Lesson 12
Derived Distributions

In real-life engineering applications, Systems are fed with random inputs, whose distributions are usually known.

We need to know the distribution of the outputs of various systems, given the distribution of their output. Previously we studied such derived distributions for discrete random variables.

In this lesson, we focus on more general cases, involving ~~of~~ continuous random variables.

Let us focus on $Y = g(X)$,

where X is a continuous random variable and $g(\cdot)$ is a measurable function.

Every r.v. has a cdf.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y)$$

$$= \mathbb{P}(X \in B_y) = \mathbb{P}_X(B_y)$$

where

$$B_y = \{x \in \mathbb{R} \mid g(x) \leq y\}$$

B_y is a Borel Set (why?)

If the pdf of X is f_X then

$$F_Y(y) = \int_{B_y} f_X(x)\,dx$$

And the pdf of Y is:

$$f_Y(y) = \frac{d\,F_Y(y)}{dy}$$

Another equivalent approach would be using the probability law of Y:

$\forall B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(g(X) \in B)$$

$$= \mathbb{P}(X \in \overleftarrow{g}(B)) = \mathbb{P}_X(\overleftarrow{g}(B))$$

$f_X(x) = \frac{1}{2} e^{-|x|}$

Example: Assume $X \sim$ Laplace(1).

$Y = |X|$. What is the cdf of Y?

What is the pdf of Y?

$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(|X| \leq y)$

$Y \in \underbrace{(-\infty, y]}_{B_y}$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \mathbb{P}(-y \leq X \leq y) & y \geq 0 \end{cases}$$

$(Y \sim Exp(1))$

$y \geq 0 \qquad F_Y(y) = \mathbb{P}(-y \leq X \leq y)$

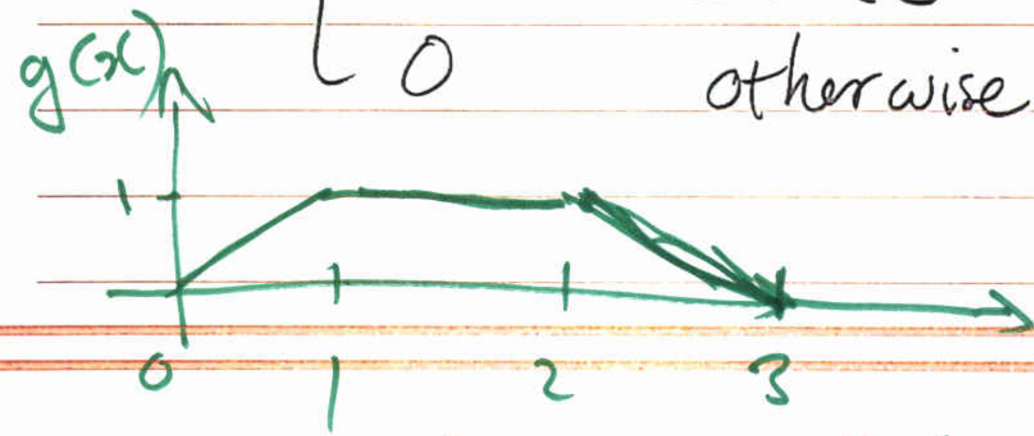
$= F_X(y) - F_X(-y)$

$f_Y(y) = \frac{dF_Y}{dy} = f_X(y) - (-f_X(-y))$

$= f_X(y) + f_X(-y) = 2 f_X(y) = e^{-|y|} = e^{-y}$

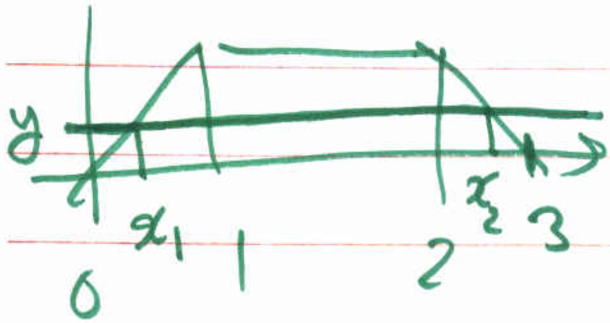
Example: Find the cdf and density of $Y = g(X)$ if $X \sim U[0,4]$ and

$$g(x) = \begin{cases} x & 0 \le x < 1 \\ 1 & 1 \le x < 2 \\ 3-x & 2 \le x < 3 \\ 0 & \text{otherwise} \end{cases}$$



$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$$

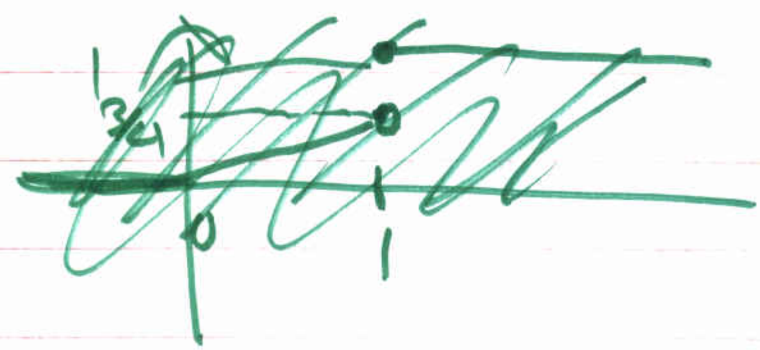
$$= \begin{cases} \mathbb{P}(\emptyset) = 0 & y < 0 \\ ? & 0 \le y < 1 \\ \mathbb{P}(\Omega) = 1 & y \ge 1 \end{cases}$$

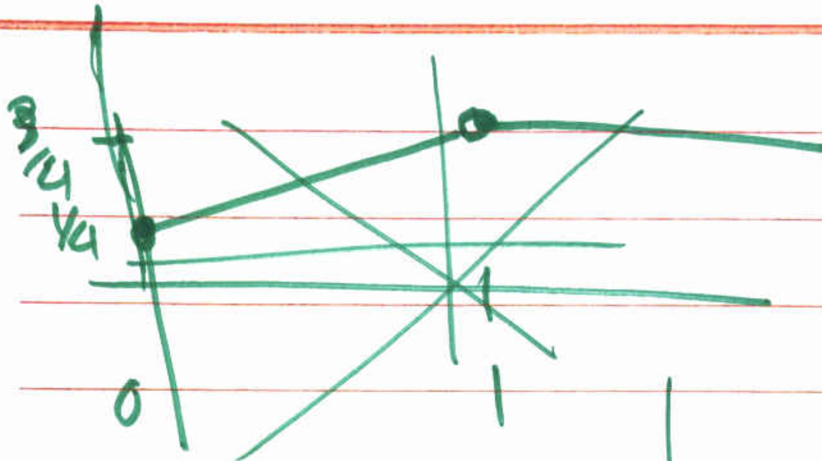


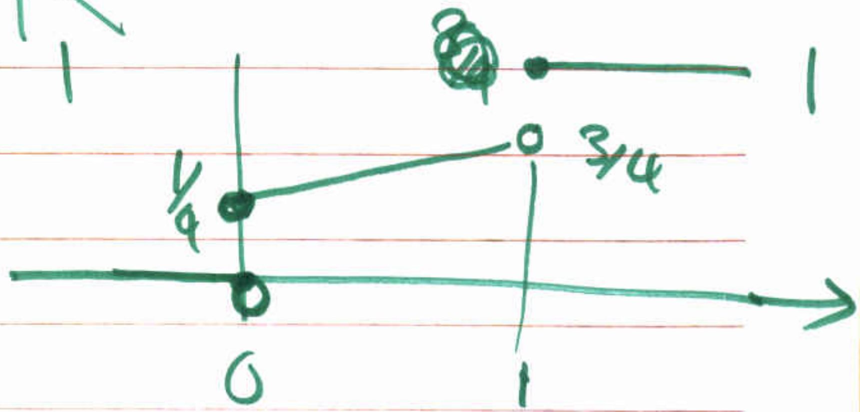
$$\mathbb{P}(g(X) \le y) = \mathbb{P}(X \le x_1) + \mathbb{P}(X \ge x_2) = \mathbb{P}(X \le y) + \mathbb{P}(X \ge 3-y) = F_X(y) + 1 - F_X(3-y)$$

