50008 - Probability and Statistics - Lecture 2

Oliver Killane

17/01/22

Introduction

Lecture Recording

Lecture recording is available here

Definition: Probability Space

$$(S, \mathcal{F}, P)$$

Models a random experiment where probability measure $P(E)$ is defined on subsets $E \subseteq S$ belonging to sigma algebra $\mathcal{F}$.

Within a sample space we can study quantities that are a function of randomly occurring events (e.g temperature, exchange rates, gambling scores).

Definition: Random Variable

A **random variable** is a mapping from the sample space to the real numbers, for example **random variable** X:

$$X : S \to \mathbb{R}$$

Each element in the sample space $s \in S$ is assigned to a numerical value by $X(s)$.

When referring to the value of a random variable we use its name, e.g X in $P(5 < X \leq 30)$

- **Simple** Finite set of possible outcomes. (e.g dice faces)
- **Discrete** Countable outcomes/support/range. (e.g distance (m))
- **Continuous** Can be a continuous range (e.g temp)

Example: Single Fair Dice Roll

$S = \{1, 2, 3, 4, 5, 6\}$, for any $s \in S . P(\{s\}) = \frac{1}{6}$.
We can define random variable X such that:

$$X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 4, X(5) = 5, X(6) = 6$$

Then we can use X:

$$P_X(1 < X \leq 5) = P(\{2, 3, 4, 5\}) = {}^2\!/_3$$

$$P_X(X \in \{2, 3\}) = P(\{2, 3\}) = {}^1\!/_3$$

We can also define random variable Y such that:

$$Y(\epsilon) = \begin{cases} 0 & \epsilon \text{ is odd} \\ 1 & \epsilon \text{ is even} \end{cases}$$

And hence:

$$P_Y(Y = 0) = P(\{1, 3, 5\}) = {}^1\!/_2$$

Induced Probability

The probability measure P defined on a sample space S induces a probability distribution on the random variable in $\mathbb{R}$ (distribution of its outcomes).

$$S_X = \{s \in S | X(s) \leq x\}$$

Such that:

$$P_X(X \geq x) = P(S_X)$$

Note that unless there is ambiguity, $P_X(\dots)$ will often be written as $P(\dots)$.

Example: Heads and Tails

We define random variable $X : \{H, T\} \to \mathbb{R}$ over the **continuum** $\mathbb{R}$ such that:

$$X(T) = 0 \text{ and } X(H) = 1$$

$$S_X = \begin{cases} \emptyset & \text{if } x < 0 \\ \{T\} & \text{if } 0 \leq x < 1 \\ \{H, T\} & \text{if } x \geq 1 \end{cases}$$

X represents the number of heads flipped.

$$P_X(X \leq x) = P(S_X) = \begin{cases} P(\emptyset) = 0 & \text{if } x < 0 \\ P(\{T\}) = 1/2 & \text{if } 0 \leq x < 1 \\ P(\{H, T\}) = 1 & \text{if } x \geq 1 \end{cases}$$

Now we can use X to compactly show probabilities.

$$P_X(X = 1) = 1/2$$

Example: Multiple Coin Flips

$$S = \{TTT, TTH, THT, HTT, THH, HHT, HTH, HHH\}$$

We can define X (number of heads):

$$X(s) = \begin{cases} 0 & s = TTT \\ 1 & s \in \{TTH, THT, HTT\} \\ 2 & s \in \{THH, HHT, HTH\} \\ 3 & s = HHH \end{cases}$$

Hence given 3 coin tosses:

$P_X(X > 1)$	More than one head
$P_X(X < 3)$	Not all heads
$P_X(X \leq 1)$	At least one head

Definition: Support/Range

The set of all possible values of a random variable X:

$$\mathbb{X} \equiv supp(X) \equiv X(S) = \{x \in \mathbb{R} | \exists s \in S.X(s) = x\}$$

As S contains all possible experiment outcomes, $supp(X)$ contains all possible values/outcomes for the random variables X.

$$P_X(X \leq x) \text{ is defined for all } x \in supp(X)$$

Cumulative Distributions

Definition: Cumulative Distribution Function (F_X)

