36225
Probability

ZhenTing Liu

Contents

Laws of probability

Chapter 1

1.1 independence

> **Note:-**
>
> Two events A,B are said to be independent if P(A|B) = P(A) **and** P(B|A) = P(B), note that this is a conjunction not a disjunction, if only one of the two conditions is satisfied, it is not sufficient enough to conclude A,B's independence. Naturally, to tell if two evens are dependent,we simply negate above statement, yielding A,B dependent if P(A|B) ≠ P(A) **or** P(B|A) ≠ P(B)

> **Definition 1.1.1**
>
> consider two events A,B then P(B) = P(B|A)P(A) + P(B|A^c)P(A^c), apply bayes rule we have
>
> $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$
>
> more generally
>
> $$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)}$$

Random Variables

Chapter 2

2.1 Probability and Mass functions

A **discrete random variable** X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined to be a mapping $X : \Omega \mapsto \mathbb{R}$ such that

$$\text{the image } X(\Omega) \text{ is a countable subset of } \mathbb{R} \text{ and}$$
$$\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$$

A discrete random variable X takes value in $\mathbb{R}$, but we cannot predict the actual value of X with certainty since the underlying experiment involves chance. Instead, we would like to measure the probability that X takes a given value, x say.

The most interesting things about a discrete random variable are the values that it may take and the probability associated with these values. If X is a discrete random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then its **image** imX is the image of Ω under X, that is, the set of values taken by X.

Henceforth, we abbreviate events of the form $\{\omega \in \Omega : X(\omega) = x\}$ to the more convenient form $\{X = x\}$.

Definition 2.1.1: The (probability) mass function

the pmf of the discrete random variable X is the function $p_X : \mathbb{R} \mapsto [0, 1]$ defined by

$$p_X(x) = \mathbb{P}(X = x)$$

Thus, $p_X(x)$ is the probability that the mapping X takes the value x. Note that imX is countable for any discrete random variable X, and

$$p_X(x) = 0 \quad \text{if } x \notin \text{im } X, \tag{2.1}$$

$$\sum_{x \in \text{im} X} p_X(x) = \mathbb{P}(\bigcup_{x \in \text{im} X} \{\omega \in \Omega : X(\omega) = x\}) = \mathbb{P}(\Omega) = 1 \tag{2.2}$$

$$0 \leqslant p_X(x) \leqslant 1 \forall x \tag{2.3}$$

To simplify things, here are my understanding of random variable with less rigrous math definitions.

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

---

Note:-

The realization of a random variable is the result of applying the random variable (i.e., function) to an observed outcome of a random experiment. This is what the experimenter actually observes. The realization of a random variable is typically denoted using lowercase italicized Roman letters, e.g., x is a realization of X.

Example 2.1.1

Suppose we flip a fair (two-sided) coin $n \geqslant 2$ 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 $x \in \{0, 1, ..., n\}$, i.e., we could observe any number of heads between 0 and n.

2.2 Probability Mass and Density Functions

Definition 2.2.1: Probability mass function

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:

$$f(x) = P(X = x) \geqslant 0 \quad \forall x \in S \tag{2.4}$$

$$\sum_{x \in S} f(x) = 1 \tag{2.5}$$

Example 2.2.1

See below for the PMF corresponding to the coin flipping example with $n = 5$ and $n = 10$ independent flips. As we might expect, the most likely realizations of X are those in which we observe approximately $n/2$ heads. Note that it is still possible to observe $x = 0$ or $x = n$ heads, but these extreme results are less likely to occur.
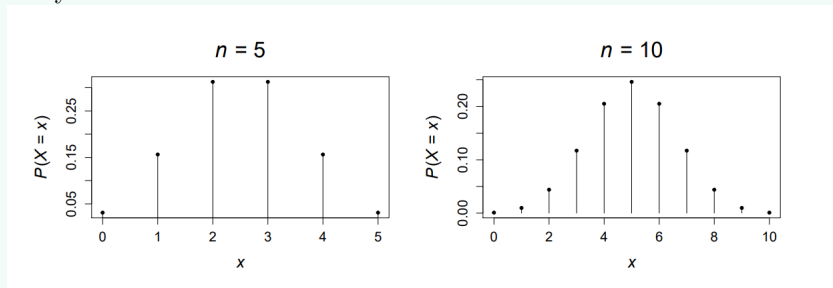
Definition 2.2.2: Probability density function

The probability density function (PDF) of a continuous random variable X is the function $f(\cdot)$ 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 have properties:

$$f(x) \geqslant 0 \quad \forall x \in S$$

$$\int_a^b f(x)dx = P(a < X < b) \geqslant 0 \quad \forall a, b \in S \quad \text{satisfying} \quad a < b$$

$$\int_{x \in S} f(x)dx = 1$$

Note that PMF is inclusively bounded above by 1 while PDF does not specify an upper bound.

