Introduction to probability and statistics

Master in Cognitive Science 2025-2026

Lecture 2

October, 2025

Objectives

- Introduce the concept of random variable, of distribution and its characteristics.
- Present some common discrete / continuous distributions.
- Define covariance, correlation.
- What are the joint, conditional, and marginal distributions?

Readings: Kass et al (2014): Chapters 3.2.1; 3.2.2; 3.2.3; 4.1; 4.2; 4.3.3; 5.1; 5.2.1; 5.3.1; 5.4.2.

Definition of a random variable

- A random variable X is a (measurable) function $X : \Omega \to \mathbb{R}$.
- Quantities measured from random events.

Example 1: Two rolls of six-sided dice
In this case, the space of possible events is

$$\Omega = \{(x, y); x = 1, \cdots, 6, y = 1, \cdots, 6\}.$$

For an outcome $\omega = (x, y) \in \Omega$, define the random variables:
$S(\omega) = x + y; M(\omega) = \max(x, y); X_1(\omega) = x; X_2(\omega) = y$.

Example 2: Imagine you are playing a game where you toss a coin three times. You gain one point for each head and lose one point for each tail. In this case, we can define the random variable X as your total score:

$$X = (\text{number of heads}) - (\text{number of tails}).$$

What is Ω in this case ?

Notation: we write X in capital letter to denote a random variable, and $X = x$ for the event that happened (or a realization of x), i.e; the random variable took the value x.

Example 3: Spike Count

In this example, we analyze the dataset `e060824citral` from the R package `STAR`. The raster plot below illustrates the spike times across repeated trials.

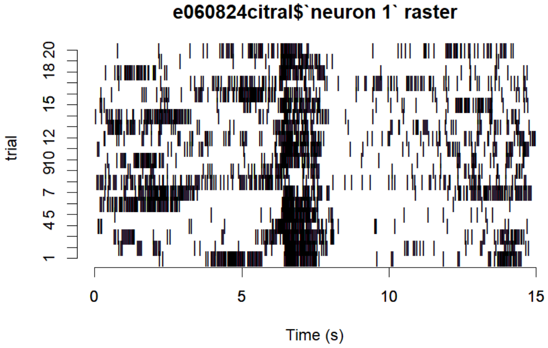


Figure 1: Raster plot of spike trains for dataset `e060824citral`. Each row corresponds to one trial, and tick marks indicate spike times.

Spike Count over 15 Seconds

Let X denote the number of spikes observed during a 15-second interval. The plot below shows the spike counts across trials.

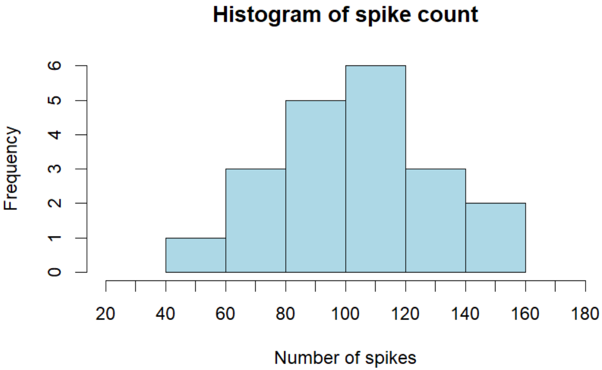


Figure 2: Spike counts X over 15-second intervals for each trial. Each bar (or point) represents the total spikes recorded in that trial.

A Variety of Random Objects

Random object	**Set**	**Example**
Number	$\mathbb{N}$	Spike count
Number	$\mathbb{R}$	Interspike time interval
Vector	$\mathbb{R}^n$	Interspike time intervals of n neurons
Matrix	$\mathbb{R}^{n\times n}$	Covariance matrix
String	A^n	Random DNA sequence $\{A, G, G, C, T\}$
Process	$\mathbb{R}^I$	Real-valued functions on the time interval I
Graph	$\{0,1\}^{V\times V}$	Graph on a set of vertices

- A random variable is **discrete** if it takes values in $\mathcal{X}$, a finite set (e.g., dice roll) or a countably infinite set (e.g., positive integers).
- A random variable is **continuous** if it can take all values in some interval (a, b), $a, b \in \mathbb{R}$.

