

Stats for Engineers Lecture 5

Summary From Last Time

Discrete Random Variables

Binomial Distribution $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Probability of number of k success when you do n Bernoulli trials

Mean and variance $\mu = np$ $\sigma^2 = np(1 - p)$

Poisson distribution $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$

Probability of k randomly occurring events, given average number is λ

Mean and variance $\mu = \lambda$ $\sigma^2 = \text{var}(X) = \lambda$

Is approximation to Binomial when n is large and p is small

Continuous Random Variables

Probability Density Function (PDF) $f(x)$ $P(a \leq X \leq b) = \int_a^b f(x') dx'$

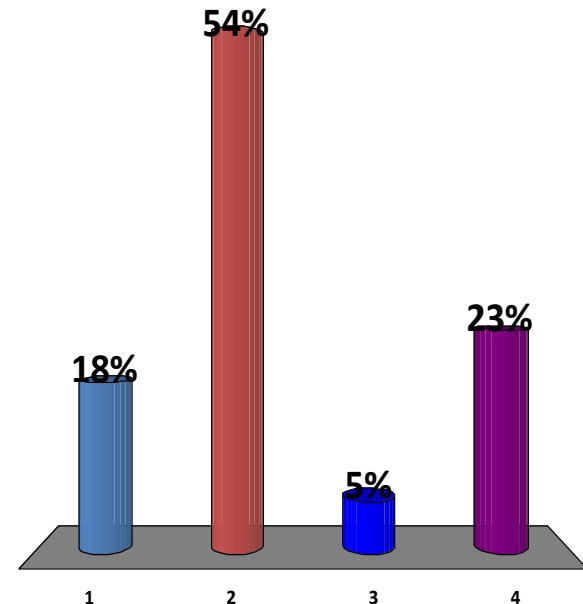
Uniform distribution $f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \theta_1 \leq x \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$



Poisson or not?

Which of the following is most likely to be well modelled by a Poisson distribution?

1. Number of trains arriving at Falmer every hour
2. Number of lottery winners each year that live in Brighton
3. Number of days between solar eclipses
4. Number of days until a component fails



Are they Poisson? *Answers:*

1. Number of trains arriving at Falmer every hour

NO, (supposed to) arrive *regularly* on a timetable not at random

2. Number of lottery winners each year that live in Brighton

Yes, is number of random events in fixed interval

3. Number of days between solar eclipses

NO, solar eclipses are not random events and this is a time *between* random events, not the number in some fixed interval

4. Number of days until a component fails

NO, random events, but this is time until a random event, not the *number* of random events

Time between random events / time till first random event ?

If a Poisson process has constant average rate ν , the mean after a time t is $\lambda = \nu t$.

What is the probability distribution for the time to the first event?

⇒ **Exponential distribution**

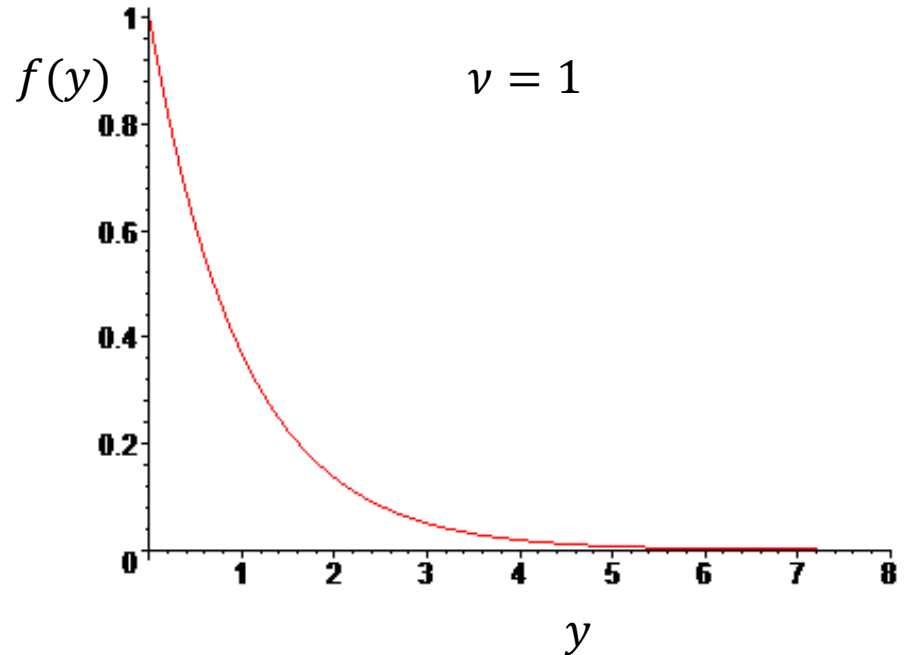
Poisson - *Discrete* distribution: P(number of events)

