Functions of Random Variables

Statistics 110

Summer 2006

Functions of Random Variables

As we've seen before, if $X \sim N(\mu, \sigma^2)$, then $Y = aX + b$ is also normally distributed. However what is the distribution of X^2, $\log(X)$, or $\sin(X)$? Similarly, what is the distribution of Y if X isn't normal, say uniform?

In the discrete case, things are easily dealt with. If X has PMF $p_X(x)$, then the PMF of $Y = g(X)$ is

$$p_Y(y) = \sum_{i:g(x_i)=y} p_X(x_i)$$

that is, add up the probabilities for all x's in the sample space that get transformed to value y.

So the more difficult situation is the continuous RV case.

As we saw last class, assume that RV X has density $f_X(x)$ and CDF $F_X(x)$. Then if $Y = aX + b$,

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right) & a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right) & a < 0 \end{cases} \quad ; \qquad f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

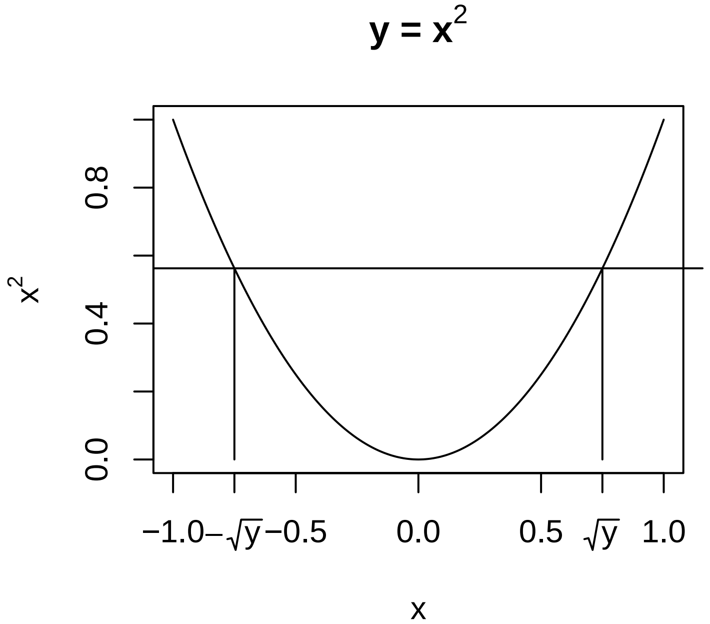
The idea behind the proof of this can be used to find the density and CDF for the transformation $Y = g(X)$ of any RV. Let the event $A = \{x : g(x) \leq y\}$. Then

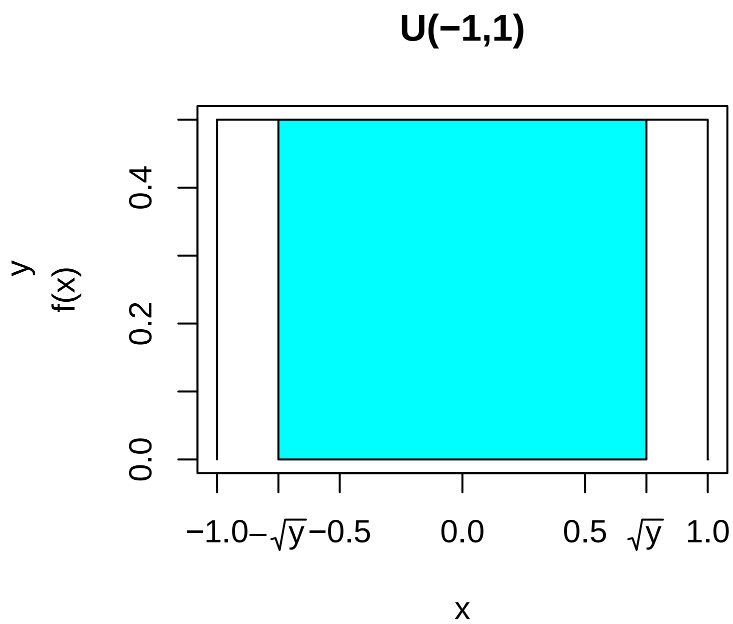
$$F_Y(y) = P[Y \leq y] = P[A] = \int_A f_X(x)dx$$

This is like the discrete case. However instead of adding up all the probabilities for all x's that get transformed to y, we "add up" the probabilities of all x's that get transformed to something less than or equal to y.

