Transformation of Variables

Daniel B. Rowe, Ph.D.

Professor
Department of Mathematical and Statistical Sciences

Outline

- Continuous Distributions
- PDF, Moments, CDF
- Transformation of Variable
- Uniform, Normal Distribution

Continuous RVs, PDFs, and CDFs

Assume that the continuous random variable (RV) can take on values

$$x \in [a,b]$$

then, the probability distribution function (PDF) is given by

$f(x \mid \theta)$ defined for $x \in [a,b]$

where x can be defined within an infinite interval

and θ are any parameters that the PDF depends on.

Continuous RVs, PDFs, and CDFs

Further, the cumulative distribution function (CDF) is given by

$$F(x \mid \theta) = \int_{t=-\infty}^{x} f(t \mid \theta) dt$$

Additionally, any PDF must satisfy

1) $0 \leq f(x \mid \theta)$

2) $\int_{x} f(x \mid \theta) dx = 1$.

Continuous Expectation

Given an arbitrary continuous probability distribution $f(x|\theta)$, we want to

compute quantitative population summaries of it such as

population mean, $\mu = \int_{x=-\infty}^{\infty} x f(x \mid \theta) dx$

population variance, $\sigma^2 = \int_{x=-\infty}^{\infty} (x-\mu)^2 f(x \mid \theta) dx$

population standard deviation, $\sigma = \sqrt{\sigma^2}$

Continuous Summaries

Given an arbitrary continuous probability distribution $f(x|\theta)$, we want to

compute quantitative population summaries of it such as

population median $\tilde{x}$, $\int_{x=-\infty}^{\tilde{x}} f(x|\theta)dx = \frac{1}{2}$

population mode $\hat{x}$, $\frac{\partial}{\partial x} f(x|\theta)\Big|_{\hat{x}} = 0$

Provided f is differentiable.
Max if 2^{nd} der neg at point.
Check boundary points for max.

Continuous Expectation

These population moment numerical summaries are found by expectation

$$E[g(X)\,|\,\theta] = \int_{x=-\infty}^{\infty} g(x) f(x\,|\,\theta) dx$$

The mean is

$$\mu = E(X\,|\,\theta)$$

and the variance is

$$\sigma^2 = E[(X-\mu)^2\,|\,\theta] \quad .$$

Change of Variable

Given a random variable x, with probability distribution function $f_X(x|\theta)$, we often would like to know the probability distribution of a random variable y, that is a function $y(x)$ of x, $y=y(x)$.

Change of Variable

Let $y=y(x)$ be a one-to-one transformation

with inverse transformation $x=x(y)$.

Then, if $f_X(x|\theta)$ is the PDF of x, the PDF of y can be found as

$$f_Y(y|\theta) = f_X(x(y)|\theta) \times |J(x \to y)|$$

where $J(x \to y) = \dfrac{dx(y)}{dy}$.

Suppress PDF subscripts.

Change of Variable

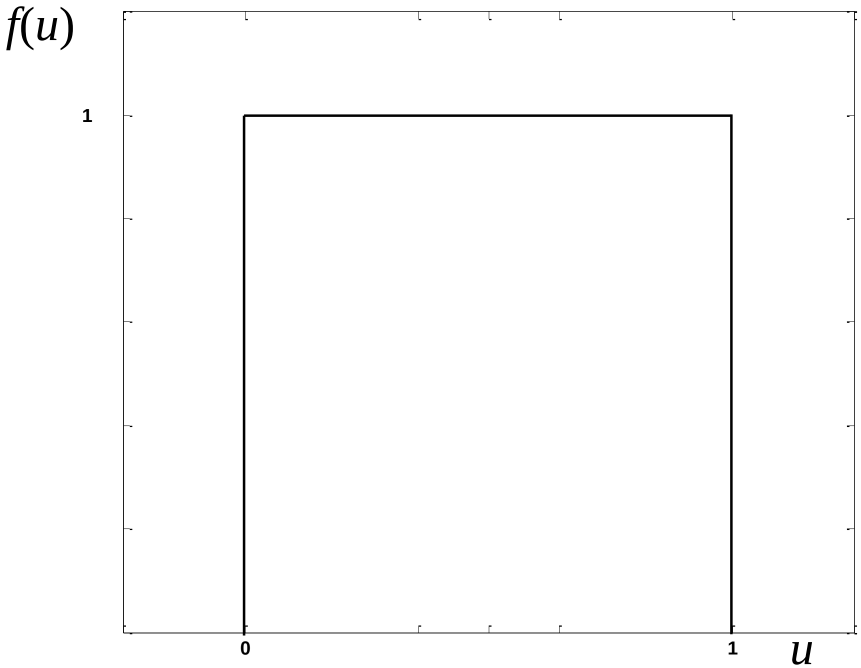
Uniform:

