \frac{v}{3}\ln 3 & 6 < v \le 8 \\ -8 + 8\ln 2 + \frac{2v}{3}\ln 2 + \frac{2}{3}v + 4\ln 3 - 4\ln v + \frac{v}{3}\ln 3 - \frac{v}{3}\ln v & 8 < v < 12 \end{cases}$$

Example 2: Product of two independent random variables (continued)

The APPL code is

```
> X := TriangularRV(1, 2, 4);
> Y := TriangularRV(1, 2, 3);
> V := Product(X, Y);
```

$$f_V(v) = \begin{cases} -\frac{4}{3}v + \frac{2}{3}\ln v + \frac{2v}{3}\ln v + \frac{4}{3} & 1 < v \le 2 \\ -8 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + \frac{10}{3}v - 4\ln v - \frac{5v}{3}\ln v & 2 < v \le 3 \\ -4 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + 2v - 2\ln v - v\ln v - 2\ln 3 - \frac{2v}{3}\ln 3 & 3 < v \le 4 \\ \frac{44}{3} - 14\ln 2 - \frac{7v}{3}\ln 2 - \frac{8}{3}v - 2\ln 3 + \frac{22}{3}\ln v - \frac{2v}{3}\ln 3 + \frac{4v}{3}\ln v & 4 < v \le 6 \\ \frac{8}{3} - 8\ln 2 - \frac{4v}{3}\ln 2 - \frac{2}{3}v + \frac{4}{3}\ln v + \frac{v}{3}\ln v + 4\ln 3 + \frac{v}{3}\ln 3 & 6 < v \le 8 \\ -8 + 8\ln 2 + \frac{2v}{3}\ln 2 + \frac{2}{3}v + 4\ln 3 - 4\ln v + \frac{v}{3}\ln 3 - \frac{v}{3}\ln v & 8 < v < 12 \end{cases}$$

Example 2: Product of two independent random variables (continued)

The APPL code is

```
> X := TriangularRV(1, 2, 4);
> Y := TriangularRV(1, 2, 3);
> V := Product(X, Y);
```

$$
f_V(v) = \begin{cases}
-\frac{4}{3}v + \frac{2}{3}\ln v + \frac{2v}{3}\ln v + \frac{4}{3} & 1 < v \le 2 \\
-8 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + \frac{10}{3}v - 4\ln v - \frac{5v}{3}\ln v & 2 < v \le 3 \\
-4 + \frac{14}{3}\ln 2 + \frac{7v}{3}\ln 2 + 2v - 2\ln v - \\
\quad v\ln v - 2\ln 3 - \frac{2v}{3}\ln 3 & 3 < v \le 4 \\
\frac{44}{3} - 14\ln 2 - \frac{7v}{3}\ln 2 - \frac{8}{3}v - 2\ln 3 + \\
\quad \frac{22}{3}\ln v - \frac{2v}{3}\ln 3 + \frac{4v}{3}\ln v & 4 < v \le 6 \\
\frac{8}{3} - 8\ln 2 - \frac{4v}{3}\ln 2 - \frac{2}{3}v + \frac{4}{3}\ln v + \\
\quad \frac{v}{3}\ln v + 4\ln 3 + \frac{v}{3}\ln 3 & 6 < v \le 8 \\
-8 + 8\ln 2 + \frac{2v}{3}\ln 2 + \frac{2}{3}v + 4\ln 3 - \\
\quad 4\ln v + \frac{v}{3}\ln 3 - \frac{v}{3}\ln v & 8 < v < 12
\end{cases}
$$

Example 3: Order statistics

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*a*) Sampling without replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4, "wo");
> PDF(Y, 5);
```

$$\frac{32}{429}$$

Example 3: Order statistics

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*a*) Sampling without replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4, "wo");
> PDF(Y, 5);
```

$$\frac{32}{429}$$

Example 3: Order statistics

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*a*) Sampling without replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4, "wo");
> PDF(Y, 5);
```

$$\frac{32}{429}$$

Example 3: Order statistics

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*a*) Sampling without replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4, "wo");
> PDF(Y, 5);
```

$$\frac{32}{429}$$

Example 3: Order statistics

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*a*) Sampling without replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4, "wo");
> PDF(Y, 5);
```

$$\frac{32}{429}$$

Example 3: Order statistics (continued)

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*b*) Sampling with replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4);
> PDF(Y, 5);
```

$$\frac{2949971}{34171875}$$

Example 3: Order statistics (continued)

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*b*) Sampling with replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4);
> PDF(Y, 5);
```

$$\frac{2949971}{34171875}$$

Example 3: Order statistics (continued)

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*b*) Sampling with replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4);
> PDF(Y, 5);
```

$$\frac{2949971}{34171875}$$

Example 3: Order statistics (continued)

