Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 5
1 February 2012

The story of the film so far...

- Partition rule: $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$

The story of the film so far...

- Partition rule: $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$
- Generalises to a partition $\{B_i\}$ of the sample space:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

The story of the film so far...

- Partition rule: $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$
- Generalises to a partition $\{B_i\}$ of the sample space:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

- It also applies to conditional probability:

$$\mathbb{P}(A|C) = \sum_i \mathbb{P}(A|B_i \cap C)\mathbb{P}(B_i|C)$$

The story of the film so far...

- Partition rule: $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$
- Generalises to a partition $\{B_i\}$ of the sample space:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

- It also applies to conditional probability:

$$\mathbb{P}(A|C) = \sum_i \mathbb{P}(A|B_i \cap C)\mathbb{P}(B_i|C)$$

- Bayes's rule allows us to compute $\mathbb{P}(A|B)$ from a knowledge of $\mathbb{P}(B|A)$ via

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

Conditional independence

- We discussed the notion of independent events: events A and B such that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Conditional independence

- We discussed the notion of independent events: events A and B such that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A typical example might be tossing a coin: the events of "getting a head in the first toss" and "getting a head in the second toss" are independent

Conditional independence

- We discussed the notion of independent events: events A and B such that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A typical example might be tossing a coin: the events of "getting a head in the first toss" and "getting a head in the second toss" are independent
- We also have the notion of events A, B which become independent once a third event C has occured

Conditional independence

- We discussed the notion of independent events: events A and B such that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A typical example might be tossing a coin: the events of “getting a head in the first toss” and “getting a head in the second toss” are independent
- We also have the notion of events A, B which become independent once a third event C has occured

Definition

Let A, B and C be events.

Conditional independence

- We discussed the notion of independent events: events A and B such that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A typical example might be tossing a coin: the events of "getting a head in the first toss" and "getting a head in the second toss" are independent
- We also have the notion of events A, B which become independent once a third event C has occured

Definition

Let A, B and C be events. We say that A and B are **conditionally independent** (given C), if

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice.

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice. Let H_1 (resp. H_2) denote the event of getting heads in the first (resp. second) toss.

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice. Let H_1 (resp. H_2) denote the event of getting heads in the first (resp. second) toss.

The events H_1 and H_2 are **not** independent, but if we condition them to the chosen coin, then they are.

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice. Let H_1 (resp. H_2) denote the event of getting heads in the first (resp. second) toss.
The events H_1 and H_2 are **not** independent, but if we condition them to the chosen coin, then they are. In other words, let C stand for the event of having chosen a given coin.

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice. Let H_1 (resp. H_2) denote the event of getting heads in the first (resp. second) toss.

The events H_1 and H_2 are **not** independent, but if we condition them to the chosen coin, then they are. In other words, let C stand for the event of having chosen a given coin. Then

$$\mathbb{P}(H_1 \cap H_2|C) = \mathbb{P}(H_1|C)\mathbb{P}(H_2|C)$$

Example

Suppose that we have a bag containing two coins: a fair coin and a double-headed coin. We choose a coin at random and toss it twice. Let H_1 (resp. H_2) denote the event of getting heads in the first (resp. second) toss.

The events H_1 and H_2 are **not** independent, but if we condition them to the chosen coin, then they are. In other words, let C stand for the event of having chosen a given coin. Then

$$\mathbb{P}(H_1 \cap H_2|C) = \mathbb{P}(H_1|C)\mathbb{P}(H_2|C)$$

yet

$$\mathbb{P}(H_1 \cap H_2) \neq \mathbb{P}(H_1)\mathbb{P}(H_2) .$$

Example (Continued)

Indeed, by the partition rule and letting F denote the event of having picked the fair coin,

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(H_1 \cap H_2|F)\mathbb{P}(F) + \mathbb{P}(H_1 \cap H_2|F^c)\mathbb{P}(F^c) ,$$

Example (Continued)

Indeed, by the partition rule and letting F denote the event of having picked the fair coin,

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(H_1 \cap H_2 | F)\mathbb{P}(F) + \mathbb{P}(H_1 \cap H_2 | F^c)\mathbb{P}(F^c) ,$$

where $\mathbb{P}(H_1 \cap H_2 | F) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ and $\mathbb{P}(H_1 \cap H_2 | F^c) = 1$, whence

$$\mathbb{P}(H_1 \cap H_2) = (\tfrac{1}{4} \times \tfrac{1}{2}) + (1 \times \tfrac{1}{2}) = \tfrac{5}{8} .$$

