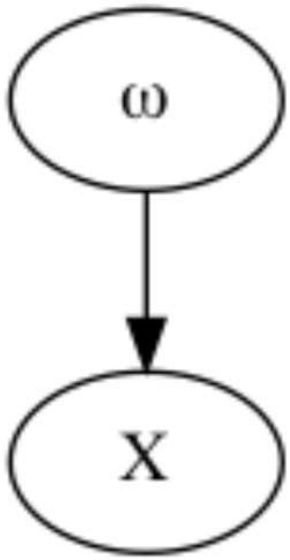
Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

What is a random variable?

PREDICTIVE
SCIENCE LABORATORY

Mathematical definition of random variables



A discrete random variable is a function $X(\omega)$ giving the result of an uncertain experiment.

- *Discrete random* variable if takes values 0, 1, …
- *Continuous random* variable if it takes real values.

Even though a random variable is always a function of some ω, we can often get away with not explicitly showing it.

PREDICTIVE
SCIENCE LABORATORY

Mathematical notation

- Upper case letters to represent random variables, like X, Y, Z.
- Lower case letters to represent the values of random variables, like x, y, z.
- But we are not going to be too strict about this if there is no danger of ambiguity.

Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

The probability mass function

PREDICTIVE
SCIENCE LABORATORY

Probability mass function

Let X be a discrete random variable. The *probability mass function (pmf) of* X *is:*

$p(X = x) =$ Probability that the random variable X takes the value x

PREDICTIVE SCIENCE LABORATORY

Probability mass function

Let X be a discrete random variable. The *probability mass function (pmf) of* X *is:*

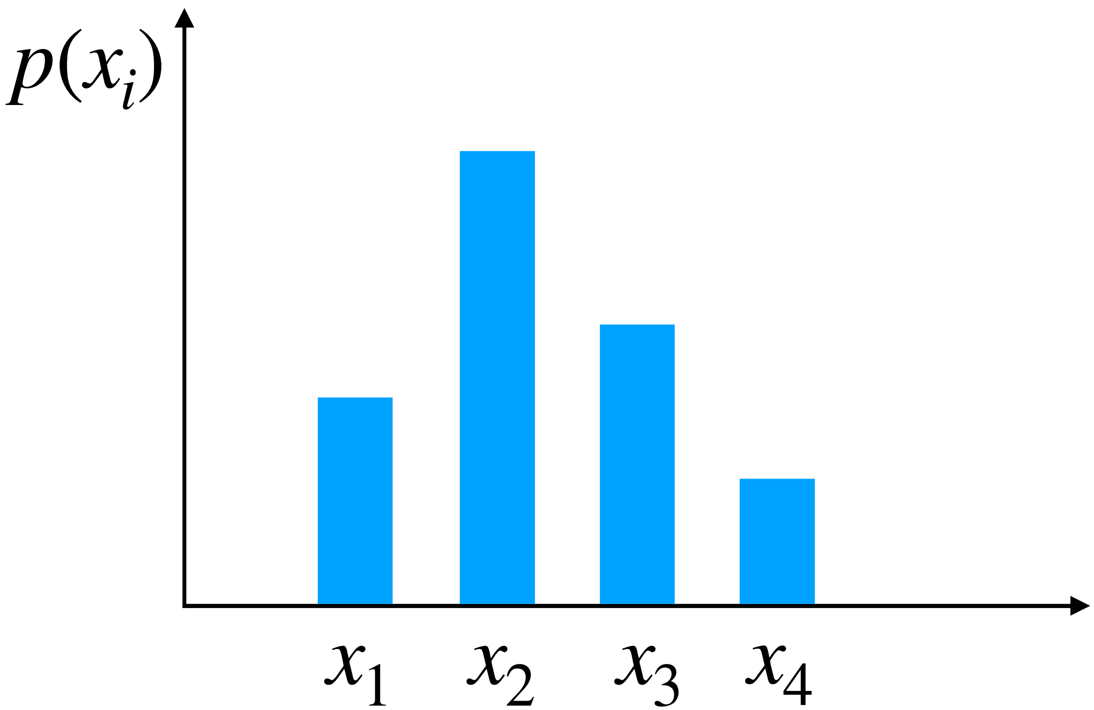
$p(X = x) =$ Probability that the random variable X takes the value x

When there is no ambiguity:

$$\boxed{p(x) \equiv p(X = x)\,.}$$

PREDICTIVE
SCIENCE LABORATORY

Visualization of the probability mass function



Properties of the probability mass function

- The probability mass function is nonnegative:

$$p(x) \geq 0.$$

- The probability mass function is normalized:

$$\sum_x p(x) = 1,$$

where the summation is over all the possible values of X.

