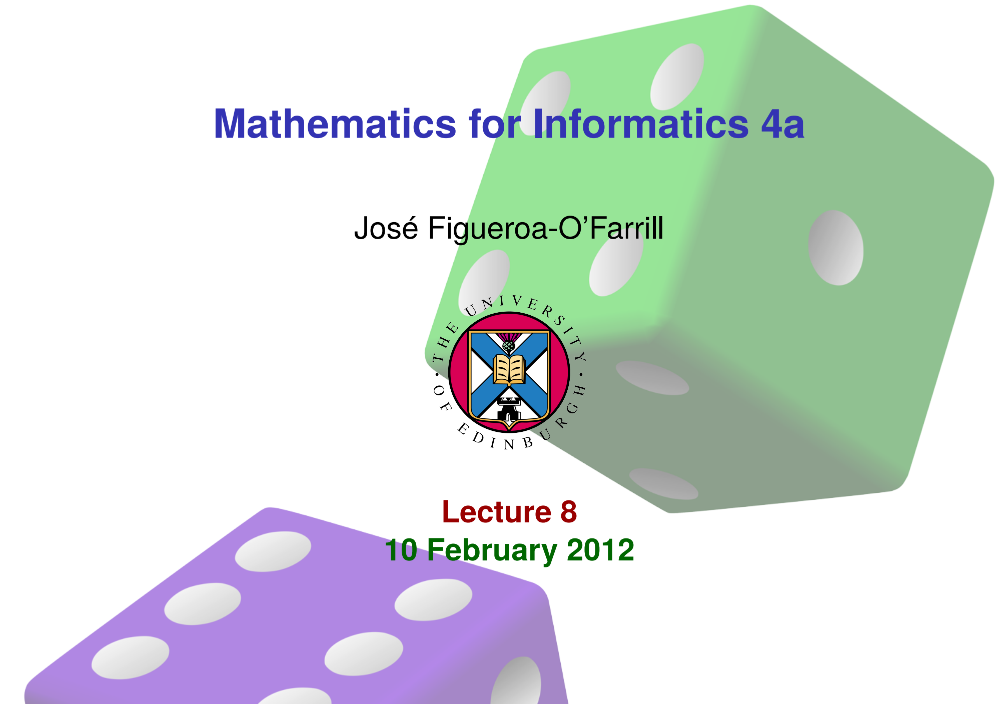


Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 8

10 February 2012

The story of the film so far...

Let X be a discrete random variable with mean $\mathbb{E}(X) = \mu$.

- For any function h, $Y = h(X)$ is a discrete random variable with mean $\mathbb{E}(Y) = \sum_x h(x) f_X(x)$.
- X has a **moment generating function** $M_X(t) = \mathbb{E}(e^{tX})$ from where we can compute the mean μ and standard deviation σ by
  - $\mu = \mathbb{E}(X) = M_X'(0)$
  - $\sigma^2 = \mathbb{E}(X^2) - \mu^2 = M_X''(0) - M_X'(0)^2$
- For binomial (n, p): $\mu = np$ and $\sigma^2 = np(1-p)$
- For Poisson λ: $\mu = \sigma^2 = \lambda$
- The Poisson distribution with mean λ approximates the binomial distribution with parameters n and p in the limit $n \to \infty$, $p \to 0$, but $np \to \lambda$
- "Rare" events occurring at a constant rate are distributed according to a Poisson distribution

Two random variables

- It may happen that one is interested in two (or more) different numerical outcomes of the same experiment.
- This leads to the simultaneous study of two (or more) random variables.
- Suppose that X and Y are discrete random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- The values of X and Y are distributed according to f_X and f_Y, respectively.
- But whereas $f_X(x)$ is the probability of $X = x$ and $f_Y(y)$ that of $Y = y$, they generally do *not* tell us the probability of $X = x$ *and* $Y = y$.
- That is given by their *joint distribution*.

Joint probability mass function

Let X and Y be two discrete random variables in the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the subsets $\{X = x\}$ and $\{Y = y\}$ are events and hence so is their intersection.

Definition

The **joint probability mass function** of the two discrete random variables X and Y is given by

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

Notation: often written just $f(x, y)$ if no ambiguity results.

Being a probability, $0 \leqslant f(x, y) \leqslant 1$.
But also $\sum_{x,y} f(x, y) = 1$, since every outcome $\omega \in \Omega$ belongs to precisely one of the sets $\{X = x\} \cap \{Y = y\}$. In other words, those sets define a partition of Ω, which is moreover countable.

Examples (Fair dice: scores, max and min)

We roll two fair dice.

