STAT 2

Random Variables

1- **What is a Random Variable**?

The concept of "randomness" is fundamental to the field of statistics. As mentioned in the probability theory notes, the science of statistics is concerned with assessing the uncertainty of inferences drawn from random samples of data. Now that we've defined some basic concepts related to set operations and probability theory, we can more formally discuss what it means for things to be random.

The concept of uncertainty is inherent to the definition of randomness. For any random process, we cannot be certain about the action's outcome because there is some element of chance involved. Note that the opposite of a random process is a "deterministic process", which is some action that always results in the same outcome.

Example 1. Flipping a two-sided (heads-tails) coin is a random process because we do not know if we will observe a heads or tails. Flipping a one-sided (heads-heads) coin is a deterministic process because we know that we will always observe a heads.

Definition. In probability and statistics, a random variable is an abstraction of the idea of an outcome from a randomized experiment. More formally, a random variable is a function that maps the outcome of a (random) simple experiment to a real number. Random variables are typically denoted by capital italicized Roman letters such as X.

A random variable is an abstract way to talk about experimental outcomes, which makes it possible to flexibly apply probability theory. Note that you cannot observe a random variable X itself, i.e., you cannot observe the function that maps experimental outcomes to numbers. The experimenter defines the random variable (i.e., function) of interest, and then observes the result of applying the defined function to an experimental outcome.

Example 2. Suppose we flip a fair (two-sided) coin $n \geq 2$ times, and assume that the n flips are independent of one another. Define X as the number of coin flips that are heads. Note that X is a random variable given that it is a function (i.e., counting the number of heads) that is applied to a random process (i.e., independently flipping a fair coin n times). Possible realizations of the random variable X include any $x \in \{0, 1, \ldots, n\}$, i.e., we could observe any number of heads between 0 and n.

2- Discrete and continuons random variable :

2-1- Discrete random variable :

A random variable is discrete if its domain consists of a finite (or countably infinite) set of values.

Example 3. Suppose we flip a fair (two-sided) coin $n \geq 2$ times, and assume that the n flips are independent of one another. Define X as the number of coin flips that are heads. The random variable X is a discrete random variable given that the domain $S = \{0, . . . , n\}$ is a finite set (assuming a fixed number of flips n). Thus, we could associate a specific probability to each $x \in S$. See Example 13 from the "Introduction to Probability Theory" notes for an example of the probability distribution with $n = 3$.

2-2- Continuons random variable :

A random variable is continuous if its domain is uncountably infinite.

Example. Consider the face of a clock, and suppose that we randomly spin the second hand around the clock face. Define X as the position where the second hand stops spinning (see Figure 1). The random variable X is a continuous random variable given that the domain $S = \{x \mid x \text{ is a point on a circle}\}$ is an uncountably infinite set. Thus, we cannot associate a specific probability with any given $x \in S$, i.e., $P(X = x) = 0$ for any $x \in S$, but we can calculate the probability that X is in a particular range, e.g., $P(3 < X < 6) = 1/4$.

3- Probability Mass and Density Functions

3-1 Probability Mass Functions :

For discrete random variables, we can enumerate all of the possible realizations of the random variable, and associate a specific probability with each possible realization.

Definition. The probability mass function (PMF) of a discrete random variable X is the function $f(\cdot)$ that associates a probability with each $x \in S$. In other words, the PMF of X is the function that returns $P(X = x)$ for each x in the domain of X.

Any **PMF** must define a valid probability distribution, with the properties:

STAT 2

- $f(x) = P(X = x) \geq 0$ for any $x \in S$
- $\Sigma_{x \in S} f(x) = 1$

Example : Let the random experiment E be represented by "rolling two dice."

We define the random variable X as the sum of the two outcomes.