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

(*b*) Sampling with replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4);
> PDF(Y, 5);
```

$$\frac{2949971}{34171875}$$

Example 3: Order statistics (continued)

A bag contains 15 billiard balls, numbered 1 to 15. If 7 balls are drawn from the bag at random, find the probability that the median number drawn is 5 when (*a*) sampling is performed without replacement; (*b*) sampling is performed with replacement.

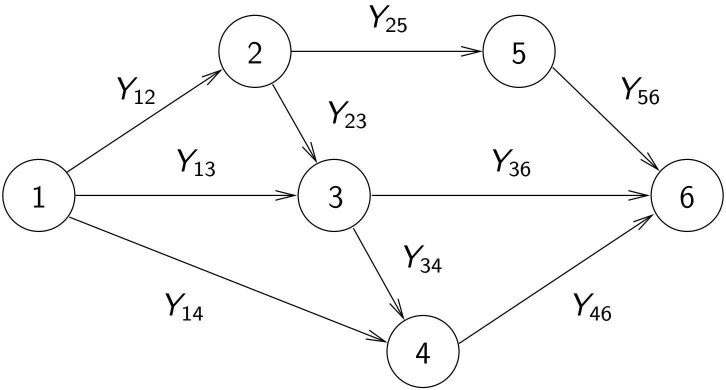
(*b*) Sampling with replacement

