Recitation 1 – Basic Concepts in Probability

Problem 1:

Given are two coins, one fair and one having "head" on both of its sides. One of the coins is chosen randomly with equal probability and it is then flipped twice. Assuming we got "head" on both of the flips, what are the chances that also in the third flip we get "head"?

Solution 1:

First, let us define the events:

F – the fair coin was chosen
U – the unfair coin was chosen
A – we got "head" on the two first flips
B – we got "head" on the third flip

The requested probability is the probability of event B given that event A occurred, namely P(B|A).

However, it is easier to calculate the said probability **given that we know** which coin was chosen:

$$P(B \mid A, F) = P(\text{HHH} \mid \text{HH, and the coin is fair}) = \tfrac{1}{2}$$
$$P(B \mid A, U) = P(\text{HHH} \mid \text{HH, and the coin is unfair}) = 1$$

To get rid of the dependency, let us use the Law of Total Probability:

(*) $$P(B \mid A) = P(F \mid A)P(B \mid A, F) + P(U \mid A)P(B \mid A, U) = P(F \mid A)\cdot\tfrac{1}{2} + P(U \mid A)\cdot 1$$

The probabilities of events F, U given event A we will find using Bayes' Theorem:

(**) $$P(F \mid A) = \frac{P(A \cap F)}{P(A)} = \frac{P(A \mid F)P(F)}{P(A)}$$
$$P(U \mid A) = \frac{P(A \cap U)}{P(A)} = \frac{P(A \mid U)P(U)}{P(A)}$$

The following holds:

$$P(F \mid A) = \frac{\frac{1}{4}\cdot\frac{1}{2}}{\frac{5}{8}} = \tfrac{1}{5};\;\; P(U \mid A) = \frac{1\cdot\frac{1}{2}}{\frac{5}{8}} = \tfrac{4}{5}$$

In other words, given that we got "head" twice, there is an 80% chance that the coin is the unfair one.

If we insert this into (*), we get $P(B \mid A) = \frac{1}{5} \cdot \frac{1}{2} + \frac{4}{5} \cdot 1 = \frac{9}{10}$. Thus, if both of the first flips resulted in "head", the third flip will result in "head" with a 90% probability.

Problem 2:

Let us designate X to be a random variable which represents the total amount of rain measured in Israel throughout a given year (in millimeters). Consider the probability density function (PDF) of X:

$$f_X(x) = \frac{1}{2} \cdot \delta(x) + \frac{1}{2} \cdot e^{-x} \cdot u(x)$$

1. Draw the PDF and the cumulative distribution function (CDF).

2. What is the probability that in a given year 100ml of rain will fall?

3. What is the probability that in a given year no rain will fall?

4. What is the probability that in a given year it will rain up to 100ml?

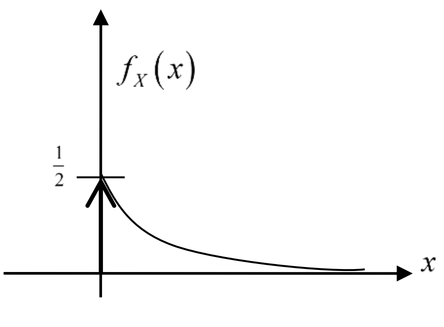
Solution 2:

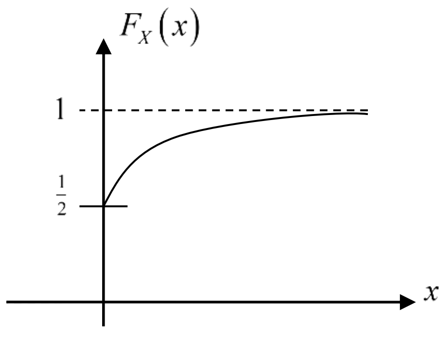
1. Let us first find the CDF:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{x} \frac{1}{2} \cdot \delta(t) + \frac{1}{2} \cdot e^{-t} \cdot u(t)dt = \int_{-\infty}^{x} \frac{1}{2} \cdot \delta(t)dt + \int_{-\infty}^{x} \frac{1}{2} \cdot e^{-t} \cdot u(t)dt =$$

$$\frac{1}{2} \cdot u(x) + \int_{0}^{x} \frac{1}{2} \cdot e^{-t} dt \cdot u(x) = \frac{1}{2} \cdot u(x) - \frac{1}{2} \cdot \left(e^{-x} - 1\right) \cdot u(x) = \left(1 - \frac{1}{2} \cdot e^{-x}\right) \cdot u(x)$$

We can now draw both functions:





2. The probability that in a given year, exactly 100ml of rain will fall is:

$$P(X = 100) = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} P(100 - \varepsilon \le X \le 100 + \varepsilon) = \int_{100^-}^{100^+} f_X(t)dt = 0$$

And this is because around $x = 100$, the PDF does not include a delta function, which also means that the CDF is a continuous function.