Example (Continued)

Indeed, by the partition rule and letting F denote the event of having picked the fair coin,

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(H_1 \cap H_2|F)\mathbb{P}(F) + \mathbb{P}(H_1 \cap H_2|F^c)\mathbb{P}(F^c) ,$$

where $\mathbb{P}(H_1 \cap H_2|F) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ and $\mathbb{P}(H_1 \cap H_2|F^c) = 1$, whence

$$\mathbb{P}(H_1 \cap H_2) = (\tfrac{1}{4} \times \tfrac{1}{2}) + (1 \times \tfrac{1}{2}) = \tfrac{5}{8} .$$

On the other hand, the probability of getting a head is $\frac{3}{4}$ since there are four faces in total, three of which are heads, whence

$$\mathbb{P}(H_1)\mathbb{P}(H_2) = \tfrac{3}{4} \times \tfrac{3}{4} = \tfrac{9}{16} .$$

Numerical outcomes

It is often the case that the outcomes of a trial are numbers or can be converted into numbers.

Numerical outcomes

It is often the case that the outcomes of a trial are numbers or can be converted into numbers.

Notation

We will denote such numerical outcomes by capital letters $X, Y, \ldots$ and their values by lowercase $x, y, \ldots$.

Numerical outcomes

It is often the case that the outcomes of a trial are numbers or can be converted into numbers.

Notation

We will denote such numerical outcomes by capital letters $X, Y, \ldots$ and their values by lowercase $x, y, \ldots$.

Please observe this convention very carefully!!!

Numerical outcomes

It is often the case that the outcomes of a trial are numbers or can be converted into numbers.

Notation

We will denote such numerical outcomes by capital letters $X, Y, \ldots$ and their values by lowercase $x, y, \ldots$.

Please observe this convention very carefully!!!

Possible events now include

$$\{a < X \leqslant b\} \qquad \{X = x\} \qquad \{Y > 0\}$$

Numerical outcomes

It is often the case that the outcomes of a trial are numbers or can be converted into numbers.

Notation

We will denote such numerical outcomes by capital letters $X, Y, \ldots$ and their values by lowercase $x, y, \ldots$.

Please observe this convention very carefully!!!

Possible events now include

$$\{a < X \leqslant b\} \qquad \{X = x\} \qquad \{Y > 0\}$$

and we will denote their probabilities by

$$\mathbb{P}(a < X \leqslant b) \qquad \mathbb{P}(X = x) \qquad \mathbb{P}(Y > 0) ,$$

respectively.

Discrete probability distributions

- Let us consider an experiment whose outcomes X are integers. The **probability distribution of** X is the function $p : \mathbb{Z} \to \mathbb{R}$ defined by $p(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$.

Discrete probability distributions

- Let us consider an experiment whose outcomes X are integers. The **probability distribution of X** is the function $p : \mathbb{Z} \to \mathbb{R}$ defined by $p(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$.
- It obeys $0 \leqslant p(x) \leqslant 1$ and (see later) $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Discrete probability distributions

- Let us consider an experiment whose outcomes X are integers. The **probability distribution of X** is the function $p : \mathbb{Z} \to \mathbb{R}$ defined by $p(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$.
- It obeys $0 \leqslant p(x) \leqslant 1$ and (see later) $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Dice)

Consider rolling a fair die. The possible outcomes are ⚀, ⚁, . . . , ⚅, which we convert to a numerical outcome $X \in \{1, 2, \ldots, 6\}$ in the obvious way.

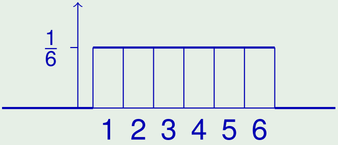
Discrete probability distributions

- Let us consider an experiment whose outcomes X are integers. The **probability distribution of** X is the function $p : \mathbb{Z} \to \mathbb{R}$ defined by $p(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$.
- It obeys $0 \leqslant p(x) \leqslant 1$ and (see later) $\sum_{x\in\mathbb{Z}} p(x) = 1$.

