More Continuous Probability

Mark Muldoon

Departments of Mathematics and

Optometry & Neuroscience

UMIST

http://www.ma.umist.ac.uk/mrm/Teaching/2P1/

Overview

Today we'll continue our discussion of continuous probability

- **Continuous distributions:** probability densities
- **Expectations:** integrating to get μ and σ
- **Famous distributions:** the normal (again) and the exponential distribution
- **Why is the normal famous?** averaging and the Central Limit Theorem.

The life table

Number of men remaining alive at intervals of ten years

Age in years, x	Number surviving, l_x	Age in years, x	Number surviving, l_x
0	1000	60	758
10	959	70	524
20	952	80	211
30	938	90	22
40	920	100	0
50	876		

From *English Life Table No. 11, Males*

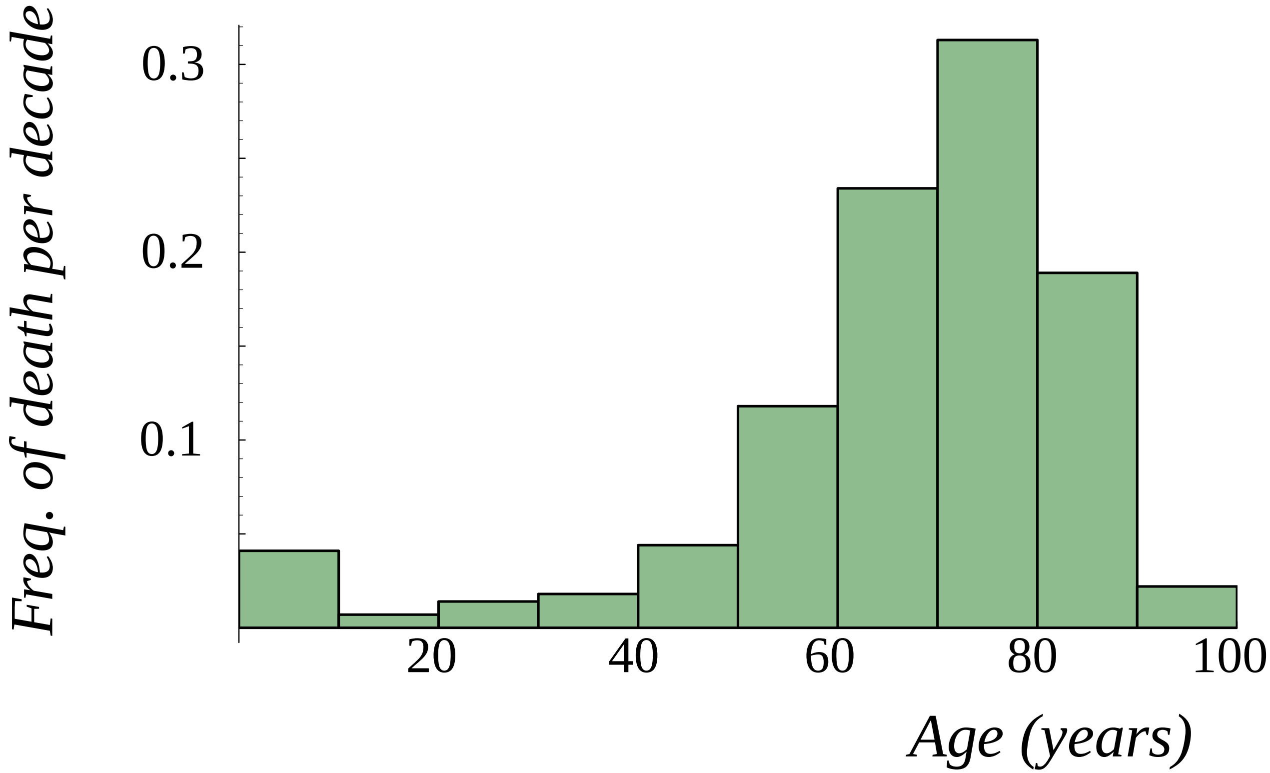
Prob. of death by decade

Probability of dying in each decade

Age at death x	Prob.	Age at death x	Prob.
$0 \leq x < 10$	0.041	$50 \leq x < 60$	0.118
$10 \leq x < 20$	0.007	$60 \leq x < 70$	0.234
$20 \leq x < 30$	0.014	$70 \leq x < 80$	0.313
$30 \leq x < 40$	0.018	$80 \leq x < 90$	0.189
$40 \leq x < 50$	0.044	$90 \leq x < 100$	0.022