3. The probability that in a given year no rain will fall is given by:

$$P(X=0)=F_X(0^+)-F_X(0^-)=\frac{1}{2}$$

Another way of solving this is by noticing that at $x=0$ there is a delta function with height of 0.5 in the PDF and, therefore, the probability of getting this value is ½ .

4. The probability that in a given year it will rain up to 100ml is:

$$F_X(100)=P(X\le 100)=\int_{-\infty}^{100} f_X(t)dt=1-\frac{1}{2}\cdot e^{-100}$$

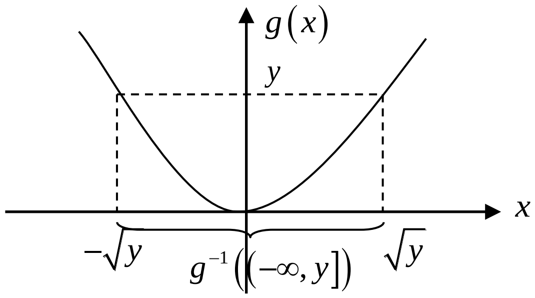
Problem 3:

Let X be a **continuous** random variable with known PDF and CDF $f_X(x)$ and $F_X(x)$, respectively. Find the PDF of $Y=g(X)=X^2$.

Solution 3:

Let us solve in two ways:

1. We will begin by finding the CDF. Let us draw the function $g(x)$:



The CDF of Y is, therefore:

$$\begin{aligned} F_Y(y) &= P(Y\le y) \\ &= \begin{cases} 0 & y<0 \\ P(-\sqrt{y}\le X\le\sqrt{y}) & y\ge 0 \end{cases} = \begin{cases} 0 & y<0 \\ F_X(\sqrt{y})-F_X(-\sqrt{y}) & y\ge 0 \end{cases} \\ &= \left[F_X(\sqrt{y})-F_X(-\sqrt{y})\right]u(y) \end{aligned}$$

Note that this function is continuous and differentiable and, thus, Y is a continuous random variable (its PDF does not include delta functions).

In order to find $f_Y(y)$, let us derive $F_Y(y)$:

$$f_Y(y) = \begin{cases} 0 & y \le 0 \\ \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right] & y > 0 \end{cases}$$

Please note: any value for $f_Y(0)$ may be chosen.

2. Recall the following proposition:

<u>Proposition 1:</u> Given $Y = g(X)$ where X is a random variable. Let us assume the equation $g(x) = y$ has a finite or countable amount of solutions, which we will represent as $\{x_i\}_{i=1}^{n}$ (where $n = \infty$ stands for a countable amount of solutions). Assuming also $\forall i \quad g'(x_i) \ne 0$, then $f_y(y)$ can be expressed as:

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}$$

Note the following:

- $f_Y(y) = 0$ if there are no solutions.
- When the number of solutions is not countable or $g'(x_i) = 0$, this proposition may not be used, rather they must be solved using the original definitions.

This proposition allows us to obtain the PDF of Y directly from the PDF of X.

In our case: $Y = g(X) = X^2$, $g'(x) = 2x$.

Let us divide the problem into three cases:

- $y < 0$: $g^{-1}(y) = \varnothing \quad \Rightarrow \quad f_Y(y) = 0$
- $y > 0$: we get two solutions $x_1 = -\sqrt{y}, x_2 = \sqrt{y}$.
$$f_Y(y) = \frac{f_X(x_1)}{|2x_1|} + \frac{f_X(x_2)}{|2x_2|} = \frac{f_X\left(\sqrt{y}\right)}{2\sqrt{y}} + \frac{f_X\left(-\sqrt{y}\right)}{2\sqrt{y}}$$
- $y = 0$: there is one solution $x_1 = 0$, for which $g'(x_1) = g'(0) = 0$ so we cannot use the above proposition. However, since X is a continuous random variable, $P(Y = 0)$ therefore, we may choose any value for $f_Y(0)$.