Now often this can be written in terms of $F_X(x)$. For example, let $X \sim U(-1,1)$ and $Y = X^2$, $A = [-\sqrt{y}, \sqrt{y}]$. Thus

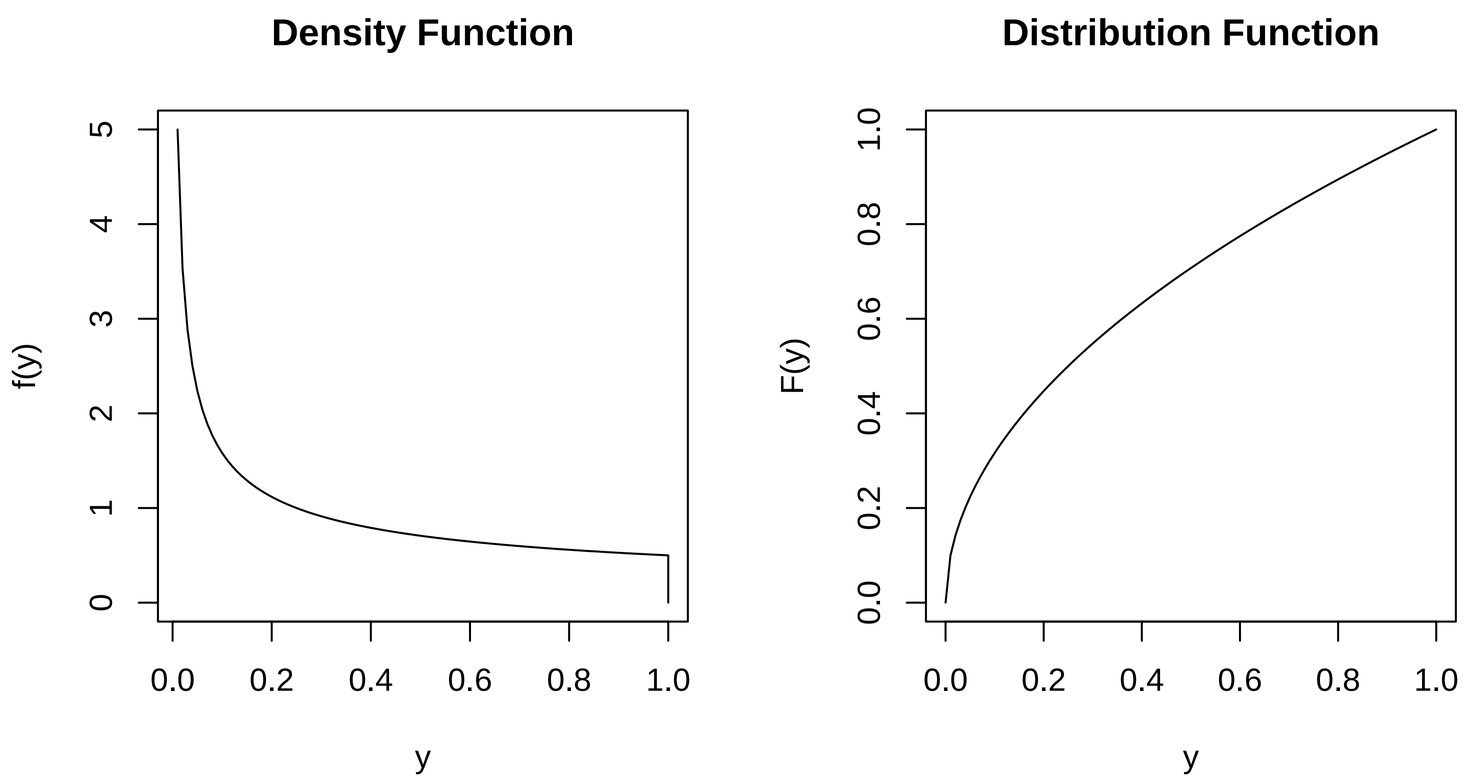
$$
\begin{aligned}
F_Y(y) &= P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\
&= \sqrt{y}
\end{aligned}
$$