Relative frequency histo

Relative frequency histogram for the distribution of age-at-death by decade



Review: expectation

Q: Based on the information in the previous two slides, estimate the mean lifetime of british males.

Review: expectation

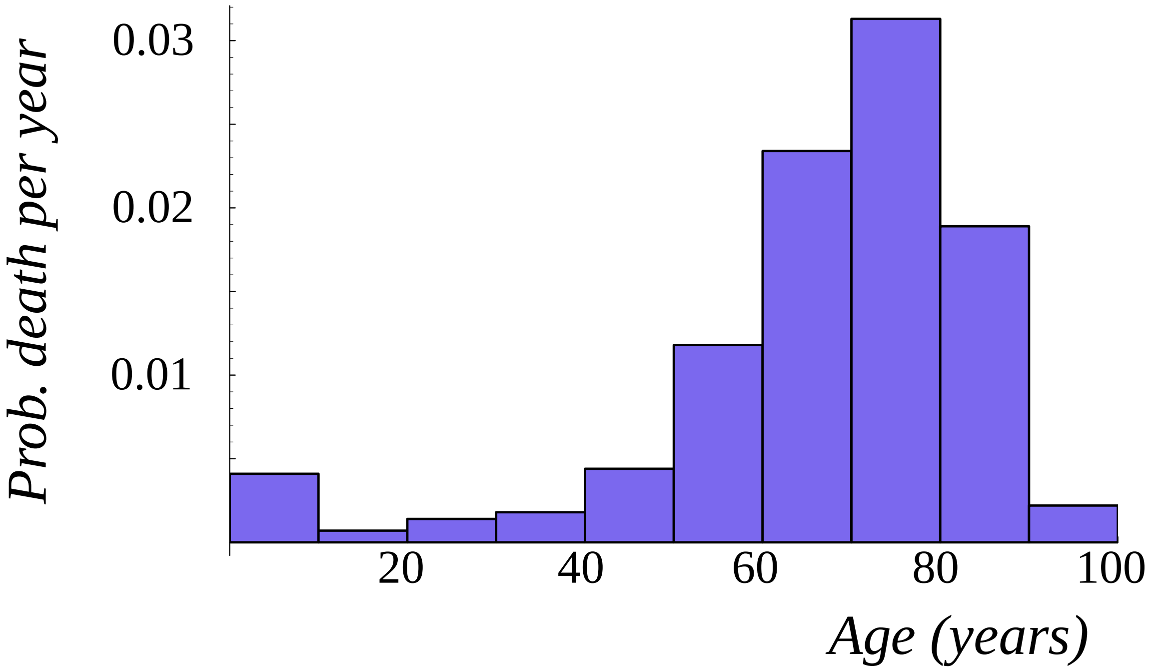
Q: Based on the information in the previous two slides, estimate the mean lifetime of british males.

A: There's not enough detail here to give a very precise answer, but if we assume that the average bloke lives 5 years into the decade of his death, then the expected lifespan at birth is:

$$\begin{aligned} P(\ 0 \leq x < 10\ ) \times 5 &+ \\ P(\ 10 \leq x < 20\ ) \times 15 &+ \\ &\vdots \\ P(\ 90 \leq x < 100\ ) \times 95 &= 66.6 \end{aligned}$$

A probability density function

Scale each probability by the width of the interval (that is, 10 years) to get probability of death per year. This is an example of a *probability density function*, or pdf for short.