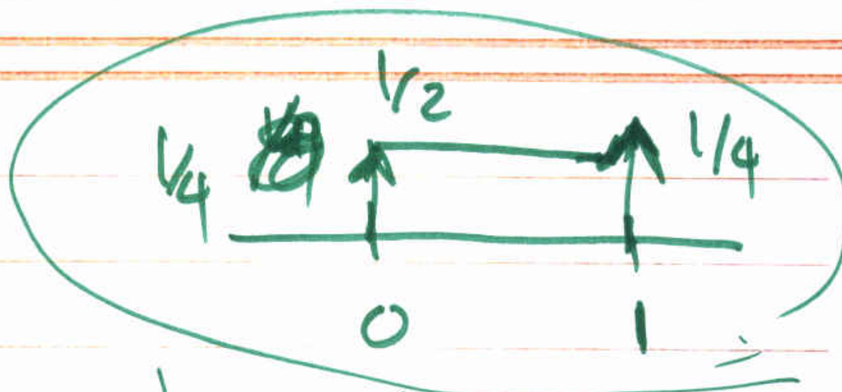
$$= \frac{1}{4} y + 1 - \frac{3-y}{4} = \frac{y + 4 - 3 + y}{4}$$

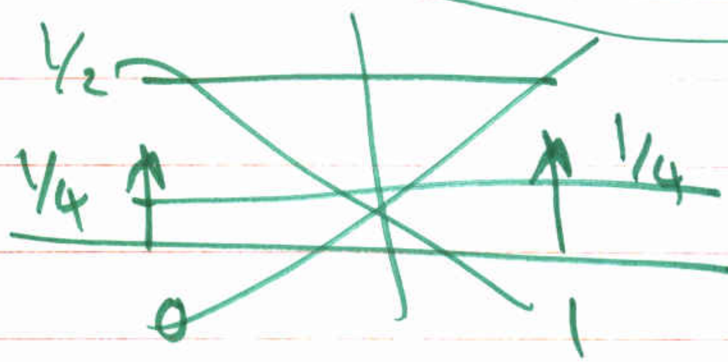
$$= \frac{2y+1}{4}$$




3/4
1/4
0
1


1
3/4
1/4
0
1


1/2
1/4
1/4
0
1


1/2
1/4
1/4
0
1

Example: Let $Y = g(X)$, where X is a continuous r.v. with known pdf. Assume $g(x) = x^2$. Find the pdf and cdf of $Y = g(X)$.

$$F_Y(y) \triangleq \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y)$$

$$= \begin{cases} 0 & y < 0 \\ \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) & y > 0 \end{cases}$$

$y \geq 0 \quad F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$y > 0 \quad f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$

$$= \frac{1}{2\sqrt{y}} \left(f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right)$$

Exercise: Assume that X is a continuous r.v. and $X \geq 0$ with probability 1. Also assume that $Y = \exp(X^4)$. Find the pdf and cdf of Y.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(e^{X^4} \leq y)$$

$$= \begin{cases} 0 & y < 1 \\ \mathbb{P}(e^{X^4} \leq y) & y \geq 1 \end{cases}$$

$y \geq 1$ $\quad F_Y(y) = \mathbb{P}(e^{X^4} \leq y)$

$$= \mathbb{P}(X^4 \leq \ln y)$$

10a

$$= \mathbb{P}(0 \le X^4 \le \ln y)$$

$$= \mathbb{P}(0 \le X \le \sqrt[4]{\ln y})$$ $\to$ $\mathbb{P}_y : [0, \sqrt[4]{\ln y})$

$$= F_X(\sqrt[4]{\ln y}) - F_X(0)$$

$$\cancel{f_Y(y) = \frac{1}{4\sqrt[4]{(\ln y)^3}}}$$

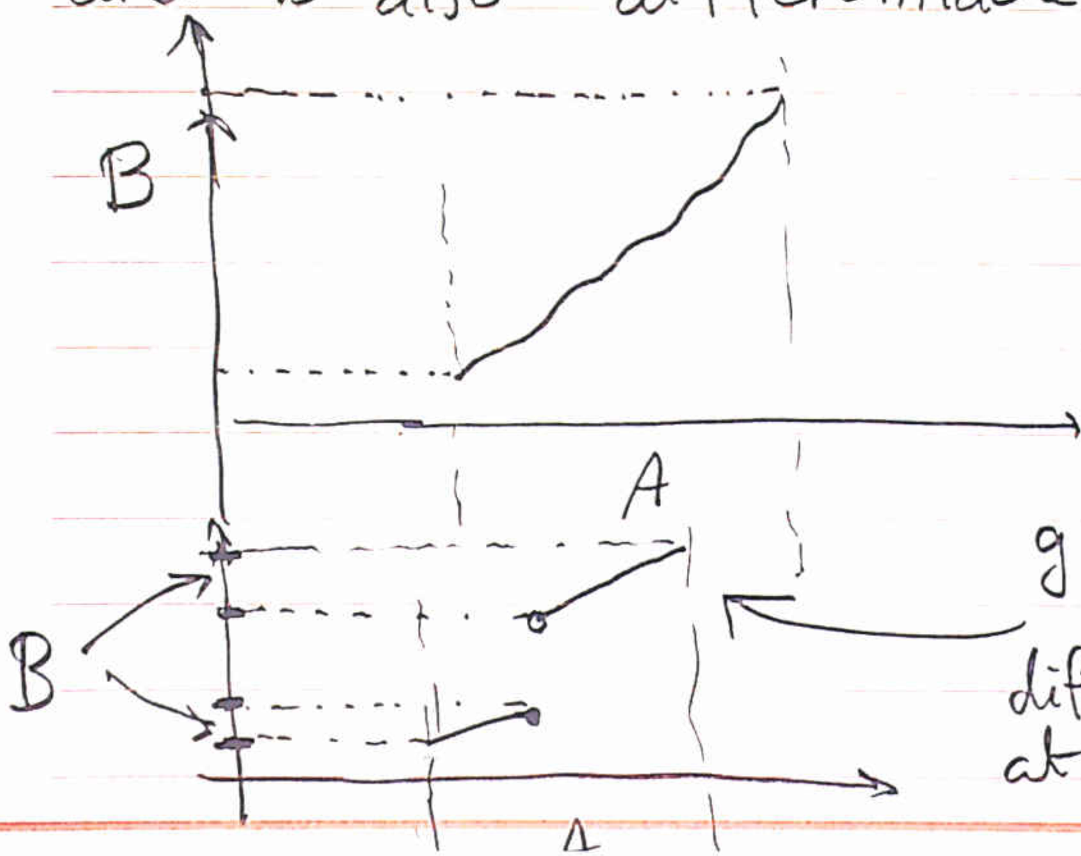
$$\frac{d}{dy}(\ln y)^{1/4} = \frac{1}{4}(\ln y)^{-3/4} \, \frac{1}{y}$$

$$f_Y(y) = \frac{1}{4y(\ln y)^{3/4}} f_X(\sqrt[4]{\ln y})$$

Monotonic Functions

Assume that the range of $\vec{X}$ contains an open interval, A, i.e. $A \subseteq \vec{X}(\Omega)$. Also assume that g is strictly monotonic on A

and is also differentiable on A



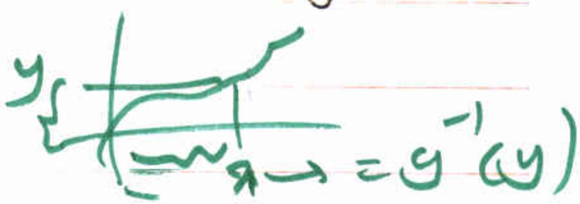
g is non differentiable at one point

Let B be the set of all values of $g(x)$, when $x \in A$, i.e. $B = \vec{g}(A)$. First let's assume g is increasing. Since g is monotonic, it has an inverse function. Let g^{-1} be the inverse function of g.

Then $g(g^{-1}(y)) = y$

$\forall y \in B$

$\forall y \in B$,



$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$$
$$= \mathbb{P}(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

Then

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \frac{d}{dy} F_X(g^{-1}(y))$$

Using the chain rule:

$$f_Y(y) = \frac{dg^{-1}(y)}{dy} f_X(g^{-1}(y))$$

/

Next, one needs to calculate

$\frac{d g^{-1}(y)}{dy}$. Because $g(g^{-1}(y)) = y$,

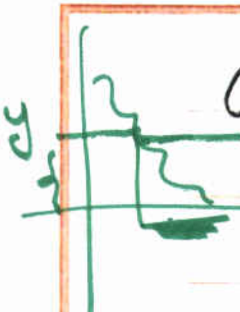
Using the chain rule:

$$g(g^{-1}(y)) = y \Rightarrow \frac{dg^{-1}(y)}{dy} g'(g^{-1}(y)) = 1$$

$$\Rightarrow \frac{dg^{-1}(y)}{dy} = \frac{1}{g'(g^{-1}(y))}$$

substitution of the expression for $\frac{dg^{-1}(y)}{dy}$ in the formula

$$\rightarrow f_Y(y) = \frac{1}{g'(g^{-1}(y))} f_X(g^{-1}(y))$$



When g is monotonic and decreasing

$$F_Y(y) = \mathbb{P}(g(X) \le y)$$

$$= \mathbb{P}(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

therefore

$$\rightarrow f_Y(y) = \frac{-1}{g'(g^{-1}(y))} f_X(g^{-1}(y))$$

Combining these two cases,

one has

$$f_Y(y) = f_X(g^{-1}(y)) \frac{1}{|g'(g^{-1}(y))|}$$

$y \in B$

Linear Functions

Consider the special case

$a \neq 0$, $g(x) = ax + b$. $\underline{Y} = aX + b$

Then: $g(x) = ax + b \quad g'(x) = a$

$$y = ax + b \Rightarrow \frac{y-b}{a} = x = g^{-1}(y)$$

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y))$$

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

Conclusion: A linear function shifts and scales the pdf of a r.v.

