Mathematics for Informatics 4a

José Figueroa-O’Farrill

Lecture 6
3 February 2012

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0, 1]$, satisfying $\sum_{x \in \mathbb{Z}} p(x) = 1$.

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0, 1]$, satisfying $\sum_{x \in \mathbb{Z}} p(x) = 1$.
- We met several famous discrete probability distributions:

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0, 1]$, satisfying $\sum_{x \in \mathbb{Z}} p(x) = 1$.
- We met several famous discrete probability distributions:
  - **uniform** on $E = \{1, 2, \ldots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0,1]$, satisfying $\sum_{x\in\mathbb{Z}} p(x) = 1$.
- We met several famous discrete probability distributions:
  - **uniform** on $E = \{1, 2, \dots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **2-digit Benford**: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 10 \leqslant x \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0, 1]$, satisfying $\sum_{x \in \mathbb{Z}} p(x) = 1$.
- We met several famous discrete probability distributions:
  - **uniform** on $E = \{1, 2, \dots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **2-digit Benford**: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 10 \leqslant x \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$
  - **binomial** with parameters n, p:

    $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise} \end{cases}$

    the probability of exactly x successes in n independent Bernoulli trials with success probability p

The story of the film so far...

- Experiments with integer outcomes give rise to probability distributions $p : \mathbb{Z} \to [0, 1]$, satisfying $\sum_{x\in\mathbb{Z}} p(x) = 1$.
- We met several famous discrete probability distributions:
  - **uniform** on $E = \{1, 2, \ldots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **2-digit Benford**: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 10 \leqslant x \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$
  - **binomial** with parameters n, p:
    $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise} \end{cases}$
    the probability of exactly x successes in n independent Bernoulli trials with success probability p
- We also introduced the distribution function $F : \mathbb{Z} \to [0, 1]$ associated to p, defined by $F(x) = \sum_{t\leqslant x} p(t)$: monotonically increasing from 0 to 1.

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*

Let $p(k)$ be the probability of stopping at the kth toss.

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*
Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \dots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \dots \end{cases}$$

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*
Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \dots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \dots \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{2}$.

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*

Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \dots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \dots \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{2}$. Of course,

$$\sum_{k \in \mathbb{Z}} p(k)$$

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*

Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \dots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \dots \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{2}$. Of course,

$$\sum_{k \in \mathbb{Z}} p(k) = \sum_{k=1}^{\infty} (\tfrac{1}{2})^k$$

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*
Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \dots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \dots \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{2}$. Of course,

$$\sum_{k\in\mathbb{Z}} p(k) = \sum_{k=1}^{\infty} (\tfrac{1}{2})^k = \sum_{k=0}^{\infty} (\tfrac{1}{2})^k - 1$$

The mathematics of waiting

Example (Alice and Bob's favourite game)

We toss a fair coin until it comes up H. *How long must we wait for the game to end?*
Let $p(k)$ be the probability of stopping at the kth toss. Clearly,

$$p(k) = \begin{cases} 0, & k = 0, -1, -2, \ldots \\ (\frac{1}{2})^k, & k = 1, 2, 3, \ldots \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{2}$. Of course,

$$\sum_{k\in\mathbb{Z}} p(k) = \sum_{k=1}^{\infty} (\tfrac{1}{2})^k = \sum_{k=0}^{\infty} (\tfrac{1}{2})^k - 1 = \frac{1}{1-\frac{1}{2}} - 1 = 1 \, .$$

Example

- Suppose we decide to toss the coin at most N times, whether or not a head appears.

Example

- Suppose we decide to toss the coin at most N times, whether or not a head appears.
- Stopping at the Nth toss is equiprobable to getting tails in the first $N-1$ tosses: $p(N) = (\frac{1}{2})^{N-1}$.

Example

- Suppose we decide to toss the coin at most N times, whether or not a head appears.
- Stopping at the Nth toss is equiprobable to getting tails in the first $N-1$ tosses: $p(N) = (\frac{1}{2})^{N-1}$.
- The resulting probability distribution is now

$$p(k) = \begin{cases} 0, & k \leqslant 0 \text{ or } k > N \\ (\frac{1}{2})^k, & k = 1, 2, \dots, N-1 \\ (\frac{1}{2})^{N-1}, & k = N \end{cases}$$

  and is called the **truncated geometric distribution** with parameters N and $\frac{1}{2}$.