Remarks about densities

The previous slide showed a sort of relative-frequency histogram, but arranged so that

- height of bar above jth decade is

$$(1/10) \times P(\text{ death in } j\text{th decade })$$

- width of bar above the jth decade is ten, so *area* of the bar is $P($ death in jth decade $)$;
- total area covered by bars is one;
- more detailed data could produce histogram with narrower bins, smoother density

Densities: probabilities are integrals

Think of the density as a (piecewise-constant) function $f(x)$. Then, for example,

$$P(\text{ death in 2nd decade }) = \int_{10}^{20} f(x)\,dx$$

We calculated the expected lifespan-from-birth using a sum, but we could also have thought of it as an integral

$$\begin{aligned} E(x) &= \int_0^{100} x f(x)\,dx \\ &\approx 66.6 \end{aligned}$$

Using densities in general (review)

- The probability that a random variable x with density $f(x)$ falls in a range $a \leq x \leq b$ is:

$$\int_a^b f(x)\,dx$$

- Expectations are computed by doing integrals rather than sums

$$\mu \;=\; E(x) \;=\; \int_{-\infty}^{\infty} x\,f(x)\,dx$$

and

$$\begin{aligned}
\sigma^2 &= E(x^2) - (E(x))^2 \\
&= \left[\int_{-\infty}^{\infty} x^2\,f(x)\,dx\right] - \mu^2 \\
&\approx 365.6 \text{ for the lifetime distrib.}
\end{aligned}$$

Cumulative densities

If $f(X)$ is a pdf then it has a useful companion, the Cumulative Density Function (a.k.a the cdf) $F(X)$ given by

$$F(X) = \int_{-\infty}^{X} f(x)\, dx.$$

In words

$$F(x) = P(\text{ random variable with density } f(X) \text{ is} \leq x).$$

Example: the exponential distribution

Consider events that occur randomly, but at a steady average rate: big floods in York, spontaneous dimerization of visual pigments or failure of electrical components. The intervals between such events often follow an *exponential* distribution with density function

$$f(T) = re^{-rT}$$

where r is the steady rate and we consider only positive inter-event intervals: $0 \leq T < \infty$.