```
> X := UniformDiscreteRV(1, 15);
> Y := OrderStat(X, 7, 4);
> PDF(Y, 5);
```

$$\frac{2949971}{34171875}$$

Application 3: Stochastic activity networks

Stochastic activity networks arise in *project management*



Our goal: find the distribution of T_6, the time to complete the network

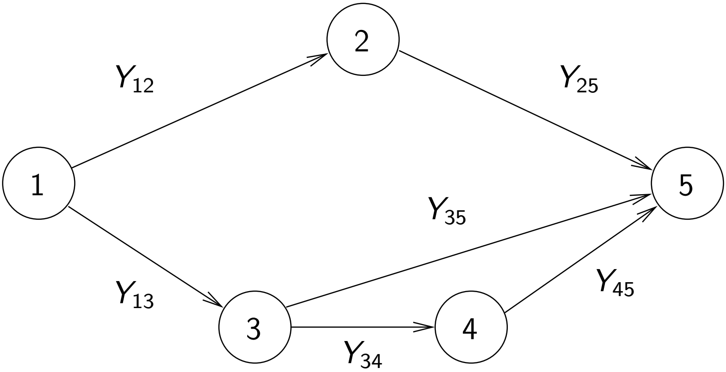
Application 3: Stochastic activity networks (continued)

Popular analysis techniques

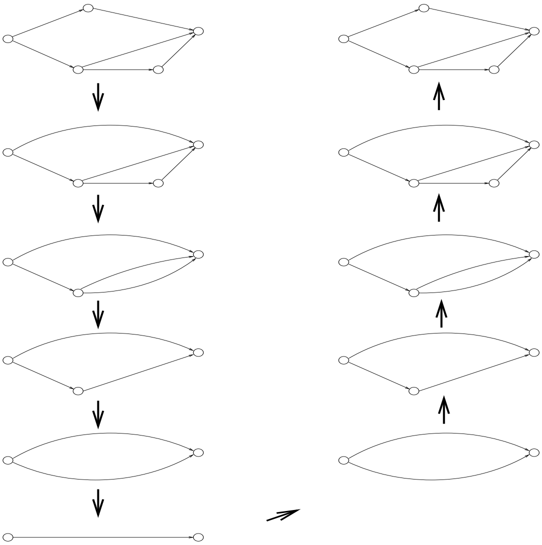
- CPM
- PERT
- Simulation

Application 3: Stochastic activity networks (continued)

Series–parallel networks constitute a class of stochastic activity networks that are easy to analyze. This sample series–parallel network is from Elmaghraby (1977, p. 261).



Application 3: Stochastic activity networks (continued)



Application 3: Stochastic activity networks (continued)

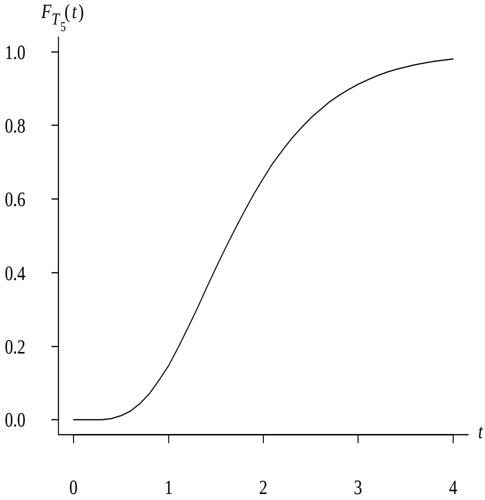
When all arc durations are independent exponential(b) random variables, where b is a rate, the time to complete the network T_5 has cdf

$$F_{T_5}(t) = 1 - 3bte^{-bt} - \frac{b^2t^2}{2}e^{-bt} - 3e^{-2bt} + \frac{5b^2t^2}{2}e^{-2bt} + \frac{b^3t^3}{2}e^{-2bt}$$

$$+2e^{-3bt} + 3bte^{-3bt} + b^2t^2e^{-3bt} \qquad t > 0$$

Application 3: Stochastic activity networks (continued)

The cdf is shown below for $b = 2$



Application 3: Stochastic activity networks (continued)

Bonus material: for exponential(2) arc durations

Paths π_k and critical path probabilities $p(\pi_k)$

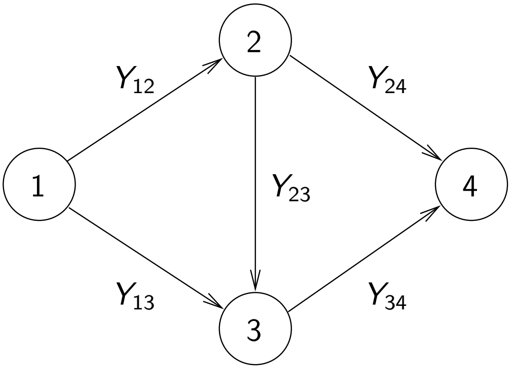
k	Node sequence	π_k	$p(\pi_k)$
1	$1 \rightarrow 2 \rightarrow 5$	$\{a_{12}, a_{25}\}$	$115/432 \cong 0.266$
2	$1 \rightarrow 3 \rightarrow 5$	$\{a_{13}, a_{35}\}$	$317/1728 \cong 0.183$
3	$1 \rightarrow 3 \rightarrow 4 \rightarrow 5$	$\{a_{13}, a_{34}, a_{45}\}$	$317/576 \cong 0.550$

Criticalities ρ_{ij}

Arc	Paths	ρ_{ij}
a_{12}	π_1	$115/432 \cong 0.266$
a_{13}	π_2, π_3	$317/432 \cong 0.734$
a_{25}	π_1	$115/432 \cong 0.266$
a_{35}	π_2	$317/1728 \cong 0.183$
a_{34}	π_3	$317/576 \cong 0.550$
a_{45}	π_3	$317/576 \cong 0.550$

Application 3: Stochastic activity networks (continued)

A non-series–parallel network is much more difficult to analyze. Here is the well-known *bridge network*.



Solution: conditional probability

Application 3: Stochastic activity networks (continued)

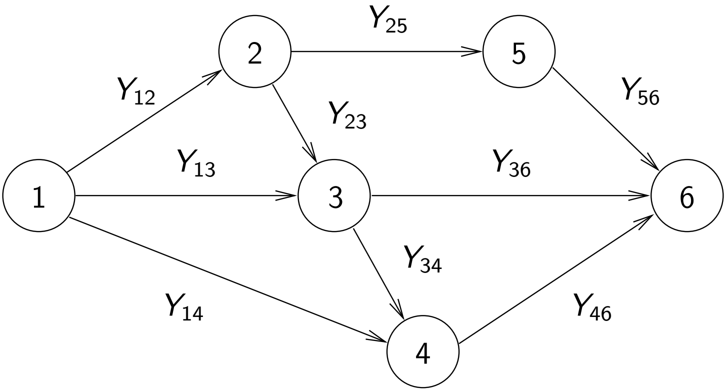
When the arc durations Y_{ij} are independent $U(0,1)$ random variables

$$F_{T_4}(t) = \begin{cases} 0 & t \le 0 \\ \frac{11}{120}\, t^5 & 0 < t \le 1 \\ -\frac{1}{120}\, t^5 - \frac{1}{6}\, t^4 + \frac{2}{3}\, t^3 - \frac{1}{3}\, t^2 - \frac{1}{6}\, t + \frac{1}{10} & 1 < t \le 2 \\ \frac{1}{6}\, t^3 - \frac{3}{2}\, t^2 + \frac{9}{2}\, t - \frac{23}{6} & 2 < t \le 3 \\ 1 & t > 3 \end{cases}$$

Next step: develop an algorithm to automate this process

Application 3: Stochastic activity networks (continued)

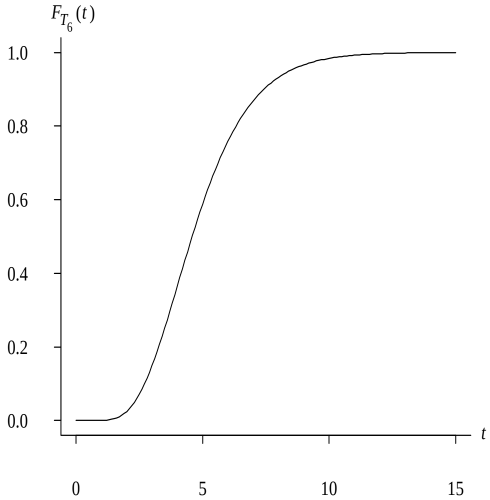
Consider the network



with independent exponential(1) arc durations

Application 3: Stochastic activity networks (continued)

The cdf of the completion time for the network is



Application 3: Stochastic activity networks (continued)

$$
\begin{aligned}
F_{T_6}(t) &= 1 + \frac{107}{4}e^{-2t} - \frac{71}{4}e^{-4t} - 8e^{-2t}t^2 - \frac{45}{2}e^{-2t}t \\
&- \frac{1}{6}e^{-2t}t^3 - \frac{1}{6}e^{-t}t^3 - 2e^{-t}t^2 - 2e^{-4t}t^2 \\
&- \frac{71}{2}e^{-3t}t + \frac{1}{8}e^{-2t}t^4 - \frac{1}{8}e^{-3t}t^4 - 9e^{-3t}t^2 \\
&+ \frac{2}{3}e^{-3t}t^3 - 12e^{-4t}t - \frac{85}{4}e^{-3t} + \frac{45}{4}e^{-t} \qquad t > 0
\end{aligned}
$$

The mean network completion time is

$$
E[T_6] = \int_0^\infty (1 - F_{T_6}(t))\, dt = \frac{4213}{864} \approx 4.8762
$$

Bootstrapping in systems reliability

Use bootstrapping to determine a 95% lower confidence bound on the system reliability for a series system of three independent components using the binary failure data (y_i, n_i), where

- y_i is the number of components of type i that pass the test
- n_i is the number of components of type i on test

for $i = 1, 2, 3$

Component number	$i = 1$	$i = 2$	$i = 3$
Number passing (y_i)	21	27	82
Number on test (n_i)	23	28	84

Point estimate for the system reliability:

$$\frac{21}{23} \cdot \frac{27}{28} \cdot \frac{82}{84} = \frac{1107}{1288} \cong 0.8595$$

Application 4: Lower bound on system reliability

Bootstrapping in systems reliability

Use bootstrapping to determine a 95% lower confidence bound on the system reliability for a series system of three independent components using the binary failure data (y_i, n_i), where

- y_i is the number of components of type i that pass the test
- n_i is the number of components of type i on test

for $i = 1, 2, 3$

Component number	$i = 1$	$i = 2$	$i = 3$
Number passing (y_i)	21	27	82
Number on test (n_i)	23	28	84

Point estimate for the system reliability:

$$\frac{21}{23} \cdot \frac{27}{28} \cdot \frac{82}{84} = \frac{1107}{1288} \cong 0.8595$$

Bootstrapping in systems reliability

Use bootstrapping to determine a 95% lower confidence bound on the system reliability for a series system of three independent components using the binary failure data (y_i, n_i), where

- y_i is the number of components of type i that pass the test
- n_i is the number of components of type i on test

for $i = 1, 2, 3$

Component number	$i = 1$	$i = 2$	$i = 3$
Number passing (y_i)	21	27	82
Number on test (n_i)	23	28	84

Point estimate for the system reliability:

$$\frac{21}{23} \cdot \frac{27}{28} \cdot \frac{82}{84} = \frac{1107}{1288} \cong 0.8595$$

Application 4: Lower bound on system reliability