A random variable u has a continuous uniform distribution, u~uniform(0,1) if

$$f(u) = \begin{cases} 1 & \text{if} \quad u \in [0,1] \\ 0 & \text{if} \quad u \notin [0,1] \end{cases} ,$$

and

$$\mu_u = \frac{1}{2} \qquad \sigma_u^2 = \frac{1}{12} .$$



Change of Variable

Uniform:

We can generate 10^6 random uniform(0,1) variates and compare theoretical PDF to empirical histogram

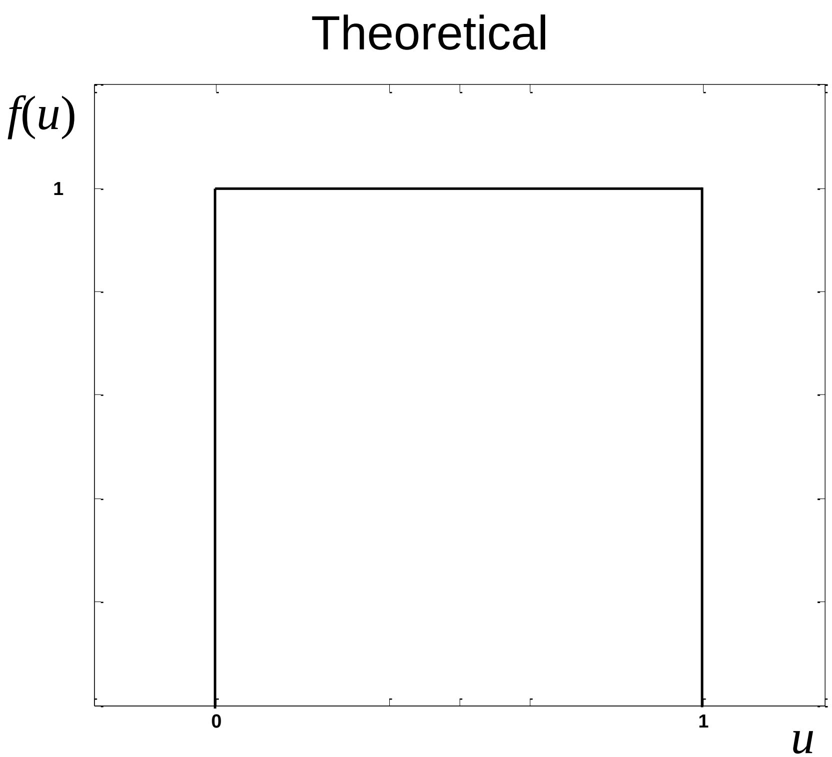
$$f(u) = \begin{cases} 1 & \text{if} \quad u \in [0,1] \\ 0 & \text{if} \quad u \notin [0,1] \end{cases}$$

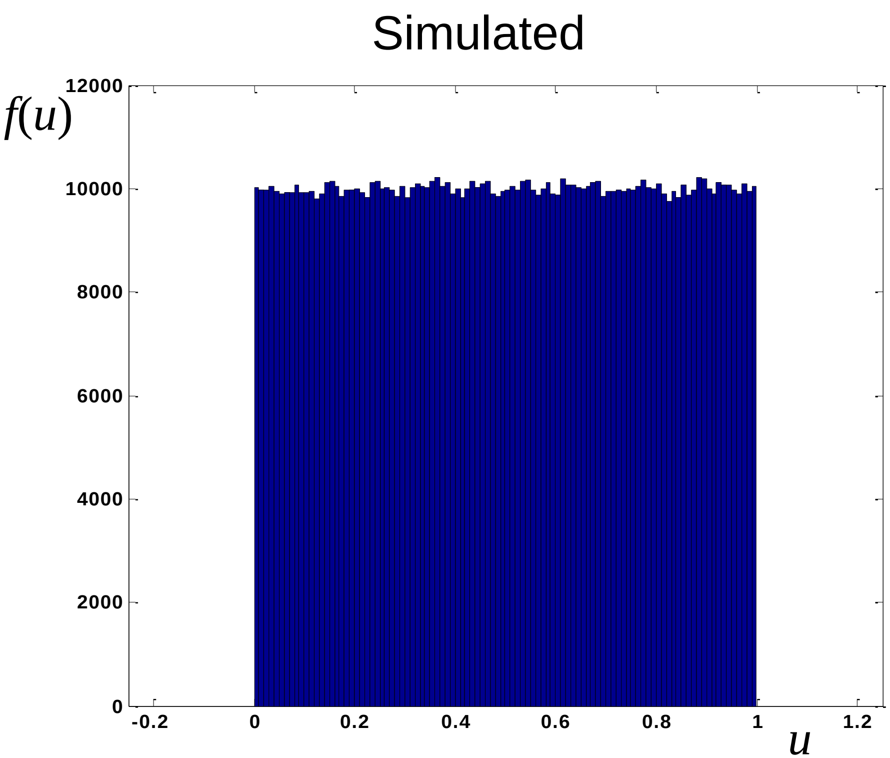
along with mean and variance

```
u=rand(10^6,1);
hist(u,100)
mean(u)
var(u)
```