1. Let X and Y denote their scores. The joint probability mass function is given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{36}, & 1 \leqslant x, y \leqslant 6 \\ 0, & \text{otherwise} \end{cases}$$

2. Let U and V denote the minimum and maximum of the two scores, respectively. The joint probability mass function is given by

$$f_{U,V}(u,v) = \begin{cases} \frac{1}{36}, & 1 \leqslant u = v \leqslant 6 \\ \frac{1}{18}, & 1 \leqslant u < v \leqslant 6 \\ 0, & \text{otherwise} \end{cases}$$

Marginals

The joint probability mass function $f(x, y)$ of two discrete random variables X and Y contains the information of the probability mass functions of the individual discrete random variables. These are called the **marginals**:

$$f_X(x) = \sum_y f(x,y) \qquad \text{and} \qquad f_Y(y) = \sum_x f(x,y) .$$

This holds because the sets $\{Y = y\}$, where y runs through all the possible values of Y, are a countable partition of Ω. Therefore,

$$\{X = x\} = \bigcup_y \{X = x\} \cap \{Y = y\} .$$

and computing $\mathbb{P}$ of both sides:

$$f_X(x) = \mathbb{P}(\{X = x\}) = \sum_y \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \sum_y f_{X,Y}(x,y) .$$

A similar story holds for $\{Y = y\}$.

Examples

1. Toss a fair coin. Let X be the number of heads and Y the number of tails:

$$f_X(0) = f_X(1) = f_Y(0) = f_Y(1) = \tfrac{1}{2}$$
$$f_{X,Y}(0,0) = f_{X,Y}(1,1) = 0 \qquad f_{X,Y}(1,0) = f_{X,Y}(0,1) = \tfrac{1}{2}$$

2. Toss two fair coins. Let X be the number of heads shown by the first coin and Y the number of heads shown by the second:

$$f_X(0) = f_X(1) = f_Y(0) = f_Y(1) = \tfrac{1}{2}$$
$$f_{X,Y}(0,0) = f_{X,Y}(1,1) = f_{X,Y}(1,0) = f_{X,Y}(0,1) = \tfrac{1}{4}$$

Moral: the marginals do not determine the joint distribution!

More than two random variables

There is no reason to stop at two discrete random variables: we can consider a finite number $X_1, \dots, X_n$ of discrete random variables on the same probability space. They have a joint probability mass function $f_{X_1,\dots,X_n} : \mathbb{R}^n \to [0, 1]$, defined by

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\})$$

and obeying

$$\sum_{x_1,\dots,x_n} f_{X_1,\dots,X_n}(x_1, \dots, x_n) = 1 .$$

It has a number of marginals by summing over the possible values of any k of the X_i.

Independence

In the ‣ second of the above examples, we saw that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. This is explained by the fact that for all x, y the events $\{X = x\}$ and $\{Y = y\}$ are independent:

$$\begin{aligned} f_{X,Y}(x,y) &= \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}) \qquad \text{(independent events)} \\ &= f_X(x)f_Y(y) \;. \end{aligned}$$

Definition

Two discrete random variables X and Y are said to be **independent** if for all x, y,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Example (Bernoulli trials with a random parameter)

Consider a Bernoulli trial with probability p of success. Let X and Y denote the number of successes and failures. Clearly they are not generally independent because $X + Y = 1$: so $f_{X,Y}(1,1) = 0$, yet $f_X(1)f_Y(1) = p(1-p)$.
Now suppose that we repeat the Bernoulli trial a random number N of times, where N has a Poisson probability mass function with mean λ. I claim that X and Y are now independent!
We first determine the probability mass functions of X and Y.
Conditioning on the value of N,

$$\begin{aligned} f_X(x) &= \sum_{n=1}^{\infty} \mathbb{P}(X = x | N = n)\mathbb{P}(N = n) = \sum_{n=x}^{\infty} \binom{n}{x} p^x q^{n-x} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \frac{(\lambda p)^x}{x!} e^{-\lambda} \sum_{m=0}^{\infty} \frac{q^m}{m!}\lambda^m = \frac{(\lambda p)^x}{x!} e^{-\lambda} e^{\lambda q} = \frac{(\lambda p)^x}{x!} e^{-\lambda p} \end{aligned}$$

So X has a Poisson probability mass function with mean λp.