```
> X1 := BinomialRV(23, 21 / 23);
> X1 := Transform(X1, [[x -> x / 23], [0, 23]]);
> X2 := BinomialRV(28, 27 / 28);
> X2 := Transform(X2, [[x -> x / 28], [0, 28]]);
> X3 := BinomialRV(84, 82 / 84);
> X3 := Transform(X3, [[x -> x / 84], [0, 84]]);
> T := Product(X1, X2, X3);
```

- There are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for T
- Of these, only 6633 are distinct because the Product procedure combines repeated values
- The lower 95% bootstrap confidence interval bound is the 0.05 fractile of the distribution of T, which is

$$6723/9016 \cong 0.7457$$

Application 4: Lower bound on system reliability

```
> X1 := BinomialRV(23, 21 / 23);
> X1 := Transform(X1, [[x -> x / 23], [0, 23]]);
> X2 := BinomialRV(28, 27 / 28);
> X2 := Transform(X2, [[x -> x / 28], [0, 28]]);
> X3 := BinomialRV(84, 82 / 84);
> X3 := Transform(X3, [[x -> x / 84], [0, 84]]);
> T := Product(X1, X2, X3);
```

- There are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for T
- Of these, only 6633 are distinct because the Product procedure combines repeated values
- The lower 95% bootstrap confidence interval bound is the 0.05 fractile of the distribution of T, which is

$$6723/9016 \cong 0.7457$$

Application 4: Lower bound on system reliability

```
> X1 := BinomialRV(23, 21 / 23);
> X1 := Transform(X1, [[x -> x / 23], [0, 23]]);
> X2 := BinomialRV(28, 27 / 28);
> X2 := Transform(X2, [[x -> x / 28], [0, 28]]);
> X3 := BinomialRV(84, 82 / 84);
> X3 := Transform(X3, [[x -> x / 84], [0, 84]]);
> T := Product(X1, X2, X3);
```

- There are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for T
- Of these, only 6633 are distinct because the Product procedure combines repeated values
- The lower 95% bootstrap confidence interval bound is the 0.05 fractile of the distribution of T, which is

$$6723/9016 \cong 0.7457$$

Application 4: Lower bound on system reliability