Definition of a Distribution

The distribution of a random variable X is a table or a function that specifies its possible values and their probabilities.

Example 1: Sum of two rolls of a fair die.

Let X_1 be the outcome of the first roll and X_2 the outcome of the second roll. Define $S = X_1 + X_2$ as the sum.

S can take 11 possible values: $2, 3, \ldots, 12$. The probabilities are:

$$P(S = 2) = P((1,1)) = \frac{1}{36}, \quad P(S = 3) = P((1,2),(2,1)) = \frac{2}{36}, \ldots$$

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

Example 2: Consider $M = \max(X_1, X_2)$.

$$\begin{aligned} P(M=4) &= P(M \le 4) - P(M \le 3) \\ &= P(X_1 \le 4 \text{ and } X_2 \le 4) - P(X_1 \le 3 \text{ and } X_2 \le 3) \\ &= P(X_1 \le 4)P(X_2 \le 4) - P(X_1 \le 3)P(X_2 \le 3) \\ &= \left(\frac{4}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = \frac{7}{36}. \end{aligned}$$

k	1	2	3	4	5	6
$P(M=k)$	1/36	3/36	5/36	7/36	9/36	11/36

Example 3: Interspike time interval.
Consider the random variable T, representing the interspike time interval.
Suppose it is equally likely to observe a spike at any time in the interval $[0, 10]$ ms. Then:

$$P(2 \leq T \leq 3) = \frac{3-2}{10} = \frac{1}{10},$$

and since T is a continuous random variable,

$$P(T = 2) = 0.$$

Characteristics of a distribution

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

$$F(x) = P(X \leq x).$$

It completely determines the distribution of the r.v X.

Properties

The CDF satisfies:

- $0 \leq F(x) \leq 1, \ \forall x \in \mathbb{R}$.
- $\lim\limits_{x \to -\infty} F(x) = 0$.
- $\lim\limits_{x \to +\infty} F(x) = 1$.
- If $x \leq y$ then $F(x) \leq F(y)$.

Discrete distribution

The probability mass function (PMF): if X is **discrete**, its distribution can be characterized by its PMF

$$P(X \in A) = \sum_{x \in A} p_X(x),$$

where

$$p_X(x) = P(X = x), \quad \sum_{x \mathcal{X}} p_X(x) = 1.$$

Example: Maximum of Two Dice Rolls

Example: Let $M = \max(X_1, X_2)$, where X_1 and X_2 are the outcomes of two fair dice rolls.
Consider the event $A = \{1, 3, 5\}$. Then the probability is

$$P(M \in A) = P(M = 1) + P(M = 3) + P(M = 5) = \frac{5}{36} + \frac{7}{36} + \frac{9}{36} = \frac{21}{36}.$$

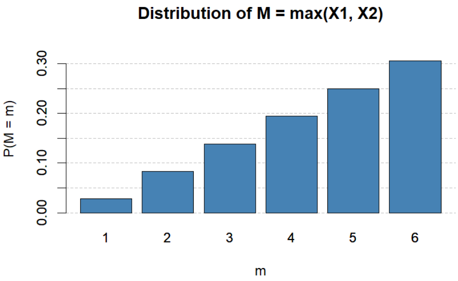


Figure 3: Bar plot of the distribution of $M = \max(X_1, X_2)$.

Continuous distribution

The probability density function (p.d.f): if X is a **continuous** random variable, the CDF F_X can be differentiated almost everywhere. In this case, we define the p.d.f as the derivative of the CDF:

$$f_X(x) := \frac{dF_X(x)}{dx} = F_X'(x).$$

In this case, for small $\Delta x, P\left(x \leq X \leq x + \Delta x\right) \approx f_X(x)\Delta x$.

Properties

The probability density function satisfies

- $f_X(x) \geq 0$.
- $P(X \in A) = \int_A f_X(x)\ dx$.
- $\int_{\mathbb{R}} f_X(x)\ dx = 1$.

Example: Continuous Uniform Distribution

Example: Continuous uniform distribution.

Consider the random variable T, representing the interspike time interval.

