Lecture 2: Random Variables

PhD Mathematics II: Probability

Juan Llavador Peralt

IIES

November 5, 2024

Outline

Random Varibles – Informal Definition

A **random variable** (RV)...

- is a numerical quantity
- is **stochastic**, i.e. the value it takes is uncertain
- if the values are known with certainty, it is a **deterministic** (or degenerate) RV

Example

- Let X be the sum of the roll of two fair dice (clearly $X \in \{2, 3, \ldots, 12\} \subset \mathbb{R}$)
- What is $\mathbb{P}(X = 10)$?

$$
\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}
$$

- $\{X = 10\} = \{\omega \in \Omega : X(\omega) = 10\} = \{(4,6), (5,5), (6,4)\}$
- So using $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$, $\mathbb{P}(X = 10) = \frac{3}{36} = \frac{1}{12}$
- Easy to extend to $X \leq 10 : \mathbb{P}(X \leq 10) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq 10\}) = \frac{33}{36} = \frac{11}{12}$

A first definition

- Given a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$
- a **random variable** is a function $X : \Omega \to \mathbb{R}$ with the property that:

$$A_x = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$$

- That is, A_x is an event for all $x \in \mathbb{R}$

Borel σ-algebra

- The **Borel σ-algebra** on $\mathbb{R}$ is the smallest σ-algebra containing all open intervals
- Denoted by $\mathcal{B}$:
$$\mathcal{B} = \sigma(\{(a,b) : a,b \in \mathbb{R}\})$$
- Note that sets of the form $[a,b]$, $(a,b]$, $[a,b)$ are also in $\mathcal{B}$
- It is hard to construct a set that is not in $\mathcal{B}$
- Extends to higher dimensions

Measurable Functions

- Consider the **measurable spaces** $(\Omega, \mathcal{F})$, and $(\mathbb{R}, \mathcal{B})$
- A function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-**measurable** if

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}$$

- **Intuition:** We can determine wich outcomes map to the elements of $B \in \mathcal{B}$

Random Variables as Measurable Functions

- For a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$, and a **measurable space** $(\mathbb{R}, \mathcal{B})$
- A **random variable** X is a $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$
- The **law of a random variable** is the probability measure induced by X:

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

- Note that $\mathbb{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B})$
- It is enought to specify $\mathbb{P}_X$ for all $B \in \mathcal{B}$ of the form $(-\infty, x]$:

Cumulative Distribution Function

- The **cumulative distribution function** (CDF) of a random variable X is defined as

$$F_X(x) = \mathbb{P}(X \leq x)$$

  and satisfies the following properties:

  1. $F_X(x)$ is **non-decreasing**: $x \leq y \Rightarrow F_X(x) \leq F_X(y)$
  2. $F_X(x)$ is **right-continuous**: $\lim_{y \downarrow x} F_X(y) = F_X(x)$
  3. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$

- **Uniqueness:** A CDF uniquely determines the RV

Proof of Right-Continuity

- Want to prove that $\lim_{y \downarrow x} F_X(y) = F_X(x)$
- Let $A_x = \{\omega \in \Omega : X(\omega) \leq x\}$
- Take a decreasing sequence $x_n \downarrow x$
- Note that $A_x = \cap_{i=1}^{\infty} A_{x_i}$
- and A_{x_n} is a decreasing sequence of sets
- By continuity of probability measures (see lecture 1),

$$\lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} \mathbb{P}(A_{x_n}) = \mathbb{P}(\cap_{i=1}^{\infty} A_{x_i}) = \mathbb{P}(A_x) = F_X(x)$$

Probabilities from CDFs

Consider $x, y \in \mathbb{R}$ with $x < y$, F_X the CDF of X, and $F_X(x^-) = \lim_{z \uparrow x} F_X(z)$

1. $\mathbb{P}(X > x) = 1 - F_X(x)$
2. $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$
3. $\mathbb{P}(X < x) = F_X(x^-)$
4. $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x)$

Support of a Random Variable

- Take a non-negative real-valued function, f
- The **support** of f is the set of the real line where f is strictly positive:

$$\{x \in \mathbb{R} : f(x) > 0\}$$

- We will refer to the **support of a random variable** X as the support of f_X
- i.e. the set of values/intervals that X can take with positive probability

