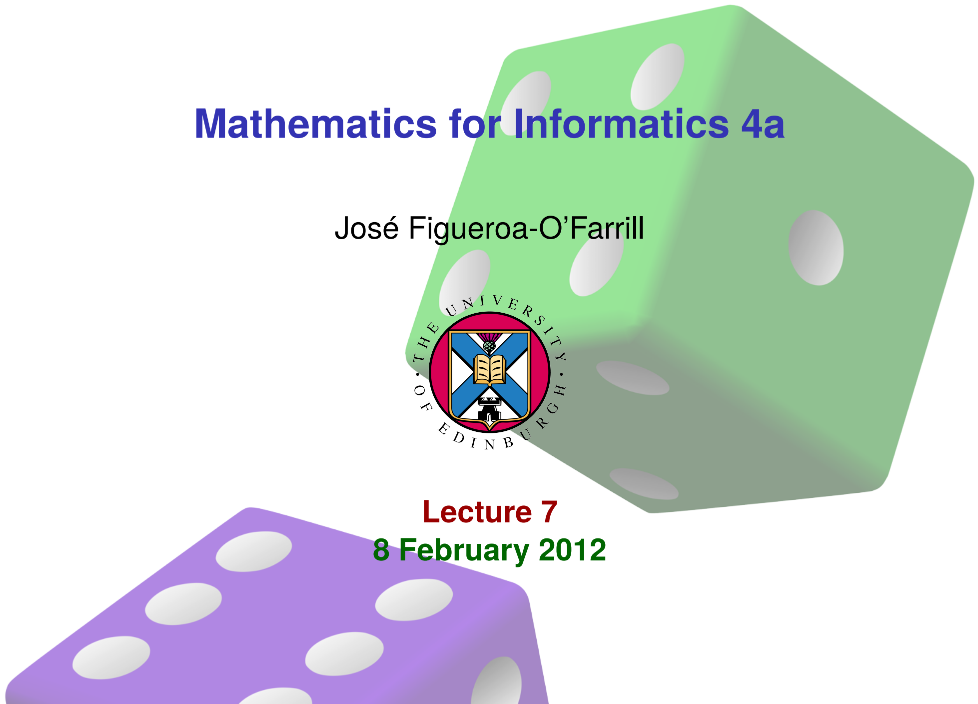


Mathematics for Informatics 4a

José Figueroa-O'Farrill

The University of Edinburgh

Lecture 7

8 February 2012

The story of the film so far...

- 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 $\mathbb{E}(X) = \sum_x x f_X(x)$.
- For f_X a uniform distribution, $\mathbb{E}(X)$ is simply the average.
- For f_X the Poisson distribution with parameter λ, $\mathbb{E}(X) = \lambda$.
- For f_X the binomial distribution with parameters n and p, $\mathbb{E}(X) = np$.

New random variables out of old

Suppose that X is a discrete random variable with probability mass function f_x and let $h : \mathbb{R} \to \mathbb{R}$ be a function; e.g., $h(x) = x^2$. Let $Y : \Omega \to \mathbb{Z}$ be defined by $Y(\omega) = h(X(\omega))$, written $Y = h(X)$.

Lemma

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

$$f_Y(y) = \sum_{\{x|h(x)=y\}} f_X(x) \ .$$

e.g., if $h(x) = x^2$, then $f_Y(4) = f_X(2) + f_X(-2)$.

Proof.

By definition $f_Y(y)$ is the probability of the event $\{\omega \in \Omega | Y(\omega) = y\} = \{\omega \in \Omega | h(X(\omega)) = y\}$, but this is the disjoint union of $\{\omega \in \Omega | X(\omega) = x\}$ for all x such that $h(x) = y$. □

What is the expectation value of $Y = h(X)$?
Luckily we don't have to determine f_Y in order to compute it.

Theorem

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

Proof.

By definition and the previous lemma,

$$\mathbb{E}(Y) = \sum_y y f_Y(y) = \sum_y y \sum_{\substack{x \\ h(x)=y}} f_X(x)$$

$$= \sum_y \sum_{\substack{x \\ h(x)=y}} y f_X(x) = \sum_x h(x) f_X(x)$$

□

Examples

Let a be a constant.