Example: A linear function of a normal r.v.:

Assume that $X \sim N(\mu, \sigma^2)$.

Find the pdf of $Y = aX + b, a \neq 0$

$$f_Y(y) = \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma|a|}\exp\left[\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right]$$

$$Y \sim N(a\mu+b,\ a^2\sigma^2)$$

If $\mu = 0$, $\sigma = 1$ $\Rightarrow Y \sim N(b, a^2)$

Example: Assume that X is distributed according to a standard Cauchy. What is the pdf of $Y = g(X) = X^3 + 1$.

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

$$g(x) = x^3 + 1 \qquad g'(x) = 3x^2$$

$$y = x^3 + 1 \Rightarrow y - 1 = x^3 \Rightarrow x = g^{-1}(y) = \sqrt[3]{y-1}$$

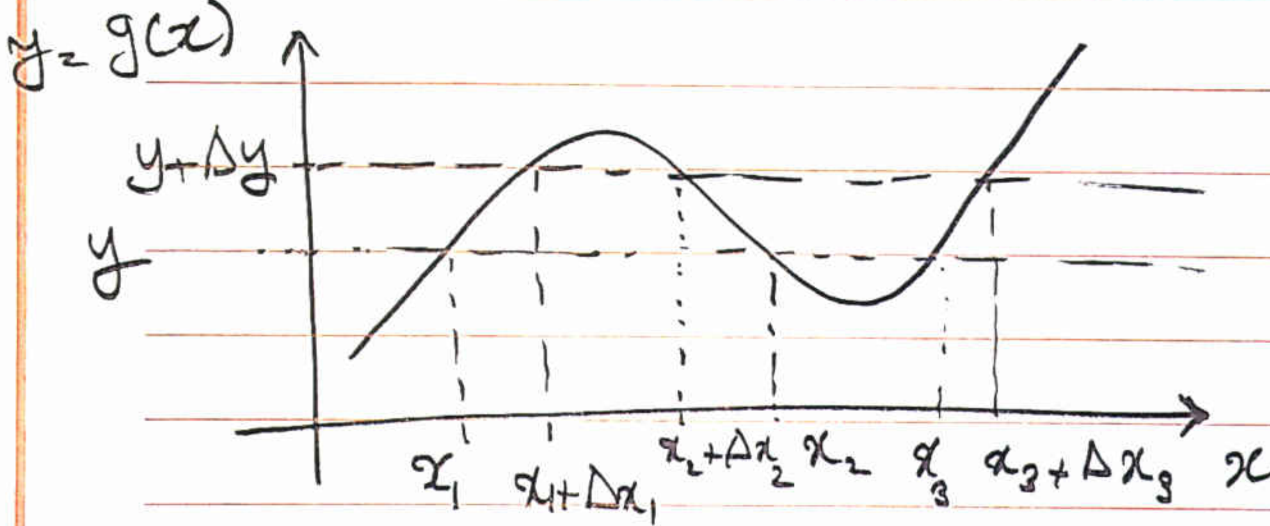
$$f_Y(y) = \frac{1}{\pi}\frac{1}{1+\left(\sqrt[3]{y-1}\right)^2} \times \frac{1}{\left|3\left(\sqrt[3]{y-1}\right)^2\right|}$$

$$f_Y(y) = \frac{1}{\pi}\frac{1}{1+\sqrt[3]{(y-1)^2}} \times \frac{1}{3\sqrt[3]{(y-1)^2}}$$

Non-monotonic Functions

Usually, we deal with non-monotonic functions of a r.v., that ~~are~~ may not even be continuous.

Consider the following function:



Assume that one wishes to determine the pdf of $Y=g(X)$

Observe that if

$$y \leq Y \leq y + \Delta y$$

then for X

$$x_1 \leq X \leq x_1 + \Delta x_1 \text{ or } x_2 + \Delta x_2 \leq X \leq x_2$$

$$\text{or } x_3 \leq X \leq x_3 + \Delta x_3$$

where x_1, x_2, x_3 are roots of the equation $y = g(x)$

In general, for a function $g(x)$ that is locally differentiable and piece-wise monotonic if the event $\{y \le Y \le y + \Delta y\}$ occurs,

then

$X \in [x_i, x_i + \Delta x_i]$ if $g(x)$ is increasing around x_i

and $X \in [x_i + \Delta x_i, x_i]$ negative

if $g(x)$ is decreasing around x_i.

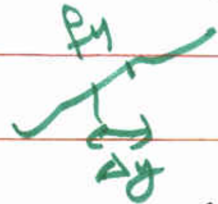
Therefore

$$P(y \le Y \le y + \Delta y) = \underbrace{\sum_{i} P(x_i \le X \le x_i + \Delta x_i)}_{\text{increasing parts}}$$

$$+ \underbrace{\sum_{j} P(x_j + \Delta x_j \le X \le x_j)}_{\text{decreasing parts}}$$

where the sum is broken

according to the regions in which $g(x)$ is decreasing or increasing. Using pdfs and cdfs



$$f_Y(y)\,|\Delta y| = \sum_k f_X(x_k)\,|\Delta x_k|$$

$$\Rightarrow f_Y(y) = \sum_k \frac{f_X(x)}{\left|\frac{\Delta y}{\Delta x}\right|}\Bigg|_{x=x_k}$$

$\Delta x \to 0$

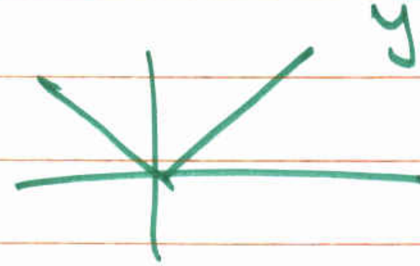
$$\boxed{f_Y(y) = \sum_k \frac{f_X(x)}{|g'(x)|}\Bigg|_{x=x_k}}$$

$g(x)=y$ $\quad |x|=y \Rightarrow \begin{matrix} x=y \\ x=-y \end{matrix}$

Example: Assume that $X \sim \text{Laplace}(1)$ and $Y=|X|$. Find the pdf of Y.

$$f_X(x)=\frac{1}{2}e^{-|x|}$$

$g(x)=|x|$



$$g'(x)=\begin{cases} -1 & x<0 \quad (y=-x) \\ 1 & x>0 \quad (y=x) \end{cases}$$

$$f_Y(y)=\frac{f_X(y)}{|1|}+\frac{f_X(-y)}{|-1|}$$

$$=f_X(y)+f_X(-y) \qquad y>0$$

$$=\frac{1}{2}e^{-|y|}+\frac{1}{2}e^{-|-y|}=e^{-y} \qquad y>0$$

Monday 3/26/18

Multivariable Functions

(Functions of Random Vectors)

Suppose $\underline{X} = [X_1, \dots, X_n]^T$ is

a ~~ra~~ random vector, and that X_i's

are jointly continuous with

joint pdf $f_{\underline{X}}(\underline{x}) = f_{\underline{X}}(x_1, \dots, x_n)$

Consider a function $g: \mathbb{R}^n \to \mathbb{R}^n$

and the random vector $\underline{Y} = g(\underline{X})$

$\underline{Y} = [Y_1, Y_2, \dots, Y_n]$.

Let g_i be the i^{th} component of

g, so that $Y_i = g_i(\underline{X}) = g_i(X_1, \ldots, X_n)$.

Example: Let $g: \mathbb{R}^2 \to \mathbb{R}^2$

$$g(x_1, x_2) = g\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} g_1(x_1, x_2) \\ \\ g_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_1^2 + x_2^2 \\ x_1 x_2 \end{bmatrix}$$

$y_1 = g_1(x_1, x_2) = x_1^2 + x_2^2$

$y_2 = g_2(x_1, x_2) = x_1 x_2$

We are interested in finding the joint pdf of $Y_1, \ldots, Y_n$ based on the joint pdf of $X_1 \ldots X_n$.

