

# Review and Background

Andrew Nobel

August, 2025

## Review of Probability

# Probability Model for Experiment with Uncertain Outcomes

1. A non-empty set  $\Omega$  whose elements  $\omega$  are the possible *outcomes* of the experiment
2. Subsets  $A \subseteq \Omega$  are called events. Event  $A$  occurs if outcome  $\omega \in A$
3. A probability measure  $\mathbb{P}$  assigns probabilities to events. Two properties
  - ▶  $\mathbb{P}(\emptyset) = 0$  and  $\mathbb{P}(\Omega) = 1$
  - ▶ If  $A_1, A_2, \dots$  are disjoint then  $\mathbb{P}(\cup_{i \geq 1} A_i) = \sum_{i \geq 1} \mathbb{P}(A_i)$

Interpretation:  $\mathbb{P}(A)$  = likelihood that  $A$  occurs when the experiment is performed

## Probability Models, cont.

**Fact:** Probabilities have the following properties

1.  $0 \leq \mathbb{P}(A) \leq 1$  for all events  $A$
2.  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$  for all events  $A$
3. If  $A, B$  are events and  $A \subseteq B$  then  $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. If  $A_1, A_2, \dots$  are any events then  $\mathbb{P}(\cup_{i \geq 1} A_i) \leq \sum_{i \geq 1} \mathbb{P}(A_i)$
5. Increasing events: if  $A_1 \subseteq A_2 \subseteq \dots$  then  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\cup_{i \geq 1} A_i)$
6. Decreasing events: if  $A_1 \supseteq A_2 \supseteq \dots$  then  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\cap_{i \geq 1} A_i)$

# Random Variables

**Setting:** Experiment with outcomes  $\Omega$ , events  $A \subseteq \Omega$ , and probability  $\mathbb{P}$

**Definition:** A random variable is a function  $X : \Omega \rightarrow \mathbb{R}$ . For  $B \subseteq \mathbb{R}$  we define

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

**Idea:**

- ▶  $X$  is a deterministic (non-random) function
- ▶  $X(\omega)$  is a number that depends on the outcome  $\omega$  of the experiment
- ▶  $X(\omega)$  inherits uncertainty from uncertainty in the outcome  $\omega$  of the experiment
- ▶  $\mathbb{P}(X \in B)$  is the probability of the event  $A = \{\omega : X(\omega) \in B\}$

## Specifying the Distribution of a Random Variable

# Distribution of a Random Variable

**Recall:** The distribution  $P$  of a random variable  $X$  is defined by

$$P(B) = \mathbb{P}(X \in B) \quad \text{for all } B \subseteq \mathbb{R}$$

## Note

- ▶  $P$  is a probability distribution on  $\mathbb{R}$
- ▶  $P$  models uncertainty in the observed values of  $X$  while  $\mathbb{P}$  models uncertainty in the outcomes of the underlying experiment
- ▶ In many inference problems the underlying probability space is implicit, e.g.,

$$X_1, \dots, X_n \text{ iid } \sim \mathcal{N}(0, 1)$$

# Cumulative Distribution Function

**Recall:** The cumulative distribution function (CDF) of a random variable  $X \sim P$  is the function  $F : \mathbb{R} \rightarrow [0, 1]$  defined by

$$F(x) = \mathbb{P}(X \leq x) = P((-\infty, x])$$

**Fact:** If  $F$  is a CDF then the following hold

1.  $F$  is non-decreasing: if  $x \leq y$  then  $F(x) \leq F(y)$
2.  $F(x) \rightarrow 1$  as  $x \rightarrow \infty$
3.  $F(x) \rightarrow 0$  as  $x \rightarrow -\infty$
4. If  $x_n$  decreases to  $x$  then  $F(x_n)$  decreases to  $F(x)$

A CDF can have jumps and flat bits. Last fact says that  $F$  is "continuous from the right"

## Indicator Functions

**Definition:** If  $\mathcal{X}$  is a set and  $A \subseteq \mathcal{X}$  the indicator function of  $A$  is given by

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Sometimes  $\mathbb{I}_A(x)$  is written  $\mathbb{I}(x \in A)$ . Basic properties