Discrete Random Variables

- A random variable X is **discrete** if it takes countably many values $\{x_1, x_2, \ldots\}$

- The **probability mass function** (PMF) of X is defined as

$$\begin{aligned} f_X(x) &= \mathbb{P}(X = x) \\ &= \mathbb{P}(X \leq x) - \mathbb{P}(X < x) \\ &= F_X(x) - F_X(x^-) \end{aligned}$$

- The CDF can be recovered as

$$F_X(x) = \sum_{u \leq x} f_X(u)$$

- The PMF satisfies the following properties:
  1. $f_X(x) \in [0, 1]$
  2. $\sum_x f_X(x) = 1$

Example: Two fair dice

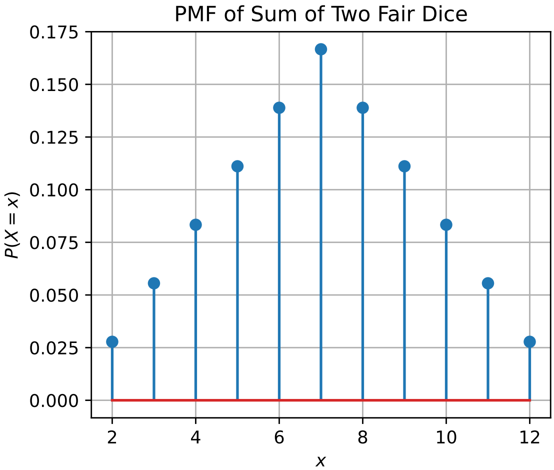
- Let X be the sum of the roll of two fair dice

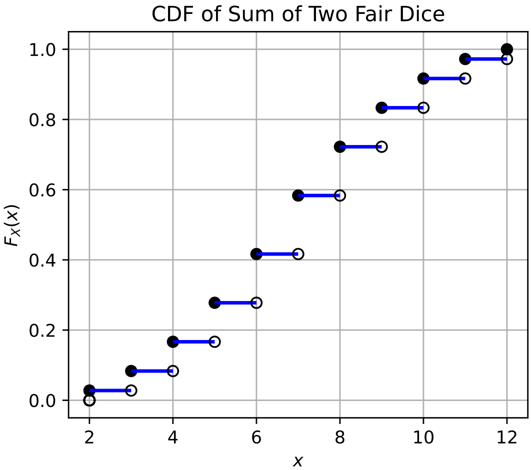
$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

- What are the PMF and CDF of X?

Example: Two fair dice

- Let X be the sum of the roll of two fair dice





Continuous Random Variables: Motivation

- So far, we have considered random variables that take countably many values
- In many cases, it is more natural to consider RVs that can take any value in an interval
- In practice, measurement is limited by precision
- Consider measuring the height of a person with an ever more precise ruler
- We also model as continuous, variables that take many discrete values like income, wealth, etc.

Continuous Random Variables

- A random variable X is **continuous** if it has a CDF $F_X(x)$ that can be written as

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

for some function integrable function $f_X(x)$

- The **probability density function** (PDF) of X is defined as

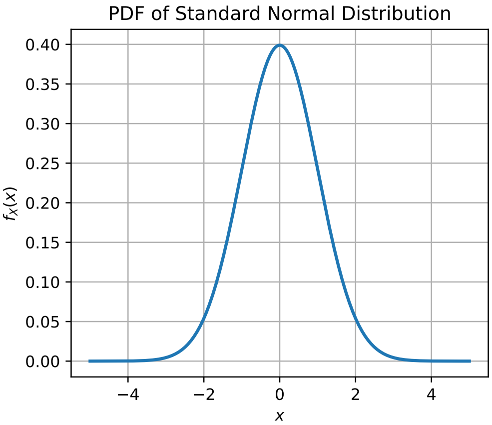
$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \lim_{h \to 0} \frac{F_X(x+h) - F_X(x)}{h} \\ &= \lim_{h \to 0} \frac{\mathbb{P}(x < X \leq x+h)}{h} \end{aligned}$$

