Review and Background

Andrew Nobel

August, 2024

Review of Probability

Probability Model for Random Experiment

1. A non-empty set Ω whose elements ω are the possible *outcomes* of the experiment

2. Sets of possible outcomes $A \subseteq \Omega$ are called events

3. A probability measure $\mathbb{P}$ assigns probabilities to events. Two properties
   - $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$
   - If $A_1, A_2, \ldots$ are disjoint then $\mathbb{P}(\cup_{i \geq 1} A_i) = \sum_{i \geq 1} \mathbb{P}(A_i)$

Probability Models, cont.

Fact: Probabilities have the following properties

1. $0 \le \mathbb{P}(A) \le 1$ 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 $\mathbb{P}(A) \le \mathbb{P}(B)$
4. If $A_1, A_2, \ldots$ are any events then $\mathbb{P}(\cup_{i \ge 1} A_i) \le \sum_{i \ge 1} \mathbb{P}(A_i)$
5. If events $A_1 \subseteq A_2 \subseteq \cdots$ are increasing then $\mathbb{P}(A_n) \to \mathbb{P}(\cup_{i \ge 1} A_i)$
6. If events $A_1 \supseteq A_2 \supseteq \cdots$ are decreasing then $\mathbb{P}(A_n) \to \mathbb{P}(\cap_{i \ge 1} A_i)$

Random Variables

Setting: Experiment with possible outcomes Ω and a probability measure $\mathbb{P}$

Definition: A random variable is a function $X : \Omega \to \mathbb{R}$. We regard X as a numerical measurement based on the outcome of the experiment

A random variable X is a deterministic function, but its value $X(\omega)$ inherits uncertainty from the outcome ω of the experiment, which is modeled by $\mathbb{P}$. For $B \subseteq \mathbb{R}$ we define

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

In other words, the probability that X is in B is the probability of the set of outcomes ω for which $X(\omega)$ is in B

Distribution of a Random Variable

Definition: Let P be a probability measure on $\mathbb{R}$. We say that a random variable X has distribution P, written $X \sim P$, if

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

Note

- P describes uncertainty in the observed values of X while $\mathbb{P}$ governs uncertainty in the outcomes of the underlying experiment

- In many inference problems the underlying probability space is not explicit, e.g.,

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

Example: Three Coin Flips

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

Events: Any subset of Ω, for example

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

Probabilities

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

Random Variables

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

Cumulative Distribution Function

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

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

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. $F(x) \to 1$ as $x \to \infty$
3. $F(x) \to 0$ as $x \to -\infty$
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

- $\mathbb{I}_{A^c} \;=\; 1 - \mathbb{I}_A$
- $\mathbb{I}_{A \cap B} \;=\; \mathbb{I}_A \, \mathbb{I}_B$
- $\mathbb{I}_{A \cup B} \;=\; \max(\mathbb{I}_A, \mathbb{I}_B)$
- $\int_A h(x)\, dx \;=\; \int h(x) \mathbb{I}_A(x)\, dx$

Discrete Distributions

Definition: A distribution P on $\mathbb{R}$ is discrete if there is a probability mass function (pmf) $p : \mathbb{R} \to [0, 1]$ such that for every Borel 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 $X \sim P$ then $p(x) = \mathbb{P}(X = x)$ and we write $X \sim p$
- There can be at most countably many x for which $p(x) > 0$, so the sums above are well defined
- Examples: Bernoulli, binomial, Poisson, geometric, hypergeometric distributions

Continuous Distributions

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

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

- If $X \sim P$ then $\mathbb{P}(x - \delta/2 < X < x + \delta/2) \approx f(x)\,\delta$, and we write $X \sim f$
- Changing f at countably many points leaves integrals defining $P(B)$ unchanged
- Examples: normal, uniform, exponential, gamma, beta

Support Sets and Proportional Densities

Definition

- The support of a pmf p is $\text{Supp}(p) = \{x : p(x) > 0\}$
- The support of a pdf f is $\text{Supp}(f) = \{x : f(x) > 0\}$

Roughly speaking, $\text{Supp}(p)$ is the set of possible values of X

Definition: Two pdf's f and g are proportional, written $f \propto g$, if there is a constant $c > 0$ such that $f(x) = cg(x)$ for all x

Fact: If $f \propto g$ then f, g have the same support and $f = g$

Jointly Distributed Random Variables

Let $X, Y : \Omega \to \mathbb{R}$ be real-valued measurements on the experiment $(\Omega, \mathcal{F}, \mathbb{P})$. The *joint distribution* P of (X, Y) is defined by

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

for all 2-dimensional Borel sets $B \subseteq \mathbb{R}^2$. In this case we write $(X, Y) \sim P$. If P is continuous then there is a joint density $f(x, y)$ such that

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

Recall: If $(X, Y) \sim f$ then the marginal densities of X and Y are given by

$$f_X(x) = \int f(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int f(x, y)\, dx$$

The CDF Method

Common task: We are given

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

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 $Y = X^2$

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 $\omega \in \Omega$

More generally, we only require that $X = Y$ with probability one, that is $\mathbb{P}(X = Y) = 1$, and similarly $\mathbb{P}(X \leq Y) = 1$

Comparing Distributions

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

1. Equality in distribution: $X \overset{\mathrm{d}}{=} Y$ means that $P = Q$, that is, for all Borel sets B

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

This is equivalent to $F_X = F_Y$

2. Stochastic order: $X \overset{\mathrm{d}}{\leq} Y$ is defined by the relation $F_Y \leq F_X$, in other words

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

for each number u

Expected Values, Variances, Covariances

Fubini's Theorem

Setting: Let $h : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ 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 $X \perp\!\!\!\perp Y$ 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

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

The covariance of X and Y is

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

Basic Properties

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