The cumulative distribution function (cfd) of a random variable X is the probability where X takes some value less than or equal to some x:

$$F_X : \mathbb{R} \to [0, 1] \text{ such that } F_X(x) = P_x(X \leq x)$$

To be a valid cfd, 3 criteria must be met:

1. **Probability between 0 and 1** $\forall x \in \mathbb{R}. 0 \leq F_X(x) \leq 1$
2. **Monotonicity** $\forall x_1, x_2 \in \mathbb{R} x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
3. **Infinite Bounds** $F_X(-\infty) = 0, F_X(\infty) = 1$

For any random variable a **cfd** is right-continuous (a result of monotonicity).

$$x_1 > x_2 > x_3 ... > x \Rightarrow F_X(x_1) >= F_X(x_2) >= ... >= F_X(x)$$

We can determine the probability over finite intervals using the cumulative distribution:

$$\text{for } (a, b] \subseteq \mathbb{R} \ P_X(a < X \leq b) = F_X(b) - F_X(a)$$

Distributions

Lecture Recording

Lecture recording is available here

Definition: Probability Mass Function (p_X)

Also called **probability function** gives the probability that a discrete random variable is exactly equal to a value.

The sample space S is mapped onto elements in the **support** of X (one-to-one).

We can then partition the sample space into a countable, disjoint collection od event subsets:

$$s \in E_i \Leftrightarrow X(s) = x_i, i = 1, 2 \ldots$$

A probability mass function is valid if and only if:

1. **No negative probabilities** $\forall x \in supp(X).\ p_X(x) \geq 0$
2. **Probabilities sum to 1** $\sum_{x \in supp(x)} p_X(x) = 1$

Discrete Random Variable

For a **discrete random variable** we define the probability mass function as:

$$p_X(x_i) = P(X = x_i) = P(E_i) \text{ where } x_i \in supp(X) \text{ and } x_i \text{ is the outcome of event } E_i$$

We can also define using **cfds**:

$$F_X(x_i) = \sum_{j=1}^{i} p_X(x_j) \Leftrightarrow p_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \quad \text{where } i = 2, 3 \ldots$$

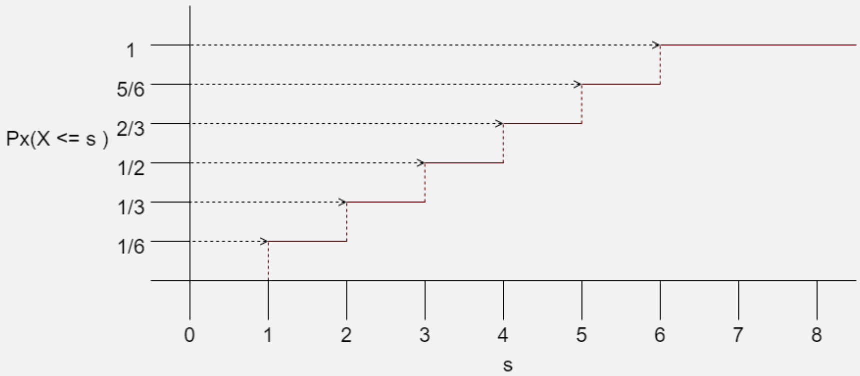
Or more simply:

$$p_X(x_i) = P_X(X = x_i) = P(X \leq x_i) - P(X \leq x_{i-1}) = F_X(x_i) - F_X(x_{i-1})$$

When graphed, F_X is a monotonically increasing, stepped function with jumps at points in $S(X)$.

Example: Six Sided Dice

Here we have X representing the value of the dice roll. We can plot the cumulative distribution (showing probability a dice roll is less than or equal to a given value).



