

# Probability Notes

April 2024

**Definition 1.** Let  $X$  be a set.  $\Sigma \subseteq P(X)$  is a  **$\sigma$ -algebra** if it satisfies the following three properties:

- $X \in \Sigma$ .
- $\Sigma$  is closed under complementation.
- $\Sigma$  is closed under countable unions.

**Lemma 0.1.** Sigma algebras are closed under countable intersections.

*Proof.* Follows by De Morgan's Laws. □

The smallest possible  $\sigma$ -algebra is  $\{\emptyset, X\}$ . It is also the intersection of all possible  $\sigma$ -algebras on  $X$ .

The largest possible  $\sigma$ -algebra is  $P(X)$ .

Elements of  $\sigma$ -algebras are called **measurable sets**.

**Lemma 0.2.** The intersection of a collection of  $\sigma$ -algebras is a  $\sigma$ -algebra.

**Lemma 0.3.** For any collection  $A$  of subsets of  $\Omega$ , there's a smallest  $\sigma$ -algebra containing  $A$ . This is called the  **$\sigma$ -algebra** generated by  $A$  and denoted  $\sigma(A)$ .

*Proof.* Let  $S = \{\mathcal{F} \subseteq P(X) : \mathcal{F} \text{ is a } \sigma\text{-algebra containing } A\}$ . We'll show that

$$\sigma(A) = \bigcap_{\mathcal{F} \in S} \mathcal{F}$$

□

**Lemma 0.4.** Let  $\Omega = [0, 1]$ . Let  $\Sigma$  be a  $\sigma$ -algebra that contains all closed intervals  $[a, b]$  in  $[0, 1]$ . Then,  $\Sigma$  contains all open and closed sets in  $[0, 1]$ .

In particular, the  $\sigma$ -algebra generated by all open sets is equivalent to  $\Sigma$ .

*Proof.* It suffices to show that  $\Sigma$  contains all open sets since closed sets are just complements of open sets and  $\Sigma$  is closed under complementation.

Since open intervals are a basis for the standard topology on  $[0, 1]$ , it suffices to show that  $\Sigma$  contains all open intervals.

Let  $(a, b) \subseteq [0, 1]$ . *Rest of the argument is straightforward.* □

**Definition 2.** If  $X$  is a metric/topological space, the  $\sigma$ -algebra generated by all open sets is called a **Borel  $\sigma$ -algebra**.

When is office hours?

1.  $\Pr[\Omega] = 1$
2.  $\forall A \in \mathcal{F} : \Pr[A] \geq 0$
3. Let  $A_n$  be a disjoint sequence of events. Then,  $\sum_{n=1}^{\infty} \Pr(A_n) = \Pr(\bigcup_{n=1}^{\infty} A_n)$

If you replace the countable additivity axiom with a finite additivity axiom, things go crazy. Uncountable additivity also doesn't work, things break.

**Definition 3.** A **random variable** is a function  $X : \Omega \rightarrow \mathbb{R}$  that satisfies

$$\forall \lambda \in \mathbb{R} : \{\omega \in \Omega : X(\omega) \leq \lambda\} \in \mathcal{F}$$

Since  $\mathcal{F}$  is closed under complements, equivalently, we can require  $\forall \lambda \in \mathbb{R} : \{\omega \in \Omega : X(\omega) > \lambda\} \in \mathcal{F}$ .

**Lemma 0.5.** Let  $X, Y$  be random variables. Then,  $X + Y$  defined by pointwise addition is a random variable.

*Proof.* The fact that  $X + Y$  is a function is immediate. Thus, it suffices to show that  $\forall \lambda \in \mathbb{R} : \{\omega \in \Omega : X(\omega) + Y(\omega) > \lambda\} \in \mathcal{F}$ .

What we'd like to have is the following:

$$\{X + Y > \lambda\} = \bigcup_{\mu \in \mathbb{R}} (\{X > \mu\} \cap \{Y > \lambda - \mu\})$$

This is true, but not helpful, since the uncountable union doesn't help us. We'll use the density of the rationals to bring the cardinality down. We'd like to show the following:

$$\{X + Y > \lambda\} = \bigcup_{q \in \mathbb{Q}} (\{X > q\} \cap \{Y > \lambda - q\})$$

Let's show the containment both ways.