Example (Dice)

Consider rolling a fair die. The possible outcomes are ⚀, ⚁, . . . , ⚅, which we convert to a numerical outcome $X \in \{1, 2, \ldots, 6\}$ in the obvious way. Then

$$p(x) = \begin{cases} \frac{1}{6}, & x \in \{1,2,3,4,5,6\} \\ 0, & \text{otherwise} \end{cases}$$



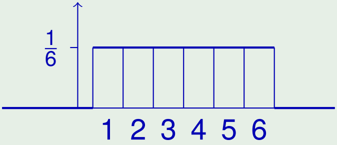
Discrete probability distributions

- Let us consider an experiment whose outcomes X are integers. The **probability distribution of X** is the function $p : \mathbb{Z} \to \mathbb{R}$ defined by $p(x) = \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$.
- It obeys $0 \leqslant p(x) \leqslant 1$ and (see later) $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Dice)

Consider rolling a fair die. The possible outcomes are ⚀, ⚁, . . . , ⚅, which we convert to a numerical outcome $X \in \{1, 2, \ldots, 6\}$ in the obvious way. Then

$$p(x) = \begin{cases} \frac{1}{6}, & x \in \{1,2,3,4,5,6\} \\ 0, & \text{otherwise} \end{cases}$$

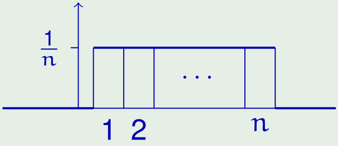


Notice that $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Uniform distribution)

Generalising the above, we define the **uniform distribution** on $\{1, 2, \ldots, n\}$ by

$$p(x) = \begin{cases} \frac{1}{n}, & x \in \{1, 2, \ldots, n\} \\ 0, & \text{otherwise} \end{cases}$$

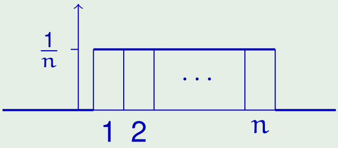


Again notice that $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Uniform distribution)

Generalising the above, we define the **uniform distribution** on $\{1, 2, \ldots, n\}$ by

$$p(x) = \begin{cases} \frac{1}{n}, & x \in \{1, 2, \ldots, n\} \\ 0, & \text{otherwise} \end{cases}$$



Again notice that $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Bernoulli trials)

Consider a Bernoulli trial with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. Let $X \in \{0, 1\}$ denote the number of successes, so that $p(0) = q$ and $p(1) = p$ and $p(x) = 0$ for $x \neq 0, 1$. Of course, $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Independent Bernoulli trials)

We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \ldots, n\}$ denote the number of successes.

Example (Independent Bernoulli trials)

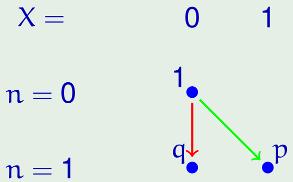
We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \ldots, n\}$ denote the number of successes.

$X = \quad 0$

$n = 0 \quad 1$ •

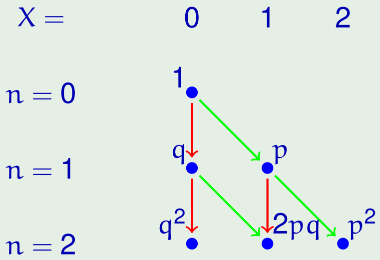
Example (Independent Bernoulli trials)

We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \dots, n\}$ denote the number of successes.



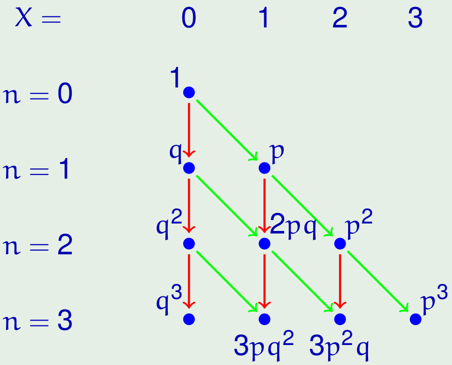
Example (Independent Bernoulli trials)

We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \ldots, n\}$ denote the number of successes.



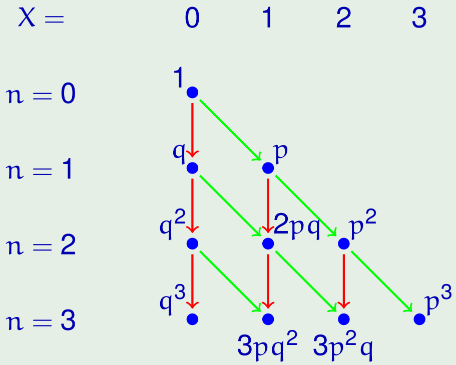
Example (Independent Bernoulli trials)

We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \ldots, n\}$ denote the number of successes.



Example (Independent Bernoulli trials)

We could also consider a sequence of n independent Bernoulli trials, each one with $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$. We let $X \in \{0, 1, \ldots, n\}$ denote the number of successes.



cf. Pascal's triangle!