Example

- Suppose we decide to toss the coin at most N times, whether or not a head appears.
- Stopping at the Nth toss is equiprobable to getting tails in the first $N-1$ tosses: $p(N) = (\frac{1}{2})^{N-1}$.
- The resulting probability distribution is now

$$p(k) = \begin{cases} 0, & k \leqslant 0 \text{ or } k > N \\ (\frac{1}{2})^k, & k = 1, 2, \dots, N-1 \\ (\frac{1}{2})^{N-1}, & k = N \end{cases}$$

  and is called the **truncated geometric distribution** with parameters N and $\frac{1}{2}$.
- Again one has $\sum_k p(k) = \sum_{k=1}^{N-1} (\frac{1}{2})^k + (\frac{1}{2})^{N-1} = 1$.

Example (Dice instead of coins)

Suppose that now Alice and Bob roll a fair die instead and the game ends when one of them rolls a ⚅. *What is the probability* $p(k)$ *that the game ends with the* k*th roll?*

Example (Dice instead of coins)

Suppose that now Alice and Bob roll a fair die instead and the game ends when one of them rolls a ⚅. *What is the probability* $p(k)$ *that the game ends with the* k*th roll?*

Let S denote the event of rolling a ⚅. Then $\mathbb{P}(S) = \frac{1}{6}$ and hence $\mathbb{P}(S^c) = \frac{5}{6}$.

Example (Dice instead of coins)

Suppose that now Alice and Bob roll a fair die instead and the game ends when one of them rolls a ⚅. *What is the probability* $p(k)$ *that the game ends with the* k*th roll?*

Let S denote the event of rolling a ⚅. Then $\mathbb{P}(S) = \frac{1}{6}$ and hence $\mathbb{P}(S^c) = \frac{5}{6}$. The game ends with the kth roll if the first $k-1$ rolls do not show ⚅ and the kth roll does.

Example (Dice instead of coins)

Suppose that now Alice and Bob roll a fair die instead and the game ends when one of them rolls a ⚅. *What is the probability* $p(k)$ *that the game ends with the* k*th roll?*

Let S denote the event of rolling a ⚅. Then $\mathbb{P}(S) = \frac{1}{6}$ and hence $\mathbb{P}(S^c) = \frac{5}{6}$. The game ends with the kth roll if the first $k-1$ rolls do not show ⚅ and the kth roll does. The probability of such a sequence of rolls is then

$$p(k) = \begin{cases} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}, & k \geqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

Example (Dice instead of coins)

Suppose that now Alice and Bob roll a fair die instead and the game ends when one of them rolls a ⚅. *What is the probability* $p(k)$ *that the game ends with the* k*th roll?*

Let S denote the event of rolling a ⚅. Then $\mathbb{P}(S) = \frac{1}{6}$ and hence $\mathbb{P}(S^c) = \frac{5}{6}$. The game ends with the kth roll if the first $k-1$ rolls do not show ⚅ and the kth roll does. The probability of such a sequence of rolls is then

$$p(k) = \begin{cases} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}, & k \geqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

This is called the **geometric distribution** with parameter $\frac{1}{6}$.

Geometric distribution

Definition

The **geometric distribution** with parameter p is given by

$$p(k) = \begin{cases} (1-p)^{k-1}p, & k \geqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

Geometric distribution

Definition

The **geometric distribution** with parameter p is given by

$$p(k) = \begin{cases} (1-p)^{k-1}p, & k \geqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

The number $p(k)$ is the probability that in independent Bernoulli trials with success probability p, the first success occurs at the kth trial.