```
> X1 := BinomialRV(23, 21 / 23);
> X1 := Transform(X1, [[x -> x / 23], [0, 23]]);
> X2 := BinomialRV(28, 27 / 28);
> X2 := Transform(X2, [[x -> x / 28], [0, 28]]);
> X3 := BinomialRV(84, 82 / 84);
> X3 := Transform(X3, [[x -> x / 84], [0, 84]]);
> T := Product(X1, X2, X3);
```

- There are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for T
- Of these, only 6633 are distinct because the Product procedure combines repeated values
- The lower 95% bootstrap confidence interval bound is the 0.05 fractile of the distribution of T, which is

$$6723/9016 \cong 0.7457$$

Application 4: Lower bound on system reliability


```
> X1 := BinomialRV(23, 21 / 23);
> X1 := Transform(X1, [[x -> x / 23], [0, 23]]);
> X2 := BinomialRV(28, 27 / 28);
> X2 := Transform(X2, [[x -> x / 28], [0, 28]]);
> X3 := BinomialRV(84, 82 / 84);
> X3 := Transform(X3, [[x -> x / 84], [0, 84]]);
> T := Product(X1, X2, X3);
```

- There are a possible $24 \cdot 29 \cdot 85 = 59,160$ potential mass values for `T`
- Of these, only 6633 are distinct because the `Product` procedure combines repeated values
- The lower 95% bootstrap confidence interval bound is the 0.05 fractile of the distribution of `T`, which is

$$6723/9016 \cong 0.7457$$

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing

Example 1. Consider an $M/M/1$ queue with arrival rate λ and service rate μ starting empty and idle at time $t = 0$. Calculate the fourth customer's sojourn time distribution, mean sojourn time and sojourn time variance.

```
> X := ExponentialRV(lambda);
> Y := ExponentialRV(mu);
> T := Queue(X, Y, 4, 0, 1);
> Mean(T);
> Variance(T);
```

Application 6: Queueing (continued)

The probability density function of the fourth customer's sojourn time is

$$f_4(t) = \frac{1}{6(\lambda+\mu)^5}\mu^4 e^{-\mu t}\Big(30\lambda^2 + 30\lambda^3 t + 24\lambda\mu + 24\lambda^2\mu t + 6\mu^2 + 6\mu^2\lambda t + 9t^2\lambda^4 + 12t^2\lambda^3\mu + 3t^2\lambda^2\mu^2 + t^3\lambda^5 + 2t^3\lambda^4\mu + t^3\lambda^3\mu^2\Big)$$

for $t > 0$ which has mean

$$E[T_4] = \frac{\mu^5 + 6\lambda\mu^4 + 26\mu^2\lambda^3 + 16\mu^3\lambda^2 + 17\mu\lambda^4 + 4\lambda^5}{\mu(\lambda+\mu)^5}$$

and variance

$$V[T_4] = \Big(181\mu^2\lambda^8 + 484\mu^3\lambda^7 + 816\mu^4\lambda^6 + 868\mu^5\lambda^5 + 574\mu^6\lambda^4 + 244\mu^7\lambda^3 + 40\mu\lambda^9 + 68\mu^8\lambda^2 + 12\mu^9\lambda + \mu^{10} + 4\lambda^{10}\Big) \Big/ \Big(\mu^2(\lambda+\mu)^{10}\Big)$$

Application 6: Queueing (continued)

Example 2. Calculate the mean, variance, skewness, and kurtosis of the initial customer sojourn times for an $M/M/1$ queue with $\lambda = 1$, $\mu = 2$, and $k = 0, 4, 8$ customers initially present.

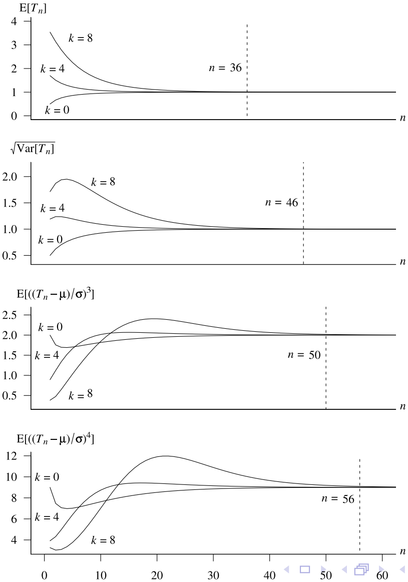
```
> X := ExponentialRV(1);
> Y := ExponentialRV(2);
> for i from 2 to 60 by 1 do
>   T := Queue(X, Y, i, k, 1):
>   print(i, evalf(Mean(T)), evalf(Variance(T))
      evalf(Skewness(T)), evalf(Kurtosis(T))):
> od:
```

Application 6: Queueing (continued)

Example 2. Calculate the mean, variance, skewness, and kurtosis of the initial customer sojourn times for an $M/M/1$ queue with $\lambda = 1$, $\mu = 2$, and $k = 0, 4, 8$ customers initially present.