Exponential - *Continuous* distribution: P(time till first event)

Exponential distribution

The continuous random variable Y has the Exponential distribution, with **constant** rate parameter ν if:

$$f(y) = \begin{cases} \nu e^{-\nu y}, & y > 0 \\ 0, & y < 0 \end{cases}$$



Occurrence

- 1) Time until the failure of a part.
 - 2) Separation between randomly happening events
- Assuming the probability of the events is constant in time: $\nu = \text{const}$

Relation to Poisson distribution

If a Poisson process has constant average rate ν , the mean after a time t is $\lambda = \nu t$.

The probability of no-occurrences in time t is

$$P(k = 0) = \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} = e^{-\nu t}.$$

If $f(t)$ is the pdf for the first occurrence, then the probability of no occurrences is

$$P(\text{no occurrence by } t) = 1 - P(\text{first occurrence has happened by } t)$$

$$= 1 - \int_0^t f(t) dt$$

$$\Rightarrow 1 - \int_0^t f(t) dt = e^{-\nu t} \quad \Rightarrow \int_0^t f(t) dt = 1 - e^{-\nu t}$$

Solve by differentiating both sides respect to t assuming constant ν ,

$$\frac{d}{dt} \int_0^t f(t) dt = \frac{d}{dt} (1 - e^{-\nu t})$$

$$\Rightarrow f(t) = \nu e^{-\nu t}$$

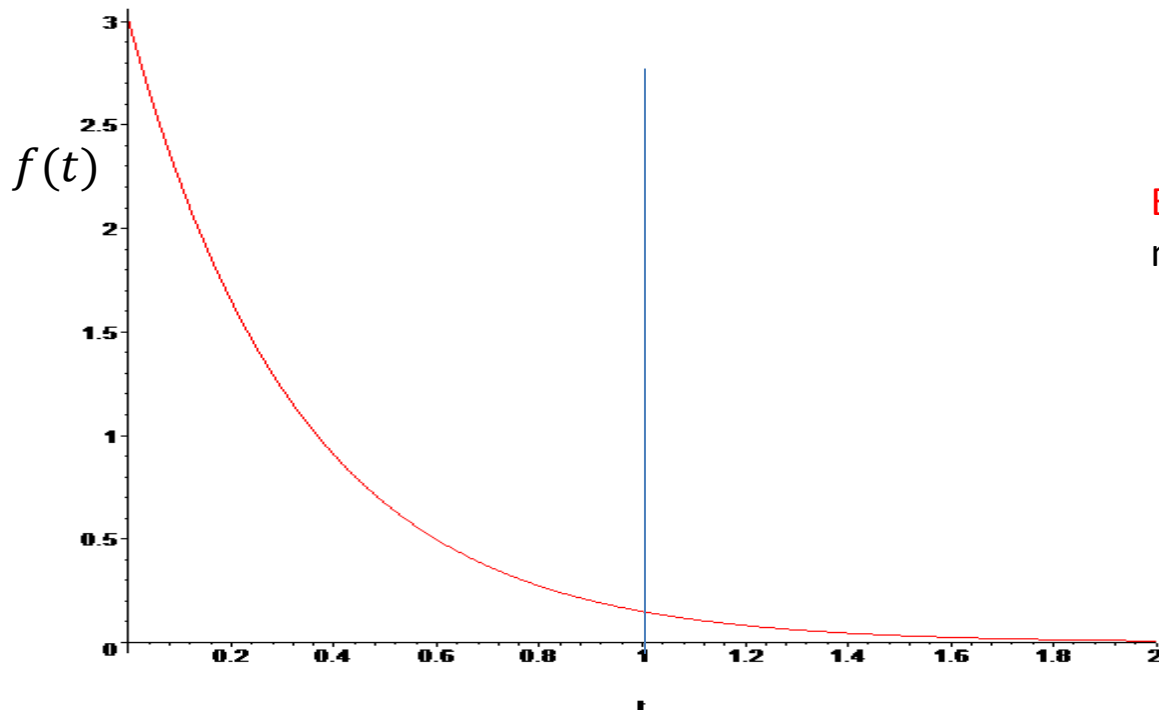
The time until the first occurrence (and between subsequent occurrences) has the Exponential distribution, parameter ν .