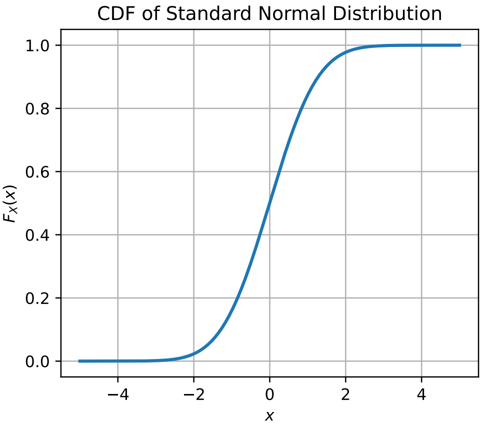
- The PDF satisfies the following properties:
  1. $f_X(x) \geq 0$
  2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Example: Gaussian Random Variable

- Let $X \sim N(0,1)$ be a standard normal random variable

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$





Any non-negative integrable function can be a PDF

- Let $g : \mathbb{R} \to \mathbb{R}_+$ be a continuous, integrable, and non-negative function
- Then $\int_{-\infty}^{\infty} g(x)dx = K > 0$
- Define $f_X(x) = \frac{1}{K} g(x)$
- Then $f_X(x)$ is a PDF

Expectation

- Suppose X is a **discrete** random variable with PMF $f_X(x)$
- The **expectation** of X is defined as

$$\mathbb{E}[X] = \sum_x x f_X(x)$$

- For a **continuous** random variable with PDF $f_X(x)$, the expectation is defined as

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

- Technically, the expecation might not exist
- When writing $\mathbb{E}[X]$, we assume $\mathbb{E}[|X|] < \infty$

Law of the Unconscious Statistician (LOTUS)

- Let $g : \mathbb{R} \to \mathbb{R}$ be a nice function[1]
- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$
- Then $g(X)$ is a random variable
- **(LOTUS)** The expectation of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \sum_x g(x) f_X(x) \quad \text{if } X \text{ is discrete}$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{if } X \text{ is continuous}$$

- Again, we are assuming that $\mathbb{E}[|g(X)|] < \infty$

[1] $g : \mathbb{R} \to \mathbb{R}$ is a function for which the inverse image of any Borel set is a Borel set, i.e. g is $\mathcal{B}$-measurable

Unified Notation: The Riemann-Stieltjes Integral

- The **Riemann-Stieltjes integral** is a generalization of the Riemann integral
- Let X be a random variable with CDF $F_X(x)$
- The expectation of function $g(X)$ is defined as (if it exists):

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

- For the case where the distribution is either discrete or continuous:

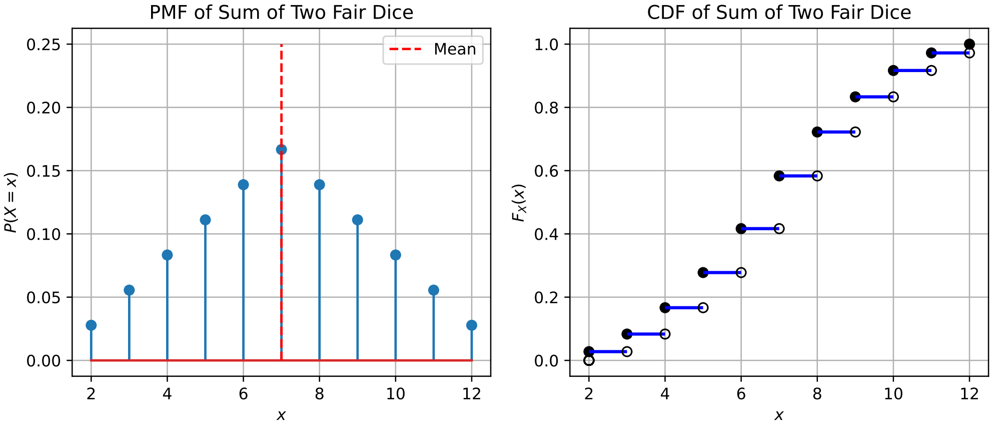
$$\int_{-\infty}^{\infty} g(x) dF_X(x) = \begin{cases} \sum_x g(x) f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

- Also works for mixed distributions, and is a precursor to the **Lebesgue integral**

Mean of Two Dice

The mean of the sum of two dice is given by

$$\mathbb{E}[X] = \sum_x x f_X(x) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 12 \cdot \frac{1}{36} = 7$$



Linearity of Expectation

- Expectations are just integrals!
- Let X and Y be random variables, and $a, b \in \mathbb{R}$
- Then $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- More generally, for any random variables $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots, a_n$
- $\mathbb{E}[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] + \ldots + a_n\mathbb{E}[X_n]$