- ▶  $\mathbb{I}_{A^c} = 1 - \mathbb{I}_A$
- ▶  $\mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B$
- ▶  $\mathbb{I}_{A \cup B} = \max(\mathbb{I}_A, \mathbb{I}_B)$

**Key use:** Restricting range of integration or summation. For  $A \subseteq \mathcal{X}$  define

$$\int_A h(x) dx = \int_{\mathcal{X}} h(x) \mathbb{I}_A(x) dx \quad \sum_{x \in A} h(x) = \sum_{x \in \mathcal{X}} h(x) \mathbb{I}_A(x)$$

# Discrete Distributions

**Definition:** A distribution  $P$  on  $\mathbb{R}$  is discrete if there is a probability mass function (pmf)  $p : \mathbb{R} \rightarrow [0, 1]$  such that for every set  $B \subseteq \mathbb{R}$

$$P(B) = \sum_{x \in B} p(x) = \sum_{x \in \mathbb{R}} p(x) \mathbb{I}(x \in B)$$

- ▶ If  $X \sim P$  we write  $X \sim p$ . In this case  $p(x) = \mathbb{P}(X = x)$
- ▶ Examples: Bernoulli, binomial, Poisson, geometric, hypergeometric distributions

Technical note: There can be at most countably many  $x$  for which  $p(x) > 0$ , so the infinite sums above are well defined

# Continuous Distributions

**Definition:** A distribution  $P$  on  $\mathbb{R}$  is (absolutely) continuous if there is a probability density function (pdf)  $f : \mathbb{R} \rightarrow [0, \infty)$  such that for every set  $B \subseteq \mathbb{R}$

$$P(B) = \int_B f(x) dx = \int_{-\infty}^{\infty} f(x) \mathbb{I}(x \in B) dx$$

- ▶ If  $X \sim P$  we write  $X \sim f$ .
- ▶ Examples: normal, uniform, exponential, gamma, beta

## Technical notes

1. Changing density  $f$  at countably many points leaves probabilities unchanged
2. If  $X \sim f$  then  $\mathbb{P}(x - \delta < X < x + \delta) \approx f(x)(2\delta)$  when  $\delta > 0$  is small
3. If  $X \sim f$  then  $\mathbb{P}(X = x) = 0$  for any number  $x$
4. A pdf is just a non-negative function that integrates to one

## Jointly Distributed Random Variables

**Setting.** Let  $X, Y : \Omega \rightarrow \mathbb{R}$  be random variables defined on same experiment with underlying probability  $\mathbb{P}$ . The *joint distribution* of  $(X, Y)$  is defined by

$$P(B) := \mathbb{P}((X, Y) \in B) \quad B \subseteq \mathbb{R}^2$$

- ▶  $P$  is continuous if there is a joint density  $f(x, y)$  such that

$$P(B) := \int_B f(x, y) dx dy$$

- ▶  $P$  is discrete if there is a joint probability mass function  $p(x, y)$  such that

$$P(B) := \sum_{(x, y) \in B} p(x, y)$$

## Example: Three Coin Flips

Sample space  $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

Events: Any subset of  $\Omega$ , for example

1. Even number of Heads,  $A = \{HHT, HTH, THH, TTT\}$
2. No adjacent Heads  $B = \{HTH, TTH, THT, HTT, TTT\}$

Consider two probabilities

1. Uniform:  $\mathbb{P}_0(HHH) = \dots = \mathbb{P}_0(TTT) = 1/8$ . Flips independent
2. Sticky:  $\mathbb{P}_1(HHH) = \mathbb{P}_1(TTT) = 1/2$ . Flips fully dependent

Random Variables

1.  $X(\omega) =$  number of heads in  $\omega$
2.  $Y(\omega)$  is 1 if the number of heads in  $\omega$  is even, and is 0 otherwise

## Comparing Random Variables

Let  $X$  and  $Y$  be random variables defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$

Standard equality and inequality

- ▶  $X = Y$  means that  $X(\omega) = Y(\omega)$  for all  $\omega \in \Omega$
- ▶  $X \leq Y$  means that  $X(\omega) \leq Y(\omega)$  for all  $\omega \in \Omega$