The cumulative distribution function (CDF) is:

$$F_T(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{10}, & 0 < x \leq 10, \\ 1, & x > 10. \end{cases}$$

The probability density function (PDF) is:

$$f_T(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10, \\ 0, & \text{otherwise.} \end{cases}$$

Think of the PDF as the probability per unit of T.

Expectation

For a random variable X, its expected value (*mean*), denoted as $\mathbb{E}[X]$ is

- The probability-weighted average of the possible values of X,
- The centered location of a distribution.

If X takes values in a discrete set $\mathcal{X}$, then:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p_X(x).$$

Example: A roll of a fair six-sided die. Let X be the random variable giving the number given by a roll of a die. Then

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

If X is a continuous random variable, then

$$\mathbb{E}[X] = \int x f(x)\, dx.$$

Example: Interspike interval.

$$\mathbb{E}[T] = \int_0^{10} \frac{t}{10}\, dt = \frac{1}{10}\left[\frac{t^2}{2}\right]_0^{10}$$
$$= \frac{1}{20}(100 - 0) = 5.$$

Properties

Let $a, b \in \mathbb{R}$ and X, Y two random variables. Then

- $\mathbb{E}[a] = a$.
- $\mathbb{E}[aX] = a\mathbb{E}[X]$.
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- $\mathbb{E}[1_{\{X=k\}}] = P(X = k)$.

Variance

The variance is a measure of dispersion of a random variable. If X is a r.v, then its **variance** is defined as

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

We often denote the mean by the symbol μ and the variance as σ^2. The positive number σ is called the standard deviation of X.

Quantiles

For $p \in [0, 1]$ the p-th quantile of a distribution with CDF $F(x)$ is the value η_p such that $F(\eta) = p$.

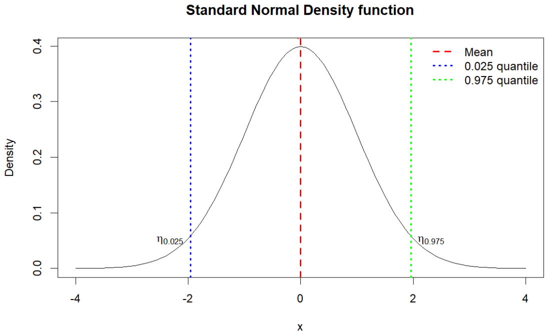


Figure 4: Visualizing the 0.025 and 0.975 Quantiles of a Normal Distribution

Bernoulli distribution

Let X be a random variable that takes value 1 if an event occurs 0 otherwise (e.g., tossing a coin, having Covid or not ...)

$$\begin{cases} P(X=1)=\theta, \\ P(X=0)=1-\theta. \end{cases}$$

We write $P(X=x)=\theta^x(1-\theta)^{1-x}$. This defines a **Bernoulli probability model**. We write shortly $X \sim Be(\theta)$.

Exercise

Let $X \sim Be(\theta)$.

- Obtain the PMF and CDF of X.
- Compute the expectation and variance of X.

Binomial distribution

Denote as $X_1, \ldots, X_n$, n independent realizations of the same random experiment with Bernoulli outcome. Consider the random variable

$$Y = \sum_{i=1}^{n} X_i.$$

Property

$P(Y = k) = \binom{k}{n}\theta^k(1-\theta)^{n-k}$.

We say that Y follows a **binomial distribution**, and we write shortly $Y \sim Bin(n, \theta)$.

Exercise

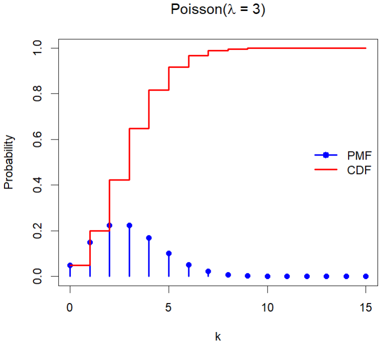
Let $X \sim Bin(20, 0.3)$.

- Using ggplot2, plot the PDF and CDF of X.
- Compute the expectation and the variance of X.