Example 2.2.2

Suppose that we randomly spin the second hand around a clock face n independent times. Define Z_i as

the position where the second hand stops spinning on the i-th replication, and define $X = \frac{1}{n} \sum_{i=1}^{n} Z_i$ as the average of the n spin results. Note that the realizations of X are any values $x \in [0, 12]$, which is the same domain as Z_i for $i = 1, \ldots, n$. See figure below for the PDF with $n = 1$ and $n = 5$ independent spins. Note that with $n = 1$ spin, the PDF is simply a flat line between 0 and 12, which implies that $P(x < X < x + 1)$ is equal for any $x \in \{0, \ldots, 11\}$. This makes sense given that, for any given spin, the second hand could land anywhere on the clock face with equal probability. With $n = 5$ spins, the PDF has a bell shape, where values around the midpoint of $x = 6$ have the largest density. With $n = 5$ spins, we have that $\int_4^8 f(x)dx = P(4 < X < 8) = 0.80$ (i.e., about 80% of the realizations of X are between 4 and 8).
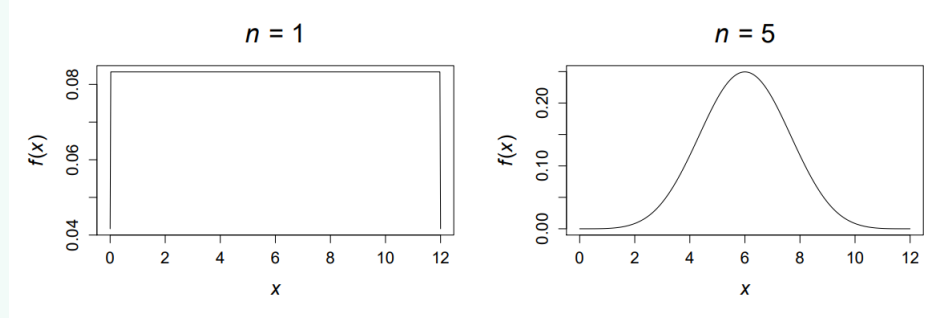
2.3 Cumulative Distribution Function

> **Definition 2.3.1: CDF**
>
> The cumulative distribution function (CDF) of a random variable X is the function $F(\cdot)$ that returns the probability $P(X \leqslant x)$ for any $x \in S$. Note that CDF is a function from S to $[0, 1]$, i.e., $F : S \mapsto [0, 1]$.

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

$$F(x) = \sum_{z \in S, z \leqslant x} f(x) \quad \text{for discrete random variables}$$

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

Furthermore, note that probabilities can be written in terms of the CDF, such as

$$P(a < X \leqslant b) = F(b) - F(a)$$

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

$$f(x) = F(x) - \lim_{a \to x^-} F(a) \quad \text{for discrete case}$$

$$f(x) = \frac{dF(x)}{dx} \quad \text{for continuous random var}$$

2.4 Expectation

Consider a fair die. If it were thrown a large amount of times, each of the possible outcomes $1, 2, \ldots, 6$ would appear on about one-sixth of the throws, and the average of the numbers observed would be approximately

$$\tfrac{1}{6} + \tfrac{1}{6}(2) + \tfrac{1}{6}(3) + \cdots + \tfrac{1}{6}(6) = \tfrac{7}{2}$$

which we call **mean value**. This notion of mean value is easily extended to more general distributions as follows.

> **Definition 2.4.1**
>
> If X is a **discrete random variable**, the expectation of X is denoted by $\mathbb{E}(X)$ and defined by
>
> $$\mathbb{E}(X) = \sum_{x \in \text{im } X} x \mathbb{P}(X = x)$$

and is often rewritten as

$$\mathbb{E}(X) = \sum_x x \mathbb{P}(X = x) = \sum_x x p_X(x)$$

2.5   Random

> **Definition 2.5.1: Normed Linear Space and Norm $\| \cdot \|$**
>
> Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A norm on V is function $\| \cdot \| \ V \to \mathbb{R}_{\geqslant 0}$ satisfying
>
> ① $\|x\| = 0 \iff x = 0 \ \forall \ x \in V$
>
> ② $\|\lambda x\| = |\lambda| \|x\| \ \forall \ \lambda \in \mathbb{R}(\text{or } \mathbb{C}), x \in V$
>
> ③ $\|x + y\| \leqslant \|x\| + \|y\| \ \forall \ x, y \in V$ (Triangle Inequality/Subadditivity)
>
> And V is called a normed linear space.
> ● Same definition works with V a vector space over $\mathbb{C}$ (again $\| \cdot \| \to \mathbb{R}_{\geqslant 0}$) where ② becomes $\|\lambda x\| = |\lambda| \|x\|$
> $\forall \ \lambda \in \mathbb{C}, x \in V$, where for $\lambda = a + ib$, $|\lambda| = \sqrt{a^2 + b^2}$

