Recap: Random Variable

- Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable is a real-valued function on Ω.
- It essentially results in an induced probability space

$$(\Omega, \mathcal{F}, P) \stackrel{X}{\rightarrow} (\Re, \mathcal{B}, P_X)$$

where $\mathcal{B}$ is the Borel σ-algebra and

$$P_X(B) = P[X \in B] = P\left(\{\omega \in \Omega \ : \ X(\omega) \in B\}\right)$$

- For X to be a random variable

$$\{\omega \in \Omega \ : \ X(\omega) \in B\} \ \in \mathcal{F}, \ \ \forall B \in \mathcal{B}$$

Recap: Distribution Function

- Let X be a random variable. It distribution function, $F_X : \Re \to \Re$, is defined by

$$F_X(x) = P[X \leq x] = P(\{\omega \in \Omega \; : \; X(\omega) \leq x\})$$

- The distribution function, F_X, completely specifies the probability measure, P_X.
- The distribution function satisfies
  1. $0 \leq F_X(x) \leq 1, \; \forall x$
  2. $F_X(-\infty) = 0; \; F_X(\infty) = 1$
  3. F_X is non-decreasing: $x_1 \leq x_2 \; \Rightarrow \; F_X(x_1) \leq F_X(x_2)$
  4. F_X is right continuous and has left-hand limits.
- We also have
  $F_X(x^+) - F_X(x^-) = F_X(x) - F_X(x^-) = P[X = x]$
  $P[a < X \leq b] = F_X(b) - F_X(a)$.

Recap: Discrete Random Variable

- A random variable X is said to be discrete if it takes only finitely many or countably infinitely many distinct values.
- Let $X \in \{x_1, x_2, \cdots\}$
- Its distribution function, F_X is a stair-case function with jump discontinuities at each x_i and the magnitude of the jump at x_i is equal to $P[X = x_i]$

Recap: probability mass function

- Let $X \in \{x_1, x_2, \cdots\}$.
- The probability mass function (pmf) of X is defined by

$$f_X(x_i) = P[X = x_i]; \quad f_X(x) = 0, \quad \text{for all other } x$$

- It satisfies
  1. $f_X(x) \geq 0, \forall x$ and $f_X(x) = 0$ if $x \neq x_i$ for some i
  2. $\sum_i f_X(x_i) = 1$
- We have
  $F_X(x) = \sum_{i: x_i \leq x} f_X(x_i)$
  $f_X(x) = F_X(x) - F_X(x^-)$
- We can calculate the probability of any event as

$$P[X \in B] = \sum_{\substack{i: \\ x_i \in B}} f_X(x_i)$$

Recap: continuous random variable

- X is said to be a continuous random variable if there exists a function $f_X : \Re \rightarrow \Re$ satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx$$

  The f_X is called the probability density function.
- Same as saying F_X is absolutely continuous.
- Since F_X is continuous here, we have

$$P[X = x] = F_X(x) - F_X(x^-) = 0, \ \forall x$$

- A continuous rv takes uncountably many distinct values. However, not every rv that takes uncountably many values is a continuous rv

Recap: probability density function

- The pdf of a continuous rv is defined to be the f_X that satisfies
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \ \ \forall x$$
- It satisfies
  1. $f_X(x) \geq 0, \ \ \forall x$
  2. $\int_{-\infty}^{\infty} f_X(t)\, dt = 1$
- We can, in principle, compute probability of any event as
$$P[X \in B] = \int_B f_X(t)\, dt, \ \ \forall B \in \mathcal{B}$$
- In particular,
$$P[a \leq X \leq b] = \int_a^b f_X(t)\, dt$$

Recap: some discrete random variables

- Bernoulli: $X \in \{0,1\}$; parameter: $p,\ 0 < p < 1$

$$f_X(1) = p;\ f_X(0) = 1 - p$$

- Binomial: $X \in \{0, 1, \cdots, n\}$; Parameters: n, p

$$f_X(x) = {}^nC_x p^x (1-p)^{n-x},\ x = 0, \cdots, n$$

- Poisson: $X \in \{0, 1, \cdots\}$; Parameter: $\lambda > 0$.