All in all, $f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right] & y > 0 \end{cases}$

Extra Questions

Problem 4:

Consider the following continuous random variable which has a uniform distribution:

$$X \sim U[-\pi, \pi)$$

Let us define the random variable:

$$Y = \cos(X)$$

1. Find without calculating what is the maximum value that the function $f_Y(y)$ can receive in the section (-1,1), and for which values of y it is obtained.

2. Find the PDF of Y.

Solution 4:

1. Let us define $g(x) = \cos(x)$. For $-1 < y < 1$, the PDF of Y is:

$$f_Y(y) = \sum_{x=\pm\arccos(y)} \frac{f_X(x)}{|g'(x)|}$$

Since the numerator has a constant value (X is uniformly distributed), the maximum value of the PDF is reached when the denominator is minimal (namely, when y approaches -1 or 1), which results in an infinite value (seeing as for y=1 or -1 we get that the denominator is zero).

2. For $-1 < y < 1$:

$$f_Y(y) = \sum_{x=\pm\arccos(y)} \frac{f_X(x)}{|g'(x)|} = \sum_{x=\pm\arccos(y)} \frac{1/2\pi}{|\sin(x)|} = \sum_{x=\pm\arccos(y)} \frac{1/2\pi}{\sqrt{1-\cos^2(x)}}$$

$$= \sum_{x=\pm\arccos(y)} \frac{1/2\pi}{\sqrt{1-y^2}} = \frac{1}{\pi\sqrt{1-y^2}}$$

For $y = -1 \; or \; y = 1$, the denominator equals to zero and, thus, we may not use the proposition. However, since these values are reached with a probability of 0, we can set an arbitrary value for $f_Y(-1), f_Y(1)$ (in other words, the PDF does not contain any delta functions).

For $y < -1 \; or \; y > 1$, there are no solutions for $y = g(x)$ and, thus, the PDF is zero.

Problem 5:

1. Given the continuous random variable X with known PDF and CDF, and given is the following transformation:

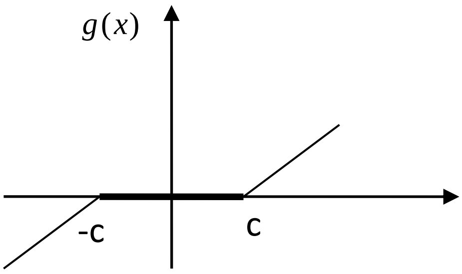
$$g(x)=\begin{cases} x-c & x>c \\ 0 & -c\le x\le c \\ x+c & x<-c \end{cases}$$

Find the PDF and CDF of Y.

2. Repeat the previous section for the transformation:

$$g(x)=sign(x)=\begin{cases} 1, & x>0 \\ -1, & x\le 0 \end{cases}$$

Solution 5:



The CDF of Y is:

$$F_Y(y)=P\{Y\le y\}=\begin{cases} P\{X\le y+c\}=F_X(y+c) & y\ge 0 \\ P\{X\le y-c\}=F_X(y-c) & y<0 \end{cases}$$

This function is continuous in every point except at $y=0$ and, therefore, the PDF of Y contains a delta function in the origin:

$$f_Y(y)=f_X(y+c)u(y)+f_X(y-c)u(-y)+(F_X(c)-F_X(-c))\delta(y)$$

Thus, the only value that can be received with a probability higher than 0 is $Y=0$:

$$P(Y=0)=F_Y(0^+)-F_Y(0^-)=F_X(c)-F_X(-c)$$

2. $Y=g(X)$ has only two solutions, which means it is a discrete random variable:

$$P\{Y=-1\}=P(X\le 0\}=F_X(0);\ P\{Y=1\}=P(X>0\}=1-F_X(0)$$

The resulting PDF is a sum of delta functions and the CDF is constituted by steps:

$$f_Y(y)=F_X(0)\delta(y+1)+(1-F_X(0))\delta(y-1)\Rightarrow F_Y(y)=F_X(0)u(y+1)+(1-F_X(0))u(y-1)$$