In most cases we only require that  $\mathbb{P}(X = Y) = 1$  and  $\mathbb{P}(X \leq Y) = 1$

# Comparing Distributions

Consider:  $X \sim P$  with CDF  $F_X$  and  $Y \sim Q$  with CDF  $F_Y$

**1. Equality in distribution:**  $X$  and  $Y$  have the same distribution, written  $X \stackrel{d}{=} Y$ , means that  $P = Q$ , that is,

$$\mathbb{P}(X \in B) = P(B) = Q(B) = \mathbb{P}(Y \in B)$$

for all sets  $B \subseteq \mathbb{R}$ . This is equivalent to  $F_X = F_Y$  (also same pdf or pmf)

**2. Stochastic order:**  $X$  is stochastically smaller than  $Y$ , written  $X \stackrel{d}{\leq} Y$ , if  $F_Y \leq F_X$ . This is equivalent to

$$\mathbb{P}(X > u) \leq \mathbb{P}(Y > u)$$

for each number  $u$ .

# The CDF Method

**Common task:** We are given

- ▶ Continuous random variable  $X$  with CDF  $F$  and density  $f$
- ▶ Function  $h : \mathbb{R} \rightarrow \mathbb{R}$

and wish to find the density of  $Y = h(X)$

## CDF Method

- ▶ Find the CDF  $G$  of  $Y$  in terms of the CDF  $F$  of  $X$
- ▶ Differentiate  $G$  to find the density  $g$  of  $Y$

## Examples

- ▶  $X \sim U(0, 1)$  and  $Y = -\log X$
- ▶  $X \sim \mathcal{N}(0, 1)$  and  $Y = X^2$

Expected Values, Variances, Covariances

## Fubini's Theorem

Setting: Let  $h : \mathbb{R}^d \times \mathbb{R}^k \rightarrow \mathbb{R}$  be a function of two arguments. Note that

$$\int_{\mathbb{R}^d} h(x, y) dx = \text{function of } y \quad \text{and} \quad \int_{\mathbb{R}^k} h(x, y) dy = \text{function of } x$$

**Theorem:** If  $h$  is non-negative, then we have

$$\int_{\mathbb{R}^k} \left[ \int_{\mathbb{R}^d} h(x, y) dx \right] dy = \int_{\mathbb{R}^d} \left[ \int_{\mathbb{R}^k} h(x, y) dy \right] dx := \iint h(x, y) dx dy$$

These same equations hold for general  $h$  if  $\iint |h(x, y)| dx dy$  is finite.

## Expected Values

**Recall:** Let  $X \in \mathbb{R}$  be a random variable

- ▶ If  $X \sim p$  and  $\sum_x |x| p(x)$  is finite, then  $\mathbb{E}X = \sum_x x p(x)$
- ▶ If  $X \sim f$  and  $\int |x| f(x) dx$  is finite, then  $\mathbb{E}X = \int x f(x) dx$

**Basic properties:** Let  $X, Y$  be jointly distributed with well defined expectations

1. If  $X \geq 0$  then  $\mathbb{E}X \geq 0$
2.  $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
3.  $|\mathbb{E}X| \leq \mathbb{E}|X|$
4. If  $X \perp Y$  then  $\mathbb{E}(XY) = \mathbb{E}X \mathbb{E}Y$
5.  $\mathbb{E}\mathbb{I}_B(X) = \mathbb{P}(X \in B)$
6. If  $X \geq 0$  then  $\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geq t) dt$

## Variance and Covariance

**Recall:** Let  $X, Y$  be random variables with  $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$ . The variance of  $X$  is

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

The covariance of  $X$  and  $Y$  is

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

### Basic Properties

1.  $\text{Var}(aX + b) = a^2 \text{Var}(X)$
2.  $\text{Var}(X) = \text{Cov}(X, X)$
3.  $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$
4.  $\text{Var}(X + Y) = \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y)$
5. If  $X \perp Y$  then  $\text{Cov}(X, Y) = 0$