$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!},\ x = 0, 1, \cdots$$

- Geometric: $X \in \{1, 2, \cdots\}$; Parameter: $p,\ 0 < p < 1$.

$$f_X(x) = p(1-p)^{x-1},\ x = 1, 2, \cdots$$

Recap: Some continuous random variables

- Uniform over $[a,\ b]$: Parameters: $a, b,\ b > a$.

$$f_X(x) = \frac{1}{b-a},\ a \leq x \leq b.$$

- exponential: Parameter: $\lambda > 0$.

$$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0.$$

- Gaussian (Normal): Parameters: $\sigma > 0, \mu$.

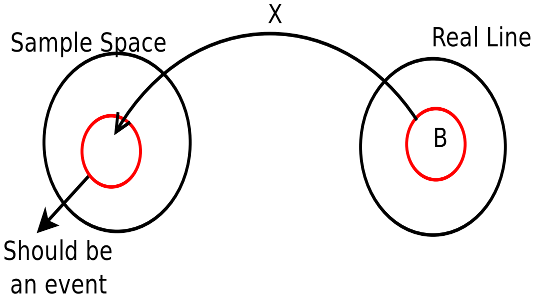
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty < x < \infty$$

Functions of a random variable

- We next look at random variables defined in terms of other random variables.

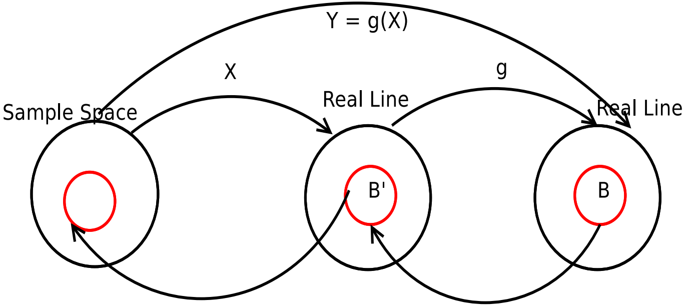
- Let X be a rv on some probability space $(\Omega, \mathcal{F}, P)$.
- Recall that $X : \Omega \to \Re$.
- Also recall that

$$[X \in B] \triangleq \{\omega \ : \ X(\omega) \in B\} \in \mathcal{F}, \ \forall B \in \mathcal{B}$$



Functions of a Random Variable

- Let X be a rv on some probability space $(\Omega, \mathcal{F}, P)$. (Recall $X : \Omega \to \Re$)
- Consider a function $g : \Re \to \Re$
- Let $Y = g(X)$. Then Y also maps Ω into real line.
- If g is a 'nice' function, Y would also be a random variable
- We need: $g^{-1}(B) \triangleq \{z \in \Re \ : \ g(z) \in B\} \in \mathcal{B}, \ \forall B \in \mathcal{B}$ (Note abuse of notation)



- Let X be a rv and let $Y = g(X)$.
- The distribution function of Y is given by

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[g(X) \leq y] \\ &= P[g(X) \in (-\infty,\ y]\ ] \\ &= P[X \in \{z\ :\ g(z) \leq y\}] \end{aligned}$$

- This probability can be obtained from distribution of X.
- Thus, in principle, we can find the distribution of Y if we know that of X