Appendix

1. Probability Space

1.1 Probability Space:

The **probability space** is $\{\Omega, F, P : F \to [0,1]\}$, where:

- Ω : sample space – the set of all possible outcomes.
- F : field of events (event is a subset of Ω). We will require it to be closed under countable intersections and unions[1], and under the complement operation. If Ω is a finite set that meets $|\Omega| = n$, then F may be chosen as the set of all the subsets of Ω and, thus, $|F| = 2^n$.
- $P : F \to [0,1]$: probability function which meets the following axioms:

Axioms of a probability function:

1. $\forall A \in F \;\; P(A) \geq 0$

2. $P(\Omega) = 1$

3. Additivity: for a set of events $\{A_i\}_{i=1}^{\infty}$, pairwise disjoint (namely $\forall i \neq j \;\; A_i \cap A_j = \phi$), the following exists:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Conclusions:

1. If $A \subseteq B$, then $P(A) \leq P(B)$.

2. $P(\phi) = 0$

3.. A for every event $0 \leq P(A) \leq 1$

4. For the complement event of A: $P(\bar{A}) = 1 - P(A)$

5. Union of non-disjoint events: given any two events A and B (which are not necessarily disjoint), we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

---

[1] Closed under intersections means that for a set of events $\{A_i\}_{i=1}^{\infty}$, such that $\forall i > 0 \quad A_i \in F$, we get $\bigcap_{i=1}^{n} A_i \in F$ (for any real value of n) and also $\bigcup_{i=1}^{n} A_i \in F$. Closed under the complement operation means that if $A \in F$ then $\bar{A} \in F$.

1.2 Conditional Probability:

Consider the event A with probability greater than 0. The probability of event B given event A is expressed as:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Claim: The conditional probability function is a probability function (proof appears later on in the appendix).

From the above relation, we can derive the following:

$$P(A \cap B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$$

And reach Bayes' Theorem:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} \quad \text{and} \quad P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

1.3 Law of Total Probability:

This law stems from the additive property of disjoint events and from Bayes' Theorem.

Given a set of disjoint events $\{A_i\}_{i=1}^{\infty}$ which meet $\bigcup_{i=1}^{\infty} A_i = \Omega$ (such a set is called partition of the probability space). Then, for any event B:

$$P(B) = P(B \cap \Omega) = P\left(B \cap \left(\bigcup_{i=1}^{\infty} A_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} (B \cap A_i)\right) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B \mid A_i) P(A_i)$$

This result is correct also for a finite partition (namely for a **partition** $\{A_i\}_{i=1}^{n}$):

$$P(B) = P(B \cap \Omega) = P\left(B \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) = P\left(\bigcup_{i=1}^{n} (A_i \cap B)\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(A_i) \cdot P(B \mid A_i)$$

1.4 Independent Events

Two events A, B are called independent if and only if $P(A \cap B) = P(A) \cdot P(B)$.

If $n > 2$ amount of events are given $A_1, A_2 ..... A_n$, it is said that they are independent if for all $2 \le k \le n$ and for all $1 \le i_1 < i_2 < .... < i_k \le n$ exists:

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P\left(A_{i_j}\right)$$

2. Random Variable

A random variable X is a mapping from the sample space to the field of real numbers, i.e. $X:\Omega\to R$.

2.1 Cumulative Distribution Function (CDF)

Every random variable X has a designated cumulative distribution function, which will be denoted as $F_X(\bullet)$, and is defined as follows:

$$F_X(x)=P(X(\omega)\le x)=P(X\le x)$$

[!] In this course, random variables will be represented as uppercase letters, whereas the values that they can receive (from the field of real numbers) will be represented as lowercase letters.
[!!] From now on, to simplify matters, instead of writing $X(\omega)$ to symbolize a random variable, we will simply write X.

Properties of the Cumulative Distribution Function[2]:

1. For all $x\in R$: $0\le F_X(x)\le 1$
2. Monotonically non-decreasing: if $x<y$ then $F_X(x)\le F_X(y)$.
3. Asymptotical limits: $\lim_{x\to-\infty}F_X(x)=0$, $\lim_{x\to\infty}F_X(x)=1$.
4. $F_X(x)$ is right-continuous: $F_X(x)=F_X(x^+)$

Calculating Probabilities Using the CDF:

1. $\boxed{P(y<X\le x)=F_X(x)-F_X(y)}$

   Proof:
   $$F_X(x)=P(X\le x)=P(\{y<X\le x\}\cup\{X\le y\})\underset{(1)}{=}P(y<X\le x)+\underbrace{P(X\le y)}_{F_X(y)}$$
   (1) disjoint events.