Geometric distribution

Definition

The **geometric distribution** with parameter p is given by

$$p(k) = \begin{cases} (1-p)^{k-1}p, & k \geqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

The number $p(k)$ is the probability that in independent Bernoulli trials with success probability p, the first success occurs at the kth trial. Notice that

$$\sum_{k \in \mathbb{Z}} p(k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{\ell=0}^{\infty} (1-p)^{\ell}$$
$$= p \frac{1}{1-(1-p)} = 1 \, .$$

Example (Weekly lottery)

Let p be the probability that a given number d is drawn in any given week. After n successive draws, let $p(k)$ be the probability that d last appeared k weeks ago. *What is* $p(k)$*?*

Example (Weekly lottery)

Let p be the probability that a given number d is drawn in any given week. After n successive draws, let $p(k)$ be the probability that d last appeared k weeks ago. *What is* $p(k)$*?*

The number d appears with probability p and does not appear with probability $1 - p$.

Example (Weekly lottery)

Let p be the probability that a given number d is drawn in any given week. After n successive draws, let $p(k)$ be the probability that d last appeared k weeks ago. *What is $p(k)$?*

The number d appears with probability p and does not appear with probability $1-p$. Then since d appeared k weeks ago and has not appeared since, we have

$$p(k) = \begin{cases} p(1-p)^k, & 0 \leqslant k \leqslant n-1 \\ 0, & \text{otherwise} \end{cases}$$

Example (Weekly lottery)

Let p be the probability that a given number d is drawn in any given week. After n successive draws, let $p(k)$ be the probability that d last appeared k weeks ago. *What is* $p(k)$*?*

The number d appears with probability p and does not appear with probability $1-p$. Then since d appeared k weeks ago and has not appeared since, we have

$$p(k) = \begin{cases} p(1-p)^k, & 0 \leqslant k \leqslant n-1 \\ 0, & \text{otherwise} \end{cases}$$

Notice that

$$\sum_{k=0}^{n-1} p(k) = 1 - (1-p)^n \ ,$$

Example (Weekly lottery)

Let p be the probability that a given number d is drawn in any given week. After n successive draws, let $p(k)$ be the probability that d last appeared k weeks ago. *What is $p(k)$?*

The number d appears with probability p and does not appear with probability $1-p$. Then since d appeared k weeks ago and has not appeared since, we have

$$p(k) = \begin{cases} p(1-p)^k, & 0 \leqslant k \leqslant n-1 \\ 0, & \text{otherwise} \end{cases}$$

Notice that

$$\sum_{k=0}^{n-1} p(k) = 1 - (1-p)^n ,$$

where $(1-p)^n$ is the probability that d does not appear in all n weeks.

Negative binomial distribution

The binomial distribution answers the question:

Given n trials, what is the chance of k successes?

Negative binomial distribution

The binomial distribution answers the question:

Given n trials, what is the chance of k successes?

Suppose, instead, that we ask:

What is the chance we need n trials to obtain k successes?

Negative binomial distribution

The binomial distribution answers the question:

Given n trials, what is the chance of k successes?

Suppose, instead, that we ask:

What is the chance we need n trials to obtain k successes?

A Bernoulli trial is repeated until we attain k successes and let us call $p_k(n)$ the probability that the total number of trials is n.

Negative binomial distribution

The binomial distribution answers the question:

Given n trials, what is the chance of k successes?

Suppose, instead, that we ask:

What is the chance we need n trials to obtain k successes?

A Bernoulli trial is repeated until we attain k successes and let us call $p_k(n)$ the probability that the total number of trials is n. If we need n trials, it is because there are $k-1$ successes in the first $n-1$ trials and the nth trial was a success.

Negative binomial distribution

The binomial distribution answers the question:

Given n trials, what is the chance of k successes?

Suppose, instead, that we ask:

What is the chance we need n trials to obtain k successes?

A Bernoulli trial is repeated until we attain k successes and let us call $p_k(n)$ the probability that the total number of trials is n. If we need n trials, it is because there are $k-1$ successes in the first $n-1$ trials and the nth trial was a success. By independence,

$$p_k(n) = \binom{n-1}{k-1} p^{k-1} q^{n-k} \times p = \binom{n-1}{k-1} p^k q^{n-k}$$

for $n \geqslant k$. This is the **negative binomial distribution**.

Discrete random variables

- We have seen how to assign a probability distribution to experiments with numerical (particularly, integer) outcomes. However not all interesting experiments are of this type.