Example (Binomial distribution)

Continuing with the previous example, it is clear that

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

Example (Binomial distribution)

Continuing with the previous example, it is clear that

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

It is called the **binomial distribution** (with parameters n and p). The quantity $p(x)$ is the probability of getting exactly x successes in n trials.

Example (Binomial distribution)

Continuing with the previous example, it is clear that

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

It is called the **binomial distribution** (with parameters n and p). The quantity $p(x)$ is the probability of getting exactly x successes in n trials. Notice that

$$\sum_{x \in \mathbb{Z}} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1 ,$$

by the binomial theorem.

Example (Tossing a fair coin)

Suppose we toss a fair coin n times.

Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads.

Example (Tossing a fair coin)

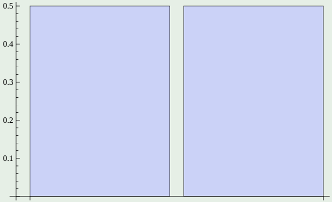
Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$

Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

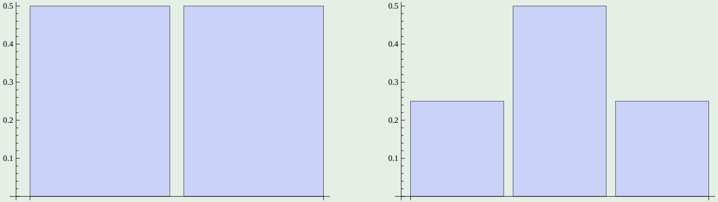
$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$



Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

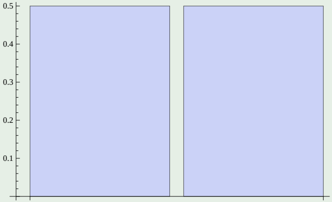
$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$

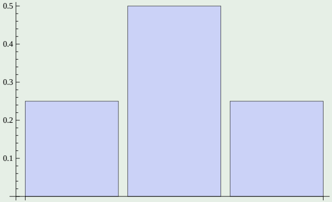


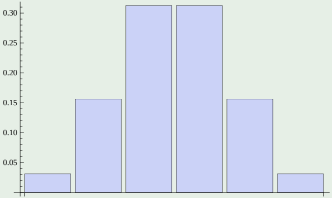
Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$



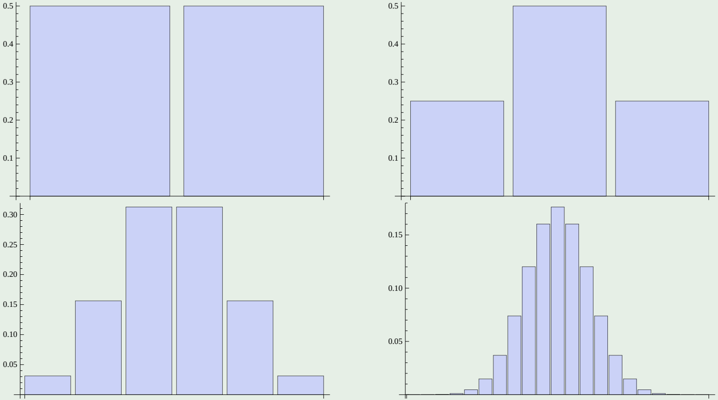




Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

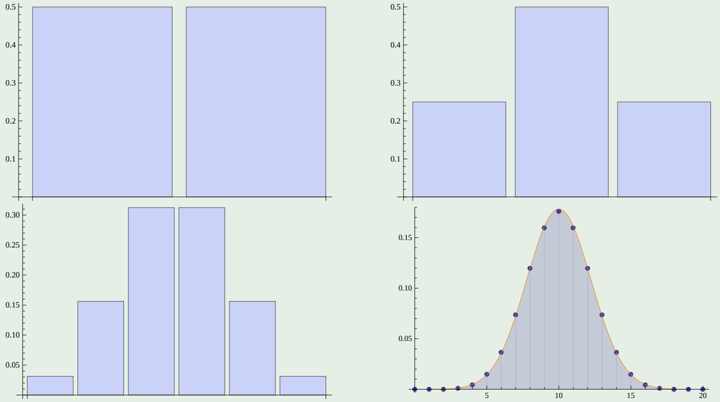
$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$



Example (Tossing a fair coin)

Suppose we toss a fair coin n times. This is just the previous example with $p = q = \frac{1}{2}$. Let X denote the number of heads. Then

$$p(x) = \begin{cases} \binom{n}{x} 2^{-n}, & 0 \leqslant x \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$



Example (The problem of the points)

In independent Bernoulli trials with success probability p, what is the probability that n successes occur before m failures?