1. Let $Y = X + a$. Then

$$\mathbb{E}(Y) = \sum_x (x+a) f_X(x) = \sum_x x f_X(x) + \sum_x a f_X(x) = \mathbb{E}(X) + a$$

2. Let $Y = aX$. Then

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

3. Let $Y = a$. Then

$$\mathbb{E}(Y) = \sum_x a f_X(x) = a$$

Moment generating function

A special example of this construction is when $h(x) = e^{tx}$, where $t \in \mathbb{R}$ is a real number.

Definition

The **moment generating function** $M_X(t)$ is the expectation value

$$M_X(t) := \mathbb{E}(e^{tX}) = \sum_x e^{tx} f_X(x)$$

(provided the sum converges)

Lemma

1. $M_X(0) = 1$
2. $\mathbb{E}(X) = M'_X(0)$, *where $'$ denotes derivative with respect to* t.

Example

Let X be a discrete random variable whose probability mass function is given by a binomial distribution with parameters n and p. Then

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} e^{tx} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} \\ &= (e^t p + 1 - p)^n \ . \end{aligned}$$

Differentiating with respect to t,

$$M'_X(t) = n(e^t p + 1 - p)^{n-1} p e^t$$

whence setting $t = 0$, $M'_X(0) = np$, as we obtained before. (This way seems simpler, though.)

Example

Let X be a discrete random variable whose probability mass function is a Poisson distribution with parameter λ. Then

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} e^{tx} \\ &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{(\lambda e^t)^x}{x!} \\ &= e^{\lambda(e^t - 1)} \ . \end{aligned}$$

Differentiating with respect to t,

$$M'_X(t) = e^{\lambda(e^t-1)} \lambda e^t \ ,$$

whence setting $t = 0$, $M'_X(0) = \lambda$, as we had obtained before. (But again this way is simpler.)

Variance and standard deviation I

The expectation value $\mathbb{E}(X)$ (also called the **mean**) of a discrete random variable is a rather coarse measure of how X is distributed. For example, consider the following three situations:

1. I give you $\pounds 1000$
2. I toss a fair coin and if it is head I give you $\pounds 2000$
3. I choose a number from 1 to 1000 and if can guess it, I give you $\pounds 1$ million

Let X be the discrete random variable corresponding to your winnings. In all three cases, $\mathbb{E}(X) = \pounds 1000$, but you will agree that your chances of actually getting any money are quite different in all three cases.
One way in which these three cases differ is by the "spread" of the probability mass function. This is measured by the *variance*.

Variance and standard deviation II

Let X be a discrete random variable with mean μ. The variance is a weighted average of the (squared) distance from the mean. More precisely,

Definition

The **variance** $\mathrm{Var}(X)$ of X is defined by

$$\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \sum_x (x-\mu)^2 f_X(x)$$

(provided the sum converges.)

Its (positive) square root is called the **standard deviation** and is usually denoted σ, whence

$$\sigma(X) = \sqrt{\sum_x (x-\mu)^2 f_X(x)}$$

One virtue of $\sigma(X)$ is that it has the same units as X.

Variance and standard deviation III

Let us calculate the variances and standard deviations of the above three situations:

1. I give you $\pounds 1000$. There is only one outcome and it is the mean, hence the variance is 0.
2. I toss a fair coin and if it is head I give you $\pounds 2000$.

   $$\mathrm{Var}(X) = \tfrac{1}{2}(2000-1000)^2 + \tfrac{1}{2}(0-1000)^2 = 10^6$$

   whence $\sigma(X) = \pounds 1,000$.
3. I choose a number from 1 to 1000 and if can guess it in one attempt, I give you $\pounds 1$ million.

   $$\mathrm{Var}(X) = 10^{-3}(10^6-10^3)^2 + 999 \times 10^{-3}(0-10^3)^2 \simeq 10^9$$

   whence $\sigma(X) \simeq \pounds 31,607$.

Another expression for the variance

Theorem

If X is a discrete random variable with mean μ, then

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2$$

Proof.

$$\begin{aligned}\mathrm{Var}(X) &= \sum_x (x-\mu)^2 f_X(x) = \sum_x (x^2 - 2\mu x + \mu^2) f_X(x)\\ &= \sum_x x^2 f_X(x) - 2\mu \sum_x x f_X(x) + \mu^2 \sum_x f_X(x)\\ &= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mu^2\end{aligned}$$

□

Properties of the variance

Theorem

Let X be a discrete random variable and α a constant. Then

$$\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X) \qquad \text{and} \qquad \mathrm{Var}(X+\alpha) = \mathrm{Var}(X)$$

Proof.

