University of BRISTOL

WEEK 2

Continuous random variables

Felipe Campelo

Adapted from Montgomery and Runger's
Applied Statistics and Probability for Engineers
lecture notes (Ch 4)

About this slide deck

These slides summarise some continuous probability distributions, which are used to model different sorts of processes.

These slides should be considered in the context of **Chapter 4** of our main textbook (*DC Montgomery, GC Runger, Applied statistics and probability for engineers, 5th ed.*), more specifically sections 4.5 - 4.8, 4.10 and 4.11).

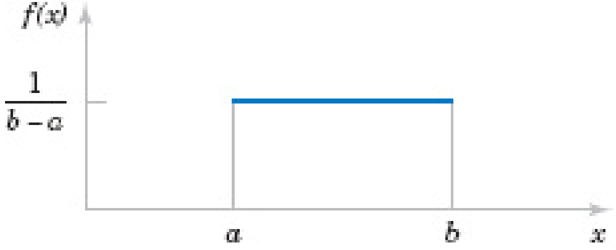
You're encouraged to read the chapter and try your hand at the worked examples.

Continuous uniform distribution

A continuous random variable X with probability density function

$$f(x) = \frac{1}{b-a}, a \leq x \leq b$$

is a **continuous uniform distribution** with parameters $\mathrm{a}, b \in \mathbb{R}$.

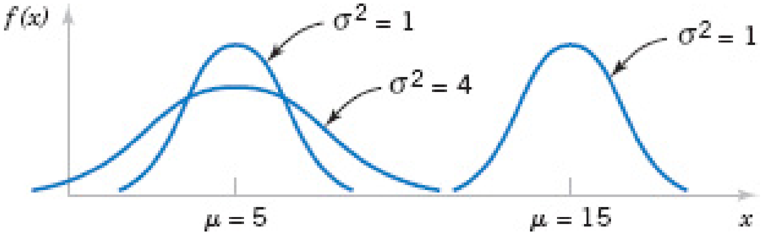


$$\mu = E[X] = \frac{a+b}{2}$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \frac{(b-a)^2}{12}$$

The continuous uniform distribution arises in some physical systems, as well as in statistical applications (such as in the distribution of *p-values* under the null hypothesis, which we'll visit in a few weeks).

Normal distribution



Without a doubt, the **normal** or **Gaussian** distribution is the most widely known probability distribution by far. As we progress in this course we'll understand a bit more why this distribution is so common (and so useful), but let's formally define it first.

A random variable X with probability density function

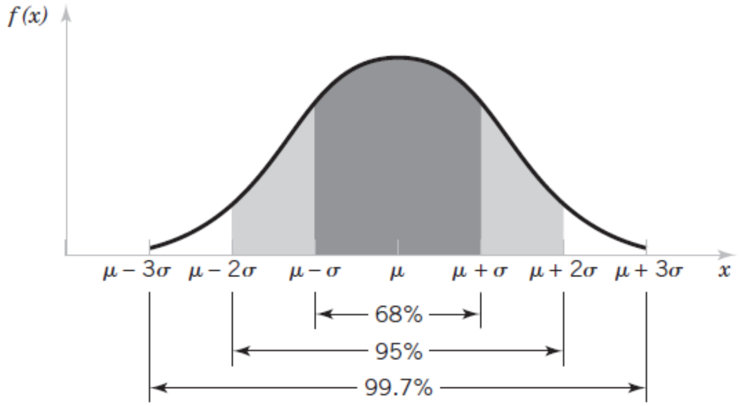
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty \leq x \leq \infty$$

is a **normal random variable** with parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$ and

$$\mu = E[X] = \mu, \qquad \sigma^2 = V(X) = \sigma^2$$

It is common to use the notation $\mathrm{N}(\mu, \sigma)$ or $\mathcal{N}(\mu, \sigma)$ for the normal distribution.

Normal distribution



For any normal random variable, the probability content of intervals defined by $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$ etc. is fixed.

By applying a *standardising transformation* to any normal random variable X with mean μ and variance σ^2 we get a **standard normal random variable,**

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

The cdf of a standard normal variable is usually denoted as $\Phi(z) = P(Z \leq z)$.

Normal distribution

Standardising normal variables is a convenient way to calculate probabilities by reducing any of an infinite amount of possible normal distributions to the same base distribution $\mathcal{N}(0, 1)$, which has easily remembered properties (e.g., that around 95% of the total probability are contained in the interval $z \in (-2,2)$.

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$F(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z) = \Phi(z)$$

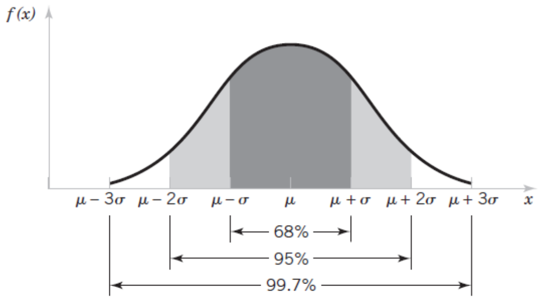
where Z is a standard normal distribution, and $\mathrm{z} = \frac{x-\mu}{\sigma}$ is the **z-value** obtained by standardising X.