1- Determine the space of initial events

2- Determine the basic set generated by the random variable X

4- Find the probability Mass function

Solution :

1- Determination of the space of initial events

$$S=\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)....,(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

2- Determination of the basic set generated by the random variable X

$$Sy=\{2,3,4,5,6,7,8,9,10,11,12\}$$

3-The probability mass function:

$$P(Y=2)=P\{(1,1)\}=1/36$$

$$P(Y=3)=P\{(2,1),(1,2)\}=2/36$$

$$P(Y=4)=P\{(2,2),(1,3),(3,1)\}=3/36$$

$$P(Y=5)=P\{(4,1),(1,4),(2,3),(3,2)\}=4/36$$

$$P(Y=6)=P\{(3,3),(5,1)(1,5),(4,2),(2,4)\}=5/36$$

$$P(Y=7)=P\{(4,3),(3,4),(6,1),(1,6),(5,2),(2,5)\}=6/36$$

$$P(Y=8)=P\{(4,4),(6,2),(2,6),(5,3),(3,5)\}=5/36$$

$$P(Y=9)=P\{(6,3),(3,6),(5,4),(4,5)\}=4/36$$

$$P(Y=10)=P\{(5,5),(4,6),\{(6,4)\}=3/36$$

$$P(Y=11)=P\{(5,6),(6,4)\}=2/36$$

$$P(Y=12)=P\{(6,6)\}=1/36$$

The probability mass function can be presented in the form of the following table:

$X=x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

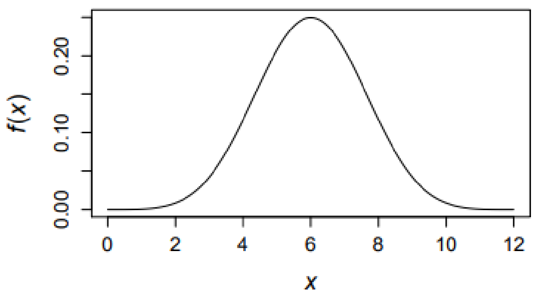
3-2- Probability Density Functions :

For continuous random variables, we cannot enumerate all of the possible realizations of the random variable, so it is impossible to associate a specific probability with each possible realization. As a result, we must define the probabilities of discrete and continuous random variable occurrences using distinct (but related) concepts.

Definition. The probability density function (PDF) of a continuous random variable X is the function f(·) that associates a probability with each range of realizations of X. The area under the PDF between a and b returns $P(a < X < b)$ for any $a, b \in S$ satisfying $a < b$.

Any PDF must define a valid probability distribution, with the properties:

- $f(x) \geq 0$ for any $x \in S$
- $\int_a^b f(x)dx = P(a < X < b) \geq 0$ for any $a, b \in S$ satisfying $a < b$
- $\int_{x \in S} f(x)dx = 1$.



Example : Let X be a random variable defined on the domain [0,4] with the following probability density:

$$f(x) = \begin{cases} \frac{1}{2} - \frac{1}{8}x & 0 \leq X \leq 4 \\ 0 & elsewhere \end{cases}$$

STAT 2

- Calculate the value of the following probability: P(2<X<4)

$$P(2 < X < 4) = \int_2^4 \frac{1}{2} - \frac{1}{8}x \; dx = \left[\frac{1}{2}x - \frac{1}{16}x^2\right]_2^4 = \frac{1}{4}$$

4-Cumulative Distribution Function

Definition. The cumulative distribution function (CDF) of a random variable X is the function F(·) that returns the probability P(X ≤ x) for any x ∈ S. Note that the CDF is the same as the probability distribution that was defined in Section 4 of the "Introduction to Probability" notes, such that the CDF is a function from S to [0, 1], i.e., F : S → [0, 1].