Discrete random variables

- We have seen how to assign a probability distribution to experiments with numerical (particularly, integer) outcomes. However not all interesting experiments are of this type.
- Even if the outcomes are numerical, we may be interested in some other numerical measure of the outcome; e.g., gamblers might be more interested in the monetary values of their winnings/losses than in the actual number of times that they win or lose. Such numerical measures are called *random variables*.

Discrete random variables

- We have seen how to assign a probability distribution to experiments with numerical (particularly, integer) outcomes. However not all interesting experiments are of this type.
- Even if the outcomes are numerical, we may be interested in some other numerical measure of the outcome; e.g., gamblers might be more interested in the monetary values of their winnings/losses than in the actual number of times that they win or lose. Such numerical measures are called *random variables*.

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is a **discrete random variable** on $(\Omega, \mathcal{F}, \mathbb{P})$ if

Discrete random variables

- We have seen how to assign a probability distribution to experiments with numerical (particularly, integer) outcomes. However not all interesting experiments are of this type.
- Even if the outcomes are numerical, we may be interested in some other numerical measure of the outcome; e.g., gamblers might be more interested in the monetary values of their winnings/losses than in the actual number of times that they win or lose. Such numerical measures are called *random variables*.

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is a **discrete random variable** on $(\Omega, \mathcal{F}, \mathbb{P})$ if

1. it takes countably many values $D = \{x_1, x_2, \dots\} \subset \mathbb{R}$, and

Discrete random variables

- We have seen how to assign a probability distribution to experiments with numerical (particularly, integer) outcomes. However not all interesting experiments are of this type.
- Even if the outcomes are numerical, we may be interested in some other numerical measure of the outcome; e.g., gamblers might be more interested in the monetary values of their winnings/losses than in the actual number of times that they win or lose. Such numerical measures are called *random variables*.

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is a **discrete random variable** on $(\Omega, \mathcal{F}, \mathbb{P})$ if

1. it takes countably many values $D = \{x_1, x_2, \ldots\} \subset \mathbb{R}$, and
2. for every $x_i \in D$, the set $\{\omega \in \Omega | X(\omega) = x_i\} \in \mathcal{F}$.

Examples

- If Ω is finite then any function $X : \Omega \to \mathbb{R}$ is a discrete random variable.

Examples

- If Ω is finite then any function $X : \Omega \to \mathbb{R}$ is a discrete random variable.
- In many practical situations, if Ω is a countable subset of $\mathbb{R}$ (e.g., $\Omega = \mathbb{Z}$) then the identity function $X(\omega) = \omega$ is a discrete random variable.

Examples

- If Ω is finite then any function $X : \Omega \to \mathbb{R}$ is a discrete random variable.
- In many practical situations, if Ω is a countable subset of $\mathbb{R}$ (e.g., $\Omega = \mathbb{Z}$) then the identity function $X(\omega) = \omega$ is a discrete random variable.
- In the game of darts, Ω is uncountable since it contains all the points in the dartboard on which the dart can land, but the score $X : \Omega \to \{0, 1, \ldots, 60\}$ is a discrete random variable.

Examples

- If Ω is finite then any function $X : \Omega \to \mathbb{R}$ is a discrete random variable.
- In many practical situations, if Ω is a countable subset of $\mathbb{R}$ (e.g., $\Omega = \mathbb{Z}$) then the identity function $X(\omega) = \omega$ is a discrete random variable.
- In the game of darts, Ω is uncountable since it contains all the points in the dartboard on which the dart can land, but the score $X : \Omega \to \{0, 1, \ldots, 60\}$ is a discrete random variable.

Notation

We will denote random variables by capital letters $T, V, X, Y, Z, \ldots$ and their values by lowercase letters $t, v, x, y, z, \ldots$.

Examples

- If Ω is finite then any function $X : \Omega \to \mathbb{R}$ is a discrete random variable.
- In many practical situations, if Ω is a countable subset of $\mathbb{R}$ (e.g., $\Omega = \mathbb{Z}$) then the identity function $X(\omega) = \omega$ is a discrete random variable.
- In the game of darts, Ω is uncountable since it contains all the points in the dartboard on which the dart can land, but the score $X : \Omega \to \{0, 1, \dots, 60\}$ is a discrete random variable.