```
> X := ExponentialRV(1);
> Y := ExponentialRV(2);
> for i from 2 to 60 by 1 do
>   T := Queue(X, Y, i, k, 1):
>   print(i, evalf(Mean(T)), evalf(Variance(T))
      evalf(Skewness(T)), evalf(Kurtosis(T))):
> od:
```

Application 6: Queueing (continued)



Application 6: Queueing (continued)

Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

Application 6: Queueing (continued)

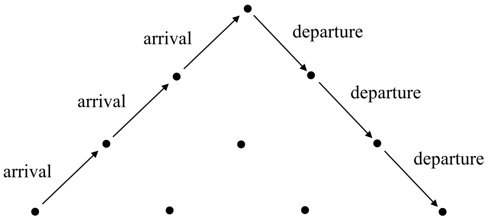
Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

Transition diagram

Application 6: Queueing (continued)

Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

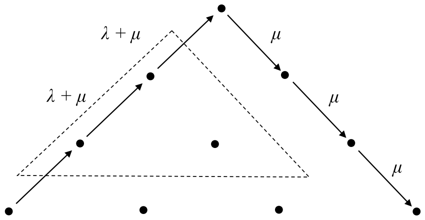
Transition diagram



Application 6: Queueing (continued)

Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

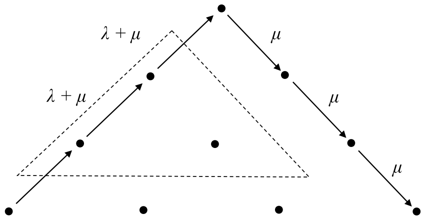
Transition diagram



Application 6: Queueing (continued)

Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

Transition diagram

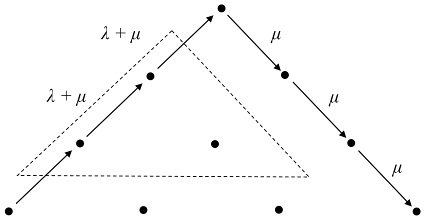


Number of paths: $\frac{(2n)!}{n!(n+1)!}$ **(Catalan number)**

Application 6: Queueing (continued)

Example 3 Find the variance–covariance matrix of the first three customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate λ and service rate μ

Transition diagram



Number of paths: $\frac{(2n)!}{n!(n+1)!}$ **(Catalan number)**

Covariance between first and second customer's sojourn times when $n = 3$