Example (The problem of the points)

In independent Bernoulli trials with success probability p, what is the probability that n successes occur before m failures?

- This is the probability of there being at least n successes in the first $n + m - 1$ trials.

Example (The problem of the points)

In independent Bernoulli trials with success probability p, what is the probability that n successes occur before m failures?

- This is the probability of there being at least n successes in the first $n + m - 1$ trials.
- The probability of there being *exactly* k successes in $n + m - 1$ trials is given by the binomial distribution

$$\binom{n+m-1}{k} p^k q^{n+m-1-k}$$

Example (The problem of the points)

In independent Bernoulli trials with success probability p, what is the probability that n successes occur before m failures?

- This is the probability of there being at least n successes in the first $n+m-1$ trials.
- The probability of there being *exactly* k successes in $n+m-1$ trials is given by the binomial distribution

$$\binom{n+m-1}{k} p^k q^{n+m-1-k}$$

- Therefore the probability we are after is

$$\sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k q^{n+m-1-k}$$

Example (Benford's distribution)

Take any large collection of numerical data (e.g., census, statistical tables, physical constants,...).

Example (Benford's distribution)

Take any large collection of numerical data (e.g., census, statistical tables, physical constants,...). *What is the probability distribution of the first significant digit?*

Example (Benford's distribution)

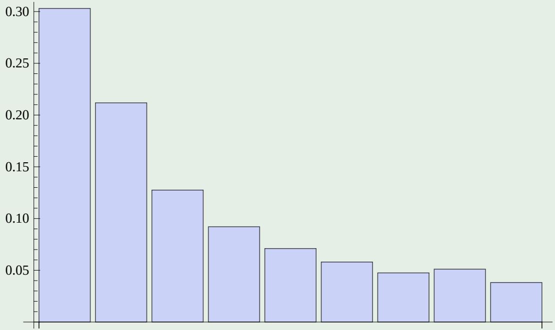
Take any large collection of numerical data (e.g., census, statistical tables, physical constants,...). *What is the probability distribution of the first significant digit?*

For example, consider the sizes of files (in 512K blocks) in my laptop (excluding directories).

Example (Benford's distribution)

Take any large collection of numerical data (e.g., census, statistical tables, physical constants,...). *What is the probability distribution of the first significant digit?*

For example, consider the sizes of files (in 512K blocks) in my laptop (excluding directories). It has over 2.5M files and the distribution of significant digits looks like this:



Example (Benford's distribution – continued)

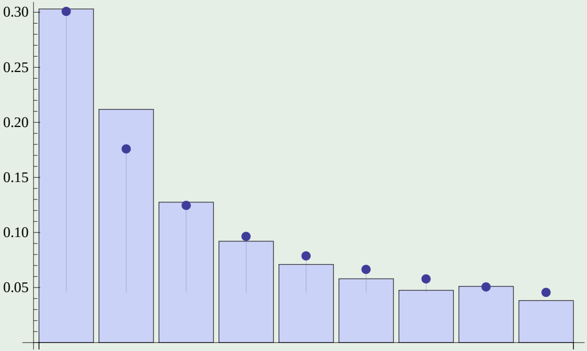
It is actually very close to **Benford's distribution**

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 1 \leqslant k \leqslant 9 \\ 0, & \text{otherwise.} \end{cases}$$

Example (Benford's distribution – continued)

It is actually very close to **Benford's distribution**

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 1 \leqslant k \leqslant 9 \\ 0, & \text{otherwise.} \end{cases}$$



Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?
Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits? Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1+\frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

$$\sum_{x\in\mathbb{Z}} p(x) = \sum_{k=10}^{99} \log_{10}(1+\frac{1}{k})$$

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?
Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

$$\sum_{x \in \mathbb{Z}} p(x) = \sum_{k=10}^{99} \log_{10}\left(\frac{k+1}{k}\right)$$

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?
Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

$$\begin{aligned} \sum_{x \in \mathbb{Z}} p(x) &= \sum_{k=10}^{99} \log_{10}\left(\frac{k+1}{k}\right) \\ &= \sum_{k=10}^{99} \left(\log_{10}(k+1) - \log_{10} k\right) \end{aligned}$$