Example

Suppose that you are building a predictive model using data from a high temperature sensor monitoring an induction furnace. The sensor readings are known to be subject to a Gaussian error due to environmental noise. In the usual operating conditions of the sensor, assume that this error results in a standard deviation of $\sigma^2 = 25℃^2$. If the actual temperature of an environment is $200℃$, what is the probability that the readings will exceed $210℃$?

Example

Suppose that you are building a predictive model using data from a high temperature sensor monitoring an induction furnace. The sensor readings are known to be subject to a Gaussian error due to environmental noise. In the usual operating conditions of the sensor, assume that this error results in a standard deviation of $\sigma^2 = 25°\mathrm{C}^2$. If the actual temperature of an environment is $200°\mathrm{C}$, what is the probability that the readings will exceed $210°\mathrm{C}$?

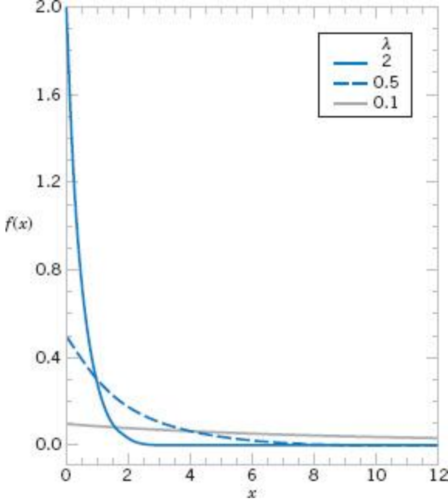


$$X = \mathcal{N}(\mu = 200, \sigma^2 = 25)$$

$$P(X \geq 220) = P\left(\frac{X - 200}{\sqrt{25}} \geq \frac{210 - 200}{\sqrt{25}}\right) = P(Z \geq 2)$$

Since we know that $P(-2 \leq Z \leq 2) = 0.95$ and that Z is symmetric about zero, it follows that $P(Z \leq -2) = P(Z \geq 2) = \frac{0.05}{2} = 0.025 = 2.5\%$.

Exponential distribution



Recall the definition of a *Poisson process*: *consider an interval T of real numbers partitioned into subintervals of small length δt, and assume that as δt tends to zero:*

1. *The probability of more than one event in any subinterval tends to zero.*
2. *The probability of one event in a subinterval tends to $\lambda \delta t / T$.*
3. *The event in each subinterval is independent of other subintervals.*

The random variable that counts the number of events in this process is a *Poisson random variable*. Analogously, the random variable X that equals the *distance between successive events* of a Poisson process with mean number of events λ per unit interval is an **Exponential random variable** with parameter λ, and pdf:

$$f(x) = \lambda e^{-\lambda x}, x \in \mathbb{R}_{\geq 0}$$

Exponential distribution

An Exponential RV has mean and variance given by:

$$\mu = E[X] = \frac{1}{\lambda}, \qquad \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

Analogously to Geometric RVs, Exponential random variables also exhibit the *lack of memory* property, i.e.:

$$P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2),$$

The Exponential distribution is the only continuous distribution with this property.

Example

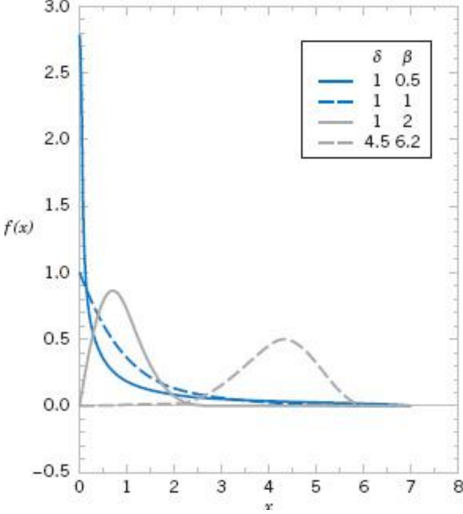
Let X denote the time between requests for a website and assume that X can be modelled as an Exponential distribution with $\mathrm{E}[X] = \frac{1}{\lambda} = 1.4ms$. The probability that we detect a request within $100\mu s$ of starting our monitoring is:

$$P(X < 100\mu s) = P(X < 0.1ms) = 1 - e^{-\left(\frac{0.1}{1.4}\right)} = 0.0689$$

Notice that all units must be consistent – if the expected value is given in milliseconds, then X must also be expressed in ms.
Now, suppose we start monitoring and wait $500ms$ minutes without detecting a request. What is the probability that a request is detected in the next $100\mu s$?
Because of the lack of memory property,

$$P(X < (500 + 0.1)ms | P > 500ms) = P(X < 0.1ms) = 1 - e^{-\left(\frac{0.1}{1.4}\right)} = 0.0689$$

Weibull distribution



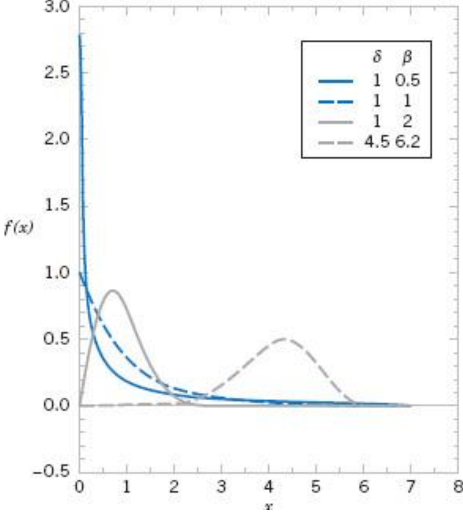
The Weibull distribution is often used to model the **time until failure** of many different physical systems. It is a distribution that appears often in fields such as survival analysis and systems reliability.