Change of Variable

Uniform:





μ_u	0.5
σ_u^2	0.083

$\bar{u}$	0.5003
s_u^2	0.0833

Change of Variable

Uniform:

We can obtain a random variable x that has a general uniform distribution in the interval a to b via the transformation

$$x = (b-a)u + a \quad .$$

The PDF of x can be obtained by

$$f(x \mid a,b) = f(u(x)) \times | J(u \to x) |$$

where $u(x)$ is u written in terms of x and $J(\cdot)$ is the Jacobian of the transformation.

Change of Variable

Uniform:

The original variable u in terms of the new variable is

$$u(x) = \frac{x-a}{b-a}$$

and the Jacobian of the transformation is

$$J(u \to x) = \frac{du(x)}{dx} = \frac{1}{b-a} \quad .$$

This yields

$$f(x\,|\,a,b) = f(u(x)) \times |\,J(u \to x)\,| = 1 \times \left|\frac{1}{b-a}\right| .$$

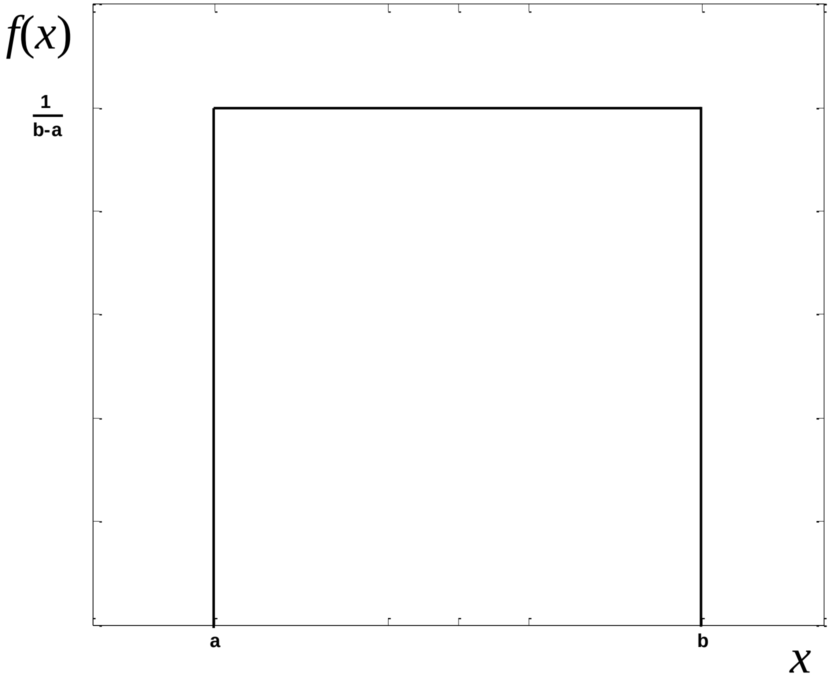
Change of Variable

Uniform:

A random variable x has a continuous uniform distribution, x~uniform(a,b) if

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{if} \quad x\in[a,b] \\ 0 & \text{if} \quad x\notin[a,b]\end{cases},$$

where, $a,b\in\mathbb{R}$, $a<b$.



Note that u=0 mapped to x=a and u=1 mapped to x=b.