Notation

We will denote random variables by capital letters $T, V, X, Y, Z, \dots$ and their values by lowercase letters $t, v, x, y, z, \dots$.

Please observe this convention very carefully!!!

Probability mass function

- Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking integer values. (There is no loss of generality in doing this, since any countable set can be labelled by integers.)

Probability mass function

- Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking integer values. (There is no loss of generality in doing this, since any countable set can be labelled by integers.)
- By definition of a discrete random variable, the subset $A_x = \{\omega \in \Omega | X(\omega) = x\}$ of Ω is an event and therefore it has a well-defined probability $\mathbb{P}(A_x) = \mathbb{P}(X = x)$.

Probability mass function

- Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking integer values. (There is no loss of generality in doing this, since any countable set can be labelled by integers.)
- By definition of a discrete random variable, the subset $A_x = \{\omega \in \Omega | X(\omega) = x\}$ of Ω is an event and therefore it has a well-defined probability $\mathbb{P}(A_x) = \mathbb{P}(X = x)$.
- This allows us to define a function f_X by $f_X(x) = \mathbb{P}(X = x)$, called the **probability mass function** of X.

Probability mass function

- Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking integer values. (There is no loss of generality in doing this, since any countable set can be labelled by integers.)
- By definition of a discrete random variable, the subset $A_x = \{\omega \in \Omega | X(\omega) = x\}$ of Ω is an event and therefore it has a well-defined probability $\mathbb{P}(A_x) = \mathbb{P}(X = x)$.
- This allows us to define a function f_X by $f_X(x) = \mathbb{P}(X = x)$, called the **probability mass function** of X.
- Being a probability, $0 \leqslant f_X(x) \leqslant 1$ for all $x \in \mathbb{R}$.

Probability mass function

- Let X be a discrete random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking integer values. (There is no loss of generality in doing this, since any countable set can be labelled by integers.)
- By definition of a discrete random variable, the subset $A_x = \{\omega \in \Omega | X(\omega) = x\}$ of Ω is an event and therefore it has a well-defined probability $\mathbb{P}(A_x) = \mathbb{P}(X = x)$.
- This allows us to define a function f_X by $f_X(x) = \mathbb{P}(X = x)$, called the **probability mass function** of X.
- Being a probability, $0 \leqslant f_X(x) \leqslant 1$ for all $x \in \mathbb{R}$.
- Since the A_x for $x \in \mathbb{Z}$ are a countable partition of Ω, the countable additivity of $\mathbb{P}$ implies that

$$\sum_{x\in\mathbb{Z}} f_X(x) = \sum_{x\in\mathbb{Z}} \mathbb{P}(A_x) = \mathbb{P}\left(\bigcup_{x\in\mathbb{Z}} A_x\right) = \mathbb{P}(\Omega) = 1 \, .$$

Remarks

- In the case of $\Omega = \mathbb{Z}$ and X being the identity function $X(\omega) = \omega$, $f_X(x)$ is what we called the probability distribution $p(x)$.

Remarks

- In the case of $\Omega = \mathbb{Z}$ and X being the identity function $X(\omega) = \omega$, $f_X(x)$ is what we called the probability distribution $p(x)$.
- Provided that we are only interested in X (and other random variables we may build out of X), we can essentially forget about $(\Omega, \mathcal{F}, \mathbb{P})$ and work with only the probability mass function f_X.

Remarks

- In the case of $\Omega = \mathbb{Z}$ and X being the identity function $X(\omega) = \omega$, $f_X(x)$ is what we called the probability distribution $p(x)$.
- Provided that we are only interested in X (and other random variables we may build out of X), we can essentially forget about $(\Omega, \mathcal{F}, \mathbb{P})$ and work with only the probability mass function f_X. We often speak about "*a discrete random variable* X *with probability mass function* f_X" without bothering to mention the probability space on which X is defined.