Then the density of Y can be derived by differentiating the CDF.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}}; \quad 0 < y \leq 1$$



Theorem. *Let X be a continuous RV with density $f_X(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ if x is not in I. Then Y has the density function*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for y such that $y = g(x)$ for some x and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g; that is $g^{-1}(y) = x$ if $y = g(x)$.

Proof. (Assuming that g is an increasing function)

$$F_Y(y) = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$$

Thus

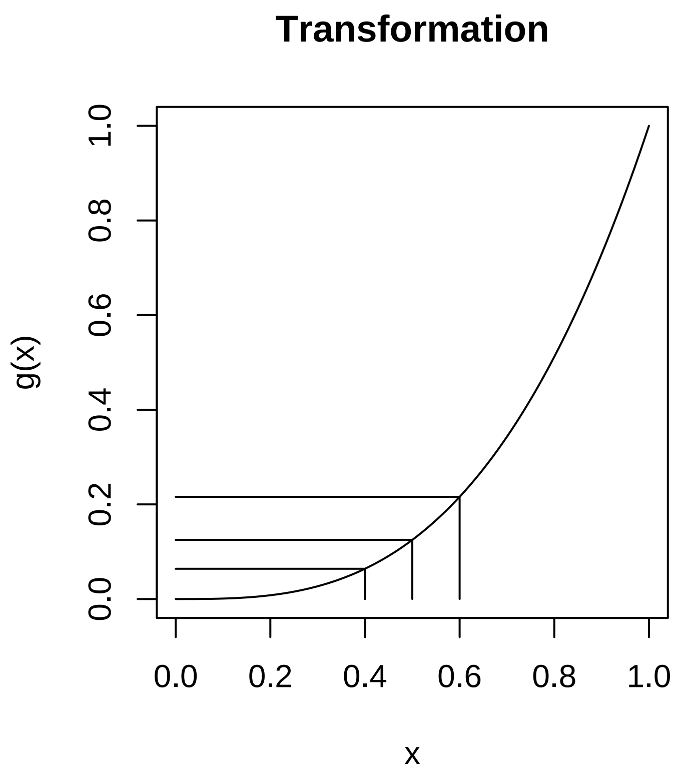
$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) \\ &= f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \end{aligned}$$

(If g is a decreasing function, $F_Y(y) = 1 - F_X(g^{-1}(y))$ and thus a couple of extra minus signs pop up which leads to the absolute value in the general case) $\square$

Intuition behind the theorem:

To find $f_Y(y)$ take a small interval (y_1, y_2) around y. Find the corresponding interval (x_1, x_2) around x, i.e. solve

$$\begin{aligned} g(x) &= y && \text{for } x \\ g(x) &= y_1 && \text{for } x_1 \\ g(x) &= y_2 && \text{for } x_2 \end{aligned}$$



Then $P[Y \in (y_1, y_2)] = P[X \in (x_1, x_2)]$, so

$$f_Y(y)(y_2 - y_1) \approx f_X(x)(x_2 - x_1)$$

or

$$f_Y(y) \approx f_X(x) \frac{\Delta x}{\Delta y}$$

The transformation g changes the scale of measurement. To keep the probabilities the same, which is needed since both are describing the same event, we must account for this change of scale. As we know from calculus, this change of scale is given by