Variance and Higher Moments

- The **variance** of X is defined as

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

- Note that

$$\mathsf{Var}(aX) = a^2 \mathsf{Var}(X)$$

- The **k-th moment** of X is defined as

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

- The **k-th central moment** of X is defined as

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

Skewness

- **Skew coefficient:** $\text{Skew}(X) = \frac{\mathbb{E}[(X-\mathbb{E}[X])^3]}{\text{Var}(X)^{3/2}}$

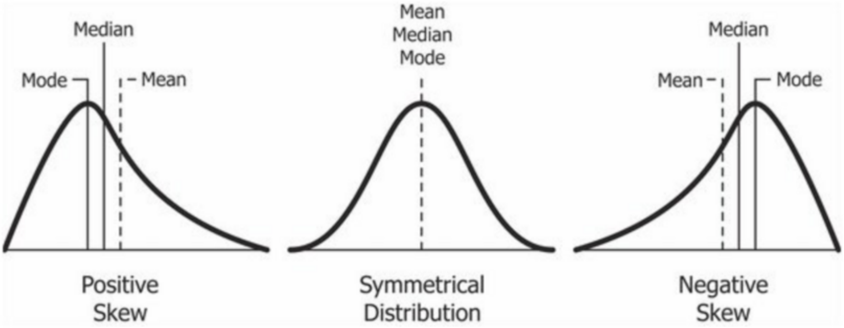


Figure: Source: Wikipedia

Kurtosis

- **Kurtosis:** $\text{Kurt}(X) = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\text{Var}(X)^2} - 3$

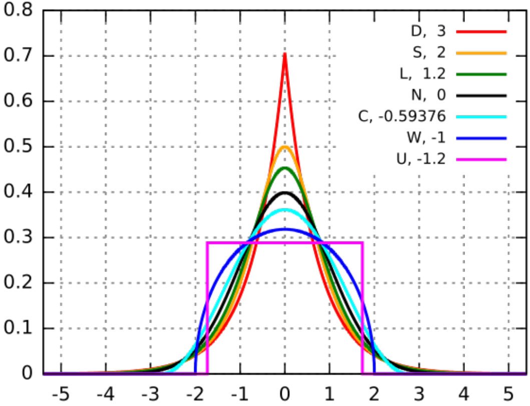


Figure: Distributions with mean zero and variance 1 distributions but different kurtosis. Source: Wikipedia

Jensen's Inequality

- Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ a convex function with $g(\mathbb{E}[X]) < \infty$
- Then $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$
- If g is concave, then the inequality is reversed

Proof

$g : \mathbb{R} \to \mathbb{R}$ is convex if for any $x_0 \in \mathbb{R}$, we can find m such that $g(x) \geq g(x_0) + m(x - x_0)$

$$\begin{aligned}
\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x) &\geq \int_{-\infty}^{\infty} \left[g(\mathbb{E}[X]) + m(x - \mathbb{E}[X]) \right] dF_X(x) \\
&= g(\mathbb{E}[X]) + m \int_{-\infty}^{\infty} (x - \mathbb{E}[X]) dF_X(x) \\
&= g(\mathbb{E}[X])
\end{aligned}$$

Examples

- **Linear**: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- **Quadratic**: $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$
- **Geometric Mean**: $\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]}$
- **Reciprocal**: $\mathbb{E}[1/X] \geq 1/\mathbb{E}[X]$

Moment Generating Function

- The **moment generating function** (MGF) of a random variable X is defined as

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} dF_X(x)$$

  if the integral exists for $t \in (-h, h)$ for some $h > 0$

- The MGF (if it exists) uniquely determines the distribution of X

- The k-th moment of X is given by:

$$\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(t) \Big|_{t=0}$$

Proof Sketch

- Recall from the definition of exponential

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right]$$
$$= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}\left[X^k\right]$$

- Hence

$$\frac{d^n}{dt^n} M_X(t) = \mathbb{E}[X^n] + \sum_{k=n+1}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k]$$

Characteristic Function

- The **characteristic function** (CF) of a random variable X is defined as

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF_X(x)$$

  where $i = \sqrt{-1}$

- The CF always exists and uniquely determines the distribution of X