> **Example 2.5.1** ($p-$Norm)
> $V = \mathbb{R}^m$, $p \in \mathbb{R}_{\geqslant 0}$. Define for $x = (x_1, x_2, \cdots, x_m) \in \mathbb{R}^m$
>
> $$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_m|^p \right)^{\frac{1}{p}}$$
>
> (In school $p = 2$)

Special Case $p = 1$: $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_m|$ is clearly a norm by usual triangle inequality.
Special Case $p \to \infty$ ($\mathbb{R}^m$ with $\| \cdot \|_\infty$): $\|x\|_\infty = \max\{|x_1|, |x_2|, \cdots, |x_m|\}$
For $m = 1$ these $p-$norms are nothing but $|x|$. Now exercise

> **Question 1**
>
> Prove that triangle inequality is true if $p \geqslant 1$ for $p-$norms. (What goes wrong for $p < 1$?)

Solution: **For Property ③ for norm-2**

When field is $\mathbb{R}$:

We have to show

$$\sum_i (x_i + y_i)^2 \leqslant \left(\sqrt{\sum_i x_i^2} + \sqrt{\sum_i y_i^2} \right)^2$$

$$\implies \sum_i (x_i^2 + 2x_i y_i + y_i^2) \leqslant \sum_i x_i^2 + 2\sqrt{\left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]} + \sum_i y_i^2$$

$$\implies \left[\sum_i x_i y_i\right]^2 \leqslant \left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]$$

So in other words prove $\langle x, y \rangle^2 \leqslant \langle x, x \rangle \langle y, y \rangle$ where

$$\langle x, y \rangle = \sum_i x_i y_i$$

> **Note:-**
>
> - $\|x\|^2 = \langle x, x \rangle$
>
> - $\langle x, y \rangle = \langle y, x \rangle$
>
> - $\langle \cdot, \cdot \rangle$ is $\mathbb{R}$−linear in each slot i.e.
>
> $$\langle rx + x', y \rangle = r\langle x, y \rangle + \langle x', y \rangle \text{ and similarly for second slot}$$
>
> Here in $\langle x, y \rangle$ x is in first slot and y is in second slot.

Now the statement is just the Cauchy-Schwartz Inequality. For proof

$$\langle x, y \rangle^2 \leqslant \langle x, x \rangle \langle y, y \rangle$$

expand everything of $\langle x - \lambda y, x - \lambda y \rangle$ which is going to give a quadratic equation in variable λ

$$
\begin{aligned}
\langle x - \lambda y, x - \lambda y \rangle &= \langle x, x - \lambda y \rangle - \lambda \langle y, x - \lambda y \rangle \\
&= \langle x, x \rangle - \lambda \langle x, y \rangle - \lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle \\
&= \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle
\end{aligned}
$$

Now unless $x = \lambda y$ we have $\langle x - \lambda y, x - \lambda y \rangle > 0$ Hence the quadratic equation has no root therefore the discriminant is greater than zero.

When field is $\mathbb{C}$:

Modify the definition by

$$\langle x, y \rangle = \sum_i \overline{x_i} y_i$$

Then we still have $\langle x, x \rangle \geqslant 0$

9

2.6 Algorithms

Algorithm 1: what

Input: This is some input
Output: This is some output
/* This is a comment */

1 some code here;
2 $x \leftarrow 0$;
3 $y \leftarrow 0$;
4 **if** $x > 5$ **then**
5 | x is greater than 5 ; // This is also a comment
6 **else**
7 | x is less than or equal to 5;
8 **end**
9 **foreach** y *in 0..5* **do**
10 | $y \leftarrow y + 1$;
11 **end**
12 **for** y *in* $0..5$ **do**
13 | $y \leftarrow y - 1$;
14 **end**
15 **while** $x > 5$ **do**
16 | $x \leftarrow x - 1$;
17 **end**
18 **return** Return something here;