$$\frac{d}{dy} g^{-1}(y) = \lim_{\Delta y \to 0} \frac{\Delta x}{\Delta y}$$

The example $Y = aX + b$ is a special case of this. For this transformation

$$g^{-1}(y) = \frac{y-b}{a}; \qquad \frac{d}{dy} g^{-1}(y) = \frac{1}{a}$$

thus giving

$$f_Y(y) = \frac{1}{|a|} f_X \left(\frac{y-b}{a} \right)$$

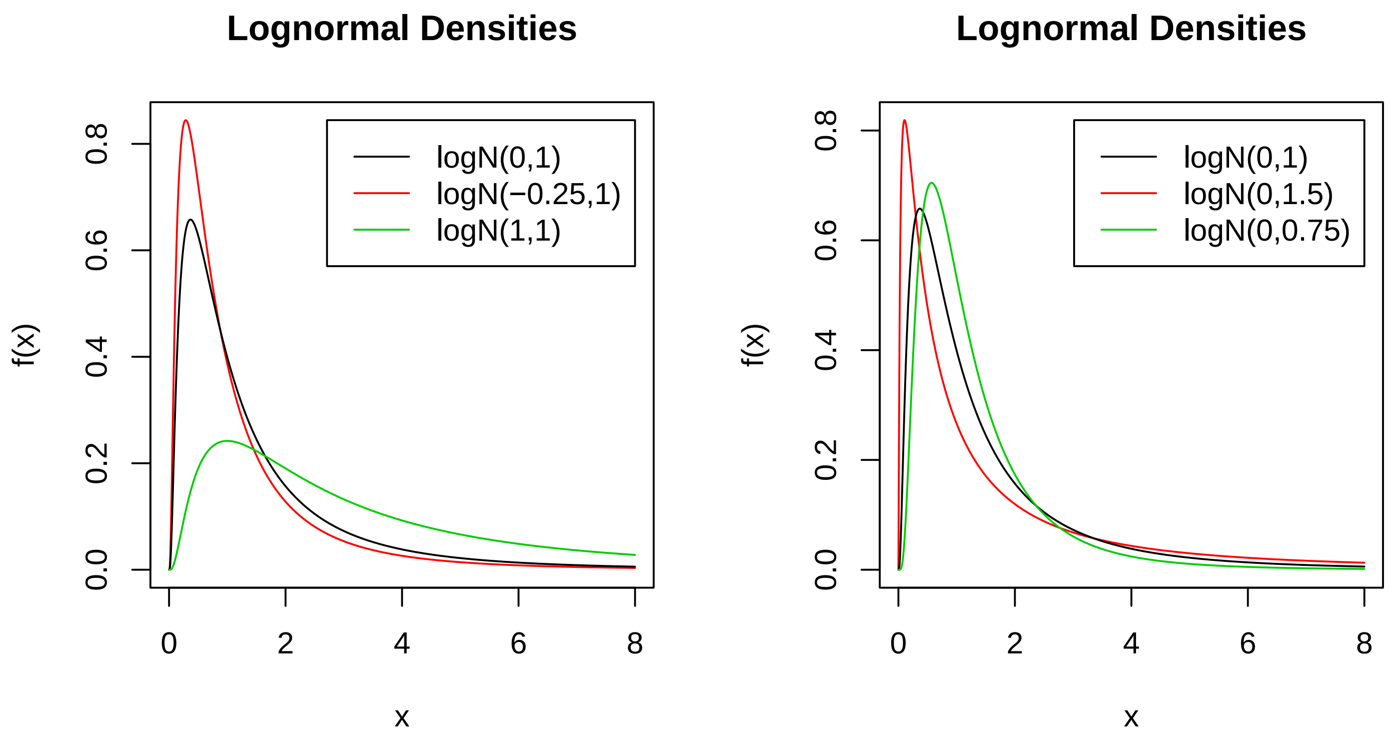
Example: Lognormal distribution

Let $X \sim N(\mu, \sigma^2)$. The density function of $Y = e^X$ is

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y - \mu)^2}{2\sigma^2}\right\}; \quad y \geq 0$$

since

$$g^{-1}(y) = \log y; \quad \frac{d}{dy} g^{-1}(y) = \frac{d}{dy} \log y = \frac{1}{y}$$



The moments of the lognormal are

$$E[X] = \exp(\mu + 0.5\sigma^2); \qquad \begin{aligned} \mathrm{Var}(X) &= \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2) \\ &= \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1) \end{aligned}$$

The lognormal is often useful when effects are multiplicative as it is possible to show that

$$\begin{aligned} \mathrm{SD}(X) &= E[X]\sqrt{\exp(\sigma^2) - 1} \\ &\propto E[X] \end{aligned}$$

where the proportionality constant depends on σ.