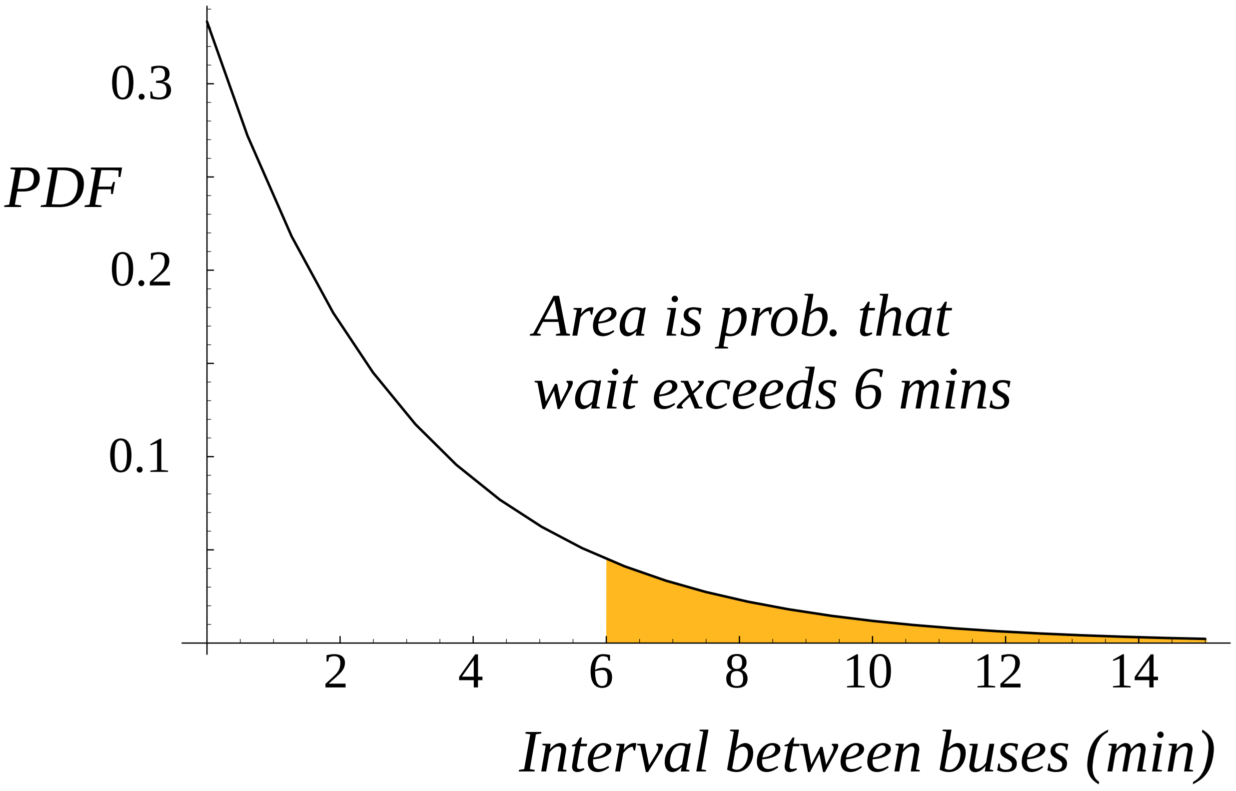
Buses on the Oxford road

During busy times buses arrive about every three minutes, so if we measure t in minutes the rate r of the exponential distribution is $r = (1/3)$.

Q: What is the probability of having to wait 6 or more minutes for a bus?

Integrate your wait away

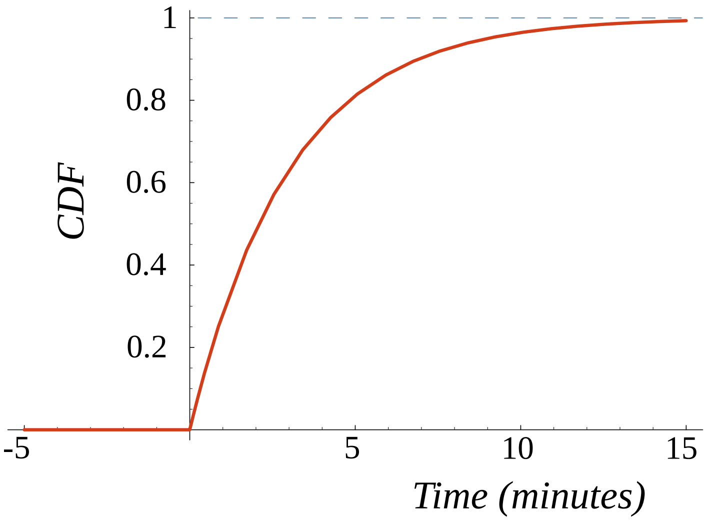
A: Integrate over the shaded area:



Details

$$
\begin{aligned}
P(\mathsf{T > 6}) &= \int_6^{\infty} f(t)\,dt \\
&= \int_6^{\infty} (1/3)e^{-t/3}\,dt \\
&= \left[-e^{-t/3}\right]_6^{\infty} \\
&= (-e^{-\infty}) - (-e^{2}) \\
&= 0 + e^{2} \;\approx\; 0.135
\end{aligned}
$$