□

**Definition 4.** The **cumulative distribution function (cdf)** of a random variable  $X : \Omega \rightarrow \mathbb{R}$  is defined to be  $F_X : \mathbb{R} \rightarrow [0, 1]$  by

$$F_X(x) = \Pr(\{\omega \in \Omega : X(\omega) \leq x\})$$

The cdf characterizes the distribution of  $X$ .

We say that  $X \sim \text{Bernoulli}(p)$  to mean "is distributed as".

In this case:

**Definition 5.** A function  $f$  is **cadlag** if it's continuous from the right and limits exist from the left. More formally,

•

**Proposition 0.5.1.** A function  $F : \mathbb{R} \rightarrow [0, 1]$  is the cumulative distribution function of a random variable if and only if

- $\lim_{x \rightarrow -\infty} F(x) = 0$
- $\lim_{x \rightarrow \infty} F(x) = 1$
- It is cadlag. (It is continuous from the right and limits exist from the left.)

Professor calls this the monotone convergence of probabilities.

If  $A_1 \supseteq A_2 \supseteq A_3$ , then

Then,  $\Pr(\bigcap A_n) = \lim_{n \rightarrow \infty} \Pr(A_n)$ .

If  $A_1 \subseteq A_2 \subseteq A_3 \dots$ , then

Then,  $\Pr(\bigcup A_n) = \lim_{n \rightarrow \infty} \Pr(A_n)$ .

These two properties are equivalent to the countable additivity of probability measures.

- If  $F$  has these three properties, then we can build a random variable with this CFG – e.g. the inverse distribution method.
- If  $X$  is a random variable, then  $\lim_{y \rightarrow x^-} F_x(y) = \Pr(X < x)$ . NOT LESS THAN OR EQUAL TO-LESS THAN.

**Definition 6.** A **discrete random variable**  $\mathbf{X}$  on the probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  is a mapping  $X : \Omega \rightarrow \mathbb{R}$  such that

- $X(\Omega)$  is at most countable.
- $\forall x \in \mathbb{R} : \{w \in \Omega : X(w) = x\} = X^{-1}(x) \in \mathcal{F}$ .

**Definition 7.** The **probability mass function (pmf)** of the discrete random variable  $\mathbf{X}$  is the function  $p_X : \mathbb{R} \rightarrow [0, 1]$  defined by  $p_X(x) = \mathcal{P}(X^{-1}(x))$ .

## 1 Birthday Paradox

## 2 Geometric Random Variables

**Definition 8.**  $X$  has a **geometric distribution** if  $\forall k \in \mathbb{Z}^+ : \mathcal{P}(X = k) = (1 - p)^{k-1}p$  for some  $p \in (0, 1)$ .

Let's now examine the probability that  $X \sim \text{Geometric}(p)$  is even.

### 3 Bernoulli Random Variables

**Definition 9.**  $X$  has a **Bernoulli distribution** if  $\mathcal{P}(X = 0) = p$  and  $\mathcal{P}(X = 1) = 1 - p$  for some  $p \in [0, 1]$ .

**I carry out a sequence of independent Bernoulli( $p$ ) trials. Independently of me, you carry out a sequence of independent Bernoulli( $q$ ) trials. Here  $p, q \in (0, 1)$ . What is the probability we both have our first success after the same number of trials?**

## 4 Binomial Distribution



## 5 Variance

Unlike the expectation, the variance of a random variable violates both conditions of linearity.

We have that

$$\text{var}(cX) = c^2 \text{var}(X)$$

## 6 Exponential Random Variable

## 7 Poisson Distribution

**Definition 10.**  $X$  has a **Poisson distribution** with parameter  $\lambda > 0$  if  $X$  takes values in  $\mathbb{Z}^+$  and

$$\mathcal{P}(X = k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

.

## 8 Normal (Gaussian) Distributions

$$f(z) = \frac{1}{2\pi} e^{-z^2/2}$$

To find the normalizing constant, we can use a trick with multivariable integration and polar coordinates. The point is that while switching to polar coordinates, a factor of  $r$  pops out due to the Jacobian which makes the integral easier.

It's in Stirling's approximation in my Notability notes. In fact, it's provably impossible to otherwise evaluate the indefinite integral with the regular techniques and elementary functions in closed form.

Notice that all moments of the normal distribution are 0 since the probability density function of the normal distribution is an odd function.

Then, we can define  $X \sim N(\mu, \sigma^2)$  if  $X = \mu + \sigma Y$  where  $Y$  is the standard normal distribution.