The name lognormal comes from the fact if Y is lognormal, then $X = \log(Y)$ is normal.

One example of where the lognormal distribution is used is the modeling of stock prices. For example, let $S(t)$ be the value of a stock at time t. Then the value at time $t + \Delta_t$ satisfies

$$\frac{S(t + \Delta_t)}{S(t)} \sim logN(\mu\Delta_t, \sigma^2\Delta_t)$$

where μ is a drift parameter and σ is the volatility of the stock.

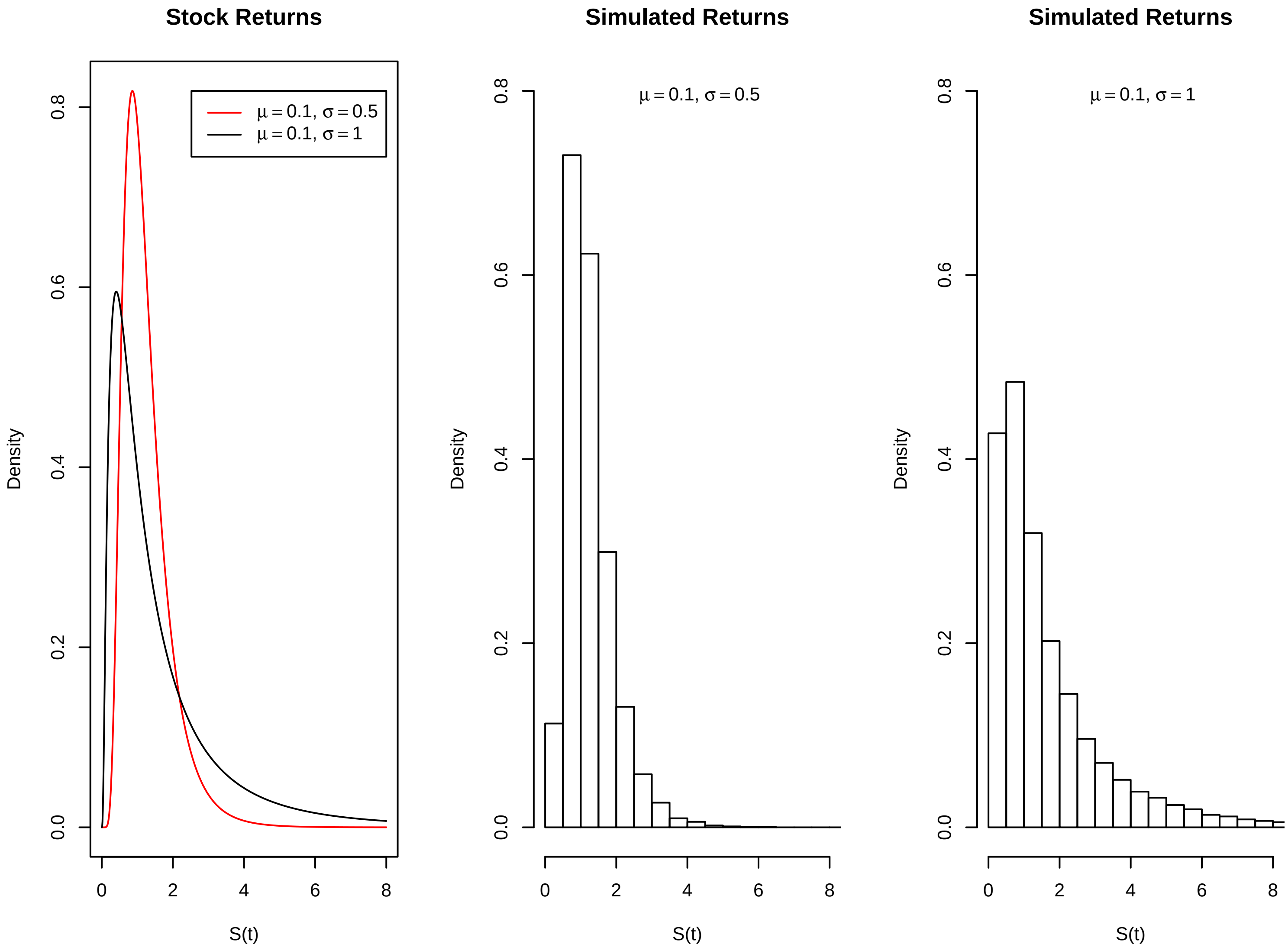
If the multiplicative increments

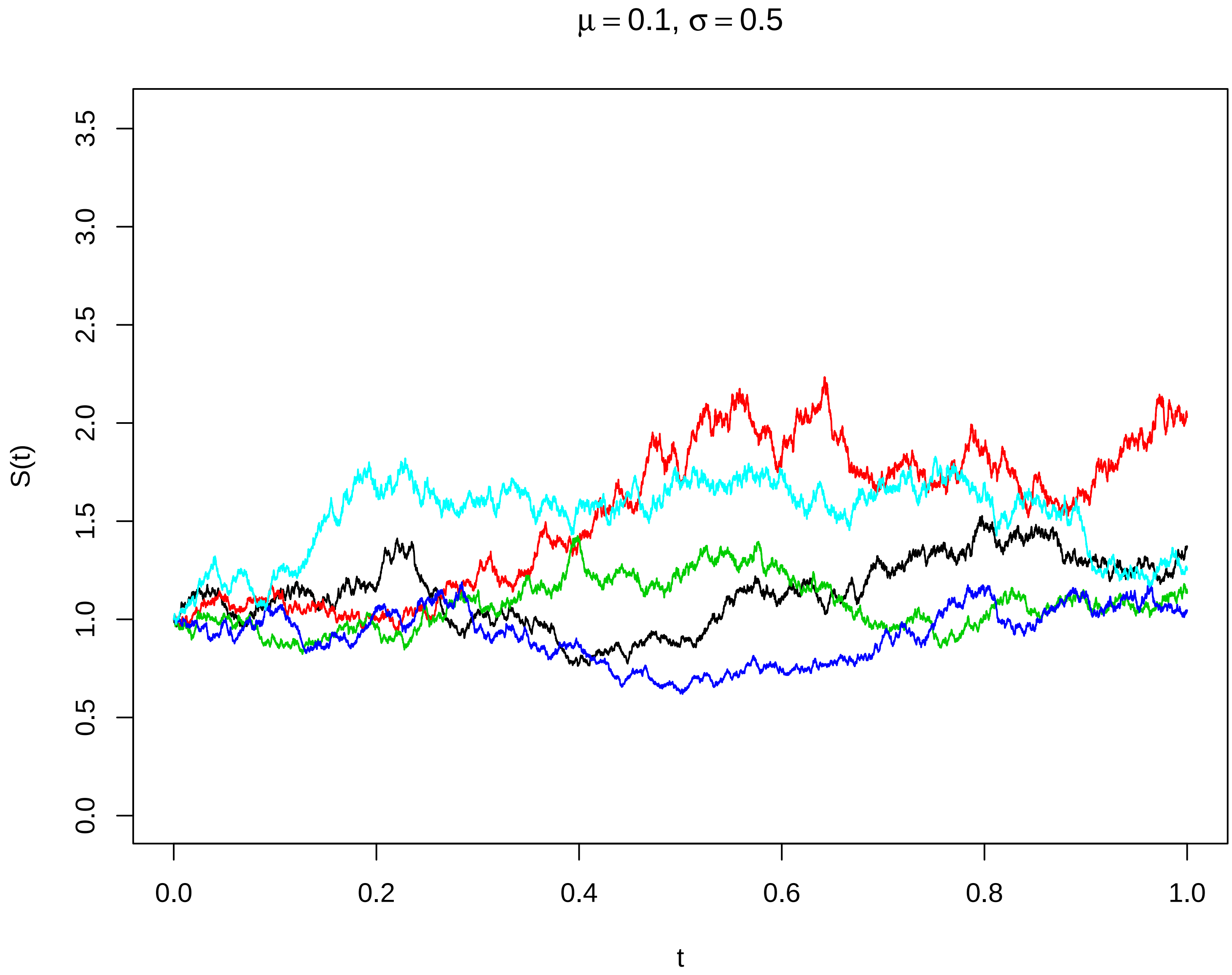
$$\frac{S(\Delta_t)}{S(0)}, \frac{S(2\Delta_t)}{S(\Delta_t)}, \frac{S(3\Delta_t)}{S(2\Delta_t)}, \ldots, \frac{S(1)}{S(1 - \Delta_t)}$$