Cdf for the exponential distribution

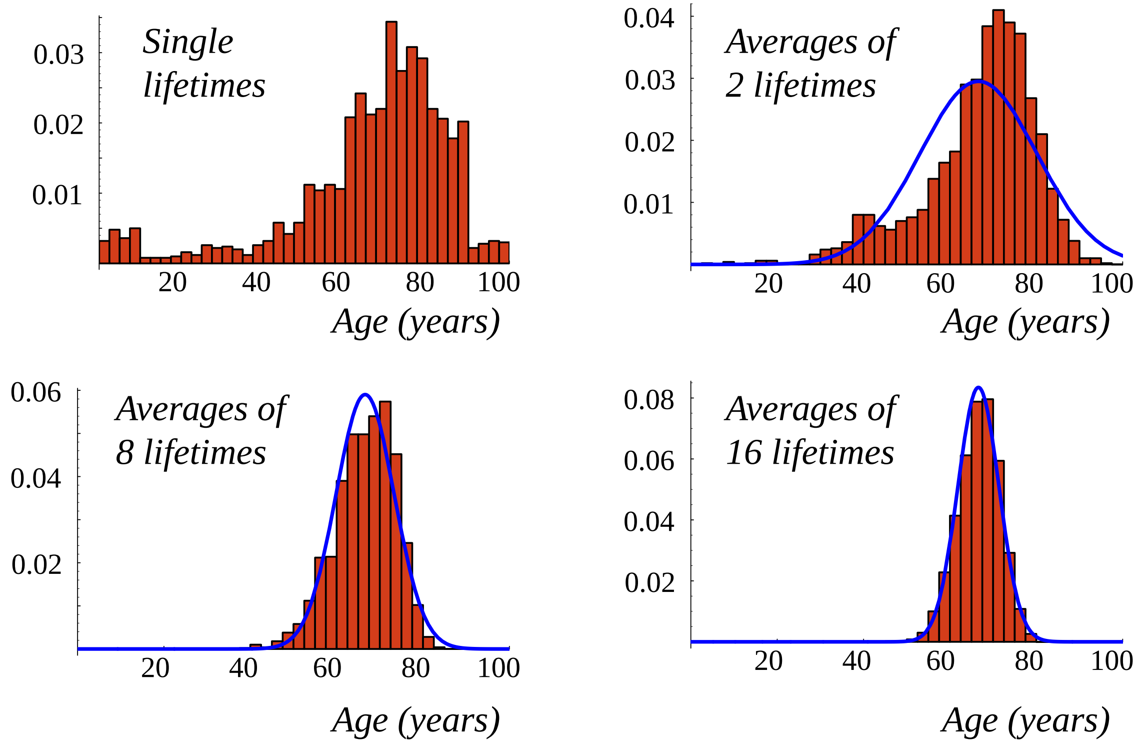


The cdf for the inter-bus interval distribution

$$
\begin{aligned}
F(T) &= \int_{-\infty}^{T} f(t)\,dt \\
&= \int_{0}^{T} \left(\frac{e^{-t/3}}{3} \right) dt \\
&= 1 - e^{-t/3}
\end{aligned}
$$

Simulation: mean ages-at-death

The figure below shows histograms of results from simulated experiments in which I averaged the lifetimes of variously-sized samples of english men:



The famous normal (again)

The blue curves plotted on top the histograms were examples of the *normal distribution*, a continuous probability distribution given by the formula

$$f(y) = \frac{\exp\left[-(y-\mu)^2/(2\sigma^2)\right]}{\sqrt{2\pi\sigma^2}}$$

The normals used to approximate the average-life histograms had the same mean μ as the age data, but a variance σ^2 that depends on N in a way we'll consider in today's final slide.

Averages and the Normal

Why does the normal arise so often? Because it is the "natural" distribution of averaged quantities. Think about an experiment in which you draw N random variables from (almost) any distribution $f(x)$ and average them: this defines a new random variable

$$y = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

The Central Limit Theorem

The *averages* Y will be approximately normally distributed and will have the same mean as $f(x)$ does, $\mu_y = \mu_x$, but will have a variance that decreases as N increases:

$$\sigma_y^2 = \frac{\sigma_x^2}{N}$$

Check: means of samples

The mean of a large sample:

a) is always greater than the median;

b) is calculated with the formula $m = (1/N) \sum_{j=1}^{N} x_j$;

c) is from an approximately Normal distribution;

d) increases as the sample size increases;

e) is always greater than the standard deviation.

Answer:

Check: means of samples

The mean of a large sample:

a) is always greater than the median;

b) is calculated with the formula $m = (1/N) \sum_{j=1}^{N} x_j$;

c) is from an approximately Normal distribution;

d) increases as the sample size increases;

e) is always greater than the standard deviation.

Answer: *Only items (b) and (c) are true.*