Example

On average lightning kills three people each year in the UK, $\lambda = 3$. So the rate is $\nu = 3/\text{year}$.

Assuming strikes occur randomly at any time during the year so ν is constant, time from today until the next fatality has pdf (using t in years)

$$f(t) = \nu e^{-\nu t} = 3 e^{-3t}$$



E.g. Probability the time till the next death is less than one year?

$$\begin{aligned} \int_0^1 f(t) dt &= \int_0^1 3 e^{-3t} dt \\ &= \left[\frac{3e^{-3t}}{-3} \right]_0^1 \\ &= -e^{-3} + 1 \approx 0.95 \end{aligned}$$

t



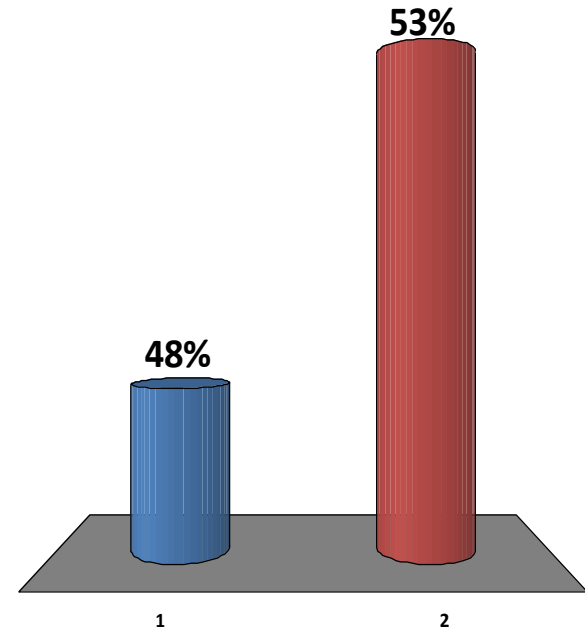
Question from Derek Bruff

Exponential distribution

A certain type of component can be purchased new or used. 50% of all new components last more than five years, but only 30% of used components last more than five years. Is it possible that the lifetimes of new components are exponentially distributed?

1. YES

✓ 2. NO





Exponential distribution

A certain type of component can be purchased new or used. 50% of all new components last more than five years, but only 30% of used components last more than five years. Is it possible that the lifetimes of new components are exponentially distributed?



Exponential distribution models time between independent randomly occurring events, where frequency of events is independent of time.

i.e. probability of failing in the first 5 years has to be same as the probability of failing in any other period of 5 years. *No memory property.*

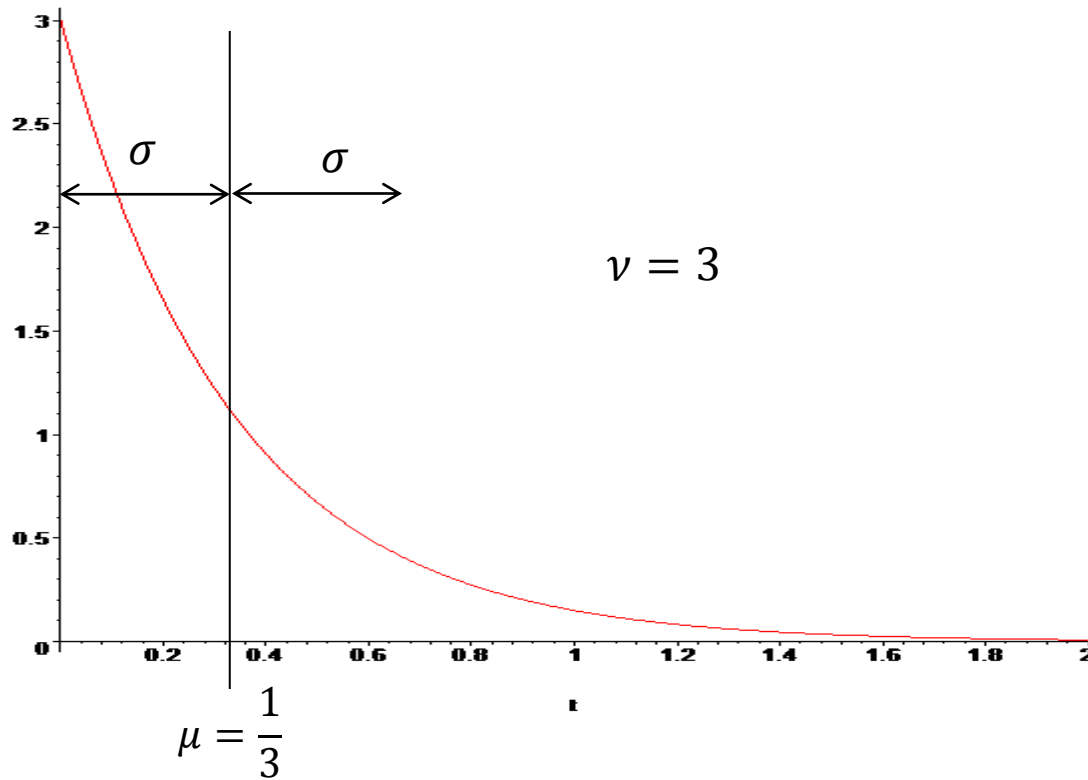
The observed lifetimes imply that instead the failure rate must increase with time