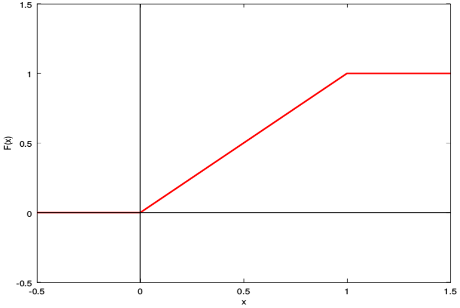
Example

- Let $Y = aX + b, \; a > 0$.
- Then we have

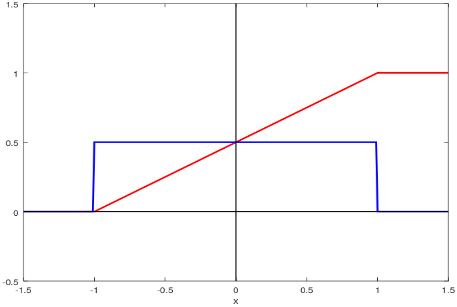
$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] \\
&= P[aX + b \leq y] \\
&= P[aX \leq y - b] \\
&= P\left[X \leq \frac{y-b}{a}\right], \quad \text{since } a > 0 \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

- This tells us how to find df of Y when it is an affine function of X.
- If X is continuous rv, then, $f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$

- In many examples we would be using uniform random variables.
- Let $X \sim U[0,\ 1]$. Its pdf is $f_X(x) = 1,\ 0 \leq x \leq 1$.
- Integrating this we get the df: $F_X(x) = x,\ 0 \leq x \leq 1$



- Let $X \sim U[-1,\ 1]$. The pdf would be $f_X(x) = 0.5,\ -1 \leq x \leq 1$.
- Integrating this, we get the df: $F_X(x) = \frac{1+x}{2}$ for $-1 \leq x \leq 1$.
- These are plotted below



- Suppose $X \sim U[0,\ 1]$ and $Y = aX + b$
- The df for Y would be

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) = \begin{cases} 0 & \frac{y-b}{a} \leq 0 \\ \frac{y-b}{a} & 0 \leq \frac{y-b}{a} \leq 1 \\ 1 & \frac{y-b}{a} \geq 1 \end{cases}$$

- Thus we get the df for Y as

$$F_Y(y) = \begin{cases} 0 & y \leq b \\ \frac{y-b}{a} & b \leq y \leq a+b \\ 1 & y \geq a+b \end{cases}$$

- Hence $f_Y(y) = \frac{1}{a},\ y \in [b,\ a+b]$ and $Y \sim U[b,\ a+b]$.
- If $X \sim U[0,\ 1]$ then $Y = aX + b$, $(a > 0)$, is uniform over $[b,\ a+b]$.

- Recall that Gaussian density is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- We denote this as $\mathcal{N}(\mu, \sigma^2)$
- Let $Y = aX + b$ where $X \sim \mathcal{N}(0, 1)$. The df of Y is

$$
\begin{aligned}
F_Y(y) &= F_X\left(\frac{y-b}{a}\right) \\
&= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \; dx
\end{aligned}
$$

we make a substitution: $t = ax + b \Rightarrow x = \dfrac{t-b}{a}$, and $dx = \dfrac{1}{a} dt$

$$
F_Y(y) = \int_{-\infty}^{y} \frac{1}{a\sqrt{2\pi}}\, e^{-\frac{(t-b)^2}{2a^2}} \; dt
$$

- This shows that $Y \sim \mathcal{N}(b, a^2)$

- Suppose X is a discrete rv with $X \in \{x_1, x_2, \cdots\}$.
- Suppose $Y = g(X)$.
- Then Y is also discrete and $Y \in \{g(x_1), g(x_2), \cdots\}$.
- Though we use this notation, we should note:
  1. these values may not be distinct (it is possible that $g(x_i) = g(x_j)$);
  2. $g(x_1)$ may not be the smallest value of Y and so on.
- We can find the pmf of Y as

$$
\begin{aligned}
f_Y(y) &= p[Y = y] = P[g(X) = y] \\
&= P[X \in \{x_i \;:\; g(x_i) = y\}] \\
&= \sum_{\substack{i: \\ g(x_i)=y}} f_X(x_i)
\end{aligned}
$$

