Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 11

29 February 2012

The real story of the film so far...

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- (Some) continuous random variables have **probability density functions** f such that

$$\mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

- $F(x) = \mathbb{P}(X \leqslant x)$ is the **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - **exponential**: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$ (has no memory!)
- The **mean** $\mu = \int_{-\infty}^{\infty} x f(x) dx$, and equals $\frac{a+b}{2}$, μ and $\frac{1}{\lambda}$ for the above p.d.f.s, respectively.

Functions of a random variable

- Let X be a continuous random variable with probability density function f
- Let $g : \mathbb{R} \to \mathbb{R}$ be a function; e.g., $g(x) = x^2$
- Let $Y = g(X)$ be defined by $Y(\omega) = g(X(\omega))$
- Then for many functions g, Y is again a continuous random variable
- It is possible to determine the probability density function of Y, by first computing the (cumulative) distribution function $\mathbb{P}(Y \leqslant y)$
- Although one can derive some general formulae for certain kinds of functions g, it is perhaps better to do a couple of examples

Example (A gamma distribution)

Let X be normally distributed with parameters $\mu = 0$ and σ^2.
What is the probability density function of $Y = X^2$*?*
We start by calculating the cumulative distribution function $F_Y(y) = \mathbb{P}(Y \leqslant y)$, which is only nonzero for $y > 0$.

$$\begin{aligned}
\mathbb{P}(Y \leqslant y) &= \mathbb{P}(X^2 \leqslant y) = \mathbb{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\
&= \mathbb{P}(X \leqslant \sqrt{y}) - \mathbb{P}(X \leqslant -\sqrt{y}) \\
&= \int_{-\infty}^{\sqrt{y}} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx
\end{aligned}$$

The probability density function $f_Y(y) = F_Y'(y)$, whence by the chain rule,

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y/2\sigma^2} \frac{1}{\sqrt{y}} \qquad \text{for } y > 0$$

This is a special case of the "gamma" distribution.

Example (The log-normal distribution)

Let X be normally distributed with parameters μ and σ^2. *What is the probability density function of* $Y = e^X$*?*
Let us calculate $\mathbb{P}(Y \leqslant y)$, which is only nonzero for $y > 0$.

$$\begin{aligned}\mathbb{P}(Y \leqslant y) &= \mathbb{P}(e^X \leqslant y)\\ &= \mathbb{P}(X \leqslant \log y)\\ &= \int_{-\infty}^{\log y} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx\end{aligned}$$

whence

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\log y-\mu)^2/2\sigma^2} \frac{1}{y} \qquad \text{for } y > 0$$

Expectation of a function of a random variable

- As before, X is a continuous random variable with probability density function f_X
- Then the expectation value $\mathbb{E}(Y)$ of $Y = g(X)$ is given by

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx \; ,$$

(assuming the integral exists)

- For example,

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

and

$$\mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

(provided the integrals exist)

Variance of a continuous random variable

Let X be a continuous random variables with mean $\mu = \mathbb{E}(X)$.
We define the **variance** of X by

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

The **standard deviation** is the (+ve) square-root of the variance.

Example (Variance of uniform distribution)

Let X be uniformly distributed in $[a, b]$, so $\mathbb{E}(X) = \frac{1}{2}(a+b)$. Then

$$\mathbb{E}(X^2) = \int_a^b \frac{x^2}{b-a}dx = \left.\frac{\frac{1}{3}x^3}{b-a}\right|_a^b = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{1}{3}(a^2+ab+b^2)$$

whence

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu^2 = \frac{1}{3}(a^2+ab+b^2) - \frac{1}{4}(a+b)^2 = \frac{1}{12}(a-b)^2$$

Example (Variance of exponential distribution)

Let X be exponentially distributed with parameter λ, so $\mathbb{E}(X) = \frac{1}{\lambda}$. Then