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?
Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

$$\begin{aligned} \sum_{x \in \mathbb{Z}} p(x) &= \sum_{k=10}^{99} \log_{10}\left(\frac{k+1}{k}\right) \\ &= \sum_{k=10}^{99} \left(\log_{10}(k+1) - \log_{10} k\right) \\ &= \log_{10} 100 - \log_{10} 10 \end{aligned}$$

Example (Benford's distribution – continued)

How about the distribution of the first two significant digits?
Again, it is empirically very close to

$$p(k) = \begin{cases} \log_{10}(1 + \frac{1}{k}), & 10 \leqslant k \leqslant 99 \\ 0, & \text{otherwise} \end{cases}$$

which is **Benford's 2-digit distribution**.

$$\begin{aligned} \sum_{x \in \mathbb{Z}} p(x) &= \sum_{k=10}^{99} \log_{10}\left(\frac{k+1}{k}\right) \\ &= \sum_{k=10}^{99} (\log_{10}(k+1) - \log_{10} k) \\ &= \log_{10} 100 - \log_{10} 10 \\ &= 1 \end{aligned}$$

Example (Benford's distribution under change of base)

Pulponio is an M-class planet, not unlike our own, whose inhabitants count in base 8.

Example (Benford's distribution under change of base)

Pulponio is an M-class planet, not unlike our own, whose inhabitants count in base 8. Their chief scientist, Dr O. Fneb, observed empirically that the distribution of the most significant digit in their statistical tables was very close to

$$p(k) = \begin{cases} \log_8(1 + \frac{1}{k}), & 1 \leqslant k \leqslant 7 \\ 0, & \text{otherwise.} \end{cases}$$

Example (Benford's distribution under change of base)

Pulponio is an M-class planet, not unlike our own, whose inhabitants count in base 8. Their chief scientist, Dr O. Fneb, observed empirically that the distribution of the most significant digit in their statistical tables was very close to

$$p(k) = \begin{cases} \log_8(1 + \frac{1}{k}), & 1 \leqslant k \leqslant 7 \\ 0, & \text{otherwise.} \end{cases}$$

Should this surprise us?

Example (Benford's distribution under change of base)

Pulponio is an M-class planet, not unlike our own, whose inhabitants count in base 8. Their chief scientist, Dr O. Fneb, observed empirically that the distribution of the most significant digit in their statistical tables was very close to

$$p(k) = \begin{cases} \log_8(1 + \frac{1}{k}), & 1 \leqslant k \leqslant 7 \\ 0, & \text{otherwise.} \end{cases}$$

Should this surprise us? It should not. In fact, if we take our own statistical tables and re-express the entries in base b instead of base 10, we still get a distribution which is close to

$$p(k) = \begin{cases} \log_b(1 + \frac{1}{k}), & 1 \leqslant k \leqslant b - 1 \\ 0, & \text{otherwise.} \end{cases}$$

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).
- The effect of changing units is simply to multiply the numbers by the relevant conversion factor

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).
- The effect of changing units is simply to multiply the numbers by the relevant conversion factor
- Under $X \mapsto \alpha X$, $\log_{10} X \mapsto \log_{10} X + \log_{10} \alpha$

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).
- The effect of changing units is simply to multiply the numbers by the relevant conversion factor
- Under $X \mapsto \alpha X$, $\log_{10} X \mapsto \log_{10} X + \log_{10} \alpha$
- This means that if we instead took the logarithms of the numbers, the distribution should not change if we add some constant ($\log_{10} \alpha$).

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).
- The effect of changing units is simply to multiply the numbers by the relevant conversion factor
- Under $X \mapsto \alpha X$, $\log_{10} X \mapsto \log_{10} X + \log_{10} \alpha$
- This means that if we instead took the logarithms of the numbers, the distribution should not change if we add some constant ($\log_{10} \alpha$).
- The only distribution which does not change under this transformation is the uniform distribution.

Example (Benford's distribution under change of units)

- If there is any "truth" to Benford's observation, it should be independent of which units are used (e.g., metric vs. imperial,...).
- The effect of changing units is simply to multiply the numbers by the relevant conversion factor
- Under $X \mapsto \alpha X$, $\log_{10} X \mapsto \log_{10} X + \log_{10} \alpha$
- This means that if we instead took the logarithms of the numbers, the distribution should not change if we add some constant ($\log_{10} \alpha$).
- The only distribution which does not change under this transformation is the uniform distribution.
- Indeed, Benford's distribution corresponds to the uniform distribution of the first significant digit of the logarithms of the numbers!