- Let $X \in \{1, 2, \cdots, N\}$ with $f_X(k) = \frac{1}{N},\ 1 \leq k \leq N$
- Let $Y = aX + b$, $(a > 0)$.
- Then $Y \in \{b + a,\ b + 2a, \cdots,\ b + Na\}$.
- We get the pmf of Y as

$$f_Y(b + ka) = f_X(k) = \frac{1}{N},\ 1 \leq k \leq N$$

- Suppose X is geometric:
  $f_X(k) = (1-p)^{k-1}p, \ k = 1, 2, \cdots.$
- Let $Y = X - 1$
- We get the pmf of Y as

$$\begin{aligned} f_Y(j) &= P[X - 1 = j] \\ &= P[X = j + 1] \\ &= (1-p)^j p, \ j = 0, 1, \cdots \end{aligned}$$

- Suppose X is geometric. $(f_X(k) = (1-p)^{k-1}p)$
- Let $Y = \max(X, 5) \;\; \Rightarrow \;\; Y \in \{5, 6, \cdots\}$
- We can calculate the pmf of Y as

$$f_Y(5) = P[\max(X,5) = 5] \;\; = \;\; \sum_{k=1}^{5} f_X(k) = 1 - (1-p)^5$$

$$f_Y(k) = P[\max(X,5) = k] \;\; = \;\; P[X = k] = (1-p)^{k-1}p, \;\; k = 6, 7, \cdots$$

- We next consider $Y = h(X)$ where

$$h(x) = \begin{cases} x & \text{if } \; x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- This is written as $Y = X^+$ to indicate the function only keeps the positive part.

- Let $X \sim U[-1,\ 1]$: $F_X(x) = \frac{1+x}{2}$ for $-1 \leq x \leq 1$.
- Let $Y = X^+$. That is,

$$Y = X^+ = \begin{cases} X & \text{if } X > 0 \\ 0 & \text{otherwise} \end{cases}$$

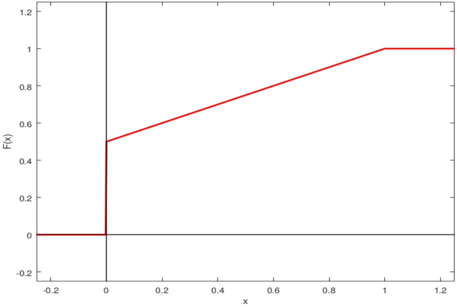
- For $y < 0$, $F_Y(y) = P[Y \leq y] = 0$ because $Y \geq 0$.
- $F_Y(0) = P[Y \leq 0] = P[X \leq 0] = F_X(0) = 0.5$.
- For $0 < y < 1$,
  $F_Y(y) = P[Y \leq y] = P[X \leq y] = F_X(y) = \frac{1+y}{2}$
- For $y \geq 1$, $F_Y(y) = 1$.
- Thus, the df of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 0.5 & \text{if } y = 0 \\ \frac{1+y}{2} & \text{if } 0 < y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

- The df of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1+y}{2} & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

- This is plotted below



- This is neither a continuous rv nor a discrete rv.

- Let $Y = X^2$.
- For $y < 0$, $F_Y(y) = P[Y \leq y] = 0$ (since $Y \geq 0$)
- For $y \geq 0$, we can get $F_Y(y)$ as

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

- If X is a continuous random variable, then we get

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\ (F_X(\sqrt{y}) - F_X(-\sqrt{y})) \\ &= \frac{1}{2\sqrt{y}}\,[f_X(\sqrt{y}) + f_X(-\sqrt{y})] \end{aligned}$$

- This is the general formula for density of X^2 when X is continuous rv.