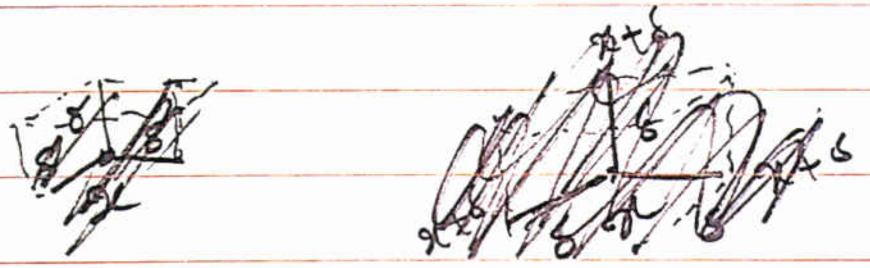
Linear Functions

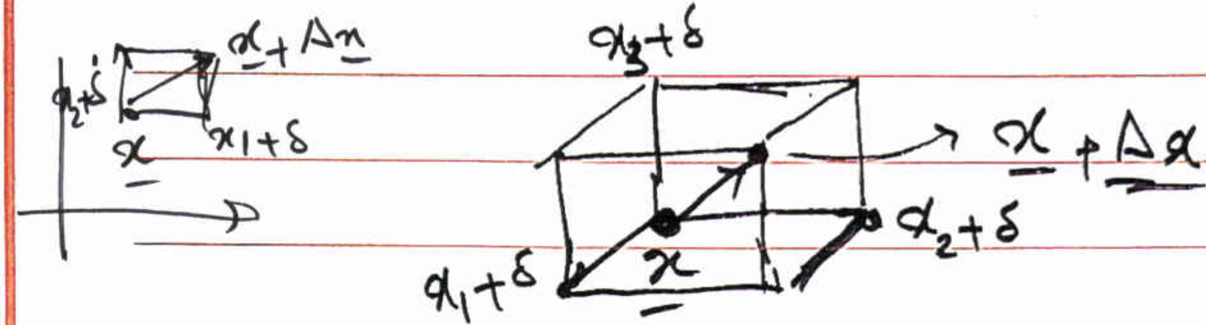
Assume that $G: \mathbb{R}^n \to \mathbb{R}^n$

$$g(\underline{x}) = G\underline{x}, \text{ where } G \in \mathbb{R}^{n\times n}$$

Assume that $A \subseteq \mathbb{R}^n$ is an open subset of $\mathbb{R}^n$. Fix

$\underline{x} \in A$ and $\Delta\underline{x} = [\delta, \dots, \delta]^T$, where δ is a small and $\underline{x} + \Delta\underline{x}$ represents a hypercube



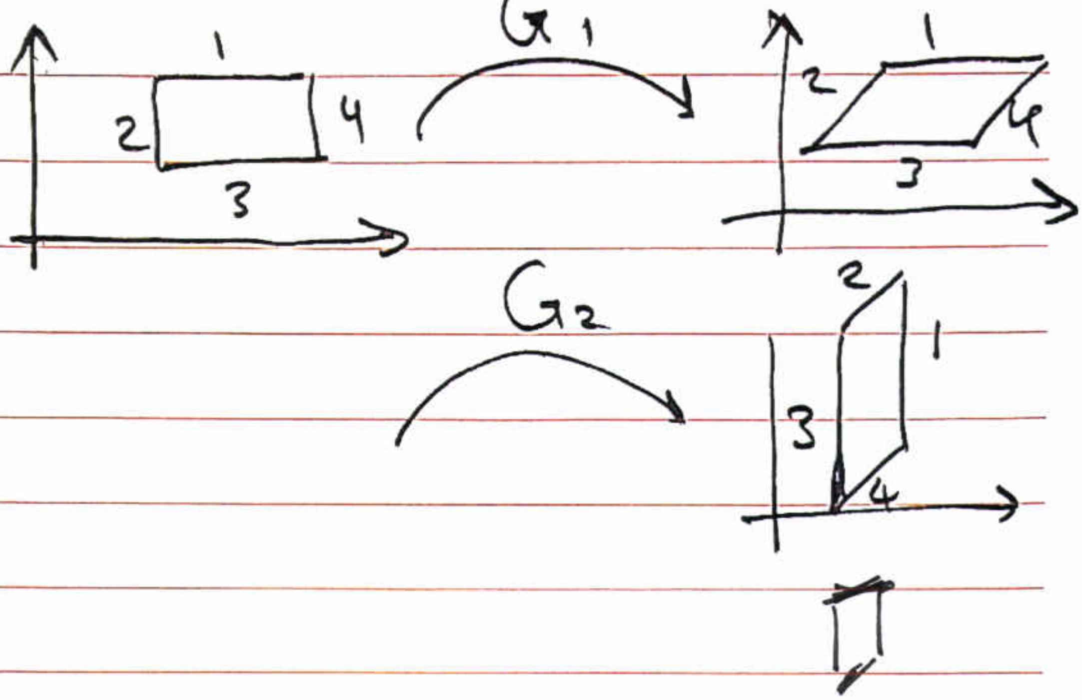


Assume that $\underline{\Delta x}$ is small enough to keep the hyper cube inside A. Let's

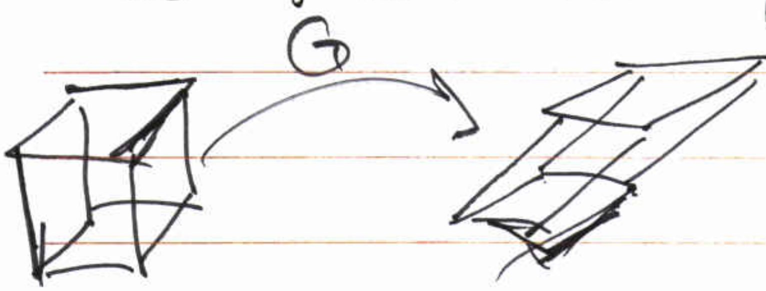
call the ~~pt~~ set of points inside the hyper cube, C.

A linear mapping "preserves" parallel lines, but not necessarily angles:

Ex:



The image of C under G is "parallelpiped," i.e. all of its faces are parallelograms.



the image of C

Let's call $\vec{g}(C)$, P.

$$\mathbb{P}(\underline{X} \in C) = \mathbb{P}(g(\underline{X}) \in \vec{g}(C))$$

$$= \mathbb{P}(g(\underline{X}) \in P)$$

On the other hand C and

$P = \vec{g}(C)$ are Borel subsets of

$\mathbb{R}^n$, therefore

$$\mathbb{P}(\underline{X} \in C) = \int_C f_{X_1, X_2, \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Δx

$$\approx f_{X_1, \ldots, X_n}(x_1, \ldots x_n)\, \delta^n$$

and

$$\mathbb{P}(g(\underline{X}) \in P) = \mathbb{P}(\underline{Y} \in P)$$

$$= \int_P f_{Y_1,\dots,Y_n}(y_1,\dots,y_n)\, dy_1\, dy_2 \dots dy_n$$

$$\simeq f_{\underline{Y}}(\underline{y}) \cdot Vol(P)$$

From analytic geometry, we know that the Volume of P is

$$Vol(P) = |\det(G)|\, Vol(C)$$

$$= |\det(G)|\, \delta^n$$

because

$$\mathbb{P}(\underline{X} \in C) = \mathbb{P}(\vec{g}(\underline{X}) \in \vec{g}(C))$$

$$= \mathbb{P}(Y \in P)$$

$$f_{\underline{X}}(\underline{x}) \cancel{\delta^n} = f_{\underline{Y}}(\underline{y}) \, |\det(G)| \cancel{\delta^n}$$

$$\Rightarrow f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{y}) \, |\det(G)|$$

$$\Rightarrow f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}(\underline{x})}{|\det(G)|}$$

$\boxed{\text{when } \det(G) \neq 0.}$

when $\det(G) \neq 0$, G is invertible,

so $\underline{y} = G\underline{x} \Rightarrow \underline{x} = G^{-1}\underline{y}$

$$\Rightarrow f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}(G^{-1}\underline{y})}{|\det(G)|}$$

$$= f_{X}(G^{-1}\underline{y})\,|\det(G^{-1})|$$

What happens if G is singular?

If G is singular, the linear transformation $\underline{y} = G\underline{x}$ cannot be onto.

(Reason: Fundamental Theorem of Linear Algebra

$$\underbrace{\text{Rank}}_{\leq n} + \underbrace{\text{Nullity}}_{>0} = n$$

when nullity >0, a subspace of $\mathbb{R}^n$ is mapped into $\underline{0}$,

so the linear transformation cannot cover all $\mathbb{R}^n$.

The image of $\mathbb{R}^n$ will be a "pure" subspace of $\mathbb{R}^n$.