General properties of discrete probability distributions

Let $p : \mathbb{Z} \to \mathbb{R}$ be a discrete probability distribution.

General properties of discrete probability distributions

Let $p : \mathbb{Z} \to \mathbb{R}$ be a discrete probability distribution.

- It is clear from the definition $p(x) = \mathbb{P}(X = x)$, that $0 \leqslant p(x) \leqslant 1$.

General properties of discrete probability distributions

Let $p : \mathbb{Z} \to \mathbb{R}$ be a discrete probability distribution.

- It is clear from the definition $p(x) = \mathbb{P}(X = x)$, that $0 \leqslant p(x) \leqslant 1$.
- If $x_1 \neq x_2$, then $\{X = x_1\}$ and $\{X = x_2\}$ are disjoint, whence

$$\mathbb{P}(X \in \{x_1, x_2\}) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) = p(x_1) + p(x_2) \;.$$

General properties of discrete probability distributions

Let $p : \mathbb{Z} \to \mathbb{R}$ be a discrete probability distribution.

- It is clear from the definition $p(x) = \mathbb{P}(X = x)$, that $0 \leqslant p(x) \leqslant 1$.
- If $x_1 \neq x_2$, then $\{X = x_1\}$ and $\{X = x_2\}$ are disjoint, whence

$$\mathbb{P}(X \in \{x_1, x_2\}) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) = p(x_1) + p(x_2) \; .$$

- More generally, using the countable additivity of $\mathbb{P}$,

$$\mathbb{P}(X \in C) = \sum_{x \in C} p(x)$$

General properties of discrete probability distributions

Let $p : \mathbb{Z} \to \mathbb{R}$ be a discrete probability distribution.

- It is clear from the definition $p(x) = \mathbb{P}(X = x)$, that $0 \leqslant p(x) \leqslant 1$.
- If $x_1 \neq x_2$, then $\{X = x_1\}$ and $\{X = x_2\}$ are disjoint, whence

$$\mathbb{P}(X \in \{x_1, x_2\}) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) = p(x_1) + p(x_2) .$$

- More generally, using the countable additivity of $\mathbb{P}$,

$$\mathbb{P}(X \in C) = \sum_{x \in C} p(x)$$

- and since $\mathbb{P}(\Omega) = 1$, it follows that $\sum_{x \in \mathbb{Z}} p(x) = 1$.

Example (Errors in a bit stream)

Consider a bit stream being transmitted across a noisy channel in which the probability of a transmission error is p independently for each bit transmitted. *What is the probability of at least one error in* n *bits transmitted?*

Example (Errors in a bit stream)

Consider a bit stream being transmitted across a noisy channel in which the probability of a transmission error is p independently for each bit transmitted. *What is the probability of at least one error in n bits transmitted?*

The complementary event is when the n bits have been transmitted error-free, whose probability is $(1-p)^n$. Therefore, the probability we are after is $1-(1-p)^n$.

Example (Errors in a bit stream)

Consider a bit stream being transmitted across a noisy channel in which the probability of a transmission error is p independently for each bit transmitted. *What is the probability of at least one error in n bits transmitted?*

The complementary event is when the n bits have been transmitted error-free, whose probability is $(1-p)^n$. Therefore, the probability we are after is $1-(1-p)^n$.

Now suppose that we transmit each bit three times ("Bellman's code") and the receiver interprets the most common bit as correct. *What is the probability of an incorrect transmission?*

Example (Errors in a bit stream)

Consider a bit stream being transmitted across a noisy channel in which the probability of a transmission error is p independently for each bit transmitted. *What is the probability of at least one error in n bits transmitted?*

The complementary event is when the n bits have been transmitted error-free, whose probability is $(1-p)^n$. Therefore, the probability we are after is $1-(1-p)^n$.

Now suppose that we transmit each bit three times ("Bellman's code") and the receiver interprets the most common bit as correct. *What is the probability of an incorrect transmission?*

The only outcomes resulting in a transmission error are those where there are at least two bits in error, whose probability is

$$p^3 + 3p^2(1-p) = p^2(3-2p)$$

Distribution function

Definition

The function $F : \mathbb{Z} \to \mathbb{R}$ defined by

$$F(x) = \sum_{t \leqslant x} p(t) = \mathbb{P}(X \leqslant x)$$

is called the **distribution function** of X.

Distribution function

Definition

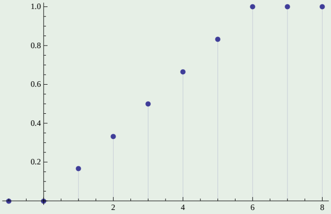
The function $F : \mathbb{Z} \to \mathbb{R}$ defined by

$$F(x) = \sum_{t \leqslant x} p(t) = \mathbb{P}(X \leqslant x)$$

is called the **distribution function** of X.

