A random variable simply quantities (assigns) a real number to the outcome of an experiment.

Random variables can be separated into two classes:
- Discrete (like flipping a coin, rolling a die, etc.)
- Continuous (time to failure, actual resistance of a resistor, etc.)

Discrete Random variables are characterized by their Probability mass function (PMF)

\[ P_X(K) = \Pr(X = K) \]

Example: Roll a 4-sided dice twice

\[ X = \text{Sum} \]

The PMF \( P_X(K) \) looks like

![Graph showing PMF for sum of two 4-sided dice rolls]
Common discrete random variables

\[ X \sim \text{Uniform} \{1, \ldots, n\} \]

Any value between 1 and \(n\) gets equal probabilities

\[ X \sim \text{Bernoulli} (p) \]

\[ P_X(k) = \begin{cases} 1-p & \text{k=0} \\ p & \text{k=1} \end{cases} \]

e.g., transmitting a bit on a communication channel

\[ X=0 \quad \text{(correct)} \]
\[ X=1 \quad \text{(incorrect)} \]

\[ X \sim \text{Binomial} (n, p) \]

\[ X = \text{number of successes in n Bernoulli trials} \]
\[ \text{each with a prob } p \quad X \in \{0, \ldots, n\} \]

\[ P_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \]

\[ X \sim \text{Poisson} (\lambda) \]

\[ P_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k \geq 0) \]

(unbounded)

Example: photon count at a light sensor
Continuous random variables are characterized by their probability density function (PDF)

\[
f_X(x) \quad \Pr(a \leq X \leq b) = \int_a^b f_X(x) \, dx
\]

\[
\int_{-\infty}^{\infty} f_X(x) \, dx = 1 \quad \text{and} \quad f_X(x) \geq 0 \quad \forall x
\]

Question: Can \( f_X(x) \) be larger than 1?

A: Yes. As long as the integral above is 1, \( f_X(x) \) can take any value in the continuous domain.

Common continuous random variables

\( X \sim \text{Uniform} \left( [a, b] \right) \)

\( X \) can take any value between \( a \) and \( b \) with equal probability.

\[ f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \]

\( X \sim \text{Exp} (\lambda) \quad \rightarrow \quad f_X(x) = \begin{cases} x e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases} \)

\[ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

Gaussian or a Bell curve

This is a very important distribution.
Both discrete and continuous random variables can be characterized by their cumulative distribution function (CDF):

\[ F_X(k) = \Pr(X \leq k) \]

Properties of CDF:
- \( 0 \leq F_X(k) \leq 1 \)
- \( F_X(-\infty) = 0 \quad F_X(+\infty) = 1 \)
- \( F_X(k) \) is nondecreasing as \( k \) increases
- \( \Pr(a \leq X \leq b) = F_X(b) - F_X(a) \)

Examples:

Discrete

Continuous

Example: Consider the random variable \( X \) with the CDF

\[ \Pr(X \leq -5) = \]

\[ \Pr(X > -5) = \]

\[ \Pr(X > 8) = \]

\[ \Pr(-5 < X < 8) = \]

\[ \Pr(X > 0) = \]
Relationship between PMF and CDF
PDF and CDF

1. PDF and CDF
\[ \int_{-\infty}^{x} f_X(u) \, du = F_X(x) \]
\[ \int_{a}^{b} f_X(u) \, du = F_X(b) - F_X(a) \quad f_X(x) = \frac{dF_X(x)}{dx} \]

2. PMF and CDF
\[ \sum_{i=-\infty}^{K} p_X(i) = F_X(K) \]
\[ P_X(K) = \Pr(X \leq K) - \Pr(X \leq K-1) = F_X(K) - F_X(K-1) \]

Example: Random variable \( X \) has a PDF
\[ f_X(x) = \begin{cases} 5e^{-kx} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \]

1. Find the value of \( k \).
2. Find \( \Pr(X > 1) \)
3. Find \( \Pr(X < \sqrt{2}) \)
Mean and variance of random variables

Mean (average): what does it mean??

We find one number that is weighted sum of all values in the distribution. The weights are given by the probability function.

Mean of a random variable $X$ is denoted by $\mu(X)$ or $E[X]$ where $E$ stands for expectation.

- For a discrete random variable with a PMF $P_X(K)$, the mean would be:

$$E(X) = \sum_K K P_X(K)$$

- For continuous random variables with PDF of $f_X(u)$

$$E[X] = \int_{-\infty}^{+\infty} u f_X(u) \, du$$

Example: Suppose $X$ has a PDF $f_X(x)$

What is $E[X] =$ ?