Discrete CFDs have several properties:

- **Limiting Cases**
$$\lim_{x \to -\infty} F_X(x) = 0 \quad \lim_{x \to \infty} F_X(x) = 1$$
At ∞ the whole set of outcomes is covered, probabilities sum to 1. At $-\infty$ none are covered.
- **Continuous from the right**
$$\text{For } x \in \mathbb{R} \lim_{h \to 0^+} F_X(x+h) = F_X(x)$$
Moving from the right to the left the probability will reduce and tend towards the value.
- **Non-Decreasing**
$$a < b \Rightarrow F_X(a) \leq F_X(b)$$
As it is cumulative, the value can only grow larger moving right.
- **Can cover a range**
$$\text{For } a < b.\ P(a < X \leq b) = F_X(b) - F_X(a)$$

Definition: Poisson Distribution

A discrete probability distribution expressing the probability of a given number of events occuring in a fixed time interval, given a constant mean.

$$Pois(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ where } k \text{ is the number of occurrences}$$

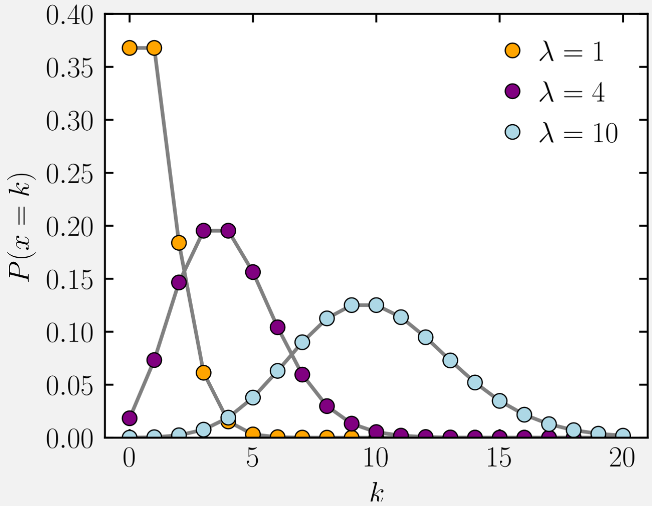
e.g What is the probability exactly 7 people buy pizzas at a stall in one hour, given on average is 4 people per hour?

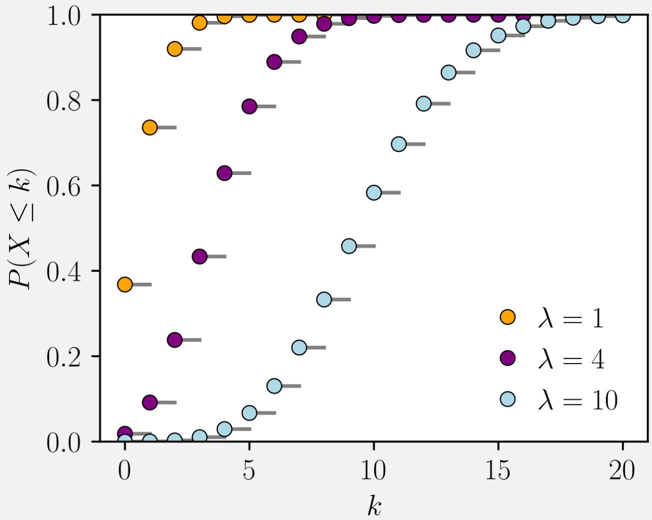
$$X \approx Poisson(4)$$

For a poisson distribution the mean (expected) and variance are equal.

$$E(X) = Var(X)$$

$$P(X = 7) = \frac{4^7 e^{-4}}{7!}$$





Link with Statistics

We can consider a set of data as realisations of a random variable defined on some underlying population of the data.

- Frequency histogram is an empirical estimate for the **pmf**.
- Cumulative histogram is an empirical estimate of the **cdf**.

Expectation

Definition: Expected Value

The expectation of a **discrete random variable** X is:

$$E_X(X) = \sum_x xp(x)$$

Also referred to as μ_X it is the mean value of the distribution.

$$E(g(X)) = \sum_x g(x)p_X(x)$$

$$E(a \times X + b) = a \times E(X) + b$$

$$E(a \times g(X) + b \times f(X)) = a \times E(g(X)) + b \times E(f(X))$$

Given another distribution Y:

$$E(X + Y) = E(X) + E(Y)$$

Example: Dice Rolls

Given random variable X representing the value of a dice roll:

$$X(n) = n \text{ where } 1 \geq n \geq 6$$

$$P(X = x) = \begin{cases} {}^1\!/_6 & 1 \geq n \geq 6 \\ 0 & otherwise \end{cases}$$

We can get the expected as:

$$E(X) = {}^1\!/_6 \times 1 + {}^1\!/_6 \times 2 + {}^1\!/_6 \times 3 + {}^1\!/_6 \times 4 + {}^1\!/_6 \times 5 + {}^1\!/_6 \times 6 = {}^{21}\!/_6 = 3.5$$

We can base scoring on the dice roll:

$$score(x) = 4 \times x + 2$$

Hence we can calculate that the expected score is $E(score(X)) = 4 \times 3.5 + 2 = 16$.