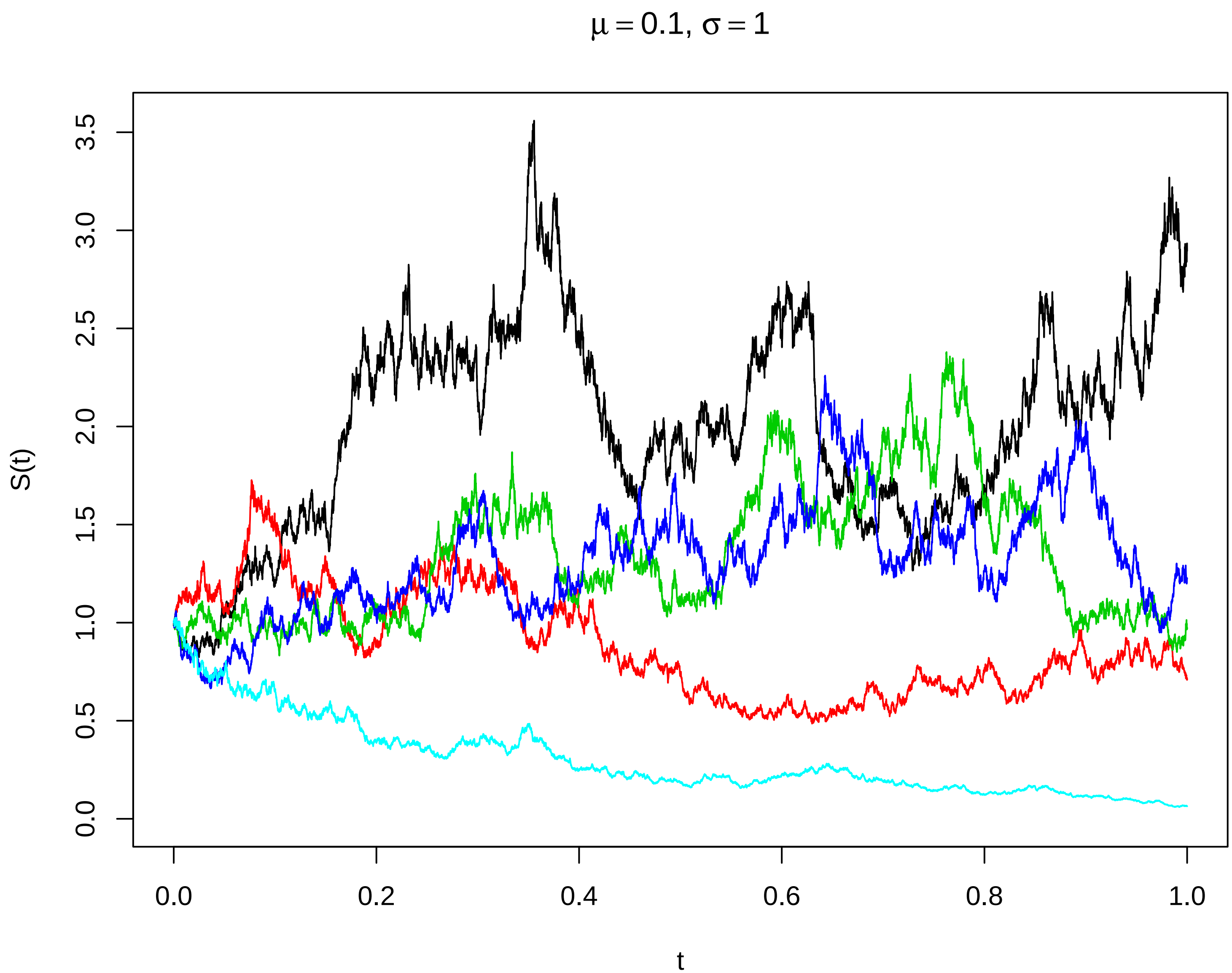
are independent, this leads to a continuous path as $\Delta_t \to 0$.