NOT exponential

Mean and variance of exponential distribution

$$\mu = \int_{-\infty}^{\infty} y f(y) dy = \int_0^{\infty} y \nu e^{-\nu y} dy = [-y e^{-\nu y}]_0^{\infty} + \int_0^{\infty} e^{-\nu y} dy = \left[-\frac{e^{-\nu y}}{\nu} \right]_0^{\infty} = \frac{1}{\nu}$$

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} y^2 f(y) dy - \mu^2 = \int_0^{\infty} y^2 \nu e^{-\nu y} dy - \frac{1}{\nu^2} \\ &= [-y^2 e^{-\nu y}]_0^{\infty} + 2 \int_0^{\infty} y e^{-\nu y} dy - \frac{1}{\nu^2} = 0 + 2 \frac{\mu}{\nu} - \frac{1}{\nu^2} = \frac{1}{\nu^2} \end{aligned}$$



Example: Reliability

The time till failure of an electronic component has an Exponential distribution and it is known that 10% of components have failed by 1000 hours.

- (a) What is the probability that a component is still working after 5000 hours?
- (b) Find the mean and standard deviation of the time till failure.

Answer

Let Y = time till failure in hours; $f(y) = \nu e^{-\nu y}$.

(a) First we need to find ν

$$P(Y \leq 1000) = \int_0^{1000} \nu e^{-\nu y} dy$$
$$= [-e^{-\nu y}]_0^{1000} = 1 - e^{-1000\nu}$$

$$P(Y \leq 1000) = 0.1 \Rightarrow 1 - e^{-1000\nu} = 0.1$$
$$\Rightarrow e^{-1000\nu} = 0.9$$
$$\Rightarrow -1000\nu = \ln 0.9 = -0.10536 \Rightarrow \nu \approx 1.05 \times 10^{-4}$$

If Y is the time till failure, the question asks for $P(Y > 5000)$:

$$\begin{aligned} P(Y > 5000) &= \int_{5000}^{\infty} \nu e^{-\nu y} dy \\ &= [-e^{-\nu y}]_{5000}^{\infty} = e^{-5000\nu} \approx 0.59 \end{aligned}$$

(b) Find the mean and standard deviation of the time till failure.

Answer: Mean = $1/\nu = 9491$ hours.

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\text{Variance}} \\ &= \sqrt{\frac{1}{\nu^2}} = 1/\nu = 9491 \text{ hours} \end{aligned}$$

