More About Continuous Random Variables
Class 5, 18.05

Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to give examples of what uniform, exponential and normal distributions are used to model.
2. Be able to give the range and pdf's of uniform, exponential and normal distributions.
3. Be able to find the pdf and cdf of a random variable defined in terms of a random variable with known pdf and cdf.
4. Be able to compute and interpret expectation, variance, and standard deviation for continuous random variables.
5. Be able to compute and interpret quantiles for discrete and continuous random variables.

2 Introduction

Here we introduce a few fundamental continuous distributions. These will play important roles in the statistics part of the class. For each distribution, we give the range, the pdf, the cdf, and a short description of situations that it models. These distributions all depend on parameters, which we specify.

As you look through each distribution do not try to memorize all the details; you can always look those up. Rather, focus on the shape of each distribution and what it models.

We call your attention to the normal distribution. It is easily the most important distribution defined here.

2.1 Parametrized distributions

When we studied discrete random variables we learned, for example, about the Bernoulli(p) distribution. The probability p used to define the distribution is called a parameter and Bernoulli(p) is called a parametrized distribution. For example, tosses of fair coin follow a Bernoulli distribution where the parameter $p = 0.5$. When we study statistics one of the key questions will be to estimate the parameters of a distribution. For example, if I have a coin that may or may not be fair then I know it follows a Bernoulli(p) distribution, but I don't know the value of the parameter p. I might run experiments and use the data to estimate the value of p.

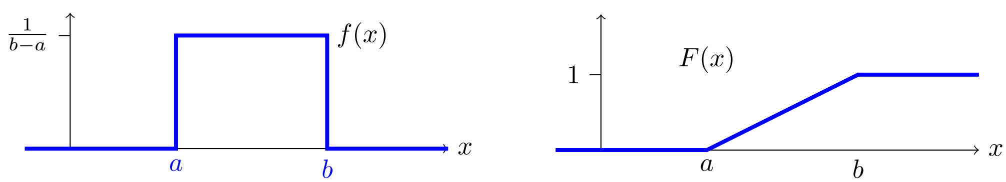
As another example, the binomial distribution Binomial(n, p) depends on two parameters n and p.

In the following sections we will look at specific parametrized continuous distributions. The applet `http://mathlets.org/mathlets/probability-distributions/` allows you to visualize the pdf and cdf of these distributions and to dynamically change the parameters.

3 Uniform distribution

1. Parameters: a, b.
2. Range: $[a, b]$.
3. Notation: uniform(a, b) or U(a, b).
4. Probability density function (pdf): $f(x) = \dfrac{1}{b-a}$ for $a \le x \le b$.
5. Cumulative distribution function (cdf): $F(x) = (x-a)/(b-a)$ for $a \le x \le b$.
6. Models: Situtations where all outcomes in the range have equal probability (more precisely all outcomes have the same probability density).

Graphs:



pdf and cdf for uniform(a,b) distribution.

Example 1. **1.** Suppose we have a tape measure with markings at each millimeter. If we measure (to the nearest marking) the length of items that are roughly a meter long, the rounding error will be uniformly distributed between -0.5 and 0.5 millimeters.

2. Many board games use spinning arrows (spinners) to introduce randomness. When spun, the arrow stops at an angle that is uniformly distributed between 0 and 2π radians.

3. In most pseudo-random number generators, the basic generator simulates a uniform distribution and all other distributions are constructed by transforming the basic generator.

4 Exponential distribution

1. Parameter: λ.
2. Range: $[0, \infty)$.
3. Notation: exponential(λ) or exp(λ).
4. Probability density function (pdf): $f(x) = \lambda \mathrm{e}^{-\lambda x}$ for $0 \le x$.
5. Cumulative distribution function (cdf): (This is an easy integral.)
$$F(x) = 1 - \mathrm{e}^{-\lambda x} \text{ for } x \ge 0$$
6. Right tail distribution: $P(X > x) = 1 - F(x) = \mathrm{e}^{-\lambda x}$. (Note: this is defined as $P(X > x)$, i.e. that X is to the right of x on the number line.)

7. Models: The waiting time for a continuous process to change state.

Example 2. If I step out to 77 Mass Ave after class and wait for the next taxi, my waiting time in minutes is exponentially distributed. We will see that in this case λ is given by 1/(average number of taxis that pass per minute).

Example 3. The exponential distribution models the waiting time until an unstable isotope undergoes nuclear decay. In this case, the value of λ is related to the half-life of the isotope.