2. $\boxed{P(y\le X\le x)=F_X(x)-F_X(y)+P(X=y)}$

   Proof:
   $$P(y\le X\le x)=P(\{y<X\le x\}\cup\{X=y\})\underset{(1)}{=}P(y<X\le x)+P(X=y)\underset{(2)}{=}F_X(x)-F_X(y)+P(X=y)$$
   (1) disjoint events
   (2) section 1.

3. $\boxed{P(X=x)=F_X(x)-F_X(x^-)}$

   Proof:
   $$P(X\le x)=P(\{X<x\}\bigcup\{X=x\})\underset{(1)}{=}P(X<x)+P(X=x)\underset{(2)}{=}F_X(x^-)+P(X=x)$$
   (1) disjoint events

---

[2] We define: $F(x^+)=\lim\limits_{\substack{\varepsilon\to0\\ \varepsilon>0}}F(x+\varepsilon)$, $\quad F(x^-)=\lim\limits_{\substack{\varepsilon\to0\\ \varepsilon>0}}F(x-\varepsilon)$

(2) $P(X < x) = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} P(X \le x - \varepsilon) = F_X(x^-)$

4. $\boxed{P(y \le X < x) = F_X(x) - F_X(y) + P(X = y) - P(X = x)}$
Proof: (exercise)

2.2 Types of Random Variables:

Continuous random variable: X is a continuous random variable if and only if $F_X(x)$ is continuous.
Discrete random variable: X is a discrete random variable if and only if $F_X(x)$ is a steps function, namely:

$$F_X(x) = \sum_i p_i \cdot U(x - x_i)$$

Mixed random variable: X is a mixed random variable if and only if exist:

- $0 < \alpha < 1$
- $F_{X_d}(\bullet)$ - CDF of a discrete random variable
- $F_{X_c}(\bullet)$ - CDF of a continuous random variable

such that $F_X(x)$ can be represented as:

$$F_X(x) = \alpha F_{X_d}(x) + (1-\alpha) F_{X_c}(x) \qquad \forall x$$

2.3 Probability Density Function (PDF)

Given a random variable X, its probability density function is defined as the derivative of its CDF, meaning:

$$f_X(x) \mathrel{\hat{=}} \frac{dF_X(x)}{dx}$$

Properties of the PDF:

1. $f_X(x) \ge 0$ for all x.
2. Calculating probabilities using the PDF (for real a, b, x, such that $a < b$):

Discrete Random Variable Mixed Random Variable	Continuous Random Variable
$F_X(x) = P(X \le x) = \int_{-\infty}^{x^+} f_X(t)dt$	$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)dt$ $P(a \le X \le b) = \int_a^b f_X(t)dt$ The following holds:

$P(a \le X \le b) = \int_{a^-}^{b^+} f_X(t)dt$ $P(a < X \le b) = \int_{a^+}^{b^+} f_X(t)dt$ $P(a < X < b) = \int_{a^+}^{b^-} f_X(t)dt$ $P(a \le X < b) = \int_{a^-}^{b^-} f_X(t)dt$ $P(X = x) = \int_{x^-}^{x^+} f_X(t)dt$	$P(a < X < b) = P(a < X \le b) =$ $= P(a \le X < b) = P(a \le X \le b)$ $P(X = x) = 0$

Points of Emphasis:

1. Notice that for a discrete random variable:

$$f_X(x) = \frac{d}{dx}\left(\sum_i p_i \cdot U(x - x_i)\right) = \sum_i p_i \delta(x - x_i)$$