Example: Assume that $G = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

and $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ X_1, X_2 jointly continuous

$$\underline{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 + X_2 \\ X_1 + X_2 \end{bmatrix}$$

$$= \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

Obviously $Y_1 = Y_2$, therefore

Y_1 and Y_2 are <u>not</u>

jointly continuous

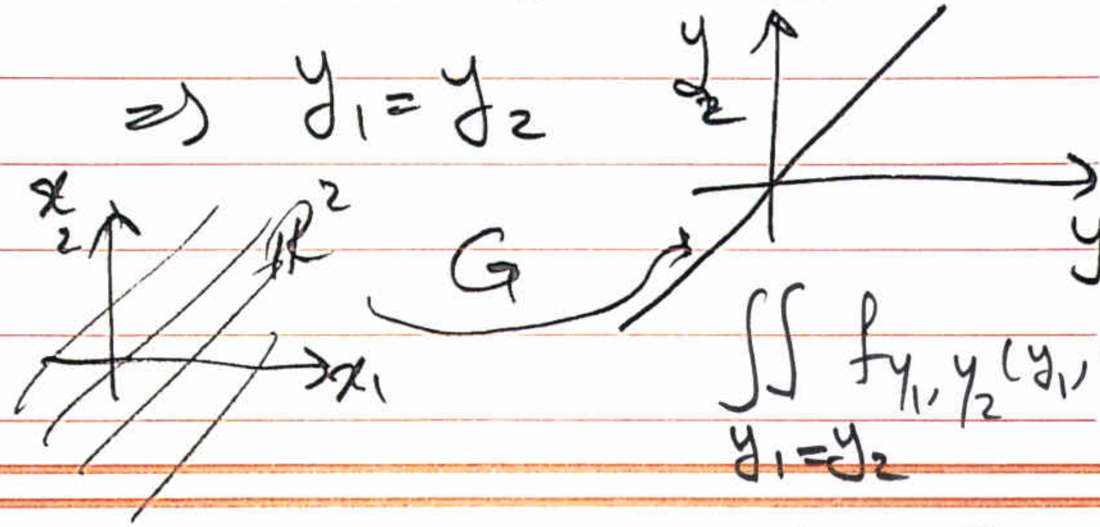
G maps $\mathbb{R}^2$ on a line:

$$y_1 = y_2$$

$$\underline{x} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \quad \underline{y} = G\underline{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 + x_2 \\ x_1 + x_2 \end{bmatrix}$$

$$\Rightarrow y_1 = y_2$$

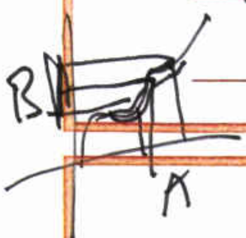


$$\iint_{y_1 = y_2} f_{y_1, y_2}(y_1, y_2)\, dy_1\, dy_2 = 1$$

In general, a singular G maps $\mathbb{R}^n$ to a subspace of $\mathbb{R}^n$ with a ~~lower~~ dimension lower than n

Monotonic Functions

Let us generalize the previous discussion to a case where $\underline{g}$ is continuously differentiable at x ~~and~~ on some open



set $A \subset \mathbb{R}^n$. Let $B = \vec{g}(A)$ be the image of A under g and also assume that g is invertible on B, i.e. g^{-1} is a well-defined function, i.e.

There exists only one

& $\forall \underline{y} \in B, \exists! \, \underline{x} \in \mathbb{R}^n$ s.t.

$$g(\underline{x}) = \underline{y}$$

Since we chose δ very small, the image of the hypercube C is approximately parallelpiped,

since in a small neighborhood of $\underline{x}$, $g(\underline{x})$ can be approximated by a linear function using Taylor Series:

$$g(\underline{x}+\underline{\Delta x}) \simeq \overset{g(\underline{x})+}{G(\underline{x})}\, \underline{\Delta x} \quad \forall \underline{x} \in C$$

matrix

where $G(\underline{x})$ is the Jacobian Matrix of g at $\underline{x}$

$$G(\underline{x}) = \left(\frac{\partial g}{\partial \underline{x}}\right)(\underline{x})$$

$$G_{ij} = \frac{\partial g_i(\underline{x})}{\partial x_j}$$

Again, the volume of $P = \vec{g}(C)$ can be approximated as

$$Vol(P) \approx |\det(G(\underline{x}))| \, \delta^n$$

Therefore:

$$f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}(\underline{g}^{-1}(\underline{y}))}{\left|\det\left(G(\underbrace{\underline{g}^{-1}(\underline{y})}_{\underline{x}})\right)\right|}$$

Recall that:

$$g(g^{-1}(\underline{y})) = \underline{y} \qquad \frac{\partial \underline{y}}{\partial y_1}$$

$$\frac{\partial}{\partial \underline{y}}\left(g(g^{-1}(\underline{y}))\right) = \frac{\partial \underline{y}}{\partial \underline{y}} = I_{n\times n}$$

$$\Rightarrow \frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} \underbrace{g'(g^{-1}(\underline{y}))}_{G(g^{-1}(y))} = I_{n\times n}$$

$$\Rightarrow \left|\det \frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}}\right| \left|\det G(g^{-1}(\underline{y}))\right| = 1$$

$$\Rightarrow \left| \det \frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} \right| = \frac{1}{\left| \det\left(G\left(g^{-1}(\underline{y})\right)\right) \right|}$$

Denote $\dfrac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} = H(\underline{y})$

then

$$f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}\left(g^{-1}(\underline{y})\right)}{\left| \det\left(G\left(g^{-1}(\underline{y})\right)\right) \right|}$$

$$= f_{\underline{X}}(g^{-1}(\underline{y})) \left| \det\left(H(\underline{y})\right) \right|$$

Example: Suppose X_1 and X_2 have the joint pdf

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the joint pdf of $Y_1 = \frac{X_1}{X_2}$ and $Y_2 = 2X_2$

$$y_1 = \frac{x_1}{x_2} \Rightarrow \quad x_1 = g_1^{-1}(\underline{y}) = x_2 y_1 = \frac{y_1 y_2}{2}$$

$$y_2 = 2x_2 \Rightarrow \quad x_2 = g_2^{-1}(\underline{y}) = \frac{y_2}{2}$$

$$H(\underline{y}) = \frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} = \begin{bmatrix} \frac{y_2}{2} & \frac{y_1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$$

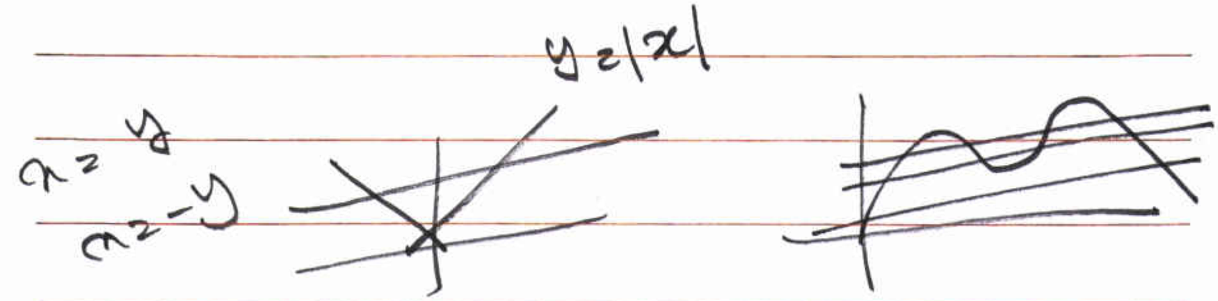
$$\Rightarrow |\cancel{\det} H(\underline{y})| = \left|\frac{y_2}{4}\right|$$

$$f_{Y_1,Y_2} = f_{X_1,X_2}\left(g^{-1}(\underline{y})\right) \left| \det\left(H(\underline{y})\right) \right|$$

$$= \begin{cases} 2\left|\dfrac{y_2}{4}\right| & 0 < \dfrac{y_1 y_2}{2} < \dfrac{y_2}{2} < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{y_2}{2} & 0 < y_1 < 1 \\ & 0 < y_2 < 2 \\ 0 & \text{otherwise} \end{cases}$$

Exercise: X_1 and X_2 are jointly continuous with joint pdf $f_{X_1,X_2}(x_1,x_2)$. Find the joint pdf of $Y_1 = \frac{X_1}{X_1+X_2}$ and $Y_2 = 1/X_2$.



Non-~~monotonic~~ invertible functions

Similar to single-variable case

$$f_{\underline{Y}}(\underline{y}) = \sum_{k} \frac{f_{\underline{X}}(\underline{x})}{\left|\det(G(\underline{x}))\right|} \Bigg|_{\underline{x}=\underline{x}_k}$$

where $\underline{x}_k$ is the root of $g(\underline{x}) = \underline{y}$
and $G(\underline{x}) = \partial g(\underline{x}) / \partial \underline{x}$

A Single Function of Multiple Random Variables

Suppose that $\underline{X} = [X_1, \ldots, X_n]^T$ is a vector of jointly continuous random variables with joint pdf $f_{\underline{X}}(\underline{x})$.