Memorylessness: There are other distributions that also model waiting times, but the exponential distribution has the additional property that it is memoryless. Here's what this means in the context of Example 2: suppose that the probability that a taxi arrives within the first five minutes is p. If I wait five minutes and, in this case, no taxi arrives, then the probability that a taxi arrives within the next five minutes is still p. That is, my previous wait of 5 minutes has no impact on the length of my future wait!

By contrast, suppose I were to instead go to Kendall Square subway station and wait for the next inbound train. Since the trains are coordinated to follow a schedule (e.g., roughly 12 minutes between trains), if I wait five minutes without seeing a train then there is a far greater probability that a train will arrive in the next five minutes. In particular, waiting time for the subway is not memoryless, and a better model would be the uniform distribution on the range [0,12].

The memorylessness of the exponential distribution is analogous to the memorylessness of the (discrete) geometric distribution, where having flipped 5 tails in a row gives no information about the next 5 flips. Indeed, the exponential distribution is precisely the continuous counterpart of the geometric distribution, which models the waiting time for a discrete process to change state. More formally, memoryless means that the probability of waiting t more minutes is independent of the amount of time already waited. In symbols,

$$P(X > s + t \,|\, X > s) = P(X > t).$$

Proof of memorylessness: We know that

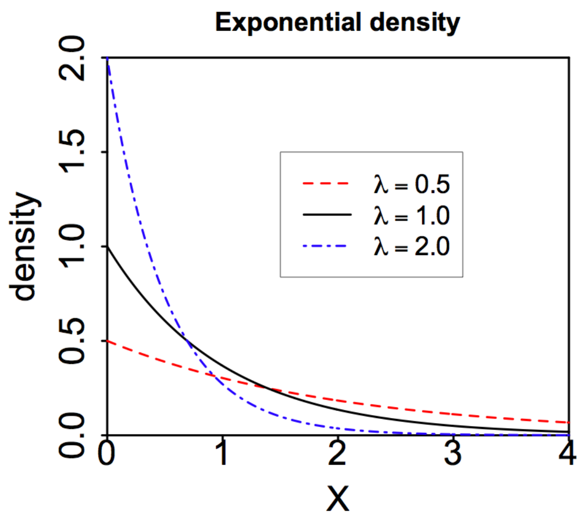
$$(X > s + t) \cap (X > s) = (X > s + t),$$

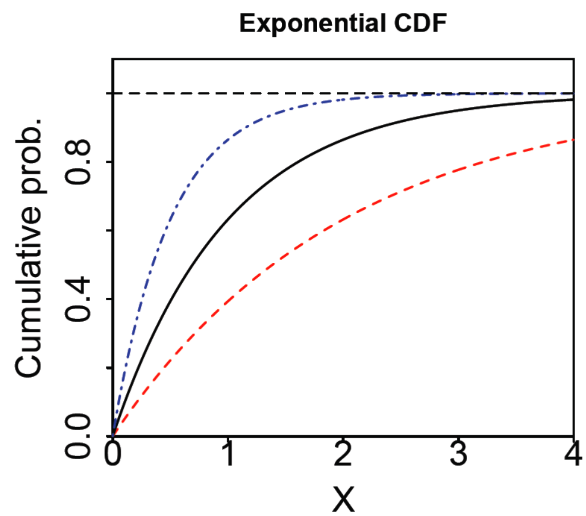
since the event 'waited at least s minutes' contains the event 'waited at least $s + t$ minutes'. Therefore the formula for conditional probability gives

$$P(X > s + t \,|\, X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}} = \mathrm{e}^{-\lambda t} = P(X > t).$$

The probability $P(X > s + t) = \mathrm{e}^{-\lambda(s+t)}$ is the formula for the right tail probability given above.

Graphs:





5 Normal distribution

In 1809, Carl Friedrich Gauss published a monograph introducing several notions that have become fundamental to statistics: the normal distribution, maximum likelihood estimation, and the method of least squares (we will cover all three in this course). For this reason, the normal distribution is also called the Gaussian distribution, and it is by far the most important continuous distribution.