Remarks

- In the case of $\Omega = \mathbb{Z}$ and X being the identity function $X(\omega) = \omega$, $f_X(x)$ is what we called the probability distribution $p(x)$.
- Provided that we are only interested in X (and other random variables we may build out of X), we can essentially forget about $(\Omega, \mathcal{F}, \mathbb{P})$ and work with only the probability mass function f_X. We often speak about "*a discrete random variable* X *with probability mass function* f_X" without bothering to mention the probability space on which X is defined.
- The probability distributions we have been discussing can play the rôle of probability mass functions.

Remarks

- In the case of $\Omega = \mathbb{Z}$ and X being the identity function $X(\omega) = \omega$, $f_X(x)$ is what we called the probability distribution $p(x)$.
- Provided that we are only interested in X (and other random variables we may build out of X), we can essentially forget about $(\Omega, \mathcal{F}, \mathbb{P})$ and work with only the probability mass function f_X. We often speak about "*a discrete random variable* X *with probability mass function* f_X" without bothering to mention the probability space on which X is defined.
- The probability distributions we have been discussing can play the rôle of probability mass functions.
- One can talk about discrete random variables with uniform, binomial, geometric, Benford,... probability mass functions.

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $f(x) \geqslant 0$ for all x and that it is nonzero only for a countable subset of $\mathbb{R}$; namely, the natural numbers.

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $f(x) \geqslant 0$ for all x and that it is nonzero only for a countable subset of $\mathbb{R}$; namely, the natural numbers. Finally

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $f(x) \geqslant 0$ for all x and that it is nonzero only for a countable subset of $\mathbb{R}$; namely, the natural numbers. Finally

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $f(x) \geqslant 0$ for all x and that it is nonzero only for a countable subset of $\mathbb{R}$; namely, the natural numbers. Finally

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Example (Poisson distribution)

Let $\lambda > 0$ be a positive real number. The **Poisson distribution** with parameter λ is defined by

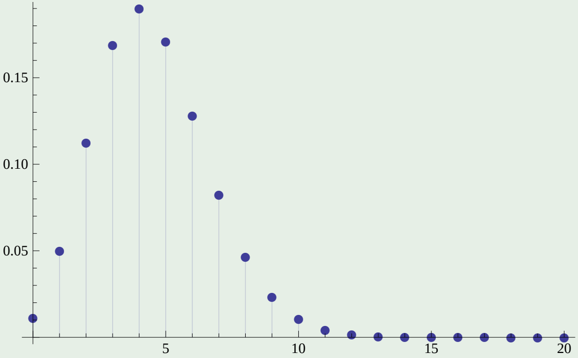
$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $f(x) \geqslant 0$ for all x and that it is nonzero only for a countable subset of $\mathbb{R}$; namely, the natural numbers. Finally

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

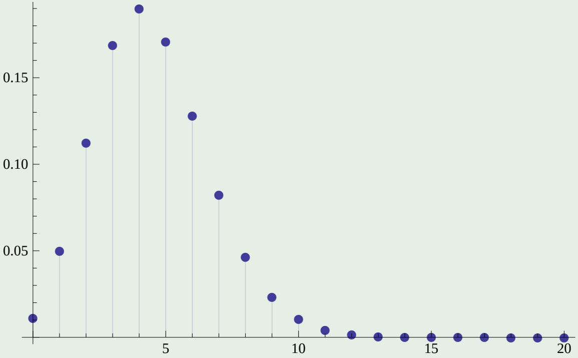
We can therefore talk about discrete random variables whose probability mass function is Poisson with parameter λ.

Example (Poisson distribution – continued)



We will see later that the Poisson distribution is a limit of the binomial distribution for large n and small p keeping np fixed.

Example (Poisson distribution – continued)



We will see later that the Poisson distribution is a limit of the binomial distribution for large n and small p keeping np fixed. This means that we can use it to approximate the binomial distribution in that limit.

Expectation value as a weighted average

Let X be a discrete random variable with probability mass function f_X. The **expectation value** $\mathbb{E}(X)$ of X is defined by

$$\mathbb{E}(X) = \sum_x x f_X(x) \ .$$

Expectation value as a weighted average

Let X be a discrete random variable with probability mass function f_X. The **expectation value** $\mathbb{E}(X)$ of X is defined by

$$\mathbb{E}(X) = \sum_x x f_X(x) \ .$$

(provided that $\sum_x |x| f_X(x) < \infty$.)