Example: Dice and Coins

Given random variable D of a fair dice, and fair coin C:

$$P(D=x) = \begin{cases} 1/6 & 1 \geq n \geq 6 \\ 0 & otherwise \end{cases} \quad \text{and} \quad P(C=x) = \begin{cases} 1/2 & x \in \{H, T\} \\ 0 & otherwise \end{cases}$$

Given $score = dice\ roll + 1$ if coin flip is heads what is the expected score?

$$E(D) = 3.5 \quad E(C) = 0.5 \quad E(score) = 3.5 + 2 * 0.5 = 4.5$$

Variance

Lecture Recording

Lecture recording is available here

Definition: Moment

A function which measures the shape of a function's graph.

The n^{th} moment of a random variable is the expected value of its n^{th} power:

$$n^{th} \text{ moment of } X = \mu_X(n) = E(X^n) = \sum_x x^n p(x)$$

- **First Moment** The expected value.
- **Central Moment** The variance $(E[(X - E(X))^2])$
- **Standardized Moment** The skew $(\frac{E(X - E(X))^3}{sd(X)^3})$

Definition: Variance

The expectation of the deviation from the expected/mean value squared.

$$Var(X) = Var_X(X) = \sigma_X^2 = E[(X - E(X))^2] = E(X^2) - (E(X))^2$$

Note that:

$$Var(a \times X + b) = a^2 Var(X)$$

Definition: Standard Deviation

The square root of the variance.

$$\sigma_X = sd_X(X) = \sqrt{Var_X(X)}$$

Example: Dice Roll

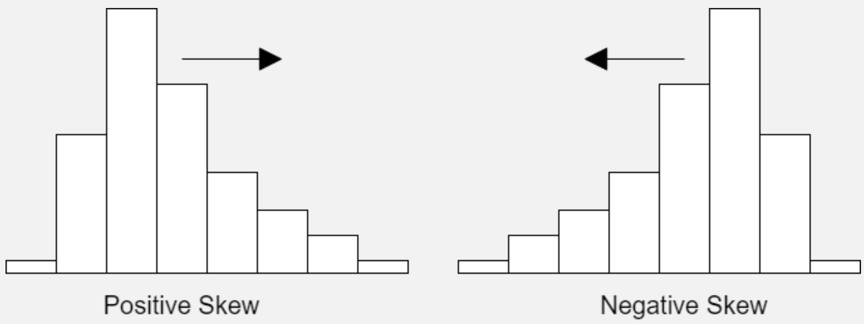
For a random variable representing a dice X:

$$Var(X) = E(X^2) - (E(X^2)) = \sum_x x^2 p(x) - (\sum_x xp(x))^2 = {}^{91}/_{6} - {}^{49}/_{4} = {}^{35}/_{12}$$

Definition: Skewness

A measure of asymmetry (the standardized moment):

$$\gamma_1 = \frac{E(X - E(X))^3}{sd(X)^3} = \frac{E(X - \mu)}{\sigma^3} \text{ where } \mu = E(X), \sigma = Sd(X)$$



Sum of Random Variables

Lecture Recording

Lecture recording is available here

Given random variables $X_1, X_2, \ldots, X_n$ (not necessarily independent, and potentially from different distributions), the sum is:

$$\text{The sum } S_n = \sum_{i=1}^{n} X_i \text{ and the average is } \frac{S_n}{n}$$

(The sum of the outcomes from all random variables)

The expected/mean value of S_n (expected value of the sum of all the random variables) is:

$$E(S_n) = \sum_{i=1}^{n} E(X_i) \text{ and } E(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} E(X_i)}{n}$$

- **All independent**

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) \text{ and } Var(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2}$$

- **All independent and Identically Distributed**
  Given that for all i, $E(X_i) = \mu_X$ and $Var(X_i) = \sigma_X^2$:

$$E(\frac{S_n}{n}) = \mu_X \quad \text{and} \quad Var(\frac{S_n}{n}) = \frac{\sigma_X^2}{n}$$

Important Discrete Random Variables

Lecture Recording

Lecture recording is available here

Definition: Bernouli Distribution

For an experiment with only two outcomes, encoded as 1 and 0.