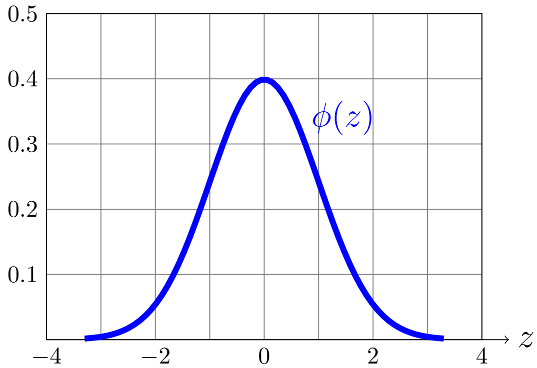
1. Parameters: μ, σ.
2. Range: $(-\infty, \infty)$.
3. Notation: Normal(μ, σ^2) or $N(\mu, \sigma^2)$.
4. Probability density function (pdf): $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}$.
5. Cumulative distribution function (cdf): $F(x)$ has no formula, so use tables or software such as `pnorm` in R to compute $F(x)$.
6. Models: Measurement error, intelligence/ability, height, averages of lots of data.

The standard normal distribution $N(0, 1)$ has mean 0 and variance 1. We reserve Z for a standard normal random variable, $\phi(z) = \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}$ for the standard normal density, and $\Phi(z)$ for the standard normal distribution.

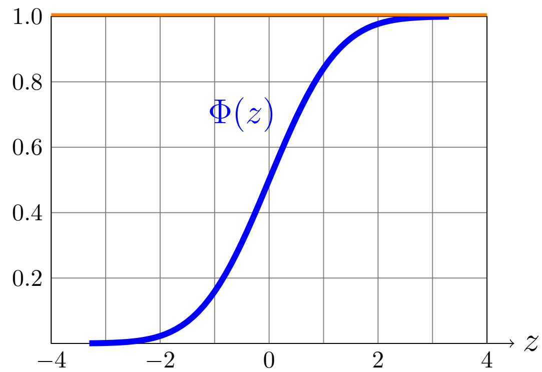
Note: we will define mean and variance for continuous random variables next time. They have the same interpretations as in the discrete case. As you might guess, the normal distribution $N(\mu, \sigma^2)$ has mean μ, variance σ^2, and standard deviation σ.

Here are some graphs of normal distributions. Note that they are shaped like a bell curve. Note also that as σ increases they become more spread out.

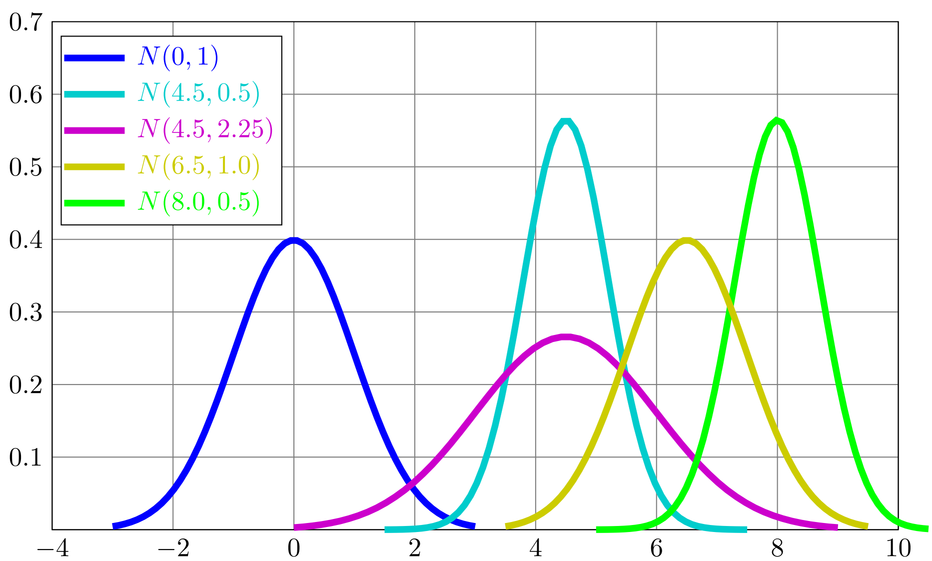
The bell curve: First we show the standard normal probability density and cumulative distribution functions. Below that is a selection of normal densities. Notice that the graph is centered on the mean and the bigger the variance the more spread out the curve.