PREDICTIVE
SCIENCE LABORATORY

Properties of the probability mass function

- Let X be a discrete random variable.

- The probability of X taking either the value x_1 or the value x_2 (assuming $x_1 \neq x_2$) is:

$$p(X = x_1 \text{ or } X = x_2) \equiv p(X \in \{x_1, x_2\}) = p(X=x_1) + p(X=x_2)$$

$$= p(x_1) + p(x_2)$$

PREDICTIVE
SCIENCE LABORATORY

Properties of the probability mass function

- More generally, the probability that the random variable X takes any value in a set A is given by:

$$p(X \in A) = \sum_{x \in A} p(x)$$

Functions of random variables

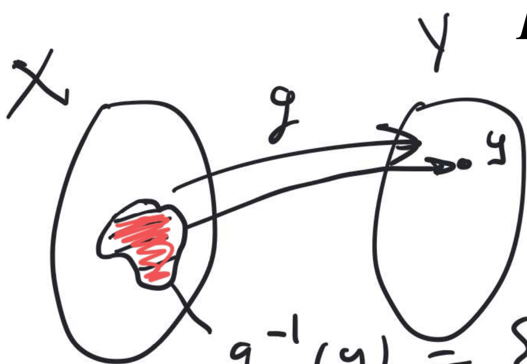
- Consider a function $g(x)$.

- We can now define a new random variable:

$$Y = g(X)\,.$$

- It has its own probability mass function (pmf):

$$p(y) = p(Y = y) = \sum_{x \in g^{-1}(y)} p(x)$$



where

$$g^{-1}(y) = \{x : g(x) = y\}$$

Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

Expectation of a discrete random variable

PREDICTIVE
SCIENCE LABORATORY

Expectation of a random variable

- The expectation of a random variable is:

$$\boxed{\mathbb{E}[X] := \sum_x x p(x)}$$

- You can think of the expectation as the value of the random variable that one should "expect" to get.

- However, take this interpretation with a grain of salt because it may be a value that the random variable has a zero probability of getting...

PREDICTIVE
SCIENCE LABORATORY

Properties of the expectation

- For any function $g(x)$:

$$\mathbb{E}[g(X)] = \sum_x g(x)p(x)$$

Properties of the expectation

- Take any constant c:

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c$$

Proof:

$$\mathbb{E}[X + c] = \sum_x (x + c)\, p(x) = \sum_x x\, p(x) + \sum_x c\, p(x)$$

$$= \mathbb{E}[X] + c \cdot \overbrace{\sum_x p(x)}^{1} \qquad \blacksquare$$

Properties of the expectation

- For any λ real number:

$$\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$$

Proof: $\mathbb{E}[\lambda X] = \sum_x \lambda x \, p(x) = \lambda \cdot \sum_x x \cdot$

Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

Variance of a discrete random variable

PREDICTIVE
SCIENCE LABORATORY

Expectation of a random variable

- The variance of a random variable is:

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sum_x (x - \mathbb{E}[X])^2 p(x)$$

- You can think of the variance as the spread of the random variable around its expectation.

- However, do not take this too literally for discrete random variables.

Properties of the variance

- Take any constant c:

$$\mathbb{V}[X+c] = \mathbb{V}[X]$$

Proof:

$$\mathbb{V}[X+c] = \mathbb{E}\left[\left(X+\cancel{c} - \underbrace{\mathbb{E}[X+c]}_{\mathbb{E}[X]+\cancel{c}}\right)^2\right]$$

$$= \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right] \quad \blacksquare$$

PREDICTIVE
SCIENCE LABORATORY

Properties of the variance

- Take any constant λ:

$$\mathbb{V}[\lambda X] = \lambda^2 \mathbb{V}[X]$$

Proof:

$$\mathbb{V}[\lambda X] = \mathbb{E}[(\lambda X - \mathbb{E}[\lambda X])^2]$$

$$\lambda \mathbb{E}[X]$$

$$= \mathbb{E}[(\lambda X - \lambda \mathbb{E}[X])^2]$$

$$= \mathbb{E}[\lambda^2 (X - \mathbb{E}[X])^2]$$

$$= \lambda^2 \mathbb{E}[(X$$

PREDICTIVE
SCIENCE LABORATORY

Properties of the variance

- It holds that:

$$\boxed{\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2}$$

Proof: $\mathbb{V}[X] = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right]$

$= \mathbb{E}\left[X^2 - 2X\,\mathbb{E}[X] + \left(\mathbb{E}[X]\right)^2\right]$