Example (Bernoulli trials with a random parameter – continued)

One person's success is another person's failure, so Y also has a Poisson probability mass function but with mean λq.
Therefore

$$f_X(x)f_Y(y) = \frac{(\lambda p)^x}{x!} e^{-\lambda p} \frac{(\lambda q)^y}{y!} e^{-\lambda q} = e^{-\lambda} \frac{\lambda^{x+y}}{x!y!} p^x q^y$$

On the other hand, conditioning on N again,

$$\begin{aligned} f_{X,Y}(x,y) &= \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \mathbb{P}(\{X = x\} \cap \{Y = y\} | N = x + y)\mathbb{P}(N = x + y) \\ &= \binom{x+y}{x} p^x q^y e^{-\lambda} \frac{\lambda^{x+y}}{(x+y)!} \\ &= e^{-\lambda} \frac{\lambda^{x+y}}{x!y!} p^x q^y = f_X(x)f_Y(y) \end{aligned}$$

Independent multiple random variables

Again there is no reason to stop at two discrete random variables and we can consider a finite number $X_1, \dots, X_n$ of discrete random variables.
They are said to be **independent** when all the events $\{X_i = x_i\}$ are independent.
This is the same as saying that for any $2 \leqslant k \leqslant n$ variables $X_{i_1}, \dots, X_{i_k}$ of the $X_1, \dots, X_n$,

$$f_{X_{i_1},\dots,X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = f_{X_{i_1}}(x_{i_1}) \dots f_{X_{i_k}}(x_{i_k})$$

for all $x_{i_1}, \dots, x_{i_k}$.

Making new random variables out of old

Let X and Y be two discrete random variables and let $h(x, y)$ be any function of two variables. Then let $Z = h(X, Y)$ be defined by $Z(\omega) = h(X(\omega), Y(\omega))$ for all outcomes ω.

Theorem

$Z = h(X, Y)$ is a discrete random variable with probability mass function

$$f_Z(z) = \sum_{\substack{x,y \\ h(x,y)=z}} f_{X,Y}(x, y)$$

and mean

$$\mathbb{E}(Z) = \sum_{x,y} h(x, y) f_{X,Y}(x, y)$$

The proof is *mutatis mutandis* the same as in the one-variable case.

Let's skip it!

Proof

The cardinality of the set $Z(\Omega)$ of all possible values of Z is at most that of $X(\Omega) \times Y(\Omega)$, consisting of pairs (x, y) where x is a possible value of X and y is a possible value of Y. Since the Cartesian product of two countable sets is countable, $Z(\Omega)$ is countable.

Now,

$$\{Z = z\} = \bigcup_{\substack{x,y \\ h(x,y)=z}} \{X = x\} \cap \{Y = y\}$$

is a countable disjoint union. Therefore,

$$f_Z(z) = \sum_{\substack{x,y \\ h(x,y)=z}} f_{X,Y}(x, y) .$$

Proof – continued

The expectation value is

$$\begin{aligned} f_Z(z) &= \sum_z z f_Z(z) \\ &= \sum_z z \sum_{\substack{x,y \\ h(x,y)=z}} f_{X,Y}(x, y) \\ &= \sum_{x,y} h(x, y) f_{X,Y}(x, y) \end{aligned}$$

□

Functions of more than two random variables

Again we can consider functions $h(X_1, \ldots, X_n)$ of more than two discrete random variables.

This is again a discrete random variable and its expectation is given by the usual formula

$$\mathbb{E}(h(X_1, \ldots, X_n)) = \sum_{x_1,\ldots,x_n} h(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

The proof is basically the same as the one for two variables and shall be left as an exercise.

Linearity of expectation I

Theorem

Let X and Y be two discrete random variables. Then

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Proof.

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_{x,y} (x+y) f(x,y) \\ &= \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y) \\ &= \sum_x x f_X(x) + \sum_y y f_Y(y) = \mathbb{E}(X) + \mathbb{E}(Y)\end{aligned}$$

□

Linearity of expectation II

Together with $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$,... this implies the linearity of the expectation value:

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$$

NB: This holds even if X and Y are not independent!

Trivial example

Consider rolling two fair dice. *What is the expected value of their sum?*
Let X_i, $i = 1, 2$, denote the score of the ith die.
We saw earlier that $\mathbb{E}(X_i) = \frac{7}{2}$, hence

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = \tfrac{7}{2} + \tfrac{7}{2} = 7\,.$$

Linearity of expectation III

Again we can extend this result to any finite number of discrete random variables $X_1, \dots, X_n$ defined on the same probability space.
If $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, then

$$\mathbb{E}(\alpha_1 X_1 + \cdots + \alpha_n X_n) = \alpha_1 \mathbb{E}(X_1) + \cdots + \alpha_n \mathbb{E}(X_n)$$

(We omit the routine proof.)

Important!

It is important to remember that this is valid for arbitrary discrete random variables **without** the assumption of independence.