Since $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$ and $\mathbb{E}(X+\alpha) = \mathbb{E}(X) + \alpha$,

$$\mathrm{Var}(\alpha X) = \mathbb{E}(\alpha^2 X^2) - \alpha^2\mu^2 = \alpha^2 \mathrm{Var}(X)$$

and

$$\mathrm{Var}(X+\alpha) = \mathbb{E}((X+\alpha-(\mu+\alpha))^2) = \mathbb{E}((X-\mu)^2) = \mathrm{Var}(X)$$

□

Variance from the moment generating function

Let X be a discrete random variable with moment generating function $M_X(t)$.

Theorem

$$\mathrm{Var}(X) = M_X''(0) - M_X'(0)^2$$

Proof.

Notice that the second derivative with respect to t of $M_X(t)$ is given by

$$\frac{d^2}{dt^2} \sum_x e^{tx} f_X(x) = \sum_x x^2 e^{tx} f_X(x) ,$$

whence $M_X''(0) = \mathbb{E}(X^2)$. The result follows from the expression $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2$ and the fact that $\mu = M_X'(0)$. □

Example

Let X be a discrete random variable whose probability mass function is a binomial distribution with parameters n and p. It has mean $\mu = np$ and moment generating function

$$M_X(t) = (e^t p + 1 - p)^n$$

Differentiating twice

$$M_X''(t) = n(n-1)(e^t p + 1 - p)^{n-2} p^2 e^{2t} + np(e^t p + 1 - p)^{n-1} e^t ,$$

Evaluating at 0, $M_X''(0) = n(n-1)p^2 + np$ and thus

$$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

Example

Let X be a discrete random variable with probability mass function given by a Poisson distribution with mean λ. Its moment generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Differentiating twice

$$M_X''(t) = e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}(\lambda e^t)^2$$

Evaluating at 0, $M_X''(0) = \lambda + \lambda^2$ and thus

$$\mathrm{Var}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$$

Approximations

- The Poisson distribution is a limiting case of the binomial distribution.
- Suppose that X is a discrete random variable whose probability mass function is a binomial distribution with parameters n and p.
- Then for $x = 0, 1, \dots, n$, $f_X(x)$ is given by

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x}$$

- We rewrite this as

$$\frac{pn(pn-p)\cdots(pn-px+p)}{x!}(1-\tfrac{np}{n})^{n-x}$$

- Now we let $np = \lambda$ and write $p = \frac{\lambda}{n}$ in the expression

$$\frac{pn(pn-p)\cdots(pn-px+p)}{x!}(1-\tfrac{np}{n})^{n-x}$$

to get

$$\frac{\lambda(\lambda-\frac{\lambda}{n})\cdots(\lambda-(x-1)\frac{\lambda}{n})}{x!}(1-\tfrac{\lambda}{n})^{n-x}$$

or equivalently

$$\frac{\lambda^x}{x!}(1-\frac{1}{n})\cdots(1-\frac{x-1}{n})(1-\frac{\lambda}{n})^{n-x}$$

which, in the limit $n \to \infty$, and using

$$\lim_{n\to\infty}(1-\frac{k}{n}) = 1 \qquad \text{and} \qquad \lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$$

becomes $\frac{\lambda^x}{x!}e^{-\lambda}$, which is the Poisson distribution.

Example (Overbooking)

A flight can carry 400 passengers. Any given passenger has a 1% probability of not showing up for the flight, so the airline sells 404 tickets. What is the probability that the flight is actually overbooked?
Overbooking results if less than 4 passengers fail to show up. With $p = 0.01$ and $n = 404$, the probability of exactly k of them failing to show up is

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda}$$

with $\lambda = np = 4.04$. The probability of overbooking is then

$$\sum_{k=0}^{3} \frac{(4.04)^k}{k!} e^{-4.04} \simeq 0.426.$$

(Using the binomial distribution the result would be $\simeq 0.425$.)

Example (Overbooking – continued)

Or in fact, exactly

0.42468363119253652820001354911679367302652425904046104945249507296865091483730020
6709158040615150407329585535240015120608219272553117981017641384828705922878440370
3215242075469960272848353133088296979751431682273196298166019175606448507563418817
4270940699381361337727727105734376654447807567617834069064865861292347589482283229
7859172633112693660439822342275313531378295457268742238146456308290233599014111615
4800343000745423704028505639402558828708863649538750495144766157478898029552419219
0912631747975464428965596189555212958443747278318077285983898463890809951167078673
8177347568229057659219954622594116676934630413343951161190275195407185240714940186
3114982185192191199682536778561407929022147875702048454991880843362752740325308776
99564281867565230149278156847391348512352077759684933445368145906389259 9

(808 decimal places)

Poisson distribution and the law of rare events I

There is a more "physical" derivation of the Poisson distribution, which has the virtue of illustrating where it is that we might expect it to arise.

