EMAT10100 Engineering Maths I
Lecture 9 of Introduction to Probability: Exponential and Normal Distributions

Rosalyn Moran & Nikolai Bode

Outline for today

On Monday we looked at discrete distributions associated to discrete random variables. We now take a closer look at two *continuous* distributions used to describe random processes in time.

- *Exponential distribution*
  - Introduced last week — now motivated in terms of random processes in time — a link to the Poisson distribution
- *Normal distribution*
  - A continuous RV that results almost universally from adding up lots of other RVs — the famous "bell-shaped" curve!

Random processes in time

Suppose that something *happens* 'randomly' *at rate* λ

- The emission of a radioactive particle
- Mutations on a DNA strand
- Rate of instrument failure etc

This means that in a very small time slice of length Δt:

- Probability of one *event* is approx $\lambda \Delta t$.
- Probability of more than one event is very small: scales like $(\Delta t)^2$
- Hence the probability of no event is approx $1 - \lambda \Delta t$.

What is the (continuous) RV that describes the time between two consecutive *events*?

Derivation of the exponential distribution for random processes in time

Suppose an *event* happens at time $t = 0$. What is the pdf $f(t)$ for the time of the next event?

- Cumulative distribution $F(t)$ is the probability that the next event happens before time t.
- Let $G(t) = 1 - F(t)$: it's the probability that the next event hasn't happened before time t.
- $G(t + \Delta t) = G(t)(1 - \lambda \Delta t)$
  (Think about this in words . . .)
- Therefore $G'(t) = -\lambda G(t)$.
- So $G(t) = \exp(-\lambda t)$ since $G(0) = 1$.
  Therefore $F(t) = 1 - \exp(-\lambda t)$.
- The pdf $f(t)$ is given by $f(t) = F'(t) = \lambda \exp(-\lambda t)$.

This is just the exponential distribution with parameter λ.

Link between Exponential and Poisson distributions

- Say λ is the mean *rate* per unit time at which some event occurs (e.g. an incoming call arrives at a switchboard or a customer enters a store)
- Assume you want to study the time between successive events (between incoming calls at a switchboard, or between successive customers entering a store).

Let X be the **continuous** RV describing the time between consecutive arrivals. Then X is exponentially distributed with parameter λ.

- pdf $f_X(t) = \lambda \exp(-\lambda t)$ for $t > 0$.
- cdf $F_x(t) = 1 - \exp(-\lambda t)$ for $t > 0$.
- mean "inter-arrival" time is $1/\lambda$, variance $1/\lambda^2$.

Link between Exponential and Poisson distributions

Now, let's consider the related problem of describing the number of arrivals (or events) in some finite time interval of interest T, e.g. number of customers entering a store in a day.

Let Y be the **discrete** RV describing the number of arrivals in a given time interval T.
Y is Poisson with parameter λT. (*)

- probability function

$$P_Y(k) = P(Y = k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}.$$

- mean λT, variance λT.

(*) Why? — divide T into N intervals of length $\Delta t = T/N$.

Normal distribution and its basic properties (1)

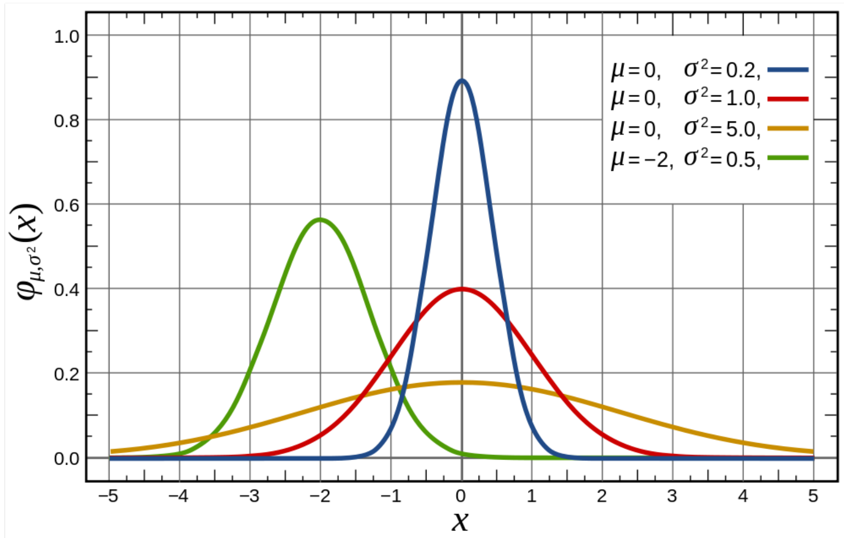
- It is the most important and most widely used continuous pdf.
- Empirical studies have shown that many physical variables have a Normal distribution.
- Examples include: meteorological data, measurements on living organisms, marks in exams, measurements of manufactured parts, instrumentation errors etc.