Consider a function $g_1 : \mathbb{R}^n \to \mathbb{R}$

We are interested in the pdf of $Y_1 = g_1(\underline{X})$. For example

$$g(x_1, x_2, x_3) = \frac{x_1 x_2}{x_3}$$

Basically, one can use the definition of cdf and pdf to calculate $f_{Y_1}(y)$:

$$F_{Y_1}(y) = \mathbb{P}\left(g(\underline{X}) \leq y\right)$$

$$= \int_{\{\underline{x} \mid g(\underline{x}) \leq y\}} f_{\underline{X}}(\underline{x})\, d\underline{x}$$

where $d\underline{x} = dx_1\, dx_2 \cdots dx_n$

and

$$f_{Y_1}(y) = \frac{dF_{Y_1}(y)}{dy}$$

Another interesting way of calculating $f_{Y_1}(y)$ is to use formulas for multivariable functions.

Let's assume $g_2, \dots, g_n : \mathbb{R}^n \to \mathbb{R}$ are functions and $Y_i = g_i(\underline{X})$ $i \in \{1, \dots, n\}$

If the function

$$g(\underline{x}) = (g_1(\underline{x}), g_2(\underline{x}), \dots, g_n(\underline{x}))$$

is invertible, we can find the joint pdf of Y_i's using previously discussed methods. Then, the marginal pdf of Y_1 is easily found by integrating $\underline{y_2}, \cdots, \overline{y_n}$ out, from the joint pdf.

A very simple choice for g_2 to g_n is the ~~identity~~ following ~~function~~

$$g(\underline{x}) = (g_1(\underline{x}), x_2, \cdots, x_n)$$

i.e. $y_i = g_i(\underline{x}) = x_i \quad \forall i \geq 2.$

~~let~~ Since $g^{-1}(\underline{y})$ is assumed ~~to~~ exist

$$g^{-1}(\underline{y}) = \begin{bmatrix} h_1(\underline{y}) \\ h_2(\underline{y}) \\ \vdots \\ h_n(\underline{y}) \end{bmatrix}$$

but $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} g_1(\underline{x}) \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

$$\Rightarrow \quad y_2 = x_2 \qquad \Rightarrow \quad x_2 = h_2(y) = y_2$$

$$y_3 = x_3 \qquad x_3 = h_3(y) = y_3$$

$$\vdots \qquad \vdots$$

$$y_n = x_n \qquad x_n = h_n(y) = y_n$$

$$\Rightarrow g^{-1}(\underline{y}) = \begin{bmatrix} h_1(\underline{y}) \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} = H(\underline{y})$$

$$= \begin{bmatrix} \frac{\partial h_1(\underline{y})}{\partial y_1} & \frac{\partial h_1(\underline{y})}{\partial y_2} & \cdots & \frac{\partial h_1(\underline{y})}{\partial y_n} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & & & 1 \end{bmatrix}$$

$$\Rightarrow \left|\det(H(\underline{y}))\right| = \left|\frac{\partial h_1(\underline{y})}{\partial y_1}\right|$$

$$f_{Y_1, Y_2, \cdots, Y_n}(y_1, \cdots y_n) = f_{X_1 \cdots X_n}(h_1(y), y_2, \cdots y_n) \times \left|\frac{\partial h_1(\underline{y})}{\partial y_1}\right|$$

and

$$f_{Y_1}(y_1) = \int f_{\underline{X}}\left(h_1(\underline{y}), y_2, \ldots, y_n\right) \left|\frac{\partial h_1}{\partial y_1}\right| dy_2\, dy_3 \cdots dy_n$$

Example: Let X_1 and X_2 be positive jointly continuous r.v.'s and suppose we wish to ~~define~~ derive the pdf of

$$Y_1 = g(X_1, X_2) = X_1 X_2$$

Introduce $Y_2 = X_2$, and

derive $f_{Y_1}(y_1)$. In particular

solve the problem when

$X_1 \sim U(0,1)$ and $X_2 \sim U(0,1)$.
and are independent.
Also, solve the problem directly

using multiple integrals

$$g(\underline{y}) = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 x_2 \\ x_2 \end{bmatrix}$$

$$g^{-1}(\underline{y}) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad x_2 = y_2$$

$$x_1 = \frac{y_1}{x_2} = \frac{y_1}{y_2}$$

$$H(\underline{y}) = \frac{\partial g^{-1}(\underline{y})}{\partial \underline{y}} = \begin{bmatrix} 1/y_2 & -y_1/y_2^2 \\ 0 & 1 \end{bmatrix}$$

$$|\det H(y)| = \left|\frac{1}{y_2}\right|$$

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}\left(\frac{y_1}{y_2}, y_2\right)\left|\frac{1}{y_2}\right|$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_{X_1,X_2}\left(\frac{y_1}{y_2}, y_2\right)\left|\frac{1}{y_2}\right| dy_2$$

In particular, for i.i.d uniforms

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 1 & 0 < x_1 < 1 \\ & 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \left|\frac{1}{y_2}\right| & 0 < \frac{y_1}{y_2} < 1 \\ & 0 < y_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 1/y_2 & 0<y_1<y_2<1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y_1}(y_1) = \int_{y_1}^{1} \frac{1}{y_2}\, dy_2 \qquad 0<y_1<1$$

$$= \log_e y_2 \Big|_{y_1}^{1} = -\ln y_1 \qquad 0<y_1<1$$

Second Solution

$$\iint_{x_1 x_2 < y} f_{X_1,X_2}(x_1,x_2)\, dx_1\, dx_2$$

Order Statistics

An important set of many to one functions $g: \mathbb{R}^n \to \mathbb{R}$ are order statistics. Assume that $X_1, X_2, \dots, X_n$ are i.i.d and jointly

Continuous. Let's call an ordered version of X_i's, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

Clearly, $X_{(1)} = \min_i X_i$ and

$X_{(n)} = \max_j X_j$.

First, let's calculate the pdf of

$Y = \min_i (X_i)$ and $Z = \max_j (X_j)$.

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(\min(X_1, \dots, X_n) \le y)$$

$$= 1 - \mathbb{P}(\min(X_1, \dots, X_n) > y)$$

$$= 1 - \mathbb{P}(X_1 > y, X_2 > y, \dots, X_n > y)$$

$$= 1 - \mathbb{P}(X_1 > y)\,\mathbb{P}(X_2 > y) \cdots \mathbb{P}(X_n > y)$$

$$= 1 - [1 - \mathbb{P}(X_1 \le y)][1 - \mathbb{P}(X_2 \le y)] \cdots [1 - \mathbb{P}(X_n \le y)]$$

$$= 1 - [1 - F_X(y)]^n$$

$$f_Y(y) = (-1)\,n\,(-1)\,f_X(y)\,[1 - F_X(y)]^{n-1}$$

$$= n\,f_X(y)\,[1 - F_X(y)]^{n-1}$$

$$F_Z(z) = \mathbb{P}(Z \le z) \qquad Z = \max(X_1 \ldots X_n)$$

$$= \mathbb{P}(\max(X_1, \ldots, X_n) \le z)$$

$$= \mathbb{P}(X_1 \le z, X_2 \le z, \ldots, X_n \le z)$$

$$= \mathbb{P}(X_1 \le z)\,\mathbb{P}(X_2 \le z) \cdots \mathbb{P}(X_n \le z)$$

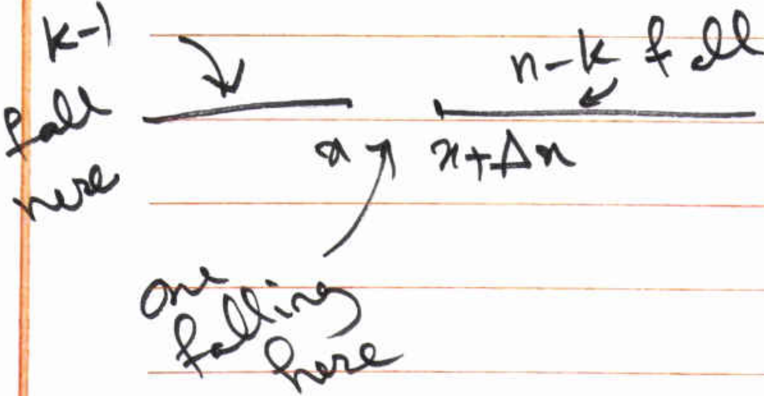
$$= [F_X(z)]^n$$

$$f_Z(z) = n f_X(z) \left[F_X(z)\right]^{n-1}$$

Now, let's calculate the pdf of $X_{(k)}$. Δx is so small that only $X_{(k)}$ falls in $[x, x+\Delta x]$

$$\mathbb{P}\left(x \leq X_{(k)} \leq x+\Delta x\right) = f_{X_{(k)}} \Delta x$$

for small Δx



$$= \binom{n}{k-1,1,n-k} \mathbb{P}\big(k-1 \text{ r.v.'s fall in } (-\infty, x]\big)$$

$$\mathbb{P}(x \leq X \leq x+\Delta x)\, \mathbb{P}\big(n-k \text{ fall in } (x,+\infty)\big)$$

$$= \frac{n!}{(k-1)!\,(n-k)!} \big[F_X(x)\big]^{k-1} f_X(x)\Delta x \big[1-F_X(x)\big]^{n-K}$$

Therefore $f_{X_{(k)}}(x)\cancel{\Delta x} = \dfrac{n!}{(k-1)!\,(n-k)!} \big[F_X(x)\big]^{k-1} f_X(x)\cancel{\Delta x} \big[1-F_X(x)\big]^{n-k}$

$$f_{X_{(k)}} = \frac{n!}{(k-1)!\,(n-k)!} \big[F_X(x)\big]^{k-1} f_X(x) \big[1-F_X(x)\big]^{n-K}$$

for $k=1$, $k=n$, see that we get the pdf of minimum and maximum.