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx\\ &= \lambda \frac{d^2}{d\lambda^2} \int_0^\infty e^{-\lambda x} dx\\ &= \lambda \frac{d^2}{d\lambda^2} \frac{1}{\lambda} = \frac{2}{\lambda^2}\end{aligned}$$

whence

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Example (Variance of normal distribution)

Let X be normally distributed with parameters $\mu = \mathbb{E}(X)$ and σ.

$$\begin{aligned}
\text{Var}(X) = \mathbb{E}((X-\mu)^2) &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2\sigma^2} dy && (y = x-\mu) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2/2} du && (u = y/\sigma) \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \frac{d}{du} e^{-u^2/2} du \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\frac{d}{du}\left(u e^{-u^2/2}\right) - e^{-u^2/2} \right] du \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = \sigma^2
\end{aligned}$$

Thus σ is the standard deviation.

Moment generating functions

Let X be a continuous random variable with probability density function f. The **moment generating function** (m.g.f.) $M_X(t)$ is defined by

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

(for those values of t for which the integral converges)

Example (M.g.f. for uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\begin{aligned}
M_X(t) &= \int_a^b \frac{e^{tx}}{b-a} dx = \left. \frac{e^{tx}}{t(b-a)} \right|_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)} \\
&= 1 + \tfrac{1}{2}(a+b)t + \tfrac{1}{6}(a^2 + ab + b^2)t^2 + \cdots
\end{aligned}$$

whence $\mathbb{E}(X) = \frac{1}{2}(a+b)$ and $\mathbb{E}(X^2) = \frac{1}{3}(a^2 + ab + b^2)$, as computed earlier.

Example (M.g.f. for exponential distribution)

Let X be exponentially distributed with mean $\frac{1}{\lambda}$.

$$\begin{aligned}
M_X(t) &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\
&= \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx \\
&= \frac{\lambda}{\lambda - t} \\
&= \frac{1}{1 - \frac{t}{\lambda}} \\
&= 1 + \frac{1}{\lambda} t + \frac{1}{\lambda^2} t^2 + \cdots
\end{aligned}$$

whence $\mathbb{E}(X) = \frac{1}{\lambda}$ and $\mathbb{E}(X^2) = \frac{2}{\lambda^2}$ as computed earlier.

Notice that $M_X(t) = 1 + \mu t + \frac{1}{2}(\mu^2 + \sigma^2)t^2 + \cdots$

Example (M.g.f. for normal distribution)

Let X be normally distributed with mean μ and variance σ^2.

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{tx} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-y^2/2\sigma^2} dy && (y = x - \mu) \\
&= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2 - 2\sigma^2 t y)/2\sigma^2} dy \\
&= \frac{e^{t\mu + \frac{1}{2}\sigma^2 t^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y - \sigma^2 t)^2/2\sigma^2} dy \\
&= \frac{e^{t\mu + \frac{1}{2}\sigma^2 t^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2\sigma^2} du && (u = y - \sigma^2 t) \\
&= e^{t\mu + \frac{1}{2}\sigma^2 t^2} \\
&= 1 + \mu t + \tfrac{1}{2}(\sigma^2 + \mu^2)t^2 + \cdots
\end{aligned}$$

Some properties of mean and variance

Theorem

Let X be a continuous random variable. Then provided that $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ exist, we have for all $a, b \in \mathbb{R}$,

1. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
2. $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$

Proof.

1. follows by linearity of integration:

$$\begin{aligned}\mathbb{E}(aX+b) &= \int_{-\infty}^{\infty} (ax+b)f(x)dx = a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx \\ &= a\mathbb{E}(X) + b\end{aligned}$$

2. follows from $\mathrm{Var}(Y) = \mathbb{E}((Y - \mu_Y)^2)$ with $Y = aX + b$

□

Standardising the normal distribution

Theorem

Let X be normally distributed with parameters μ and σ. Then $Y = \frac{1}{\sigma}(X - \mu)$ has as p.d.f. a standard normal distribution.