The pdf is a bell shaped curve described by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right].$$

where μ and σ are parameters representing the *mean* and the *standard deviation* (σ^2 is the variance). *Notation:* we write $X \sim N(\mu, \sigma^2)$

Normal distribution sketch



Normal distribution and its basic properties (2)

Given the pdf associated to a normal distribution we can compute the cumulative distribution:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2\sigma^2}(\tilde{x}-\mu)^2\right] \, \mathrm{d}\tilde{x}.$$

You can't 'do' this integral — but Matlab (or tables) will compute this for you — in terms of the so-called *error function*:

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \, \mathrm{d}t$$

Some useful rules of thumb . . .

For any Normally distributed variable:

- 68.3% of all values will lie between $\mu - \sigma$ and $\mu + \sigma$
- 95.45% of all values will lie within $\mu \pm 2\sigma$
- 99.73% of all values will lie within $\mu \pm 3\sigma$

Converting to standard Normal distribution

For any random variable, X with mean μ and variance σ^2, we know:

$$E(X - \mu) = 0.$$

The variance is not affected by this, but we also know:

$$Var(\frac{X - \mu}{\sigma}) = 1.$$

This follows from the results on functions and linear combinations of RVs we saw in lectures 3 and 4. The normal distribution has the property that the result of this is still normal, so if $X \sim N(\mu, \sigma^2)$:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

(also works the other way around: $X = \sigma Z + \mu$).

Exampercise

James 13.28, p1020 (5th ed., pp. 1046).

Given a random variable X that follows a normal distribution with mean $\mu = 4$ and variance $\sigma^2 = 4$, i.e. $X \sim N(4, 4)$, find:

(a) $P(X \leq 6.7)$

(b) the constant c such that $P(X > c) = 0.1$.

(...use tables for standard normal distribution, $N(0, 1)$)

Exampercise (table for standard normal cdf)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

University of BRISTOL
EngMaths I Probability 9
Spring Semester 2018

Modelling: origin of the Normal distribution

Key result is the *Central limit theorem*.

- Let X be *any* RV with mean μ and variance σ^2.
- Let $X_1, X_2, \ldots, X_n$ be a *large* number of independent trials of X.
- Let $Y = X_1 + X_2 + \ldots + X_n$.
- Then: $Y \sim N(n\mu, n\sigma^2)$.
- It follows that if $Z = Y/n$, then $Z \sim N(\mu, \sigma^2/n)$.

To emphasise: this is all independent of the original distribution of X.

University of BRISTOL
EngMaths I Probability 9
Spring Semester 2018

Numerical illustration of the Central Limit Theorem

Toss a (fair) coin 1000 times — what is the distribution of the number of heads?

To clarify — to understand this distribution numerically, we need to do the trial (toss a coin 1000 times) itself a large number of times — over to Matlab!!!

University of BRISTOL
EngMaths I Probability 9
Spring Semester 2018

Worked example

James 13.32 p1025 (5th ed., pp. 1051).

If 70% of airline passengers using a particular route are members of a frequent flyer club, find the probability that out of a sample of 50 chosen independently, more than 40 will be members of the club.

Solution

- Let X be a RV representing the number of passengers who are members.
- Note X is the sum of $n = 50$ Bernoulli trials, each with probability of success $p = 0.7$. So X follows a *Binomial distribution* with mean $np = 50(0.7) = 35$ and variance $np(1-p) = 50(0.7)(0.3) = 10.5$.
- The *Central limit theorem* tells us we can approximate a binomial with a normal distribution.
- So:

$$P(X > 40) \overset{CLT}{\approx} P(Y > 40) = P\left(\frac{Y-35}{\sqrt{10.5}} > \frac{40-35}{\sqrt{10.5}}\right)$$

$$= P(Z > 1.543) = 1 - F_Z(1.543) = 0.061$$

..where $Y \sim N(35, 10.5)$ and $Z \sim N(0, 1)$.

Weird and wonderful world of continuous RVs!!!

There are many other continuous distributions with real-world modelling use, e.g.:

- *Log normal* distribution [reliability analysis, friction and wear]

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right]$$

- *Gamma* distribution [size of insurance claims, interskpike intervals in neurons]

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \mathrm{e}^{-x/\theta}.$$

  for $k, \theta > 0$.

- *Chi-squared* distribution [inferential statistics]

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} \mathrm{e}^{-x/2}.$$

(All of these defined for $x \geq 0$ only.)

James homework

- Read through sections 13.4 and 13.5.
  NB the material on joint distributions, independence of random variables etc. is beyond the syllabus.
- Try a sample of exercises 13.5.7 p1027 (5th ed., pp. 1053).
- Study past exam papers.
- Use drop-in classes, online resources and good luck with the exam!