- Let $X \sim \mathcal{N}(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$
- Let $Y = X^2$. Then we know $f_Y(y) = 0$ for $y < 0$. For $y \geq 0$,

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right] \\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}\, e^{-\frac{y}{2}} + \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y}{2}}\right] \\
&= \frac{1}{2\sqrt{y}}\,\frac{2}{\sqrt{2\pi}}\, e^{-\frac{y}{2}} \\
&= \frac{1}{\sqrt{\pi}}\, \left(\frac{1}{2}\right)^{0.5}\, y^{-0.5}\, e^{-\frac{1}{2}y}
\end{aligned}$$

- This is an example of gamma density.

Gamma density

- The Gamma function is given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\, e^{-x}\, dx$$

  It can be easily verified that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.
- The Gamma density is given by

$$f(x) = \frac{1}{\Gamma(\alpha)}\, \lambda^\alpha x^{\alpha-1} e^{-\lambda x} = \frac{1}{\Gamma(\alpha)}\, (\lambda x)^{\alpha-1}\, \lambda e^{-\lambda x}, \;\; x > 0$$

- Here $\alpha, \lambda > 0$ are parameters.
- The earlier density we saw corresponds to $\alpha = \lambda = 0.5$:

$$f_Y(y) = \frac{1}{\sqrt{\pi}}\, \left(\frac{1}{2}\right)^{0.5}\, y^{-0.5}\, e^{-\frac{1}{2}y}, \;\; y > 0$$

- The gamma density with parameters $\alpha, \lambda > 0$ is given by

$$f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}, \;\; x > 0$$

- If $X \sim \mathcal{N}(0,1)$ then X^2 has gamma density with parameters $\alpha = \lambda = 0.5$.
- When α is a positive integer then the gamma density is known as the Erlang density.
- If $\alpha = 1$, gamma density becomes exponential density.

- Let $X \sim U(0,\ 1)$.
- Let $Y = \frac{-1}{\lambda} \ln(1 - X)$, where $\lambda > 0$.
- Note that $Y \geq 0$. We can find its df:

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] = P\left[\frac{-1}{\lambda} \ln(1 - X) \ \leq \ y\right] \\
&= P[-\ln(1 - X) \ \leq \ \lambda y] \\
&= P[\ln(1 - X) \ \geq \ -\lambda y] \\
&= P[1 - X \ \geq \ e^{-\lambda y}] \\
&= P[X \ \leq \ 1 - e^{-\lambda y}] \\
&= 1 - e^{-\lambda y}, \ y \geq 0 \quad \text{(since } X \sim U(0, 1))
\end{aligned}
$$

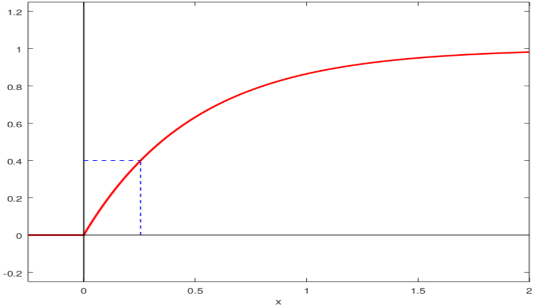
- Thus Y has exponential density
- If $X \sim U(0,\ 1)$, $\frac{-1}{\lambda} \ln(1 - X)$ has exponential density

- If $X \sim U(0,\ 1)$, $\frac{-1}{\lambda}\ln(1-X)$ has exponential density
- This is actually a special case of a general result.
- The exponential distribution fn is $F(x) = 1 - e^{-\lambda x}$.
- This is continuous, strictly monotone and hence is invertible. The inverse function maps $[0,\ 1]$ to $\Re^+$. We derive its inverse:

$$z = 1 - e^{-\lambda x} \Rightarrow e^{-\lambda x} = 1 - z \Rightarrow x = \frac{-1}{\lambda}\ln(1-z)$$

- Thus, the inverse of F is $F^{-1}(z) = \frac{-1}{\lambda}\ln(1-z)$
- So, we had $Y = F^{-1}(X)$ and the df of Y was F

- We can visualize this as shown below



- Let G be a continuous invertible distribution function.
- Let $X \sim U[0,\ 1]$ and let $Y = G^{-1}(X)$.
- We can get the df of Y as

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

- Thus, starting with uniform rv, we can generate a rv with a desired distribution.
- Very useful in random number generation. Known as the inverse function method.
- Can be generalized to handle discrete rv also. It only involves defining an 'inverse' when F is a stair-case function. (Left as an exercise!)