An important property of expectation is that it is linear! If $X_1$ and $X_2$ are random variables, then:

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

A constant, $E[aX] = a E[X]$
Generally speaking, the $n$-th moment of a random variable $X$ is:

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

The mean is the first moment:

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

The second moment is:

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x) \, dx$$

The variance of a random variable $X$ is:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x) \, dx$$

Given a distribution, its mean and variance are each a number and not a distribution.

In the class, we have shown that:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

For any given function of a random variable, say $g(X)$, we can define:

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

The average value of $g(X)$
Note: As we mentioned earlier, taking averages (mean) is a linear function, i.e., \( E(X_1 + X_2) = E(X_1) + E(X_2) \).

Summing up variances is not always linear, unless the two variables are independent of each other.

\[
\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)
\]

when \( X_1 \) and \( X_2 \) are independent!

Mean and Variance of Binomial:

suppose \( X \sim \text{Bernoulli}(p) \) \( \Rightarrow P(X = 0) = p, \ P(X = 1) = 1 - p \)

\[
E[X] = p
\]

\[
\text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)
\]

Assume we have \( n \) independent random variables

\( X_1, X_2, \ldots, X_n \)

set \( Y = \sum_{i=1}^{n} X_i \) \( \Rightarrow E[Y] = E[\sum_{i=1}^{n} X_i] = nE[X] = np \)

\( Y \sim \text{Binomial}(n, p) \)

\( P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \)

since the variables are independent, we have:

\[
\text{Var}[Y] = \text{Var}[X_1 + \ldots + X_n] = \text{Var}[X_1] + \ldots + \text{Var}[X_n] = np(1-p)
\]
Conditional probability:

Let $X$ be a random variable, and let $B$ an event involving $X$.

We can define the conditional CDF as (conditioned on event $B$)

$$F_{X|B}(x|B) = \Pr[X \leq x | B] = \frac{\Pr[X \leq x \cap B]}{\Pr[B]} = \frac{\Pr[X \leq x, B]}{\Pr[B]}$$

The conditional probability defines a new variable space, where the above CDF has the important properties it should have, i.e., monotonic between $[0,1]$. 

The conditional PDF is defined as:

$$f_{X|B}(x|B) = \frac{d}{dx} F_{X|B}(x|B)$$

the conditional PDF is a valid PDF, i.e., adds to 1, positive, etc.

The conditional mean is:

$$E[X|B] = \int_{-\infty}^{+\infty} x f_{X|B}(x|B) \, dx$$
Example:
Let \( X \sim \text{Uniform}[0, 3] \)

\( B \): event that \( \text{round}(X) \) is even

a) Find \( F_{X|B}(x|B) \)
b) Find \( f_{X|B}(x|B) \)
c) Find \( E[X|B] \)

Example:
Suppose \( X \) has PDF \( f_X(x) \) shown below:

\[
\begin{align*}
\frac{3x}{2} &= 1 \\ x &= \frac{2}{3}
\end{align*}
\]

Say \( B \) is the event \( X > 1 \)

1. What is the CDF for \( F_{X|B}(x|B) \)
2. What is the PDF for \( f_{X|B}(x|B) \)
3. What is the conditional mean \( E[X|X \geq 1] \)?

Something interesting to note:

an exponential random variable has no memory ... why?

\( X \sim f_X(x) = e^{-\lambda x}, \ x \geq 0 \)

Then:
\[
\begin{align*}
\Pr[X > t+s | X > t] &= \frac{\Pr[X > t+s]}{\Pr[X > t]} = \frac{1 - F(t+s)}{1 - F(t)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = 1 - F(s) = \Pr[X > s]
\end{align*}
\]
Multiple random variables
We want to deal with cases where we can describe/analyze multiple random variables that interact with one another.

Joint PMF
Let $X$ and $Y$ be discrete random variables.
Their joint PMF is:

$$P_{XY}(x,y) = \Pr[X=x, Y=y]$$

function of two random variables

The joint PMF has the following properties:

1. $0 \leq P_{XY}(x,y) \leq 1$
2. $\sum_x \sum_y P_{XY}(x,y) = 1$
3. $\Pr[(x,y) \in A] = \sum_{(x,y) \in A} P_{XY}(x,y)$

Example: A cooler contains 2 cokes, 3 lemonades, and a water.
We draw 2 drinks without replacement.

$X =$ # cokes we get
$Y =$ # lemonades we get

The joint PMF can be written as:
Marginals: $P_x(x)$ and $P_y(y)$ can be inferred from the joint PMF by summing over one of the dimensions:

$$P_x(x) = \sum_y P_{xy}(x,y)$$
$$P_y(y) = \sum_x P_{xy}(x,y)$$

For the cooler example above:

$$P_x(x) = \frac{6}{15} \quad \frac{8}{15} \quad \frac{1}{15} \quad \quad P_y(y) = \frac{1}{5} \quad \frac{3}{5} \quad \frac{1}{5}$$
$$x = 0 \quad 1 \quad 2 \quad \quad y = 0 \quad 1 \quad 2$$

Given two continuous random variables $X$ and $Y$, we can define the joint distribution function as

$$F_{xy}(x,y) = \Pr [X \leq x, Y \leq y]$$

The properties of CDF naturally extend to this case.