We can also reverse this and "standardize" a normal variable. Given  $Y \sim N(\mu, \sigma^2)$ , we can define the standard normal  $Z = \frac{Y-\mu}{\sigma}$ . Also notice that standardization gets rid of the units so interpreting it is straightforward.

## 9 Limsup and Liminf for Events

Let  $E_n$  be a sequence of events.

**Lemma 9.1.**

$$\{\text{infinitely many } E_k \text{ occur}\} = \left\{ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k \right\}$$

*Proof.*

□

This event is called  $\limsup_{k \rightarrow \infty} E_k$ .

## 10 Expectation

**Lemma 10.1.** Let  $X : \Omega \rightarrow \mathbb{Z}$  be a *non-negative* random variable. Then,

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathcal{P}(X \geq n)$$

Therefore  $\mathbb{E}(X)$  exists if and only if the sum on the right-hand side converges.

*Proof.* This is a simple rearrangement of sums, which converges to the same value because of absolute convergence.  $\square$

### 10.1 Problems

**I play the following game using a coin that lands heads with probability  $p$ . I start with  $X_0 = 1$  and at each stage I gamble all I have on the toss of the coin. If it lands heads I end up with twice what I started with; if it lands tails I lose everything. All coin tosses are statistically independent.**

## 11 Problems

Starting at the origin on the line, I take a step of one unit to the left or to the right with probability  $1/2$ . I do this repeatedly with independent steps. If I take  $2n$  steps, what is the probability that I find myself back at the origin?

Let  $X \sim \text{Binomial}(2n, 1/2)$ . We want to calculate  $\mathcal{P}(X = n)$ . The rest is straightforward calculation.

You and I independently complete this same random walk. What is the probability that we end up in the same location?

Suppose my knowledge/ignorance of the number of branches of a certain store (in my city) is given by the following probability law:

## 12 Poisson Processes

What if we want to compute the time of the  $n$ th success? Is there a probability distribution that models this? It turns out that this will be the Gamma distribution, i.e. the  $n$ th success will be modeled using  $\text{Gamma}(n, \lambda)$ .



## 13 Central Limit Theorem

Intuitively, here is the idea: the sum of  $n$  identical independently distributed random variables approach the normal distribution as  $n$  goes to infinity.

### 13.1 Resources

- 3Blue1Brown video

## 14 The Moment Problem

Note that some distributions (which happen to have all moments finite) are determined by finitely many moments. This is because the finitely many moments also uniquely determine the rest of the moment sequence.

**Example 14.0.1.** For example, if  $\mathcal{E}(X^2) = 0$ , then  $X$  is a "constant" random variable.

**Example 14.0.2.** For example, if  $\mathcal{E}(Y^2(1 - Y)^2) = 0$  and  $\mathcal{E}(Y) = p$ ,  $Y$  is a Bernoulli random variable with parameter  $p$  since  $f(x) = x^2(1 - x)^2$  only takes positive values and is only 0 at 0 and 1.

**Theorem 14.1.** If finitely many moments determine  $F_X$ ,  $X$  takes finitely many moments.

The proof is far beyond what we're capable of right now.

**Theorem 14.2.** Suppose  $m_k = \mathcal{E}(X^k)$  satisfy that for some  $a > 0$  the series

$$\sum_{k=0}^{\infty} \frac{a^{2k} m_{2k}}{(2k)!}$$

Then,  $\forall k \in \mathbb{Z}^{\geq 0} : \mathcal{E}(X^k) = \mathcal{E}(Y^k)$  implies  $F_X = F_Y$ .

*Proof.* Using  $|x^{2k+1}| \leq 1 + |x^{2k+2}|$ , we see that (\*) guarantees the convergence of

$$\phi_X(\xi) = \mathcal{E}(e^{i\xi X}) = \sum_{k=0}^{\infty} \frac{(i\xi)^k m_k}{(k)!}$$

for all  $|xi| \leq a$ .

Thus, moments satisfying (\*) tell us  $\phi_X(\xi)$  but only for small  $|\xi|$ .

However, knowing this for small  $\xi$  is not enough. WHY? UNDERSTAND THIS.

What follows is usually a dumb idea, but it works in this case. Let's try to find the power series of  $\phi_X(\xi)$  about  $\xi_0 = \pm a/2$ .

Well,

□

## 15 Chicken-Egg Problem

Galton-Watson Processes