- Let X be a cont rv with an invertible distribution function, say, F.
- Define $Y = F(X)$.
- Since range of F is $[0,\ 1]$, we know $0 \leq Y \leq 1$.
- For $0 \leq y \leq 1$ we can obtain $F_Y(y)$ as

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

- This means Y has uniform density.
- Has interesting applications.
  E.g., histogram equalization in image processing

- Let us sum-up the last two examples
- If $X \sim U[0,\ 1]$ and $Y = F^{-1}(X)$, then Y has df F.
- If df of X is F and $Y = F(X)$ then Y is uniform over $[0,\ 1]$.

- If $Y = g(X)$, we can compute distribution of Y, knowing the function g and the distribution of X.
- We have seen a number of examples.
- Finally, we look at a theorem that gives a formula for pdf of Y in certain special cases

- Let $g : \Re \to \Re$ be differentiable with $g'(x) > 0, \forall x$.
- Let X be a continuous rv with pdf f_X.
- Let $Y = g(X)$
- **Theorem**: With the above, Y is a continuous rv with pdf

$$f_Y(y) = f_X(g^{-1}(y)) \, \frac{d}{dy} g^{-1}(y), \; g(-\infty) \le y \le g(\infty)$$

- **Proof**: Since $g'(x) > 0$, g is strictly monotonically increasing and hence is invertible and g^{-1} would also be monotone and differentiable.
- So, range of Y is $[g(-\infty), \; g(\infty)]$.
- Now we have

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

- Since g^{-1} is differentiable, so is F_Y and we get the pdf as

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

- This completes the proof.

- Now, suppose $g'(x) < 0, \forall x$. Even then the theorem essentially holds.
- Now, g is strictly monotonically decreasing. So, we get

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

- Once again, by differentiating

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

  because g^{-1} is also monotone decreasing.
- The range of Y here is $[g(\infty), \ g(-\infty)]$
- We can combine both cases into one result.

- Let $g : \Re \rightarrow \Re$ be differentiable with $g'(x) > 0, \forall x$ or $g'(x) < 0, \forall x$.
- Let X be a continuous rv and let $Y = g(X)$.
- Then Y is a continuous rv with pdf

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

where $a = \min(g(\infty), \; g(-\infty))$ and $b = \max(g(\infty), \; g(-\infty))$

- For an example, take $g(x) = ax + b$.
- This satisfies the conditions and $g^{-1}(y) = \frac{y-b}{a}$
- Hence we get

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

- This is an example we saw earlier.
- We need to find the range of Y based on range of X.

- The function $g(x) = x^2$ does not satisfy the conditions of the theorem.
- The utility of the theorem is somewhat limited.
- However, we can extend the theorem.
- Essentially, what we need is that for a any y, the equation $g(x) = y$ would have finite solutions and the derivative of g is not zero at any of these points.
- There are multiple '$g^{-1}(y)$' and we can get density of Y by summing all the terms.

- If $Y = g(x)$ and g is monotone,

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

- Let $x_o(y)$ be the solution of $g(x) = y$; then $g^{-1}(y) = x_o(y)$.
- Also, the derivative of g^{-1} is reciprocal of the derivative of g.
- Hence, we can also write the above as

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

- However, the notation in the above may be confusing.