Poisson distribution

The Poisson p.m.f describes the probability that k events occurs in a specific unit of time or space. A random variable X has Poisson distribution if its PMF is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \; k = 0, 1, \ldots.$$



Uniform distribution

A random variable X has a uniform distribution, write $X \sim \mathcal{U}(a,b)$, then its pdf is given by:

$$f(x) = \frac{1}{b-a} 1_{\{x \in [a,b]\}}.$$

Exercise

Show that the expectation and variance of a uniform random variable in (a,b) is given by

$$\mathbb{E}(X) = \frac{a+b}{2}$$
$$Var(X) = \frac{(b-a)^2}{12}.$$

Gaussian distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$, its p.d.f is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, then $U = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. We say that U follows a **standard** normal distribution.

$$P\left(a \leq X \leq b\right) = P\left(\frac{a-\mu}{\sigma} \leq U \leq \frac{b+\mu}{\sigma}\right).$$

Expectation and variance

$$\begin{aligned}
&\mathbb{E}(X) = \mu \\
&Var(X) = \sigma^2, \\
&IQR = \eta_{0.75} - \eta_{0.25} = 1.349\sigma.
\end{aligned}$$

Other distributions

- Exponential distribution. $T \sim Exp(\lambda)$, can be used to model waiting time between two events.
- Gamma distribution, which is a generalization of the exponential distribution, we denote this distribution by $Ga(\alpha, \beta)$.
- Beta distribution denoted by $Be(\alpha, \beta)$. (We will use this distribution in Bayesian inference).
- Student's t-distribution t_ν, Chi-squared distribution χ^2, $\cdots$

There are many other interesting distributions, depending on the context and the model being used. For further reading, see the `Distributions.pdf` file in Ametice.

Covariance and correlation

Let X, Y be two random variables, and we would like to know if X and Y differs from their means toward the "same direction" and how strong is this effect. Hence, we define their covariance :

$$Cov(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

where $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$.
The covariance

- is positive, if $X - \mu_X$ and $Y - \mu_Y$ often have the same sign.
- is negative, if $X - \mu_X$ and $Y - \mu_Y$ often have opposite signs.

Exercise

Prove that the covariance is symmetric and linear in each of its arguments:

- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(aX + bY, Z) = a \cdot Cov(X, Z) + b \cdot Cov(Y, Z)$.

Correlation

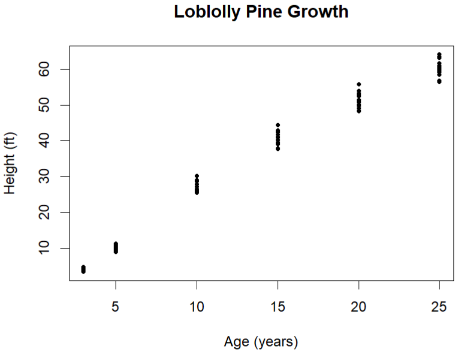
It is more meaningful to check how X and Y vary jointly in normalized units. Hence we compute the **correlation**

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

The value of the correlation coefficient always lies between -1 and 1, where values close to 1 or -1 indicate a strong relationship between the variables.

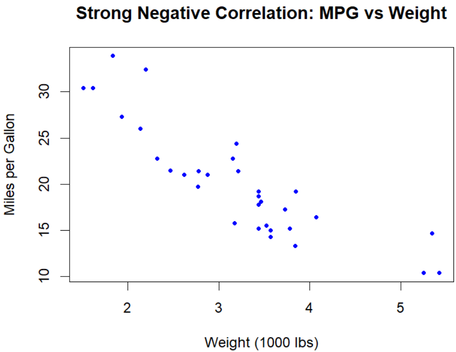
Example 1

Growth of Loblolly Pine Trees:
Consider the two random variables X, the age of the tree, and Y, the height. If we compute the correlation between X and Y, we find $Cor(X, Y) = 0.98$, which is close to 1. This suggests that X and Y are highly positively related; in other words, as the tree gets older, its height tends to increase.



Example 2:

Motor Trend Car Road Tests: In this example, we consider the dataset `datasets::mtcars`. We are interested in the weight of the car and its fuel efficiency (the distance the car can travel per unit of fuel). The correlation between this two covariates is -0.86, which is negative and close to -1.



Further Readings

- Read the definition of independence between two random variables.
- If two random variables X and Y are independent, determine the correlation between them.
- Read the definitions of joint, marginal, and conditional distributions, as well as Bayes' rule.