Remark

It follows from the previous theorem that Y has mean $\mathbb{E}(Y) = \frac{1}{\sigma}(\mathbb{E}(X) - \mu) = 0$ and variance $\mathrm{Var}(Y) = \frac{1}{\sigma^2}\mathrm{Var}(X) = 1$, just like the standard normal distribution. Moreover the moment generating function

$$M_Y(t) = \mathbb{E}(e^{t(X-\mu)/\sigma}) = e^{-\mu t/\sigma} M_X(\tfrac{t}{\sigma}) = e^{\frac{1}{2}t^2},$$

which is the moment generating function of the standard normal distribution. This makes the theorem plausible, but we wish to prove it.

Proof.

We will instead show directly that Y has the cumulative distribution function of a standard normal distribution:

$$\begin{aligned}\mathbb{P}(Y \leqslant y) &= \mathbb{P}(\tfrac{1}{\sigma}(X - \mu) \leqslant y) \\ &= \mathbb{P}(X \leqslant \sigma y + \mu) \\ &= \int_{-\infty}^{\sigma y + \mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \qquad (u = \tfrac{1}{\sigma}(x - \mu))\end{aligned}$$

whence $\mathbb{P}(Y \leqslant y) = \Phi(y)$. □

The usefulness of this result is that if X is normally distributed,

$$\mathbb{P}(|X - \mu| \leqslant c\sigma) = \mathbb{P}(|Y| \leqslant c)$$

where $c > 0$ is some constant and $Y = \frac{1}{\sigma}(X - \mu)$.

Example (The standard error)

Let X be normally distributed with mean μ and variance σ^2. *For which value of $c > 0$ is $\mathbb{P}(|X - \mu| \leqslant c\sigma) = 0.5$?*
This is the same c for which $\mathbb{P}(|Y| \leqslant c) = 0.5$, where $Y = \frac{1}{\sigma}(X - \mu)$ has a standard normal distribution:

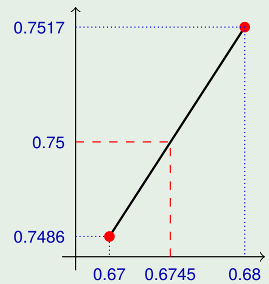
$$\begin{aligned}\mathbb{P}(|Y| \leqslant c) &= \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-y^2/2} dy \\ &= 2\frac{1}{\sqrt{2\pi}} \int_{0}^{c} e^{-y^2/2} dy \\ &= 2\frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{c} - \int_{-\infty}^{0}\right] e^{-y^2/2} dy \\ &= 2\Phi(c) - 1\end{aligned}$$

Therefore $\mathbb{P}(|Y| \leqslant c) = 0.5$ if and only if $\Phi(c) = 0.75$.

Example (The standard error – continued)

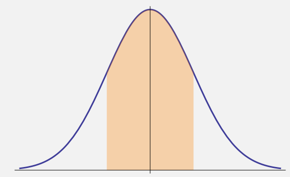
z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5363	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	[illegible]	.7823	.7852

From the tables, $\Phi(0.67) = 0.7486$ and $\Phi(0.68) = 0.7517$, and by linear interpolation $\Phi(0.6745) \simeq 0.75$. The number 0.6745σ is called the **standard error**: 50% of outcomes lie within a standard error of the mean.

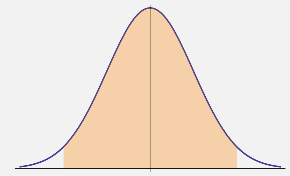


1σ, 2σ and 3σ

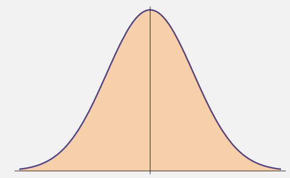
$$\mathbb{P}(|X-\mu| \leqslant \sigma) = 2\Phi(1) - 1 \simeq 0.6826$$



$$\mathbb{P}(|X-\mu| \leqslant 2\sigma) = 2\Phi(2) - 1 \simeq 0.9544$$



$$\mathbb{P}(|X-\mu| \leqslant 3\sigma) = 2\Phi(3) - 1 \simeq 0.9974$$



Maximum entropy and the normal distribution

The normal distribution is perhaps the single most important probability density function. This is due to two key results:

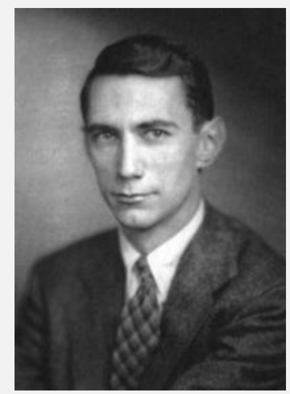
1. the central limit theorem (see later!), and
2. the maximum entropy principle.

- Suppose that all you know about a continuous random variable is its mean (μ) and variance (σ^2).
- In the absence of any more information, how are we to model this random variable?
- Is there a criterion to choose among all the probability density functions with those same mean and variance?
- There is indeed: Shannon's *maximum entropy principle*.

Shannon's maximum entropy principle

Shannon argued that the "least biased" or "most generic" p.d.f. is the one with maximum **entropy**

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx$$



It can be proved (using the variational calculus) that among all the p.d.f.s with mean μ, the one with maximum entropy is the **exponential** distribution, whereas among those which in addition have variance σ^2, it is the **normal** distribution. Their entropies are given by

$$H_{\text{exp}} = 1 + \log\mu \qquad H_{\text{normal}} = \frac{1}{2}(1 + \log(2\pi)) + \log\sigma$$

Example (The Four Sigma Society)

IQ tests are designed so that the mean is 100 and the standard deviation is 15. One of the many short-lived high-IQ societies was the Four Sigma Society, active for a few years in the late 1970s and early 1980s. As the name suggests, the entrance requirement was an IQ of at least 160. *What percentage of the population could apply for membership?*

$$\begin{aligned}\mathbb{P}(\mathrm{IQ} \geqslant 160) &= \mathbb{P}\left(\frac{\mathrm{IQ}-100}{15} \geqslant 4\right) \\ &= \int_4^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= 1 - \Phi(4) \\ &= \tfrac{1}{2} - \tfrac{1}{2}\,\mathrm{erf}(2\sqrt{2}) \simeq \tfrac{1}{31574}\end{aligned}$$

So about 1 in every 30,000 people. (cf. Mensa's 1 in 50.)

Example (Alice and Bob's first child)

Alice gives birth to her first child. Bob's joy knows no bounds, until he looks at his iCal and realises that he was away on a long trip from 283 days before the birth until 260 days before the birth. Assuming that gestation periods are normally distributed with a mean of 270 days and a standard deviation of 10 days, *what is the probability that Bob was away during conception?* Let X denote the length (in days) of gestation. Then we are after the probability that $260 \leqslant X \leqslant 283$. Let us standardise X to $Y = \frac{1}{10}(X - 270)$ and compute the probability that $-1 \leqslant Y \leqslant 1.3$:

$$\begin{aligned}\mathbb{P}(-1 \leqslant Y \leqslant 1.3) &= \mathbb{P}(Y \leqslant 1.3) - \mathbb{P}(Y \leqslant -1) \\ &= \Phi(1.3) - \Phi(-1) \\ &\simeq 0.9032 - 0.1587 \\ &\simeq 0.7445\end{aligned}$$

Summary

- If X is a continuous random variable with probability density function f, then for any function $g : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- The **variance** is $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- We calculated the variances of the uniform, exponential and normal distributions
- introduced the moment generating function and saw the usual examples: uniform, exponential and normal
- if X normally distributed with mean μ and variance σ^2, $Y = \frac{1}{\sigma}(X - \mu)$ has standard normal distribution
- introduced the **standard error** and gained some intuition for 1σ, 2σ and 3σ in a normal distribution
- motivated exponential and normal distributions from Shannon's **maximum entropy principle**