Standard normal pdf



Standard normal cdf



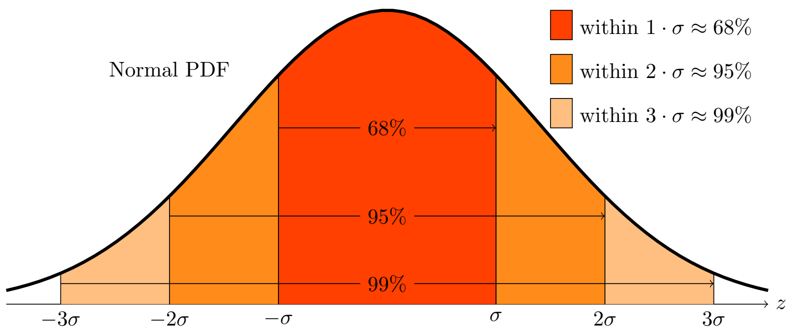
Notation note. In the figure above we use our notation $N(\mu,\sigma^2)$. So, for example, $N(8,0.5)$ has variance 0.5 and standard deviation $\sigma = \sqrt{0.5} \approx 0.7071$.

5.1 Normal probabilities

To make approximations it is useful to remember the following rule of thumb for three approximate probabilities from the standard normal distribution:

$$P(-1 \le Z \le 1) \approx 0.68, \qquad P(-2 \le Z \le 2) \approx 0.95, \qquad P(-3 \le Z \le 3) \approx 0.99.$$

The figure below shows these probabilities as areas under the graph of the standard normal pdf $\phi(z)$.

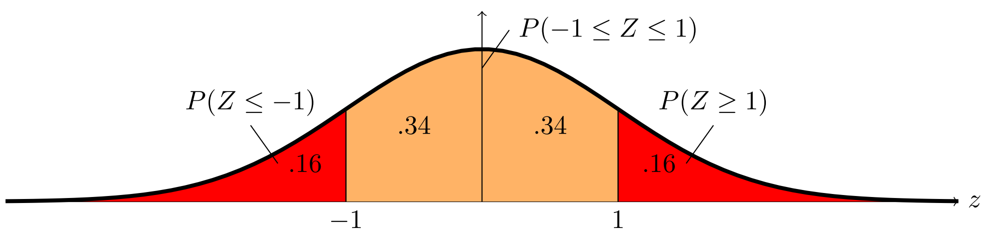


Symmetry calculations

We can use the symmetry of the standard normal distribution about $z = 0$ to make some calculations.

Example 4. The rule of thumb says $P(-1 \le Z \le 1) \approx 0.68$. Use this to estimate $\Phi(1)$.

<u>**answer:**</u> $\Phi(1) = P(Z \le 1)$. In the figure, the two tails (in red) have combined area $1-0.68 = 0.32$. By symmetry the left tail has area 0.16 (half of 0.32), so $P(Z \le 1) \approx 0.68+0.16 = 0.84$.



5.2 Using R to compute $\Phi(z)$.

```
# Use the R function pnorm(x, μ, σ) to compute F(x) for N(μ, σ^2)
pnorm(1,0,1)
[1] 0.8413447

pnorm(0,0,1)
[1] 0.5
pnorm(1,0,2)
[1] 0.6914625

pnorm(1,0,1) - pnorm(-1,0,1)
[1] 0.6826895
pnorm(5,0,5) - pnorm(-5,0,5)
[1] 0.6826895
# Of course z can be a vector of values
pnorm(c(-3,-2,-1,0,1,2,3),0,1)
[1] 0.001349898 0.022750132 0.158655254 0.500000000 0.841344746 0.977249868 0.998650102
```

Note: The R function pnorm(x, μ, σ) uses σ whereas our notation for the normal distribution $\mathrm{N}(\mu, \sigma^2)$ uses σ^2.

Here's a table of values with fewer decimal points of accuracy

z:	-2	-1	0	.3	.5	1	2	3
$\Phi(z)$:	0.0228	0.1587	0.5000	0.6179	0.6915	0.8413	0.9772	0.9987

Example 5. Use R to compute $P(-1.5 \le Z \le 2)$.

answer: This is $\Phi(2) - \Phi(-1.5)$ = `pnorm(2,0,1) - pnorm(-1.5,0,1) = 0.91044`

6 Pareto and other distributions

In 18.05, we only have time to work with a few of the many wonderful distributions that are used in probability and statistics. We hope that after this course you will feel comfortable learning about new distributions and their properties when you need them. Wikipedia is often a great starting point.

The Pareto distribution is one common, beautiful distribution that we will not have time to cover in depth.