Example (Randomised hats)

A number n of men check their hats at a dinner party. During the dinner the hats get mixed up so that when they leave, the probability of getting their own hat is $1/n$. *What is the expected number of men who get their own hat?* Let us try counting.

- If $n = 2$ then it's clear: either both men get their own hats ($X = 2$) or else neither does ($X = 0$). Since both situations are equally likely, the expected number is $\frac{1}{2}(2 + 0) = 1$.
- Now let $n = 3$. There are $3! = 6$ possible permutations of the hats: the identity permutation has $X = 3$, three transpositions have $X = 1$ and two cyclic permutations have $X = 0$. Now we get $\frac{1}{6}(3 + 3 \times 1 + 2 \times 0) = 1$... again!
- How about $n = 4$? Now there are $4! = 24$ possible permutations of the hats...

There has to be an easier way.

Example (Randomised hats – continued)

- Let X denote the number of men who get their own hats.
- We let X_i denote the *indicator variable* corresponding to the event that the ith man gets his own hat: $X_i = 1$ if he does, $X_i = 0$ if he doesn't.
- Then $X = X_1 + X_2 + \cdots + X_n$.
- (The X_i are *not* independent! Why?)
- Notice that $\mathbb{E}(X_i) = \frac{1}{n}$, so that

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \\ &= \tfrac{1}{n} + \tfrac{1}{n} + \cdots + \tfrac{1}{n} \\ &= 1\end{aligned}$$

On average one (lucky) man gets his own hat!

Example (The coupon collector problem)

A given brand of cereal contains a small plastic toy in every box. The toys come in c different colours, which are uniformly distributed, so that a given box has a $1/c$ chance of containing any one colour. You are trying to collect all c colours. *How many cereal boxes do you expect to have to buy?*

- X_i is the number of boxes necessary to collect the ith colour, having collected already $i-1$ colours
- $X = X_1 + \cdots + X_c$ is the number of boxes necessary to collect all c colours
- we want to compute $\mathbb{E}(X) = \mathbb{E}(X_1) + \ldots \mathbb{E}(X_c)$, by linearity
- having collected already $i-1$ colours, there are $c-i+1$ colours I have yet to collect
- the probability of getting a new colour is $\frac{c-i+1}{c}$
- the probability of getting a colour I already have is $\frac{i-1}{c}$

Example (The coupon collector problem – continued)

- $\mathbb{P}(X_i = k) = \left(\frac{i-1}{c}\right)^{k-1} \frac{c-i+1}{c}$ for $k = 1, 2, \ldots$

$$M_{X_i}(t) = \sum_{k=1}^{\infty} e^{kt} \left(\frac{i-1}{c}\right)^{k-1} \frac{c-i+1}{c} = \frac{(c-i+1)e^t}{c-(i-1)e^t}$$

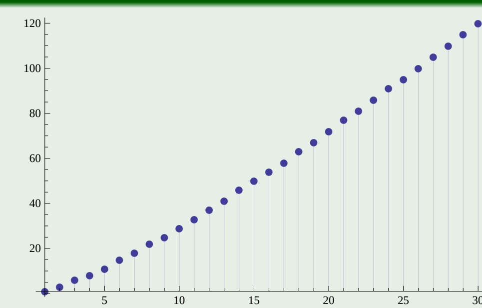
- $\mathbb{E}(X_i) = M'_{X_i}(0) = \frac{c}{c-i+1}$, whence finally

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{c} \frac{c}{c-i+1} \\ &= c\left(\frac{1}{c} + \frac{1}{c-1} + \cdots + \frac{1}{2} + 1\right) \\ &= cH_c\end{aligned}$$

where $H_c = 1 + \frac{1}{2} + \cdots + \frac{1}{c}$ is the cth **harmonic number**

Example (The coupon collector problem – continued)

c	cH_c	c	cH_c
1	1	2	3
3	6	4	8
5	11	6	15
7	18	8	22
9	25	10	29



- How many expected tosses of a fair coin until both heads and tails appear? **3**
- How many expected rolls of a fair die until we get all ⚀, . . . , ⚅? **15**
- et cetera

Summary

- Discrete random variables X, Y on the same probability space have a **joint probability mass function**:

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

- $f : \mathbb{R}^2 \to [0,1]$ and $\sum_{x,y} f(x,y) = 1$
- X, Y **independent**: $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all x, y
- $h(X,Y)$ is a discrete random variable and

$$\mathbb{E}(h(X,Y)) = \sum_{x,y} h(x,y) f_{X,Y}(x,y)$$

- Expectation is linear: $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$
- All the above generalises straightforwardly to n random variables $X_1, \dots, X_n$