Machine Learning

Probability Review for Discrete Random Variables

Fabio Vandin

October 6th, 2023

Probability Refresher

Definition

A probability space has three components:

- A sample space Z, which is the set of all possible outcomes of the random process modeled by the probability space;
- **2** A family \mathcal{F} of sets representing the allowable events, where each event A is a subset of Z: $A \subseteq Z$ (Must be a σ -field...)
- **3** A probability distribution $\mathcal{D}: \mathcal{F} \to [0,1]$ that satisfies the following conditions:
 - **1** $\mathcal{D}[Z] = 1;$
 - 2 Let E_1, E_2, E_3, \ldots be any finite or countably infinite sequence of pairwise mutually disjoint events $(E_i \cap E_j = \emptyset)$ for all $i \neq j$:

$$\mathcal{D}\left[\bigcup_{i\geq 1}E_i\right]=\sum_{i\geq 1}\mathcal{D}[E_i].$$

EXAMPLE s) fait coin flipping - sample space: $Z = \{H_1T\}$ $\left(\frac{H}{T} = \text{lead}\right)$ - events: $Z = \{ \emptyset, \{H\}, \{T\}, \{H,T\} \}$ - probability distribution: ({ 4}) = = $\mathcal{O}(\mathcal{O}) = 0$ O(3H,73)=1

3

Distributions and Probability

We use $z \sim \mathcal{D}$ to say that event $z \in Z$ is *sampled* according to \mathcal{D}

Given a function $f: Z \to \{true, false\}$, define the probability of f(z) $\mathbb{P}_{z \sim \mathcal{D}}[f(z)] = \mathcal{D}(\{z \in Z : f(z) = true\})$

Distributions and Probability

We use $z \sim D$ to say that event $z \in Z$ is sampled according to D

Given a function $f: Z \to \{true, false\}$, define the probability of f(z)

$$\mathbb{P}_{z \sim \mathcal{D}}[f(z)] = \mathcal{D}(\{z \in Z : f(z) = true\})$$

In many cases, we express an event $A \subseteq Z$ using a function $\pi_A : Z \to \{true, false\}$, that is:

$$A = \{z \in Z : \pi_A(z) = true\}$$

where $\pi_A(z) = true$ if $z \in A$ and $\pi_A(z) = false$ otherwise.

In this case we have $\mathbb{P}[A] = \mathbb{P}_{z \sim \mathcal{D}}[\pi_A(z)] = \mathcal{D}(A)$

Note: sometimes we use $\pi_A : Z \to \{0,1\}$ instead of $\pi_A : Z \to \{true, false\}$.

EXAMPLE. Lie rolling

Consider the event: A = "outcome is even" 2= {1,2,3,4,5,6} ⇒ A= {2,4,6} C Z={1,...,6} Then Try (2) = true = Try (4) = Try (6) $\begin{array}{l} \text{Tr}_{A}(3) = \text{folse} = \text{Tr}_{A}(3) = \text{Tr}_{A}(5) \\ \text{Folse} = \text{Tr}_{A}(5) = \text{Tr}_{A}(5) \\ \text{Folse} = \text{Tr}_{A}(5) = \text{Tr}_{A}(5) \\ \text{Folse} = \text{Tr}_{A}(5) = \text{Tr}_{A}(5) = \text{Tr}_{A}(5) \\ \text{Folse} = \text{Tr}_{A}(5) = \text{Tr}$ More commonly: we say Pr[A] = Pr[ontrowe is even]= {

Independent Events

Definition

Two events E and F are independent $(E \perp F)$ if and only if

$$\mathbb{P}[E \cap F] = \mathbb{P}[E] \cdot \mathbb{P}[F]$$

More generally, events $E_1, E_2, \dots E_k$ are mutually independent if and only if for any subset $I \subseteq [1, k]$,

$$\mathbb{P}\left[\bigcap_{i\in I}E_i\right] = \prod_{i\in I}\mathbb{P}[E_i].$$

Example die rolling Consider the events:

- E = "outcome is even "

- F = "outcome is \le 2" Independent? YES: 17 NO: 16 Pr [E] = 3/2 Pr (F) = 1/3 and it is <2] Pr [ENF] = Pr Soutcome is even = Pr [out one is 2] YES, they are independent = Pr[F]. Pr[F].

Random Variable (R.V.)

Definition

A (scalar) random variable X(z) on a sample space Z is a real-valued function on Z; that is, $X: z \in Z \to \mathbb{R}$.

Random Variable (R.V.)

Definition

A (scalar) random variable X(z) on a sample space Z is a real-valued function on Z; that is, $X: z \in Z \to \mathbb{R}$.