Any CDF must define a valid probability distribution, i.e., 0 ≤ F(x) ≤ 1 for any x ∈ S, which comes from the fact that F(x) = P(X ≤ x) is a probability calculation. Note that a CDF can be defined for both discrete and continuous random variables:

• $F(x) = \Sigma_{z \in S, z \leq x} f(z)$ ----→ for discrete random variables

• $F(x) = \int_{-\infty}^{x} f(z)dz$ ----→ for continuous random variables

Furthermore, note that probabilities can be written in terms of the CDF, such as P(a < X ≤ b) = F(b) − F(a)

given that the CDF is related to the PMF (or PDF), such as

• f(x) = F(x) − lima→x− F(a) for discrete random variables

• f(x) = dF(x) /dx for continuous random variables

5- Expected Value and Expectation Operator

Definition. The expected value of a random variable X is a weighed average where the weights are defined according to the PMF or PDF. The expected value of X is defined as μ = E(X) where E(·) is the expectation operator, which is defined as

$E(X) = \Sigma_{x \in S}\, x\, f(x)$ for discrete random variables

STAT 2

E(X) = $\int_{x\in S} x\, f(x)dx$ for continuous random variables.

Example : Let X be a random variable defined on the domain [0,4] with the following probability density:

$$f(x) = \begin{cases} \frac{1}{3} & si\ \ 0 \leq X \leq 3 \\ 0 & ailleurs \end{cases}$$

E(x)=$\int_0^3 x\, f(x)dx = \int_0^3 x\ \frac{1}{3}\ dx = \left[\frac{x^2}{6}\right]_0^3 = \frac{3}{2} = 1{,}5$

Example: Let X be a random variable whose probability mass function is given by the following table:

X=xi	0	1	2	3	4	5	6	7	8	9	10
P(X=xi)	0,05	0,07	0,08	0,09	0,16	0,20	0,15	0,1	0,05	0,04	0,01

$$E(X) = \sum xi\ p(X = xi) = 4{,}6$$

Expectation Rules

1. If a ∈ R is fixed constant, then E(a) = a.

2. If X is a random variable with mean µ = E(X), and if a, b ∈ R are fixed constants with b 6= 0, then E(a + bX) = E(a) + bE(X) = a + bµ.

6-**Variance**

Definition. The variance of a random variable X is a weighted average of the squared deviation between a random variable and its expectation, where the weights are defined according to the PMF or PDF

V(X)=$\sum pi[xi - E(X)]^2 = \sum xi^2 pi - \left(E(X)\right)^2 = E(X^2) - (E(X))^2$ for discrete random variables

STAT 2

$V(x) = \int_D x^2 f(x)dx - [E(x)]^2 = E(x^2) - [E(x)]^2$ for continuous random variables

Variance Rules

1. If a ∈ R is fixed constant, then Var(a) = 0.

2. If X is a random variable with variance σ^2 = Var(X), and if a, b ∈ R are fixed constants with b ≠0, then Var(a + bX) = Var(a) + b^2Var(X) = $b^2\sigma^2$.

Example : Let X be a random variable defined on the domain [0,3] with the following probability density:

$$f(x) = \begin{cases} \frac{1}{12}(2x+1) & 0 \le X \le 3 \\ 0 & elsewhere \end{cases}$$

- E(x)=$\int_0^3 x\, f(x)dx = \int_0^3 x\,\frac{1}{12}(2x+1)\,dx = \frac{1}{12}\int_0^3 (2x^2 +$ *xdx=11223x3+12x203=4524=1,875*

- V(x)=$\int_0^3 x^2 f(x)dx - E(x)^2 = \int_0^3 x^2\,\frac{1}{12}(2x+1)\,dx -$ *E(x)2 =112 032x3+x2 dx−E(x)2=112[24x4+13x3]03−1,8752=0,61*

Example: Let X be a random variable whose probability mass function is given by the following table:

X=xi	0	1	2	3	4	5	6	7	8	9	10
P(X=xi)	0,05	0,07	0,08	0,09	0,16	0,20	0,15	0,1	0,05	0,04	0,01

$E(X) = \sum xi\, p(X = xi)$ =4,6

$V(X) = \sum xi^2 pi - (E(X))^2 = 26{,}5 - (4{,}6)^2$ =5,34