The parameters in this distribution provide flexibility to model systems in which the number of failures increases with time (due to wear and tear), decreases (systems that tend to fail early), or remains constant with time (failures caused by external factors).

A **Weibull random variable** X with scale parameter $\delta \in \mathbb{R}_{\geq 0}$ and shape parameter $\beta \in \mathbb{R}_{\geq 0}$ has a pdf:

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, \text{for } x \in \mathbb{R}_{>0}$$

Weibull distribution



As mentioned before, the Weibull distribution is quite flexible to model variables of the type time-to-failure (or, more broadly, time-to-extreme-event).

- If $\beta < 1$, the distribution models a system where the rate of failures **decreases** with time – sometimes referred to as systems with a high "infant mortality".
- If $\beta = 1$, the distribution reduces to an Exponential distribution with $\lambda = \beta$, modelling a system with constant failure rate over time.
- For $\beta > 1$, the distribution models systems in which the rate of failures **increases** with time due to, e.g., wear-and-tear or "aging".

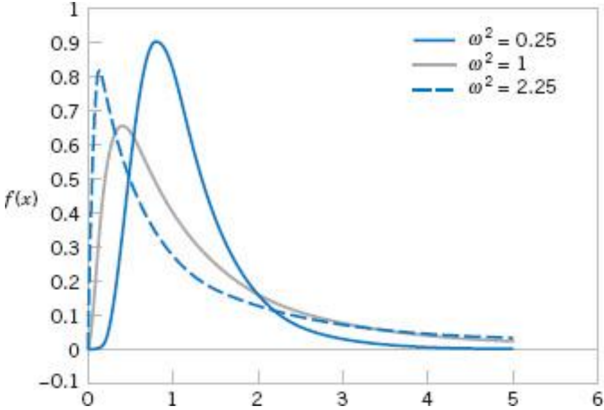
$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}$$

$$F(x) = 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}$$

$$\mu = E[X] = \delta\Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\sigma^2 = V(X) = \delta^2\Gamma\left(1 + \frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2$$

Lognormal distribution



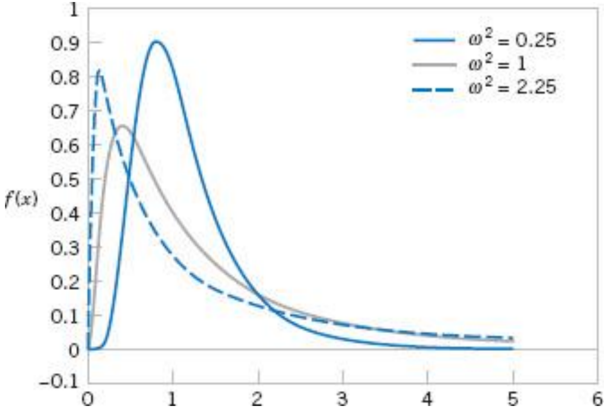
Variables in a system sometimes follow an exponential relationship as $x = \mathrm{e}^{w}$. If the exponent is a random variable W, then $X = e^{W}$ is a random variable with a distribution of interest.

An important special case occurs when W has a normal distribution. In that case, the distribution of X is called a **lognormal distribution.** The name follows from the transformation $\log(X) = W$, i.e., the natural logarithm of X is normally distributed.

Formally, let W be a normal distribution with mean θ and variance ω^2. Then the variable $X = \mathrm{e}^{\mathrm{W}}$ is a lognormal random variable with

$$f(x) = \frac{1}{xw\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\theta}{\omega}\right)^2}$$

Lognormal distribution



The parameters of a lognormal distribution are θ and ω^2, but we need to be careful to remember these as the mean and variance **of the normal random variable W.**

The mean and variance of X are the functions of these parameters, as shown to the right.

$$f(x) = \frac{1}{xw\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\theta}{\omega}\right)^2}$$

$$F(x) = \Phi\left(\frac{\log(x) - \theta}{\omega}\right)$$

$$\mu = E[X] = e^{\theta+\omega^2/2}$$

$$\sigma^2 = V[X] = (e^{\omega^2} - 1)e^{2\theta+\frac{\omega^2}{2}}$$

The time to failure of a system that degrades over time is often modelled as a lognormal RV.

As we saw in the previous slides, the Weibull distribution can also be used in this type of application – with appropriate parameters, it approaches a given lognormal RV.

However, since the Lognormal distribution is derived from a simpler exponential function of a normal random variable, it is easier to interpret and to calculate probabilities.