Expectation value as a weighted average

Let X be a discrete random variable with probability mass function f_X. The **expectation value** $\mathbb{E}(X)$ of X is defined by

$$\mathbb{E}(X) = \sum_x x f_X(x) \; .$$

(provided that $\sum_x |x| f_X(x) < \infty$.)

The expectation value agrees with our notion of *mean* or *average* in the case of a uniform distribution.

Expectation value as a weighted average

Let X be a discrete random variable with probability mass function f_X. The **expectation value** $\mathbb{E}(X)$ of X is defined by

$$\mathbb{E}(X) = \sum_x x f_X(x) \ .$$

(provided that $\sum_x |x| f_X(x) < \infty$.)

The expectation value agrees with our notion of *mean* or *average* in the case of a uniform distribution.

Example (Dice)

Consider rolling a dice and let X denote the random variable $X(⚀) = 1$, $X(⚁) = 2$, et cetera.

Expectation value as a weighted average

Let X be a discrete random variable with probability mass function f_X. The **expectation value** $\mathbb{E}(X)$ of X is defined by

$$\mathbb{E}(X) = \sum_x x f_X(x) \,.$$

(provided that $\sum_x |x| f_X(x) < \infty$.)

The expectation value agrees with our notion of *mean* or *average* in the case of a uniform distribution.

Example (Dice)

Consider rolling a dice and let X denote the random variable $X($⚀$) = 1$, $X($⚁$) = 2$, et cetera. Then $\mathbb{E}(X)$ is the average score:

$$\mathbb{E}(X) = \sum_{x=1}^{6} x f_X(x) = \sum_{x=1}^{6} x \tfrac{1}{6} = \tfrac{1}{6}(1 + 2 + \cdots + 6) = \tfrac{7}{2} \,.$$

Example (Betting)

Consider a betting game based on a Bernoulli trial with success probability p. Every bet costs $\pounds 1$: if you win you get your $\pounds 1$ back and an additional $\pounds 2$, if you lose you get nothing. *How much do you expect to win/lose on average?*

Example (Betting)

Consider a betting game based on a Bernoulli trial with success probability p. Every bet costs $\pounds 1$: if you win you get your $\pounds 1$ back and an additional $\pounds 2$, if you lose you get nothing. *How much do you expect to win/lose on average?*
Introduce the random variable X with values $X(S) = 2$ and $X(F) = -1$ which measures the amount (in $\pounds$) you win: in the case of success you win $\pounds 2$ and in the case of failure you lose $\pounds 1$, which is the same as winning $-\pounds 1$.

Example (Betting)

Consider a betting game based on a Bernoulli trial with success probability p. Every bet costs $\pounds 1$: if you win you get your $\pounds 1$ back and an additional $\pounds 2$, if you lose you get nothing. *How much do you expect to win/lose on average?*

Introduce the random variable X with values $X(S) = 2$ and $X(F) = -1$ which measures the amount (in $\pounds$) you win: in the case of success you win $\pounds 2$ and in the case of failure you lose $\pounds 1$, which is the same as winning $-\pounds 1$. We are after the expectation value of X:

$$\mathbb{E}(X) = X(S)\mathbb{P}(S) + X(F)\mathbb{P}(F) = 2p + (-1)(1 - p) = 3p - 1 .$$

Example (Betting)

Consider a betting game based on a Bernoulli trial with success probability p. Every bet costs $£1$: if you win you get your $£1$ back and an additional $£2$, if you lose you get nothing. *How much do you expect to win/lose on average?*

Introduce the random variable X with values $X(S) = 2$ and $X(F) = -1$ which measures the amount (in $£$) you win: in the case of success you win $£2$ and in the case of failure you lose $£1$, which is the same as winning $-£1$. We are after the expectation value of X:

$$\mathbb{E}(X) = X(S)\mathbb{P}(S) + X(F)\mathbb{P}(F) = 2p + (-1)(1-p) = 3p - 1\ .$$

So if $p < \frac{1}{3}$ you shouldn't play!