- $0 \leq F_{xy}(x,y) \leq 1$
- $F_{xy}(-\infty, y) = F_{xy}(x, -\infty) = F_{xy}(-\infty, -\infty) = 0$
- $F_{xy}(\infty, \infty) = 1$
- $F_{xy}(x, y)$ is a non-decreasing as either $x$ or $y$ increase
- $F_{xy}(\infty, y) = F_y(y)$
- $F_{xy}(x, \infty) = F_x(x)$
The joint density (joint PDF) is:

$$f_{xy}(x,y) = \frac{\partial^2 F_{xy}(x,y)}{\partial x \partial y}$$

It doesn't matter in which order we do the differentiation.

The properties of a 2D PDF extend naturally to this case:

$$f_{xy}(x,y) \geq 0$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{xy}(x,y) \, dx \, dy = 1$$

$$Pr[A] = Pr[(x,y) \in A] = \int_{A} \int_{-\infty}^{+\infty} f_{xy}(x,y) \, dx \, dy$$

Given the joint PDF, one can calculate the marginals

$$f_x(x) = \int_{-\infty}^{+\infty} f_{xy}(x,y) \, dy$$

$$f_y(y) = \int_{-\infty}^{+\infty} f_{xy}(x,y) \, dx$$

**Expectation**

Given a function $g(x,y)$, the expectation of $g$ is

$$E[g(x,y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{xy}(x,y) \, dx \, dy$$

**Example:**

$$E[xy] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{xy}(x,y) \, dx \, dy$$

correlation between $x$ and $y$
More on conditional probability

Suppose we have two continuous random variables \( X, Y \) with joint PDF \( f_{XY}(x,y) \).

If \( X \) and \( Y \) are related, then observing \( Y = y \) will change our knowledge of \( X \).

\[
\Pr [ X \leq x \mid Y \leq y ] = \frac{\Pr [ X \leq x, Y \leq y ]}{\Pr [ Y \leq y ]} \quad \text{joint CDF}
\]

\[
\Rightarrow F_X (x \mid Y \leq y) = \frac{F_{XY}(x,y)}{F_Y(y)} \quad \text{conditional CDF for } X
\]

Similarly,

\[
F_X (x \mid y_1 < Y < y_2) = \frac{F_{XY}(x,y_2) - F_{XY}(x,y_1)}{F_Y(y_2) - F_Y(y_1)}
\]

if one takes the above formula to the limits,

\[
f_X (x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}
\]

Bayes theorem for continuous random variables

\[
f_X (x \mid y) = \frac{f_Y(y \mid x) f_X (x)}{f_Y(y)}
\]
Example: \( X, Y \) have joint PDF
\[
f_{X,Y}(x,y) = \begin{cases} 
\frac{2}{3}(x+2y) & 0 \leq x \leq 1, \ 0 \leq y \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

Suppose that we observe \( Y = \frac{1}{2} \)

What is \( f_X(x \mid Y = \frac{1}{2}) \)?

What is \( f_X(x \mid Y = 1) \)?

Law of total probabilities applies here too:
\[
f_X(x) = \int_{-\infty}^{+\infty} f_{X \mid Y}(x \mid y) f_Y(y) \, dy
\]

---

Conditional moments

Since the distribution of \( X \) changes when we observe \( Y = y \), so does its mean, variance, etc.

We define the conditional moments as:
\[
E[X \mid Y = y] = \int x f_{X \mid Y}(x \mid y) \, dx
\]
\[
E[X^2 \mid Y = y] = \int x^2 f_{X \mid Y}(x \mid y) \, dx
\]

In general:
\[
E[g(X) \mid Y = y] = \int g(x) f_{X \mid Y}(x \mid y) \, dx
\]
Statistical Independence

We say that two random variables $X, Y$ are independent if we can factor their joint PDF as:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

since

$$f_{X,Y}(x,y) = f_X(x|y) f_Y(y) = f_Y(y|x) f_X(x)$$

We have,

$$f_X(x|y) = f_X(x) \quad \forall y$$

$$f_Y(y|x) = f_Y(y) \quad \forall x$$

$$\iff X, Y \text{ independent}$$

Independence means that knowing $Y = y$ does not tell us anything about $X$.

Also,

$$X, Y \text{ independent} \implies E[XY] = E[X] E[Y]$$

**Important Note**

$$E[XY] = E[X] E[Y] \implies X, Y \text{ independent}$$

This is not always true!

There are cases where the condition on left holds, but $X, Y$ are NOT INDEPENDENT!