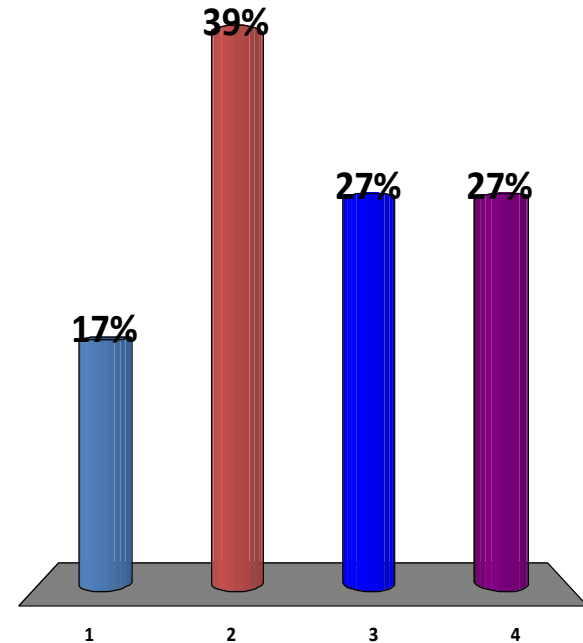


Is it exponential?

Which of the following random variables is best modelled by an exponential distribution?

*Question adapted from
Derek Bruff*

- 1. The distance between defects in an optical fibre
- 2. The number of days between someone winning the National Lottery
- 3. The number of fuses that blow in the UK today
- 4. The hours of sunshine in Brighton this week assuming an average of 7.2hrs/day





Is it exponential?

Which of the following random variables is best modelled by an exponential distribution?

1. The distance between defects in an optical fibre
 - **YES**: continuous distribution that is the separation between independent random events (the location of the defects)
2. The number of days between someone winning the National Lottery
 - **NO**: continuous (if you allow fractional days), but draws happen *regularly* on a schedule
3. The number of fuses that blow in the UK today
 - **NO**: this is a *discrete* distribution – the number of events is a Poisson distribution (exponential is the distribution of times *between* events)
4. The hours of sunshine in Brighton this week assuming an average of 7.2hrs/day
 - **NO**: This is a continuous variable, but not the time between independent random events

Normal distribution

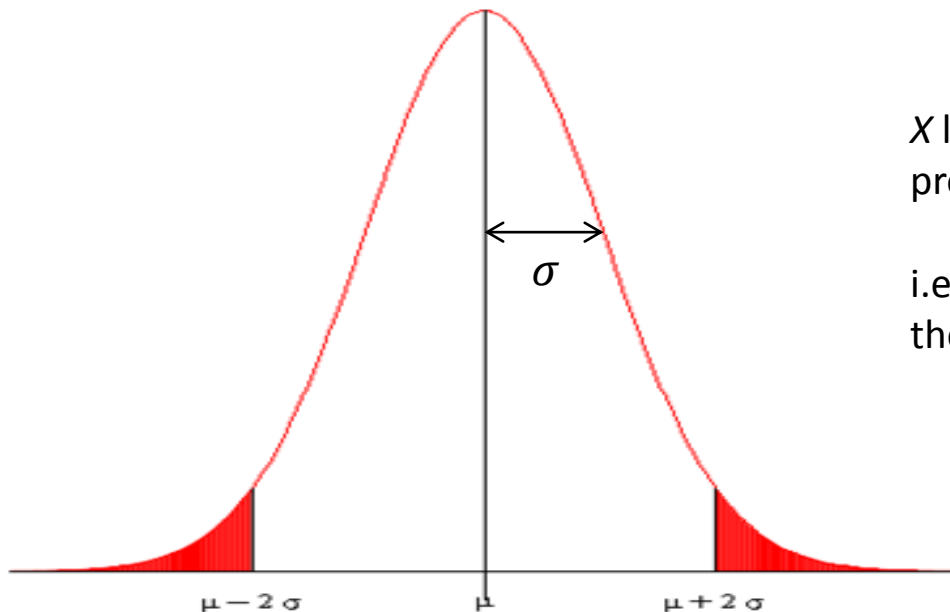
The continuous random variable X has the Normal distribution if the pdf is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)$$

μ : mean

σ : standard deviation

Note: The distribution is also sometimes called a Gaussian distribution



X lies between $\mu - 1.96$ and $\mu + 1.96$ with probability 0.95

i.e. X lies within 2 standard deviations of the mean approximately 95% of the time.

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

[see notes for proof]

If X has a Normal distribution with mean μ and variance σ^2 , write $X \sim N(\mu, \sigma^2)$

Occurrence of the Normal distribution

- 1) Quite a few variables, e.g. distributions of sizes, measurement errors, detector noise. (Bell-shaped histogram).
- 2) Sample means and totals - see later, Central Limit Theorem.
- 3) Approximation to several other distributions - see later.