Example (Betting)

Consider a betting game based on a Bernoulli trial with success probability p. Every bet costs $£1$: if you win you get your $£1$ back and an additional $£2$, if you lose you get nothing. *How much do you expect to win/lose on average?*

Introduce the random variable X with values $X(S) = 2$ and $X(F) = -1$ which measures the amount (in $£$) you win: in the case of success you win $£2$ and in the case of failure you lose $£1$, which is the same as winning $-£1$. We are after the expectation value of X:

$$\mathbb{E}(X) = X(S)\mathbb{P}(S) + X(F)\mathbb{P}(F) = 2p + (-1)(1 - p) = 3p - 1 .$$

So if $p < \frac{1}{3}$ you shouldn't play!

Notice that

$$X(S)\mathbb{P}(S) + X(F)\mathbb{P}(F) = 2f_X(2) + (-1)f_X(-1) .$$

Example (Expectation value of Poisson distribution)

Let X be a discrete random variable with probability mass function f_X given by a Poisson distribution with parameter $\lambda > 0$. *What is its expectation value?*

Example (Expectation value of Poisson distribution)

Let X be a discrete random variable with probability mass function f_X given by a Poisson distribution with parameter $\lambda > 0$. *What is its expectation value?* By definition

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = e^{-\lambda} \sum_{z=0} \frac{\lambda^{z+1}}{z!} \\ &= \lambda e^{-\lambda} \sum_{z=0} \frac{\lambda^z}{z!} = \lambda\end{aligned}$$

Example (Expectation value of binomial distribution)

Let X be a discrete random variable with probability mass function f_X given by a binomial distribution with parameters n and p. *What is* $\mathbb{E}(X)$*?*

Example (Expectation value of binomial distribution)

Let X be a discrete random variable with probability mass function f_X given by a binomial distribution with parameters n and p. *What is* $\mathbb{E}(X)$*?* By definition,

$$\mathbb{E}(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x q^{n-x} .$$

But now for $0 < x \leqslant n$,

$$x \binom{n}{x} = x \frac{n!}{(n-x)!x!} = \frac{n!}{(n-x)!(x-1)!} = n \binom{n-1}{x-1}$$

whence

$$\mathbb{E}(X) = \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x q^{n-x} = \sum_{z=0}^{n-1} n \binom{n-1}{z} p^{z+1} q^{n-1-z} = np .$$

Summary

- A **discrete random variable** X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ which can take only countably many values and such that the subsets $\{X = x\}$ are events.

Summary

- A **discrete random variable** X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ which can take only countably many values and such that the subsets $\{X = x\}$ are events.
- Since they are events, they have a probability $\mathbb{P}(X = x)$, which defines a **probability mass function** $f_X(x) = \mathbb{P}(X = x)$ obeying $0 \leqslant f_X(x) \leqslant 1$ and $\sum_x f_X(x) = 1$.

Summary

- A **discrete random variable** X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ which can take only countably many values and such that the subsets $\{X = x\}$ are events.
- Since they are events, they have a probability $\mathbb{P}(X = x)$, which defines a **probability mass function** $f_X(x) = \mathbb{P}(X = x)$ obeying $0 \leqslant f_X(x) \leqslant 1$ and $\sum_x f_X(x) = 1$.
- Given a discrete random variable X with probability mass function f_X, its **expectation value** is $\mathbb{E}(X) = \sum_x x f_X(x)$.

Summary

- A **discrete random variable** X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ which can take only countably many values and such that the subsets $\{X = x\}$ are events.
- Since they are events, they have a probability $\mathbb{P}(X = x)$, which defines a **probability mass function** $f_X(x) = \mathbb{P}(X = x)$ obeying $0 \leqslant f_X(x) \leqslant 1$ and $\sum_x f_X(x) = 1$.
- Given a discrete random variable X with probability mass function f_X, its **expectation value** is $\mathbb{E}(X) = \sum_x x f_X(x)$.
- We met the **Poisson distribution** with parameter $\lambda > 0$

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

for $x \in \mathbb{N}$ and $f(x) = 0$ otherwise. It has expectation value λ.