Consider a random process, such as radioactive decay, buses arriving at the bus stop, cars passing through a given intersection, calls arriving at an exchange, requests arriving at a server,...

All these processes have in common that whatever it is that we are interested in measuring: decays, buses, cars, calls, requests,... can happen at any time.

We are interested in the question:

how many events take place in a given time interval?

Poisson distribution and the law of rare events II

Let us model a randomly occurring event: requests arriving at a server, say. We wish to know how many requests will arrive in a given time interval $[0, t]$.

We will assume that requests arrive at a constant rate λ; that is, the probability of a request arriving in a small interval of time δt is proportional to δt: $p = \lambda \delta t$.

To find out how many requests arrive in the interval $[0, t]$, we subdivide $[0, t]$ into n subintervals of size $\delta t = t/n$. We assume that δt is so small that the probability of two or more requests arriving during the same subinterval is negligible.

Therefore the number X of requests arriving in $[0, t]$ has a binomial distribution with parameters n and $p = \lambda t/n$:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^k \approx e^{-\lambda t} \frac{(t\lambda)^k}{k!}$$

Example

Requests arrive at a server at a rate of 3 per second. Compute the probabilities of the following events:

1. exactly one request arrives in a one-second period
2. exactly ten arrive in a two-second period

We model the number of requests as a discrete random variable X with a Poisson distribution with rate $\lambda = 3$:

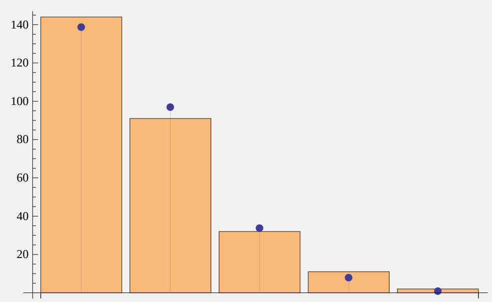
$$\mathbb{P}(X = k \text{ in } [0, t]) = e^{-3t} \frac{(3t)^k}{k!}$$

1. $\mathbb{P}(X = 1 \text{ in } [0, 1]) = 3e^{-3} \simeq 0.15$
2. $\mathbb{P}(X = 10 \text{ in } [0, 2]) = \frac{6^{10}}{10!} e^{-6} \simeq 0.04$

Poisson processes do not only model temporal distributions, but also spatial and spatio-temporal distributions!

Prussian cavalry fatalities of "death by horse"

In the 20 years from 1875 until 1894, the Prussian army kept detailed yearly records of horse-kick-induced fatalities among 14 cavalry regiments. In total there were 196 recorded fatalities distributed among $20 \times 14 = 280$ regiment-years. Ladislaus Bortkiewicz analysed this data using a Poisson distribution: The number of regiment-years with precisely k fatalities should be approximately $N(k) = 280 e^{-\lambda} \frac{\lambda^k}{k!}$, where $\lambda = \frac{196}{280} = \frac{7}{10}$.



Summary

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

- If h be any function, then $Y = h(X)$ is again a discrete random variable with
  - probability mass function $f_Y(y) = \sum_{x \mid h(x)=y} f_X(x)$, and
  - mean $\mathbb{E}(Y) = \sum_x h(x) f_X(x)$
- **moment generating function** $M_X(t) = \mathbb{E}(e^{tX})$ and $\mathbb{E}(X) = M_X'(0)$.
- **variance** $\text{Var}(X) = \mathbb{E}(X^2) - \mu^2 = M_X''(0) - M_X'(0)^2$ and **standard deviation** $\sigma = \sqrt{\text{Var}(X)}$ measure the “spread”:
  - For binomial (n, p): $\mu = np$ and $\sigma^2 = np(1-p)$
  - For Poisson λ: $\mu = \sigma^2 = \lambda$
- In the limit $n \to \infty$ and $p \to 0$, but $np \to \lambda$,

$$\text{Binomial}(n, p) \longrightarrow \text{Poisson}(\lambda)$$

- Rare events occurring at a constant rate are distributed according to a Poisson distribution.