Change of Variable

Uniform:

We can generate 10^6 random uniform(a,b) variates and compare theoretical PDF to empirical histogram

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{if} \quad x\in[a,b]\\ 0 & \text{if} \quad x\notin[a,b]\end{cases}$$

along with mean & variance by transforming random variates

a=1;,b=2;
x=a+(b-a)*u;
hist(x,100)
mean(x), var(x)

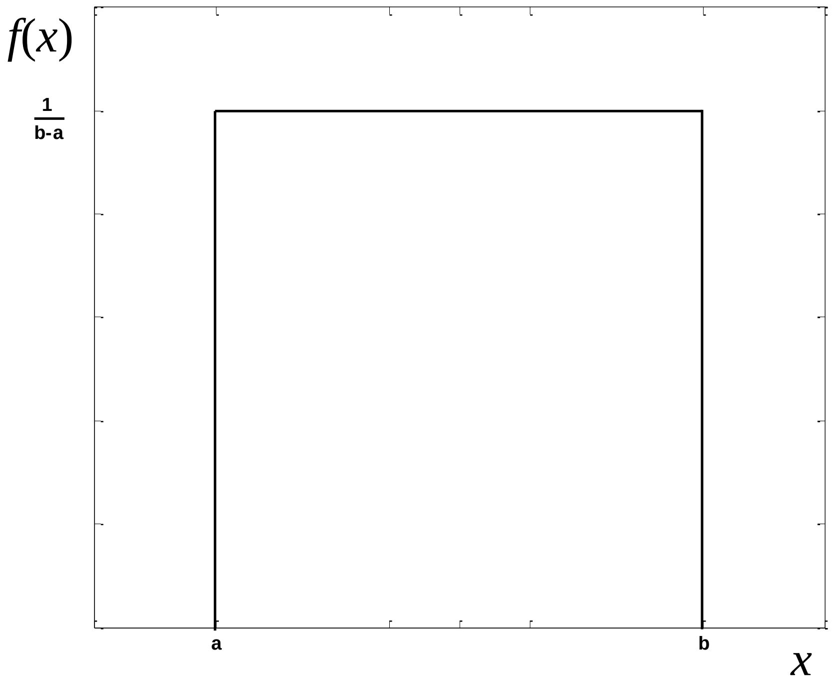
10^6 uniform variates
u=rand(10^6,1);

Change of Variable

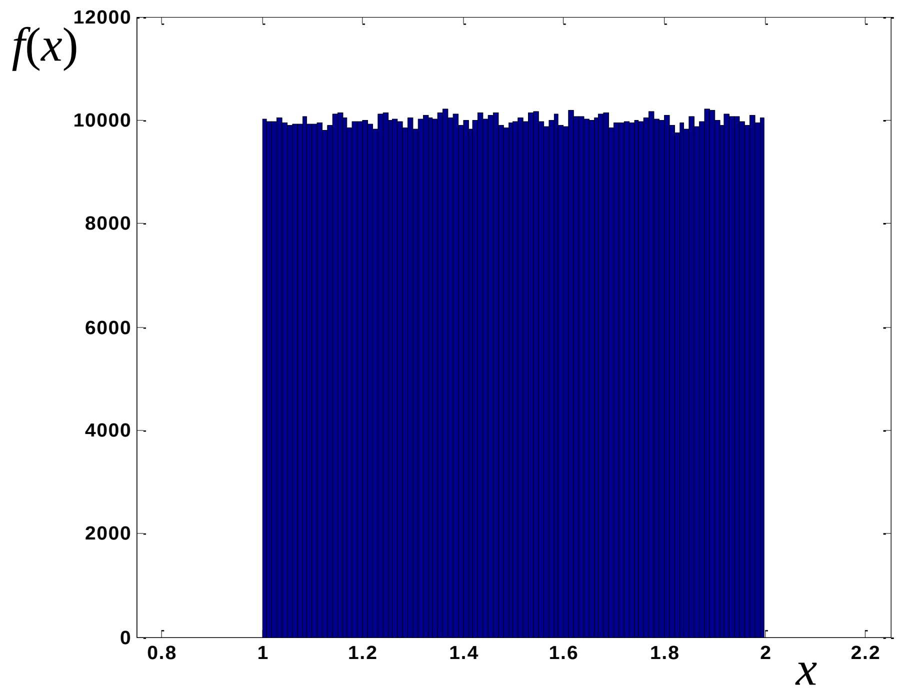
Uniform:

$$a = 1 \qquad b = 2 \qquad \mu_x = \frac{b+a}{2} \qquad \sigma_x^2 = \frac{(b-a)^2}{12}$$