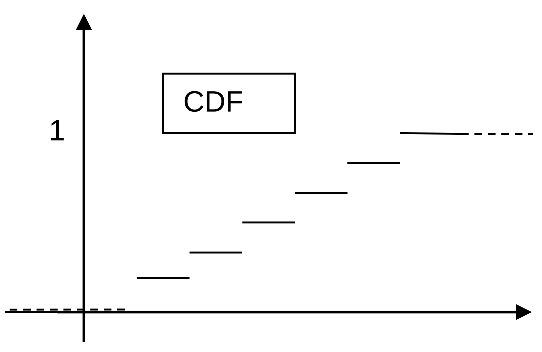
In other words, the PDF is a sequence of delta functions centered at the possible values of the discrete random variable. Each of these deltas has a weight equal to the probability of receiving said value. Thus, given $\{x_i\}_{i=1}^{n}$ values that the discrete random variable X takes with positive probability (n can be a natural number or ∞), then:

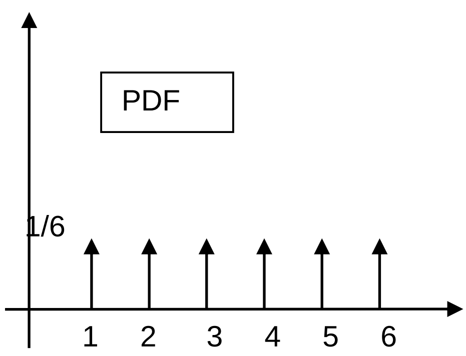
$$f_X(x) = \sum_{i=1}^{n} P(X = x_i) \cdot \delta(x - x_i)$$

2. In the discrete and mixed cases, it is necessary to be precise when writing limits of integration – $x^\pm, y^\pm$ – if there are delta functions in the end points.

Simple Examples:

1. Draw the CDF and PDF of a random variable that represents the value obtained when rolling a fair die.



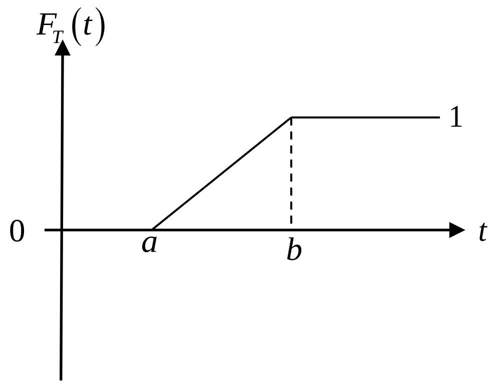


2. T is a continuous random variable, which distributes uniformly in the section (a,b): $T \sim Unif(a,b)$ ($a<b$). Find the CDF of T.

$$T \sim Unif(a,b) \Rightarrow f_T(t) = \frac{1}{b-a}\left[u(t-a)-u(t-b)\right]$$

$$F_T(t) = P(T \le t) = \int_{-\infty}^{t} \frac{1}{b-a}\left[u(x-a)-u(x-b)\right]dx$$

$$= \begin{cases} 0 & t<a \\ \int_a^t \frac{1}{b-a}dx = \frac{t-a}{b-a} & a \le t < b \\ \int_a^b \frac{1}{b-a}dx = 1 & b \le t \end{cases}$$



3. Functions of Random Variables

Consider a random variable X and a deterministic and known function $g(\bullet)$. Define:

$$Y = g(X)$$

Note that Y is a random variable: $\forall \omega \in \Omega \;\; Y(\omega) = g(X(\omega))$, namely a compound function. For $S \subset \Re$, how do we calculate $P(Y \in S)$?

Let us define for all S: $g^{-1}(S) = \{x : g(x) \in S\}$ (g is not necessarily invertible) $\Rightarrow P(Y \in S) = P(X \in g^{-1}(S))$. This probability can be calculated given the complete statistics of the random variable X ($f_X(x)$ or $F_X(x)$). Specifically, we would like to express $F_Y(y)$ using $F_X(x)$:

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

<u>Proposition:</u> Given $Y = g(X)$ where X is a random variable. Let us assume the equation $g(x) = y$ has a finite or countable amount of solutions, which we will represent as $\{x_i\}_{i=1}^{n}$ (where $n = \infty$ stands for a countable amount of solutions). Assuming also that $\forall i \;\; g'(x_i) \ne 0$, then $f_y(y)$ can be expressed as:

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}$$

- $f_Y(y) = 0$ if there are no solutions.
- For cases where the number of solutions is not countable or $g'(x_i) = 0$, this proposition cannot be used. Rather one must resort be using the original definitions.