Machine Learning

Probability Review for Discrete Random Variables

Fabio Vandin

October 4^{th}, 2022

Probability Refresher

Definition

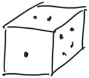
A probability space has three components:

1. 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 $i \neq j$):

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

EXAMPLES

1) die rolling



- sample space : $Z = \{1, 2, 3, 4, 5, 6\}$
- events $\mathcal{F}$: $\mathcal{F} = \{ \{1\}, \{2\}, \{1,2\}, \{2,4,6\}, \dots \}$
- probability distribution : $\mathbb{P}(\{1\}) = \frac{1}{6} = \mathbb{P}(\{i\})$, $i = 2, \dots, 6$

2) fair coin flipping

- sample space : $Z = \{H, T\}$ $\left(\begin{array}{l} H = \text{head} \\ T = \text{Tail} \end{array} \right)$
- events : $Z = \{ \emptyset, \{H\}, \{T\}, \{H, T\} \}$
- probability distrib : $\mathbb{P}(\{H\}) = \frac{1}{2} = \mathbb{P}(\{T\})$

Distributions and Probability

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

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 $f : Z \rightarrow \{\mathit{true}, \mathit{false}\}$, define the probability of $f(z)$

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

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 $f(z)$

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

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.

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 $f(z)$

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

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 $\mathbb{P}[A] = \mathbb{P}_{z\sim\mathcal{D}}[\pi_A(z)] = \mathcal{D}(A)$

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

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

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

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

where $\pi_A(z) = \mathit{true}$ if $z \in A$ and $\pi_A(z) = \mathit{false}$ 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 $\pi_A : Z \to \{\mathit{true}, \mathit{false}\}$.

EXAMPLE die rolling

Consider the event: $A =$ "outcome is even"

$$\Rightarrow A = \{2, 4, 6\} \subset Z = \{1, \dots, 6\}$$

Then: $\pi_A(2) = \text{true}$

$\pi_A(3) = \text{false}$

$$\Pr_{x \sim \mathcal{D}}\left[\pi_A(z)\right] = \frac{1}{2}$$

More commonly: we say $\Pr[A] = \Pr[\text{outcome is even}]$
$= \frac{1}{2}$

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 $E_1, E_2, \ldots E_k$ 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 ≤ 2"

Independent?

YES: 7

NO: everybody − 7

$Pr[E] = 1/2$

$Pr[F] = 1/3$

$$Pr[E \cap F] = Pr[\text{outcome is even and it is} \leq 2]$$
$$= Pr[\text{outcome is } 2]$$
$$= \frac{1}{6}$$
$$= Pr[E] \cdot Pr[F] = \frac{1}{6}$$

YES, they are independent

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: $\mathbb{Z}, \mathbb{N}, \{0; 1\}, \ldots$

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: $\mathbb{Z}, \mathbb{N}, \{0; 1\}, \ldots$

Continuous random variable: codomain is continuous.
Example: $\mathbb{R}, [a, b], \ldots$

Description of R.V.

r.v.

specific value $(x \in \mathbb{Z})$

Discrete:

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

Example: coin flipping

Consider prob. space of fair coin flipping

Let r.v. X : $X(H) = 1$

$X(T) = 0$

Then: – PMF: $p_X(0) = p_X(1) = 1/2$

– CDF: $F_X(0) = \Pr[X \leq 0] = 1/2$

$F_X(1) = \Pr[X \leq 1] = 1$

Vector Valued R.V.

Example

[bold font]

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

X_1, X_2 discrete:

value values

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

r.v. r.v.'s

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]

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 $\mathbf{X}$ 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 $= (X_2)^2$

$P_{\vec{X}}(1,3,5) \stackrel{?}{=} 0$

Vector value r.v.:

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

$P_{\vec{X}}(1,6,36) = \frac{1}{36} = \Pr[X_1 = 1, X_2 = 6, X_3 = 36]$

[There are 36 outcomes: (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 $x_i, i \in I$,

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

Example (previous example): dice rolling

Die 1, die 2, independent;

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

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

$$1 \le i, j \le 6$$

X_1, X_2 independent? YES $\Rightarrow \Pr_{X_1,X_2}(X_1 = i, X_2 = j)$

$$= \Pr(X_1 = i) \cdot \Pr(X_2 = j) = \frac{1}{36}$$

X_1, X_3 independent? YES $\Rightarrow$ HOMEWORK

X_2, X_3 independent? NO $\Pr_{X_2,X_3}[X_2 = 1, X_3 = 1] \neq \frac{1}{36}$

$$= \frac{1}{6} = \Pr(X_2 = 1) = \Pr(X_3 = 1)$$

$\Rightarrow X_1, X_2, X_3$ are not mutually independent