$= \mathbb{E}\left[X^2 - 2X\,\mathbb{E}[X]\right] + \left(\mathbb{E}[X]\right)^2$

$= \mathbb{E}\left[X^2\right] - 2\left(\mathbb{E}[X]\right)^2 + \left(\mathbb{E}[X]\right)^2$

PREDICTIVE SCIENCE LABORATORY

Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

The Bernoulli distribution

Example: The Bernoulli distribution

- Models an experiment with two outcomes.

$$X = \begin{cases} 1, & \text{with probability } \theta, \\ 0, & \text{otherwise} \,. \end{cases}$$

- Notation:

$$\boxed{X \sim \text{Bernoulli}(\theta)}$$

- You read: "X follows a Bernoulli with parameter θ."

Example: PMF of a Bernoulli

- Assume $X \sim$ Bernoulli(θ).

- We have:

$$p(X = 1) = \theta$$

- From this, because of the normalization constraint:

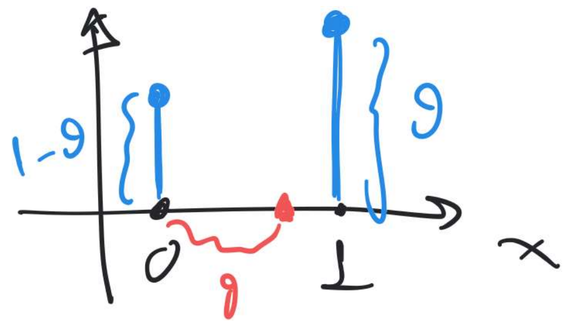
$$p(X = 0) + p(X = 1) = 1$$

we get that: $p(X = 0) =$

PREDICTIVE
SCIENCE LABORATORY

Example: Expectation and variance of a Bernoulli

- Assume $X \sim \text{Bernoulli}(\theta)$.



- The expectation is:

$$\mathbb{E}[X] = \sum_x x\,p(x) = 1 \cdot p(X=1) + 0 \cdot p(X=0) = 1 \cdot \theta + 0 \cdot (1-\theta) = \theta$$

- The variance is: $\mathbb{E}[X^2] = \sum_x x^2 p(x) = 1^2 \cdot \theta + 0^2 \cdot (1-\theta) = \theta$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \theta - \theta^2 = \theta \cdot (1-\theta)$$

PREDICTIVE
SCIENCE LABORATORY

Lecture 3: Discrete Random Variables

Professor Ilias Bilionis

The Categorical distribution

Example: The Categorical distribution

- Models an experiment with K outcomes.

$$X = \begin{cases} c_1, & \text{with probability } p_1, \\ \vdots & \\ c_K, & \text{with probability } p_K, \end{cases}$$

- Notation:

$$\boxed{X \sim \text{Categorical}(p_1, \dots, p_K)}$$

PREDICTIVE
SCIENCE LABORATORY

Example: PMF of a Categorical

- Assume $X \sim$ Categorical$(0.1, 0.3, 0.6)$.
- We have $K = 3$ possible outcomes, say c_1, c_2, c_3 .
- The PMF is:

$$p(X=c_1) = 0.1$$
$$p(X=c_2) = 0.3$$
$$p(X=c_3) = 0.6$$

Example: PMF of a Categorical

- Assume $X \sim$ Categorical$(0.1, 0.3, 0.6)$.
- We have $K = 3$ possible outcomes, say c_1, c_2, c_3 .
- The probability that X is either c_1 or c_3.

$$
\begin{aligned}
p(X = c_1 \text{ or } X = c_3) &= p(X \in \{c_1, c_3\}) \\
&= p(X = c_1) + p(X = c_3) \\
&= 0.1 + 0.6 \\
&= 0.7
\end{aligned}
$$

Example: PMF of a Categorical

- Assume $X \sim$ Categorical$(0.1, 0.3, 0.6)$.
- We have $K = 3$ possible outcomes.
- The expectation is:

$$\mathbb{E}[X] = \sum_x x p(x) = c_1 \cdot 0.1 + c_2 \cdot 0.3 + c_3 \cdot 0.6$$

Example: PMF of a Categorical

- Assume $X \sim$ Categorical$(0.1, 0.3, 0.6)$.

- We have $K = 3$ possible outcomes.

- The variance is: $\mathbb{V}[X] = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2$

$$\mathbb{E}[X^2] = \sum_x x^2 p(x) = c_1^2 \cdot 0.1 + c_2^2 \cdot 0.3 + c_3^2 \cdot 0.6$$

PREDICTIVE
SCIENCE LABORATORY