1. Parameters: $m > 0$ and $\alpha > 0$.
2. Range: $[m, \infty)$.
3. Notation: Pareto(m, α).
4. Density: $f(x) = \dfrac{\alpha\, m^\alpha}{x^{\alpha+1}}$.
5. Distribution: (easy integral)
$$F(x) = 1 - \frac{m^\alpha}{x^\alpha}, \text{ for } x \ge m$$
6. Tail distribution: $P(X > x) = m^\alpha/x^\alpha$, for $x \ge m$.
7. Models: The Pareto distribution models a power law, where the probability that an event occurs varies as a power of some attribute of the event. Many phenomena follow a power law, such as the size of meteors, income levels across a population, and population levels across cities. See Wikipedia for loads of examples:

   `http://en.wikipedia.org/wiki/Pareto_distribution#Applications`

7 Transformations of Random Variables

We frequently transform a known random variable into a new one by applying a formula. For example we might look at $Y = aX + b$ or $Y = X^2$. In this section we will see how to find the probability density and cumulative distribution of Y from those of X.

For discrete random variables it was often possible do this by looking at probability tables. For continuous random variables we will need to use systematic algebraic techniques. We will see that transforming the pdf is just the change of variables ('u-substitution') from calculus. To transform the cdf directly we will rely on its definition as a probability.

Let's remind ourselves of the basics:

- The cdf of X is $F_X(x) = P(X \le x)$.
- The pdf of X is related to F_X by $f_X(x) = F'_X(x)$.

7.1 Transforming the cdf

Example 6. Suppose X has range $[0, 2]$ and cdf $F_X(x) = x^2/4$. What is the range, pdf and cdf of $Y = X^2$?

answer: The range is easy: $[0, 4]$.

To find the cdf we work systematically from the definition. For this example we will break it down into tiny steps, so you can see the thought process in detail.

Step 1. Use definition:

$$F_Y(y) = P(Y \le y).$$

Step 2. Replace Y by its formula in X:

$$P(Y \le y) = P(X^2 \le y).$$

Step 3. Algebraically manipulate this to isolate the X:

$$P(X^2 \le y) = P(X \le \sqrt{y})$$

Step 4. Notice that this is exactly the definition of F_X:

$$P(X \le \sqrt{y}) = F_X(\sqrt{y})$$

Step 5. Use the known formula for F_X:

$$F_X(\sqrt{y}) = (\sqrt{y})^2/4 = y/4.$$

Following the chain from step 1 to step 5 we have the cdf:

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = y/4.$$

Finally, to find the pdf we can just differentiate the cdf:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{4}.$$

7.2 Transforming the pdf directly

An alternative way to find the pdf directly is by change of variables. We present this for completeness and for anyone who prefers it as a method. Our observation is that most people find the cdf easier to transform.

In calculus you learned the 'u'-substitution. We'll do a calculus example to remind you how this goes and then apply it to the pdf.

Example 7. Calculus example. Convert the integral $\int (x^2+1)^7 \, dx$ into an integral in $u = x^2 + 1$.

answer: We have to convert each part of the integral from x to u:

$$(x^2+1)^7 = u^7$$
$$du = 2x\,dx, \qquad \text{therefore} \qquad dx = \frac{du}{2x} = \frac{du}{2\sqrt{u-1}}$$

Now we replacing each piece in the integral we get

$$\int (x^2+1)^7\,dx = \int u^7 \frac{du}{2\sqrt{u-1}}.$$

Example 8. Find the pdf of Y in Example 6 directly using the method of 'u'-substitution. (In this case, 'u' will actually be 'y'.)

answer: The trick is to remember that probability is given by an integral $\int f_X(x)dx$. We are given the change of variable $y = x^2$, so we change the integral from one in x to one in y.

$$y = x^2 \;\Rightarrow\; dy = 2x\,dx, \qquad \text{therefore} \qquad dx = \frac{dy}{2\sqrt{y}}.$$

We are given $F_X(x) = x^2/4$, so we can compute $f_X(x) = F_X'(x) = x/2$. Changing this to y we have

$$f_X(x) = \sqrt{y}/2.$$

Putting the two pieces together we have the transformation

$$f_X(x)\,dx \;=\; \frac{\sqrt{y}}{2}\,\frac{dy}{2\sqrt{y}} \;=\; \frac{1}{4}\,dy$$

Since this is a probability, the factor in front of dy is the probability density. That is, $f_Y(y) = 1/4$, exactly as in Example 6.

Here are a few more examples. We do them a little more quickly than the above examples.