Exercise: Show that the joint pdf of $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ is in general

$$f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n) = \begin{cases} n!\, f(x_1, x_2, \dots, x_n) & x_1 < x_2 < \dots < x_n \\ 0 & \text{otherwise} \end{cases}$$

and if $X_1 \dots X_n$ are i.i.d

$$f_{X_{(1)},\dots,X_n}(x_1,\dots,x_n) = \begin{cases} n! \prod_{i=1}^{n} f_X(x_i) & x_1 < x_2 < \dots < x_n \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{P}(x_1 \leq X_{(1)} \leq x_1 + \Delta x, x_2 \leq X_{(2)} \leq x_2 + \Delta x,$

$\ldots, x_n \leq X_{(n)} \leq x_n + \Delta x)$

is nonzero, only when

$x_1 < x_2 < \ldots < x_n$

(Assuming that Δx is so small that each interval ~~only~~ contains only one of $X_{(i)}$'s

$$= n!\,\mathbb{P}(x_1 \leq X_1 \leq x_1 + \Delta x, \ldots, x_2 \leq X_2 \leq x_2 + \Delta x, \ldots, x_n \leq X_n \leq x_n + \Delta x)$$

$$LHS = f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n)\cancel{(\Delta x)^n}$$

$$= n!\, f_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n)\cancel{(\Delta x)^n}$$

$$x_1 < x_2 < \dots < x_n$$

If X_i's are independent and identically distributed:

$$f_{X_{(1)}\dots X_{(n)}}(x_1,\dots,x_n) = n! \prod_{i=1}^{n} f_X(x_i)$$

$$x_1 < x_2 < \dots < x_n$$

Simulation Exercise:

rand in MATLAB gives you $U(0,1)$

let's assume $n = 10$

1000 times

U_1

U_2

$\vdots$

U_n

a thousand instances of the vector

$(U_1 —— U_n)$

Order   Order

One thousand instances of $U_{(1)} \cdots — U_{(n)}$

$U_{(1)}$

$U_{(2)}$

$\vdots$

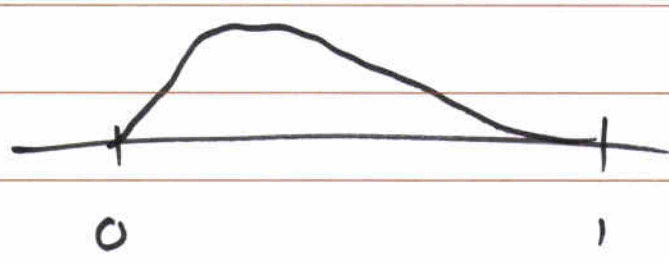
$U_{(n)}$

Draw the pdf of $U_{(1)}$ to $U_{(n)}$ by their one thousand instences.

A rough approximation: histogram or use Kernel Density Estimation (KDE)

kde

You will see $U_{(1)}$ — $U_{(n)}$ from Beta distribution



Sum of Independent Random Variables

Previously, we observed that if X and Y are independent discrete r.v.'s and $Z = X + Y$,

$$\boxed{P_Z = P_X * P_Y}$$

For continuous r.v.'s: $Z = X + Y$

$$F_Z(z) = P(X+Y \leq z) = \iint_{(x,y)\,|\,x+y \leq z} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y) \, dy \, dx$$

$$t = x + y$$
$$dy = dt$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} f_{X,Y}(x, t-x) \, dt \, dx$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \mathbb{P}(X+Y \leq z)$$

$$= \frac{d}{dz} \int_{-\infty}^{+\infty} \int_{-\infty}^{z} f_{X,Y}(x, t-x) \, dt \, dx$$

$$= \int_{-\infty}^{+\infty} \frac{d}{dz} \int_{-\infty}^{z} f_{X,Y}(x, t-x) \, dt \, dx$$

$$= \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x)\,dx$$

when X, Y are independent

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx$$

$$f_Z = f_X * f_Y$$

Exercise: Assume that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent.

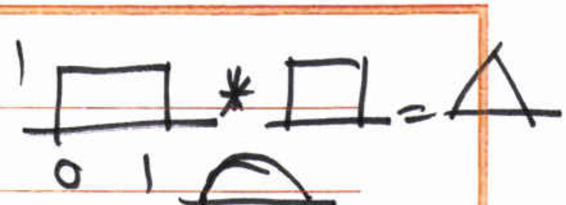
Show that $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

By induction

If $X_1, X_2, \dots, X_n$ are independent r.v.'s and $Z = X_1 + X_2 + \dots + X_n$

$$f_Z = f_{X_1} * f_{X_2} * \dots * f_{X_n}$$

Computer Exercise:



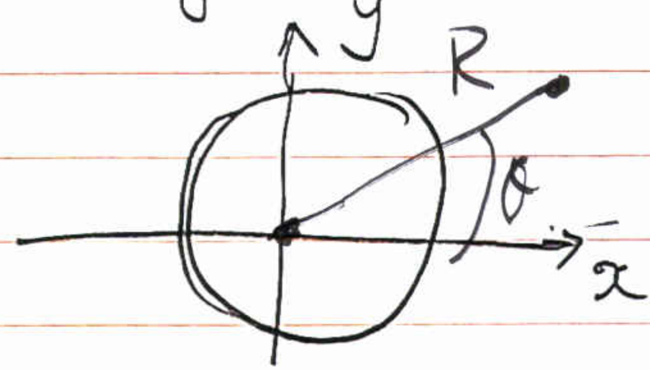
Assume that $X_1, \dots, X_n$ are i.i.d $U(0,1)$ r.v.'s. Assume that $n=5$, $n=10$, $n=30$, $n=100$.

Plot the pdf of $Z = X_1 + \dots + X_n$.

Also, repeat the exercise by sampling from $X_1, \ldots, X_n$ and creating a distribution for $Z = X_1 + \cdots + X_n$, using histograms or Kernel Density Estimation

Some Derived Distributions

Assume that a shooter shoots at a target.



Assume that the X coordinate of the shot is a $N(0,1)$ r.v. and also the Y coordinate of the shot is a $N(0,1)$ r.v. and X, Y are independent.

What are the distributions of R and θ, the distance of the shot from the origin and its angle with the x axis?

Let's define $(r, \theta) = g(x, y)$

We need the joint distribution of $(R, \Theta) = g(X, Y)$, given that $X, Y \sim N(0,1)$ and are independent.

Let's find $(x,y) = \bar{g}^{-1}(r,\theta)$:

$x = g_1^{-1}(r,\theta) = r\cos\theta$

$y = g_2^{-1}(r,\theta) = r\sin\theta$

$$H(r,\theta) = \frac{\partial \bar{g}^{-1}}{\partial(r,\theta)} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

$$\left|\det(H(r,\theta))\right| = \left|r\cos^2\theta + r\sin^2\theta\right|$$

$= |r| = r$ because $r > 0$

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$= \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}}$$

$$f_{R,\Theta}(r,\theta) = f_{X,Y}\left(g_1^{-1}(r,\theta), g_2^{-1}(r,\theta)\right)\left|\det H(r,\theta)\right|$$

$$= \frac{1}{2\pi} e^{-\frac{(r^2\cos^2\theta + r^2\sin^2\theta)}{2}} r$$

$$= \frac{1}{2\pi} r e^{-r^2/2} \qquad r \in [0, +\infty) \quad \theta \in [0, 2\pi]$$

$$f_R(r) = \int f_{R,\Theta}(r,\theta)\, d\theta =$$

$$\int_0^{2\pi} \frac{1}{2\pi} r e^{-r^2/2}\, d\theta = r e^{-r^2/2} \qquad r \geq 0$$

R is said to have a

standard Rayleigh Distribution

$f_\Theta(\theta) = \int f_{R,\Theta}(r,\theta)\, dr$

$= \int_0^\infty \frac{1}{2\pi} r e^{-r^2/2}\, dr = \frac{1}{2\pi}\left(-e^{-r^2/2}\right)_0^\infty$

$= 1/2\pi \qquad \theta \in [0, 2\pi]$

Θ has a Uniform distribution

$\Theta \sim \text{Uniform}[0, 2\pi]$

Are R, Θ independent?