In addition, it can be shown that for any $t > 0$,

$$\frac{S(t)}{S(0)} \sim logN(\mu t, \sigma^2 t)$$


Stock Returns
Simulated Returns
Simulated Returns
μ = 0.1, σ = 0.5
μ = 0.1, σ = 1
μ = 0.1, σ = 0.5
μ = 0.1, σ = 1
Density
Density
Density
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
S(t)
S(t)
S(t)


μ = 0.1, σ = 0.5
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
S(t)
0.0
0.2
0.4
0.6
0.8
1.0
t


μ = 0.1, σ = 1
S(t)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.0
0.2
0.4
0.6
0.8
1.0
t

Theorem. [Probability Integral Transform] *Let X be a continuous RV with CDF $F_X(x)$. Then $Y = F_X(X) \sim U(0,1)$. Conversely, if $Y \sim U(0,1)$, then $X = F_X^{-1}(Y)$ has CDF $F_X(x)$.*

Proof.

$$P[Y \leq y] = P[F(X) \leq y] = P[X \leq F^{-1}(y)] = F_X(F_X^{-1}(y)) = y$$

i.e. the CDF of a $U(0,1)$ RV.

Conversely,

$$P[X \leq x] = P[F_X^{-1}(Y) \leq x] = P[Y \leq F_X(x)] = F_X(x)$$

$\square$

This theorem has many useful implications.

- In statistics, it implies that p-values and confidence intervals work correctly.
- It is also the motivation for probability plots (such as the Normal Scores plot) which can be used to check distributional assumptions of an analysis.

Another useful implication is that it gives a way to generate random numbers from any distribution.

Suppose we want to generate random numbers $X_1, X_2, \ldots X_n$ from a distribution with CDF $F_X(x)$ and quantile function $F_X^{-1}(u)$. Generate $U_1, U_2, \ldots, U_n$ from $U(0,1)$ and set $X_i = F^{-1}(U_i)$.

For example, to generate a Cauchy RV, the quantile function for a standard Cauchy ($C(0,1)$) is

$$F^{-1}(u) = \tan(\pi(u - 1/2))$$

So generate $u_1, u_2, \ldots u_n$ from $U(0,1)$ RVs and plug into this function to give standard Cauchy variates $z_1, z_2, \ldots, z_n$. Then to get $x_1, x_2, \ldots, x_n$ from $C(\mu, \sigma)$, let

$$x_i = \sigma z_i + \mu$$

Note that while you can generate random numbers for any distribution using this mechanism, often they are generated by different mechanisms for computational reasons, mainly speed. This is often required since for many distributions, the CDF does not have a nice closed form expression, which implies the quantile function does not either.

An important issue I won't address here, for this mechanism to work well, we need good algorithms for generating from a $U(0,1)$ distribution. Fortunately there are good schemes for generating these psuedo-random numbers.