For$X \sim Bernoulli(p)$ where $x \in S(X) = \{0, 1\}$ and $0 \le p \le 1$:

PMF	**Expected**	**Variance**
$p_X(x) = p^x(1-p)^{1-x}$	$\mu = E(X) = p$	$\sigma^2 = Var(X) = p(1-p)$

Definition: Binomial Distribution

Given n trials with two options, binomial models the number of outcomes. (e.g 3 coin tosses, number of ways to get 2 heads out of total outcomes).

For $X \sim Bionomial(n, p)$ where X takes values $0, 1, 2, \ldots, n$ and $0 \le p \le 1$:

PMF	**Expected**	**Variance**	**Skewness**
$p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$\mu = E(X) = np$	$\sigma^2 = Var(X) = np(1-p)$	$\gamma_1 = \dfrac{1-2p}{\sqrt{np(1-p)}}$

Note that choice is: $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$

Definition: Poisson Distribution

Given a constant mean number of events per fixed itme interval, provides probabilities of different numbers of events occuring. (e.g sell on average 6 cookies an hour, what is the probability 10 cookies are sold in a given hour).

For $X \sim Poisson(\lambda)$ where λ is the mean number of events and $\lambda > 0$:

PMF	Expected	Variance	Skewness
$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$\mu = E(X) = \lambda$	$\sigma^2 = Var(X) = \lambda$	$\gamma_1 = \frac{1}{\sqrt{\lambda}}$

Note that for poisson the skew is always positive (but decreases as λ increases), and $E(X) \equiv Var(X)$.

Definition: Geometric Distribution

A potentially infinite number of trials to get an outcome (e.g attempts required to shoot a target, given probability of hit).

We can consider it infinite Bernoulli trials $X_1, X_2, \dots,$ where $X = \{i | X_i = 1\}$ (X is number of attempts to get outcome 1).

For $X \sim Geometric(p)$ where X takes all values in $\mathbb{Z}^+ = \{1, 2, \dots\}$ and $0 \leq p \leq 1$:

PMF	Expected	Variance	Skewness
$p_X(x) = p(1-p)^{x-1}$	$\mu = E(X) = \frac{1}{p}$	$\sigma^2 = Var(X) = \frac{1-p}{p^2}$	$\gamma_1 = \frac{2-p}{\sqrt{1-p}}$

Alternatively we can consider the number of trials *before* getting an outcome:
If $X \sim Geometric(P)$ consider $Y = X - 1$ where Y takes values $\mathbb{N} = \{0, 1, 2, \dots\}$:

PMF	Expected	Variance	Skewness
$p_Y(x) = p(1-p)^y$	$\mu = E(Y) = \frac{1-p}{p}$	Unchanged	Unchanged

Definition: Discrete Uniform Distribution

Where a discrete number of outcomes are equally likely (e.g fair dice, colour wheel).

For $X \sim U(\{1, 2, \dots, n\})$:

PMF	Expected	Variance	Skewness
$p_X(x) = \frac{1}{n}$	$\mu = E(X) = \frac{n+1}{2}$	$\sigma^2 = Var(X) = \frac{n^2-1}{12}$	$\gamma_1 = 0$

Poisson Limit Theorem

We can use the **Binomial Distribution** to approximate the **Poisson Distribution**:

$$Poisson(\lambda) \approx Binomial(n, p) \text{ when } \lambda = np \text{ and } n \text{ is very large, } p \text{ is very small}$$

This is as for a **Poisson distribution** mean and variance are equal and for binomial, mean is np and variance $np(1-p)$ so as p gets smaller (and n larger) $np \approx np(1-p)$.