$f_{R,\Theta}(r,\theta) = f_R(r)\, f_\Theta(\theta)$

$\Rightarrow R, \Theta$ are indep.

Question: What is the distribution of $Z = g(R) = R^2$, when R is a Rayleigh distribution.

In other words, what is the distribution of $R^2 = X^2 + Y^2$

if X and Y are independent standard normal r.v.'s.

$Z = g(r) = r^2 \Rightarrow r = g^{-1}(z) = \sqrt{z} \qquad g'(r) = 2r$

$$f_Z(z) = \frac{f_R(g^{-1}(z))}{g'(g^{-1}(z))} = \frac{\sqrt{z}\, e^{-\frac{(\sqrt{z})^2}{2}}}{2\sqrt{z}}$$

$$= \frac{e^{-z/2}}{2} \qquad Z \sim Exp(1/2)$$

Exercise: Show that if

$U \sim U(0,1)$, then ~~the~~ if

$X = \frac{1}{\lambda} \ln 1/U$, then $X \sim Exp(\lambda)$

$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(1/\lambda \ln 1/U \leq x)$

$= \mathbb{P}(\ln 1/U \leq \lambda x) = \mathbb{P}(-\ln U \leq \lambda x)$

$= \mathbb{P}(\ln U \geq -\lambda x)$

$= \mathbb{P}(U \geq e^{-\lambda x}) = 1 - \mathbb{P}(U \leq e^{-\lambda x})$

$= 1 - F_U(e^{-\lambda x}) = 1 - e^{-\lambda x}$

$\Rightarrow X \sim Exp(\lambda)$

Exercise: In general, if X is a r.v. with strictly increasing cdf F_X, ~~then~~ and $U \sim U(0,1)$, then $Y = F_X^{-1}(U)$ has the same distribution as X.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X^{-1}(U) \leq y)$$

$$= \mathbb{P}(U \leq F_X(y)) = F_U(F_X(y))$$

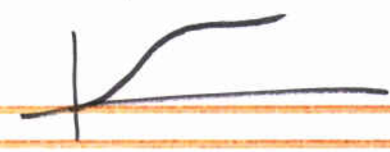
$$= F_X(y)$$

Simulation of Normal Random Variables

The previous discussion on derived distributions is very useful in simulating r.v.'s, i.e. generating samples from different distributions.

The previous exercise shows that any continuous r.v. with a strictly increasing cdf can be simulated using a uniform r.v., which can easily be generated by a simple random number generator.

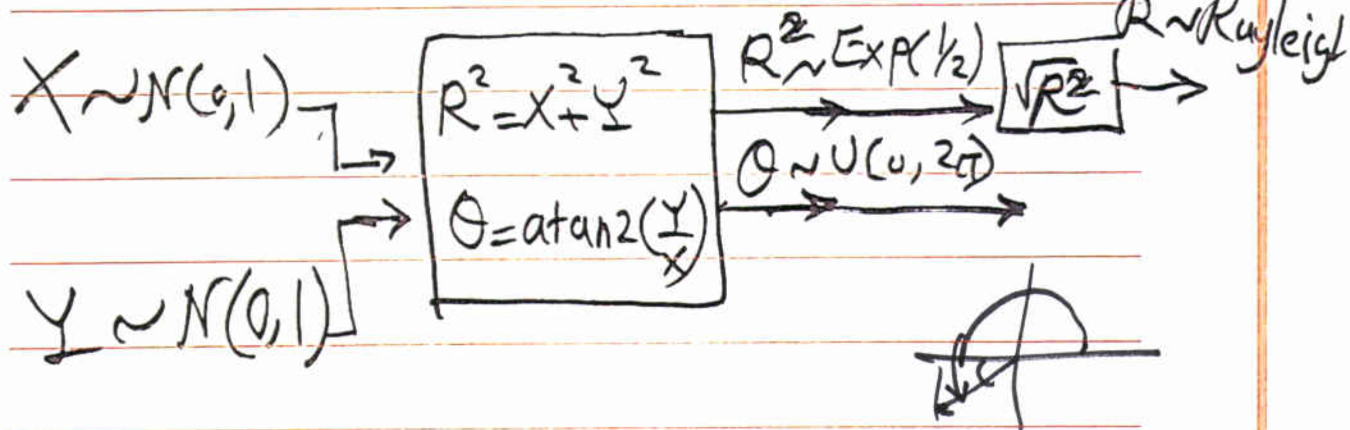
The CDF of a standard normal



r.v. is strictly increasing, but doesn't have ~~an~~ ~~analytic~~ closed form.

Usually, the so-called Box-Muller Method is used to generate ~~a~~ independent standard normal random variables.

The following diagram summarizes our discussion about the relationship between Rayleigh and Normal Distributions.



Also we learned that

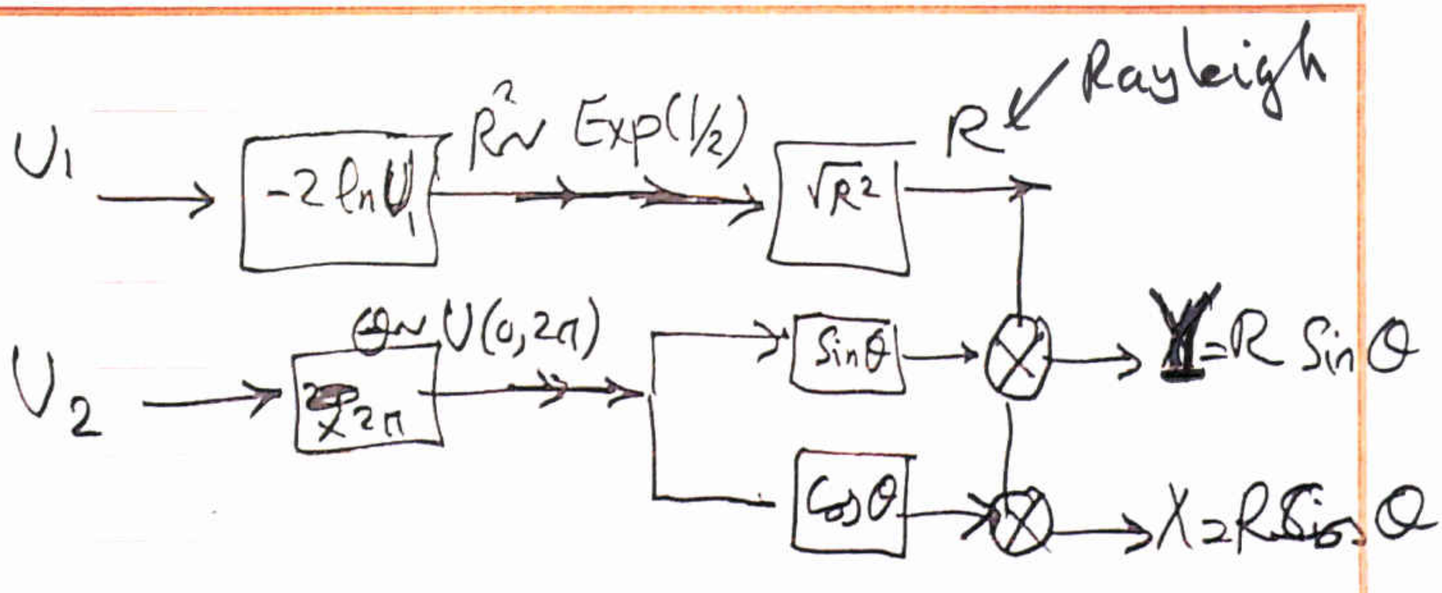
$$U(0,1) \longrightarrow \boxed{F_X(U)} \longrightarrow X \sim Exp(1/2)$$

$$-2\ln(U)$$

$$-\frac{1}{\lambda}\ln U = -2\ln U$$

The Box-Muller Method does the above process in reverse:

Assume that U_1, U_2 are independent $U(0,1)$ r.v.'s



X, Y are independent $\mathcal{N}(0,1)$ r.v.'s. (so, they are jointly normal)

Exercise: Write down the formulas for X and Y from the Box-Muller Method

$$X = \sqrt{-2\ln U_1}\, \sin(2\pi U_2)$$

$$Y = \sqrt{-2\ln U_1}\, \cos(2\pi U_2)$$

Directly show that X, Y are independent $\mathcal{N}(0,1)$ random variables.

Exercise: Let X and Y be independent and $X \sim$ Standard Rayleigh and $Y \sim N(0,1)$. Find the joint density of $U = \sqrt{X^2 + Y^2}$ and $V = \lambda Y / X$, where $\lambda > 0$, $\lambda \in \mathbb{R}$.

Are U, V independent?