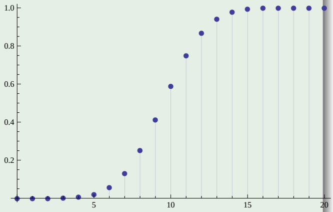
Example (Rolling a fair die)

$$F(x) = \begin{cases} 0, & x \in \{0, -1, -2, \dots\} \\ \frac{x}{6}, & x \in \{1, 2, 3, 4, 5, 6\} \\ 1, & x \in \{7, 8, 9, \dots\}. \end{cases}$$



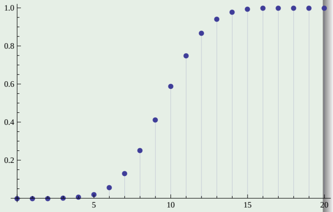
Example (Binomial distribution function with $p = \frac{1}{2}$)

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum_{k=0}^{x} \binom{n}{k} 2^{-n}, & 0 \leqslant x \leqslant n \\ 1, & x > n. \end{cases}$$



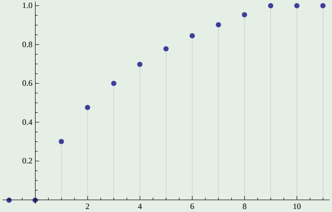
Example (Binomial distribution function with $p = \frac{1}{2}$)

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum_{k=0}^{x} \binom{n}{k} 2^{-n}, & 0 \leqslant x \leqslant n \\ 1, & x > n. \end{cases}$$



Example (Benford's distribution function)

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ \log_{10}(x+1), & 1 \leqslant x \leqslant 9 \\ 1, & x \geqslant 10. \end{cases}$$



General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

- $0 \leqslant F(x) \leqslant 1$ for all x

General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

- $0 \leqslant F(x) \leqslant 1$ for all x
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$

General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

- $0 \leqslant F(x) \leqslant 1$ for all x
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
- $F(x) - F(x-1) = p(x)$

General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

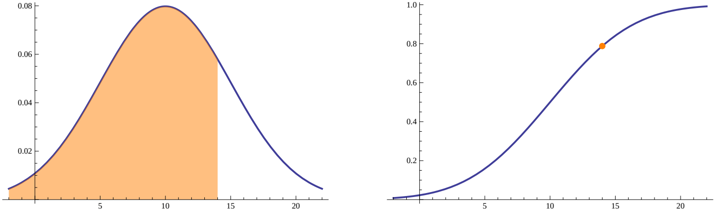
- $0 \leqslant F(x) \leqslant 1$ for all x
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- $F(x) - F(x-1) = p(x)$
- $x_1 \leqslant x_2$ if and only if $F(x_1) \leqslant F(x_2)$

General properties of distribution functions

The distribution function $F : \mathbb{Z} \to \mathbb{R}$ satisfies the following general properties:

- $0 \leqslant F(x) \leqslant 1$ for all x
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- $F(x) - F(x-1) = p(x)$
- $x_1 \leqslant x_2$ if and only if $F(x_1) \leqslant F(x_2)$

This is not unlike an area $F(x) = \int_{-\infty}^{x} p(y)dy$:



Summary

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

Summary

- 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:

Summary

- 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 $E = \{1, 2, \ldots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$

Summary

- 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 $E = \{1, 2, \ldots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **Benford's** on 1 digit: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 1 \leqslant x \leqslant 9 \\ 0, & \text{otherwise} \end{cases}$

Summary

- 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 $E = \{1, 2, \ldots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **Benford's** on 1 digit: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 1 \leqslant x \leqslant 9 \\ 0, & \text{otherwise} \end{cases}$
  - **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

Summary

- 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 $E = \{1, 2, \dots, n\}$: $p(x) = \begin{cases} \frac{1}{n}, & x \in E \\ 0, & x \notin E \end{cases}$
  - **Benford's** on 1 digit: $p(x) = \begin{cases} \log_{10}(1 + x^{-1}), & 1 \leqslant x \leqslant 9 \\ 0, & \text{otherwise} \end{cases}$
  - **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
- Another way to repackage the information in the probability distribution is in the distribution function $F : \mathbb{Z} \to [0, 1]$, defined by $F(x) = \sum_{t \leqslant x} p(t)$