Theoretical



Simulated



μ_x	1.5
σ_x^2	0.083

$\bar{x}$	1.5003
s_x^2	0.0833

Change of Variable

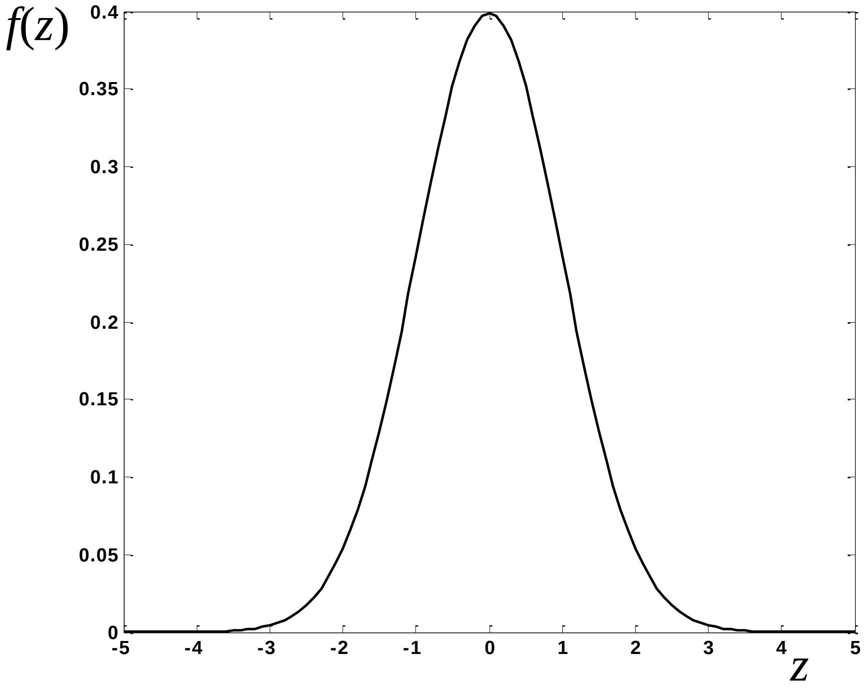
Normal: The same process can be applied.

A random variable z has a standard normal distribution, z~normal(0,1) if

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2},$$

where $z\in\mathbb{R}$ and

$$\mu_z=0 \qquad \sigma_z^2=1 \ .$$



Change of Variable

Normal:

We can generate 10^6 random normal(0,1) variates and compare theoretical PDF to empirical histogram

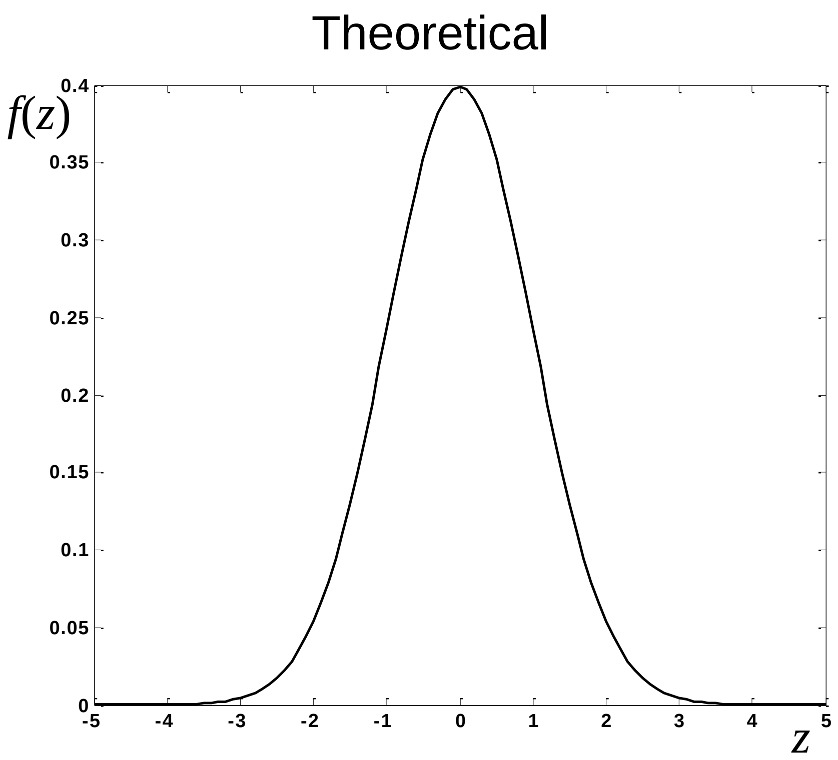
$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$