```
> Cov(1, 2, 3);
```

Application 6: Queueing (continued)

Using a Computer Algebra System in Data Analysis

Larry Leemis • The College of William & Mary

Outline

APPL survey
What is APPL?
Examples

Applications
Bootstrapping
K–S test
Stochastic activity networks
System reliability
Benford's law
Queueing
Probability distributions
Control charts
U(0,1) testing
Bivariate transformations
Wilcoxon test
ARMA models

Variance–covariance matrix of the sojourn times for $n = 3$:

$$\begin{bmatrix} \frac{1}{\mu^2} & \frac{\lambda(2\mu+\lambda)}{(\lambda+\mu)^2\mu^2} & \frac{\lambda^2(\lambda^2+4\lambda\mu+5\mu^2)}{(\lambda+\mu)^4\mu^2} \\ \bullet & \frac{2\lambda^2+4\lambda\mu+\mu^2}{(\lambda+\mu)^2\mu^2} & \frac{\lambda(2\lambda^2+8\lambda^2\mu+11\lambda\mu^2+2\mu^3)}{(\lambda+\mu)^4\mu^2} \\ \bullet & \bullet & \frac{3\lambda^6+18\lambda^5\mu+45\lambda^4\mu^2+54\lambda^3\mu^3+30\lambda^2\mu^4+8\lambda\mu^5+\mu^6}{(\lambda+\mu)^6\mu^2} \end{bmatrix}$$

For $\lambda = 1$ and $\mu = 2$, for example,

$$\Sigma = \begin{bmatrix} \frac{1}{4} & \frac{5}{36} & \frac{29}{324} \\ \bullet & \frac{7}{18} & \frac{13}{54} \\ \bullet & \bullet & \frac{1451}{2916} \end{bmatrix} \approx \begin{bmatrix} 0.2500 & 0.1389 & 0.0895 \\ \bullet & 0.3889 & 0.2407 \\ \bullet & \bullet & 0.4976 \end{bmatrix}$$

Application 6: Queueing (continued)

Example 4 Find the variance–covariance matrix of the first **nine** customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate $\lambda = 1$ and service rate $\mu = 2$

Application 6: Queueing (continued)

Example 4 Find the variance–covariance matrix of the first **nine** customer sojourn times in an initially empty and idle $M/M/1$ queue with arrival rate $\lambda = 1$ and service rate $\mu = 2$

$$\begin{vmatrix}
\frac{1}{4} & \frac{5}{36} & \frac{29}{324} & \frac{181}{2916} & \frac{1181}{26244} & \frac{2647}{78732} & \frac{18191}{708588} & \frac{127111}{6377292} & \frac{2699837}{172186884} \\
\bullet & \frac{7}{18} & \frac{13}{54} & \frac{239}{1458} & \frac{1543}{13122} & \frac{10303}{118098} & \frac{23485}{354294} & \frac{163493}{3188646} & \frac{3462503}{86093442} \\
\bullet & \bullet & \frac{1451}{2916} & \frac{8531}{26244} & \frac{53995}{236196} & \frac{356291}{2125764} & \frac{805705}{6377292} & \frac{5576849}{57395628} & \frac{39197977}{516560652} \\
\bullet & \bullet & \bullet & \frac{34514}{59049} & \frac{209794}{531441} & \frac{1357010}{4782969} & \frac{3031606}{14348907} & \frac{20810726}{129140163} & \frac{145390102}{1162261467} \\
\bullet & \bullet & \bullet & \bullet & \frac{12525605}{19131876} & \frac{77889229}{172186884} & \frac{170586983}{516560652} & \frac{1156711327}{4649045868} & \frac{8013045911}{41841412812} \\
\bullet & \bullet & \bullet & \bullet & \bullet & \frac{551583889}{774840978} & \frac{1162296371}{2324522934} & \frac{7727099083}{20920706406} & \frac{52871149859}{188286357654} \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \frac{10582107143}{13947137604} & \frac{67728246079}{125524238436} & \frac{454382575415}{1129718145924} \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \frac{225196533287}{282429536481} & \frac{1455144635743}{2541865828329} \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \frac{75890492486993}{91507169819844}
\end{vmatrix}$$

Application 9: Testing random numbers

Lehmer's (1951) linear congruential generator

$$Z_i = aZ_{i-1} \bmod m$$

$$U_i = Z_i / m$$

Marsaglia's observation (1968):

"Random numbers fall mainly in the plane"

Notorious generator: IBM's RANDU

Application 9: Testing random numbers

Lehmer's (1951) linear congruential generator

$$Z_i = aZ_{i-1} \bmod m$$

$$U_i = Z_i / m$$

Marsaglia's observation (1968):

"Random numbers fall mainly in the plane"

Notorious generator: IBM's RANDU

Application 9: Testing random numbers

Lehmer's (1951) linear congruential generator

$$Z_i = aZ_{i-1} \bmod m$$

$$U_i = Z_i / m$$

Marsaglia's observation (1968):

"Random numbers fall mainly in the plane"

Notorious generator: IBM's RANDU

Application 9: Testing random numbers

Lehmer's (1951) linear congruential generator

$$Z_i = aZ_{i-1} \bmod m$$

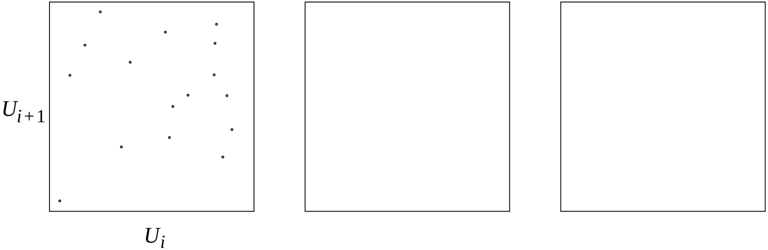
$$U_i = Z_i/m$$

Marsaglia's observation (1968):

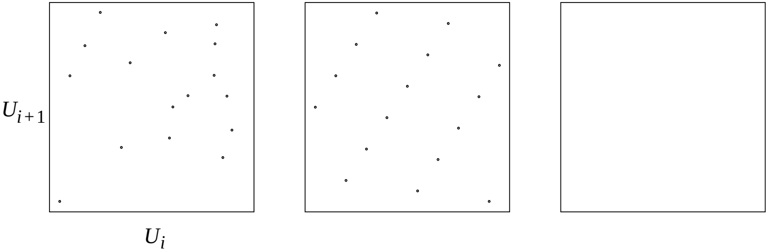
"Random numbers fall mainly in the plane"

Notorious generator: IBM's RANDU

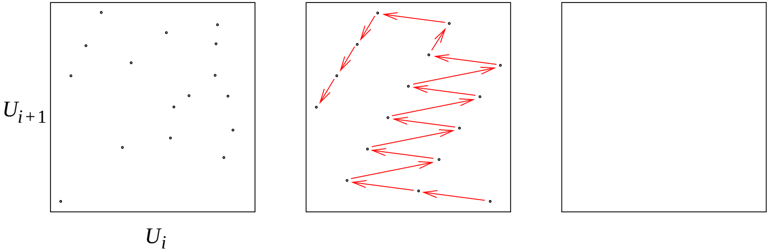
Application 9: Testing random numbers (continued)



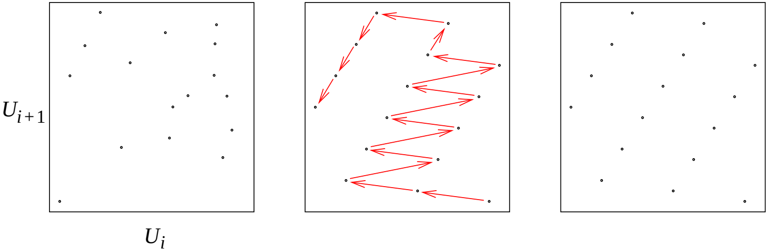
Application 9: Testing random numbers (continued)



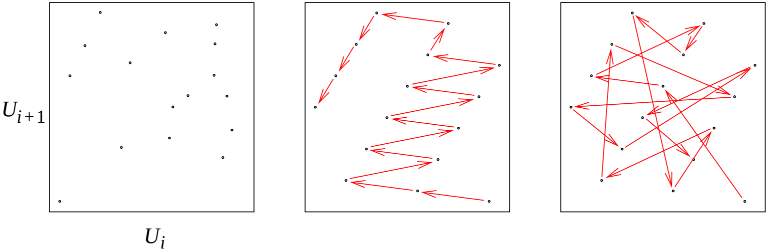
Application 9: Testing random numbers (continued)



Application 9: Testing random numbers (continued)



Application 9: Testing random numbers (continued)



Application 9: Testing random numbers (continued)

Find the distribution of the distance between two random points in the unit square $D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