- We can now extend the theorem as follows.
- Suppose, for a given y, $g(x) = y$ has multiple solutions.
- Call them $x_1(y), \cdots, x_m(y)$. Assume the derivative of g is not zero at any of these points.
- Then we have

$$f_Y(y) = \sum_{k=1}^{m} f_X(x_k(y)) \left|g'(x_k(y))\right|^{-1}$$

- If $g(x) = y$ has no solution (or no solution satisfying $g'(x) \neq 0$), then at that y, $f_Y(y) = 0$.

- Consider the old example $g(x) = x^2$.
- For $y > 0$, $x^2 = y$ has two solutions: $\sqrt{y}$ and $-\sqrt{y}$.
- At both these points, the absolute value of derivative of g is $2\sqrt{y}$ which is non-zero.
- Hence we get

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

- This is same as what we derived from first principles earlier.

Expectation and Moments of a random variable

- We next consider the important notion of expectation of a random variable

Expectation of a discrete rv

- Let X be a discrete rv with $X \in \{x_1, x_2, \cdots\}$
- We define its expectation by

$$E[X] = \sum_i x_i \, f_X(x_i)$$

- Expectation is essentially a weighted average.
- To make the above finite and well defined, we can stipulate the following as condition for existence of expectation

$$\sum_i |x_i| \, f_X(x_i) < \infty$$

Expectation of a Continuous rv

- If X is a continuous random variable with pdf, f_X, we define its expectation as

$$E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

- Once again we can use the following as condition for existence of expectation

$$\int_{-\infty}^{\infty} |x| \, f_X(x) \, dx < \infty$$

- Sometimes we use the following notation to denote expectation of both kinds of rv

$$E[X] = \int_{-\infty}^{\infty} x \, dF_X(x)$$

- Though we consider only discrete or continuous rv's, expectation is defined for all random variables.

- Let us look at a couple of simple examples.
- Let $X \in \{1, 2, 3, 4, 5, 6\}$ and $f_X(k) = \frac{1}{6},\ 1 \leq k \leq 6$.

$$EX = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$$

- Let $X \sim U[0,\ 1]$

$$EX = \int_{-\infty}^{\infty} x\ f_X(x)\ dx = \int_0^1 x\ dx = 0.5$$

- When an rv takes only finitely many values or when the pdf is non-zero only on a bounded set, the expectation is always finite.

- The way we have defined existence of expectation, implies that expectation is always finite (when it exists).
- This may be needlessly restrictive in some situations. We redefine it as follows.
- Let X be a non-negative (discrete or continuous) random variable.
- We define its expectation by

$$EX = \sum_i x_i \, f_X(x_i) \quad \text{or} \quad EX = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

  depending on whether it is discrete or continuous
  (In this course we will consider only discrete or continuous rv's)
- Note that the expectation may be infinite.
- But it always exists for non-negative random variables.

- Now let X be a rv that may not be non-negative.
- We define positive and negative parts of X by

$$X^+ = \begin{cases} X & \text{if } X > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$X^- = \begin{cases} -X & \text{if } X < 0 \\ 0 & \text{otherwise} \end{cases}$$

- Note that both X^+ and X^- are non-negative. Hence their expectations exist.
  (Also, $X(\omega) = X^+(\omega) - X^-(\omega), \ \forall \omega$).
- Now we define expectation of X by

$$EX = EX^+ - EX^-, \quad \text{if at least one of them is finite}$$

  Otherwise EX does not exist.
- Now, expectation does not exist only when $EX^+ = EX^- = \infty$

- This is the formal way of defining expectation of a random variable.
- We first note that if $\sum_i |x_i| \; f_X(x_i) < \infty$ then both EX^+ and EX^- would be finite and we can simply take the expectation as $EX = \sum_i x_i \; f_X(x_i)$.
- Also note that if X takes only finitely many values, the above always holds.
- Similar comments apply for a continuous random variable.
- This is what we do in this course because we deal with only discrete and continuous rv's.
- But to get a feel for the more formal definition, we look at a couple of examples.