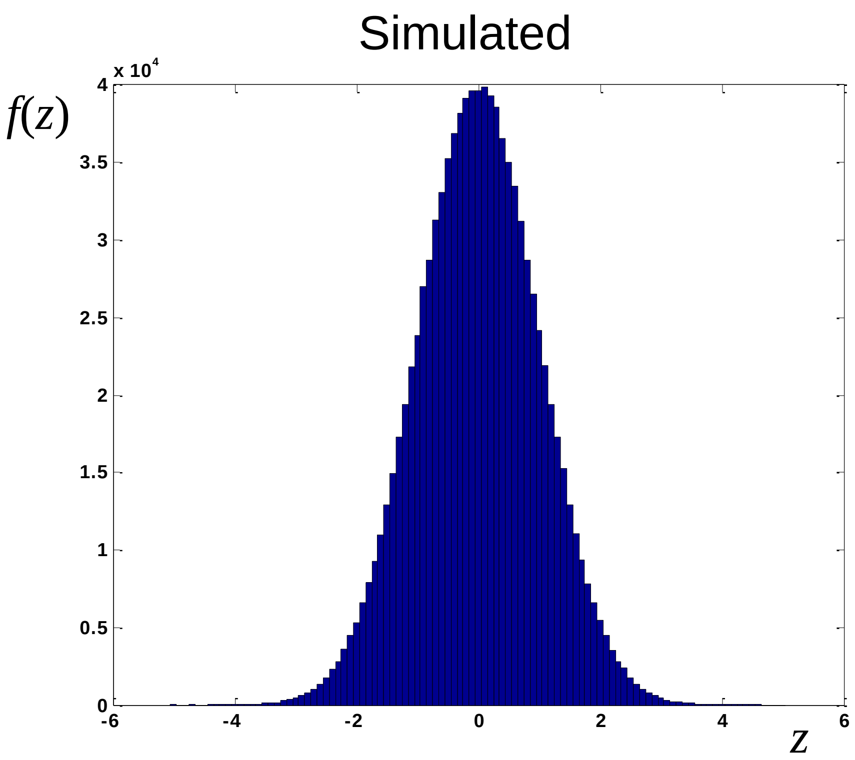
along with mean and variance

```
z=randn(10^6,1);
hist(z,(-5:.1:5))
mean(z), var(z)
xlim([-5 5])
```

Change of Variable

Normal:





μ_z	0
σ_z^2	1

$\bar{z}$	9.4576×10^{-4}
s_z^2	0.9994

Change of Variable

Normal:

We can obtain a random variable x that has a general normal distribution with mean μ and variance σ^2 via the transformation

$$x = \sigma z + \mu \quad .$$

The PDF of x can be obtained by

$$f(x \mid \mu, \sigma^2) = f(z(x)) \times | J(z \to x) |$$

where $z(x)$ is z written in terms of x and $J(\cdot)$ is the Jacobian of the transformation.

Change of Variable

Normal:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$

The original variable z in terms of the new variable is

$$z(x)=\frac{x-\mu}{\sigma}$$

and the Jacobian of the transformation is

$$J(z\to x)=\frac{dz(x)}{dx}=\frac{1}{\sigma}\ .$$

This yields

$$f(x\mid\mu,\sigma^2)=f(z(x))\times|J(z\to x)|=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\times\left|\frac{1}{\sigma}\right|.$$

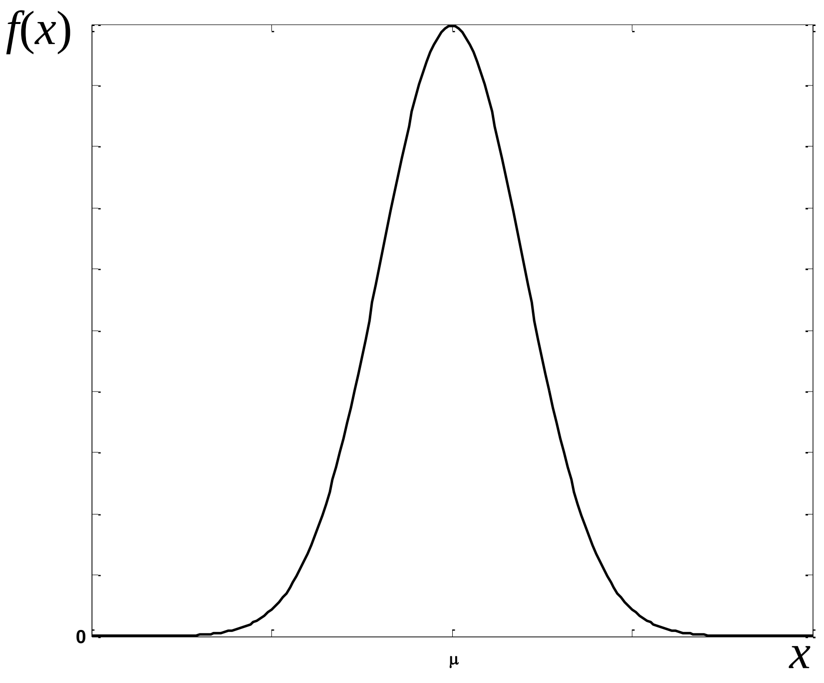
Change of Variable
Normal:

A random variable x has a general normal distribution, x~normal(μ,σ^2) if

$$f(x\,|\,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

where, $x,\mu\in\mathbb{R}\ ,0<\sigma$.

Note that z=-∞ mapped to x=-∞ and z=∞ mapped to x=∞.



Change of Variable

Normal:

We can generate 10^6 random normal(μ,σ^2) variates and compare theoretical PDF to empirical histogram

$$f(x\,|\,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

along with mean & variance by transforming random variates

```
mu=5;,sigma=2;
x=mu+sigma*z;
hist(x,(-5:.2:15) )
mean(x), var(x) , xlim([-5 15])
```

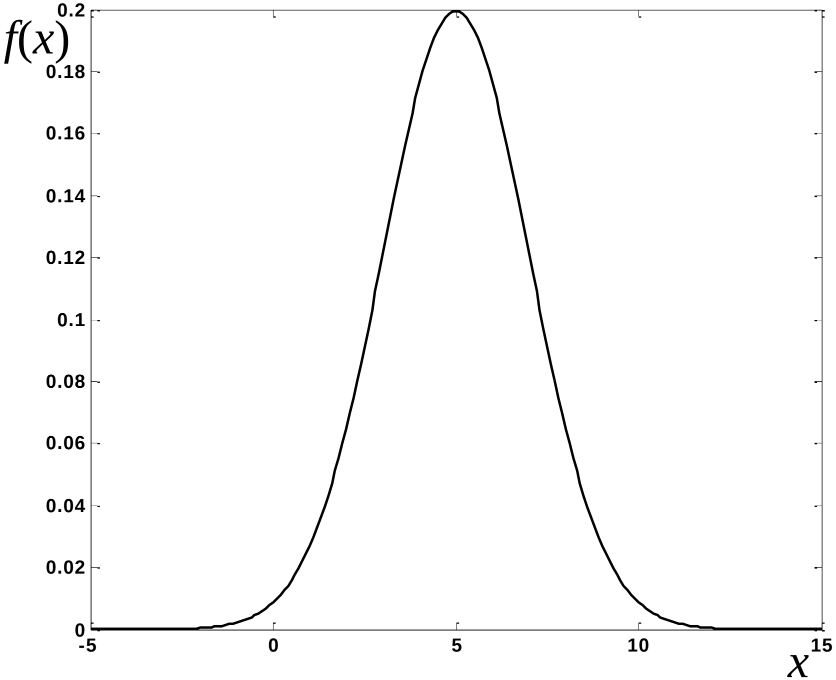
10^6 standard normal variates
z=randn(10^6,1);

Change of Variable

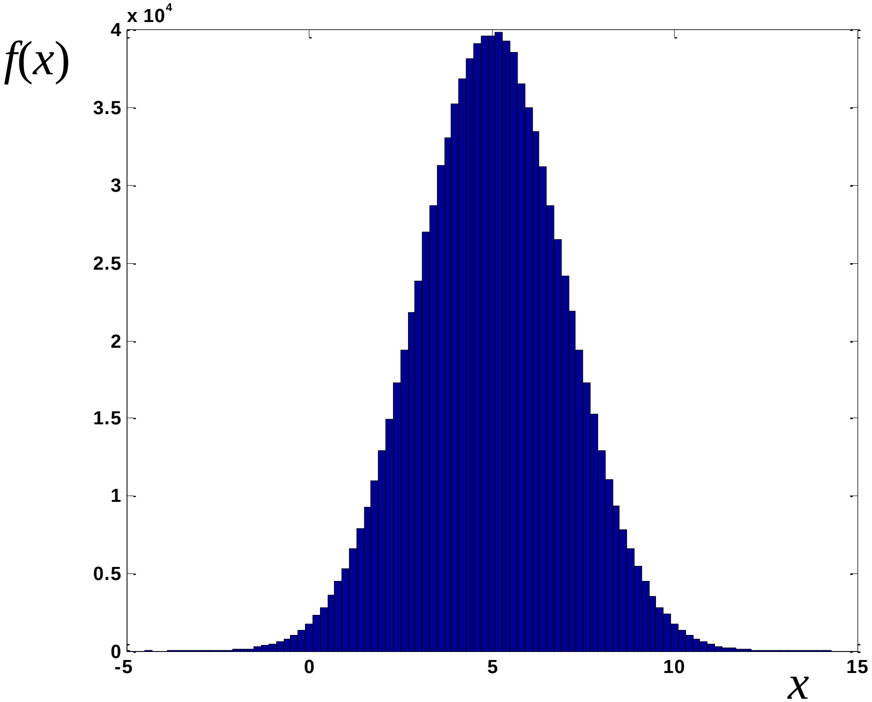
Normal:

$$\mu = 5 \quad \sigma^2 = 4$$

Theoretical



Simulated



μ_x	5
σ_x^2	4

$\bar{x}$	5.0019
s_x^2	3.9975

Change of Variable

This process can be used to find the distribution of more than linear functions $y=y(x)$ of random variables.

For example, let x~normal(μ,σ^2).

Assume we want to know the distribution of $y=\left(\frac{x-\mu}{\sigma}\right)^2$.

We can determine $f(y)$ through the transformation of variable procedure.

$$f_Y(y\,|\,\theta)=f_X(x(y)\,|\,\theta)\times|J(x\to y)|$$

Homework problem.

Change of Variable

Not one-to-one

Let $y=y(x)$ be a not one-to-one transformation,
(i.e. $y=x^2$, then $x_1(y)=+\sqrt{y}$ and $x_2(y)=-\sqrt{y}$.)

We can still perform the change of variable by breaking up the transformation into pieces that are1-to-1.

$$f_Y(y|\theta)=\sum_j f_X(x_j(y)|\theta)\times\left|\frac{dx_j(y)}{dy}\right|$$

i.e. $f_Y(y|\theta)=f_X(\sqrt{y}\,|\,\theta)\left|\frac{1}{2\sqrt{y}}\right|+f_X(-\sqrt{y}\,|\,\theta)\left|\frac{-1}{2\sqrt{y}}\right|$

Homework 6:

1) Let x~Normal(μ,σ^2).

a) Derive the distribution of $y=\left(\dfrac{x-\mu}{\sigma}\right)^2$ using the transformation of variable technique.

b) What is the name of the distribution?

c) What are the mean and variance of this distribution?

2) Generate 10^6 Normal(5,4) random variates.

a) Make a histogram, 50 bins.

b) Compute sample mean and variance.

c) Subtract 5 from each random variate, divide by 2, square.

d) Make a histogram, 50 bins.

e) Compute sample mean and variance.

Homework 6:

3) Let u~uniform(0,1).

a) Derive the distribution of y=-2ln(u) using the transformation of variable technique.

b) What is the name of the distribution?

c) What are the mean and variance of this distribution?

4) Generate 10^6 uniform(0,1) random variates.

a) Make a histogram, 50 bins.

b) Compute sample mean and variance.

c) Take natural log of each variate then multiply by -2.

d) Make a histogram, 50 bins.

e) Compute sample mean and variance.

Homework 6:

5) Let u~uniform($-\pi/2$, $\pi/2$).

a) Derive the distribution of $y=\tan(u)$ using the transformation of variable technique.

b) What is the name of the distribution?

c) What are the mean and variance of this distribution?

6) Generate 10^6 uniform($-\pi/2$, $\pi/2$) random variates.

a) Make a histogram, 50 bins.

b) Compute sample mean and variance.

c) Take tangent of each variate.

d) Make a histogram, 50 bins.

e) Compute sample mean and variance.

Homework 6:

7) Let x~normal(0,1).

   a) Use the transformation of variable technique for $y=1/x$.

   b) What can you tell us about the distribution of y?

8) Numerically integrate the $f(y)$ PDF with rectangles to find the 99th percentile.

9) Generate 10^6 normal(0,1) random variates.

   a) Make a histogram, 50 bins.

   b) Compute sample mean and variance.

   c) Take the reciprocal of each random variate for $y=1/x$.

   d) Make a histogram, 50 bins.

   e) Compute sample mean and variance.

   f) Find the $.99*10^6$ largest value x_0. (Compare to 8)