```
> U1 := UniformRV(0,1);
> U2 := UniformRV(0,1);
> V1 := Difference(U1, U2);
> g1 := [[x -> x * x, x -> x * x],
>         [-infinity, 0, infinity]];
> V2 := Transform(V1, g1);
> V3 := Convolution(V2, V2);
> g2 := [[x -> sqrt(x)], [0, 2]];
> V4 := Transform(V3, g2);
```

$$f(x) = \begin{cases} 2x(x^2 - 4x + \pi) & 0 < x \le 1 \\ \dfrac{-2x\left(2\sqrt{x^2-1}+4-4x^2+2\sqrt{x^2-1}\arcsin\left(\frac{x^2-2}{x^2}\right)+x^2\sqrt{x^2-1}\right)}{\sqrt{x^2-1}} & 1 < x < \sqrt{2} \end{cases}$$

Application 9: Testing random numbers (continued)

Find the distribution of the distance between two random points in the unit square $D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

```
> U1 := UniformRV(0,1);
> U2 := UniformRV(0,1);
> V1 := Difference(U1, U2);
> g1 := [[x -> x * x, x -> x * x],
>          [-infinity, 0, infinity]];
> V2 := Transform(V1, g1);
> V3 := Convolution(V2, V2);
> g2 := [[x -> sqrt(x)], [0, 2]];
> V4 := Transform(V3, g2);
```

$$f(x) = \begin{cases} 2x(x^2 - 4x + \pi) & 0 < x \le 1 \\ \dfrac{-2x\left(2\sqrt{x^2-1} + 4 - 4x^2 + 2\sqrt{x^2-1}\arcsin\left(\frac{x^2-2}{x^2}\right) + x^2\sqrt{x^2-1}\right)}{\sqrt{x^2-1}} & 1 < x < \sqrt{2} \end{cases}$$

Application 9: Testing random numbers (continued)

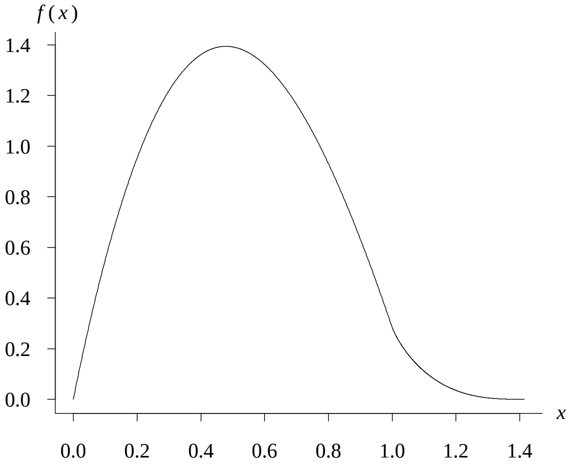
Find the distribution of the distance between two random points in the unit square $D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

```
> U1 := UniformRV(0,1);
> U2 := UniformRV(0,1);
> V1 := Difference(U1, U2);
> g1 := [[x -> x * x, x -> x * x],
>          [-infinity, 0, infinity]];
> V2 := Transform(V1, g1);
> V3 := Convolution(V2, V2);
> g2 := [[x -> sqrt(x)], [0, 2]];
> V4 := Transform(V3, g2);
```

$$f(x) = \begin{cases} 2x(x^2 - 4x + \pi) & 0 < x \le 1 \\ \dfrac{-2x\left(2\sqrt{x^2-1}+4-4x^2+2\sqrt{x^2-1}\,\arcsin\left(\frac{x^2-2}{x^2}\right)+x^2\sqrt{x^2-1}\right)}{\sqrt{x^2-1}} & 1 < x < \sqrt{2} \end{cases}$$

Application 9: Testing random numbers (continued)

Probability density function of $D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$ for two random points (X_1, Y_1) and (X_2, Y_2) in the unit square



Conclusions

- APPL is free
- APPL is easy to use
- APPL gives exact solutions to many probability problems
- APPL can be used in a variety of application areas