Discrete random variable: codomain is finite or countable (countably infinite).

Example: $\mathbb{Z}, \mathbb{N}, \{0, 1\}, \dots$

Random Variable (R.V.)

Definition

A (scalar) random variable X(z) on a sample space Z is a real-valued function on Z; that is, $X : z \in Z \to \mathbb{R}$.

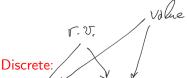
Discrete random variable: codomain is finite or countable (countably infinite).

Example: $\mathbb{Z}, \mathbb{N}, \{0, 1\}, \dots$

Continuous random variable: codomain is continuous.

Example: $\mathbb{R}, [a, b], \ldots$

Description of R.V.



- $p_{\mathbf{X}}(\mathbf{X}) = \mathbb{P}[\mathbf{X} = \mathbf{X}]$ [Probability Mass Function PMF]
- $F_X(x) = \mathbb{P}[X \le x] = \sum_{k \le x} p_X(k)$ [Cumulative Distribution Function CDF]

Example: coin flipping

Consider prob. space of fair coin flipping

Let
$$r.v.$$
 $X:$

$$X(r) = 0$$

Then: -PMF:
$$P_{X}(0) = \frac{1}{2} = P_{X}(1)$$
-CDF: $F_{X}(0) = P_{F}[X \le 0] = \frac{1}{2}$
 $F_{X}(1) = P_{F}[X \le 1] = 1$

Vector Valued R.V.

Example

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

 X_1, X_2 discrete:

$$p_{\mathbf{X}}(\mathbf{x}) \doteq p_{X_1,X_2}(x_1,x_2) = \mathbb{P}_{\mathbf{X}}[X_1 = x_1, X_2 = x_2]$$

["
$$X_1 = x_1, X_2 = x_2$$
" are joint events]

Note: If **X** is obvious, we may write $\mathbb{P}[X_1 = x_1, X_2 = x_2]$ instead of $\mathbb{P}_{\mathbf{X}}[X_1 = x_1, X_2 = x_2]$

Example: dice rolling

Consider two independent dice, die 1 and die 2. Define the random variables:

- X_1 = value of die 1
- X_2 = value of die 2
- X_3 = squared value of die $2 = (X_2)^2$

Vector value r.v.:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

$$P_X$$
 (1,6,36) = $P_F[X_1=A,X_2=6,X_3=36] = $\frac{1}{36}$
There are 36 ontrones: (outcome of die 1, outcome of die 2)$

Independence

Definition

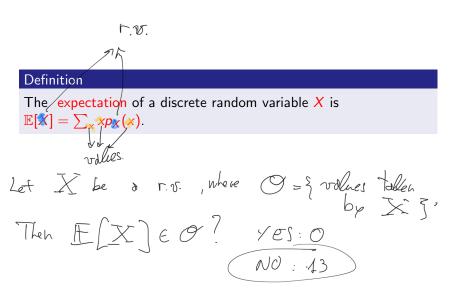
Two discrete random variables X and Y are independent $(X \perp Y)$ if and only if

$$\mathbb{P}((X = x) \cap (Y = y)) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

for all values x and y. Similarly, discrete random variables $X_1, X_2, \ldots X_k$ are mutually independent if and only if for any subset $I \subseteq [1, k]$ and any values $x_i, i \in I$,

$$\mathbb{P}_{\mathbf{X}}(\mathbf{x}) = \prod_{i \in I} \mathbb{P}(X_i = x_i) = \prod_{i \in I} p_{X_i}(x_i).$$

Expected Value and Moments



Example

1) foir oin flipping:
$$X = \begin{cases} 0 & \text{if outcome is } \\ 1 & \text{outcome is } \end{cases}$$

$$P_{v}[X=0] = \frac{1}{2} = P_{v}[X=1]$$

$$E[X] = \sum_{i=0}^{2} P_{v}[X=i] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$
2) (general) coin flipping: $X = \begin{cases} 0 & \text{if } \\ 1 & \text{if } \end{cases}$

$$P_{v}[X=1] = p ; P_{v}[X=0] = 1 - p$$

$$\Rightarrow E[X] = p$$

Theorem

Let g(X) be a function of a discrete random variable X. Then $\mathbb{E}[g(X)] = \sum_{x} g(x) p_X(x).$

For a random variable X we define:

- Mean: $m_X \doteq \mathbb{E}[X]$
- Variance: $\sigma_X^2 \doteq \mathbb{E}[(X m_X)^2] = \mathbb{E}[X^2] m_X^2 = \text{Var}[X]$ HW: prove