Example 9. Let $X \sim \exp(\lambda)$, so $f_X(x) = \lambda e^{-\lambda x}$ on $[0,\infty]$. What is the probability density of $Y = X^2$?

answer: We will do this using the change of variables for the pdf.

$$y = x^2 \;\Rightarrow\; dy = 2x\,dx, \qquad \text{therefore} \qquad dx = \tfrac{dy}{2\sqrt{y}}$$
$$f_X(x) = \lambda e^{-\lambda x} = \lambda e^{-\lambda\sqrt{y}}.$$

Combining these we get,

$$f_X(x)\,dx = \lambda e^{-\lambda\sqrt{y}}\,\frac{dy}{2\sqrt{y}} = f_Y(y)\,dy.$$

So we conclude that $\boxed{f_Y(y) = \frac{\lambda}{2\sqrt{y}}\,e^{-\lambda\sqrt{y}}}$.

Example 10. Redo the previous example using the cdf.

answer: The cdf for the exponential random variable X is $F_X(x) = 1 - \mathrm{e}^{-\lambda x}$. Therefore, for $Y = X^2$ we have

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = 1 - \mathrm{e}^{-\lambda\sqrt{y}}.$$

We have found $F_Y(y)$. If we wanted $f_Y(y)$ we could take the derivative. We would get the same answer as in the previous example.

Example 11. Assume $X \sim \mathrm{N}(5, 3^2)$ then $Z = \dfrac{X-5}{3}$ is standard normal, i.e., $Z \sim \mathrm{N}(0,1)$.

answer: Again using the change of variables and the formula for $f_X(x)$ we have

$$z = \frac{x-5}{3} \;\Rightarrow\; dz = \frac{dx}{3}, \qquad \text{therefore} \qquad dx = 3\,dz$$

For this example we will transform $f_X(x)\,dx$ in one line instead of two.

$$f_X(x)\,dx = \frac{1}{3\sqrt{2\pi}}\mathrm{e}^{-(x-5)^2/(2\cdot 3^2)}\,dx = \frac{1}{3\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,3\,dz = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,dz = f_Z(z)\,dz$$

Therefore $f_Z(z) = \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}$. Since this is exactly the density for $\mathrm{N}(0,1)$ we have shown that Z is standard normal.

This example shows an important general property of normal random variables which we state as a theorem.

Theorem. Standardization of normal random variables.
Assume $X \sim \mathrm{N}(\mu, \sigma^2)$. Show that $Z = \dfrac{X-\mu}{\sigma}$ is standard normal, i.e., $Z \sim \mathrm{N}(0,1)$.

Proof. This is exactly the same computation as the previous example with μ replacing 5 and σ replacing 3. We show the computation without comment.

$$z = \frac{x-\mu}{\sigma} \;\Rightarrow\; dz = \frac{dx}{\sigma} \;\Rightarrow\; dx = \sigma\,dz$$

$$f_X(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/(2\cdot\sigma^2)}\,dx = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\sigma\,dz = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,dz = f_Z(z)\,dz$$

Therefore $f_Z(z) = \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}$. This shows Z is standard normal.

We call the change from X to Z in this theorem standardization because it converts X from an arbitrary normal random variable to a standard normal variable.

8 Summary statistics for continuous random variables

So far we have looked at expected value, standard deviation, and variance for discrete random variables. These summary statistics have the same meaning for continuous random variables:

- The expected value $\mu = E(X)$ is a measure of location.

- The standard deviation σ is a measure of the spread or scale.
- The variance $\sigma^2 = \text{Var}(X)$ is the square of the standard deviation.

To move from discrete to continuous, we will simply replace the sums in the formulas by integrals. We will do this carefully and go through many examples in the following sections. In the last section, we will introduce another type of summary statistic, quantiles. You may already be familiar with the .5 quantile of a distribution, otherwise known as the median or 50^{th} percentile.

9 Expected value of a continuous random variable

Definition: Let X be a continuous random variable with range $[a, b]$ and probability density function $f(x)$. The expected value of X is defined by

$$E(X) = \int_a^b x f(x)\, dx.$$

Let's see how this compares with the formula for a discrete random variable:

$$E(X) = \sum_{i=1}^{n} x_i p(x_i).$$

The discrete formula says to take a weighted sum of the values x_i of X, where the weights are the probabilities $p(x_i)$. Recall that $f(x)$ is a probability density. Its units are prob/(unit of X). So $f(x)\, dx$ represents the probability that X is in an infinitesimal range of width dx around x. Thus we can interpret the formula for $E(X)$ as a weighted integral of the values x of X, where the weights are the probabilities $f(x)\, dx$.

As before, the expected value is also called the mean or average.

9.1 Examples

Let's go through several example computations. Where the solution requires an integration technique, we push the computation of the integral to the appendix.

Example 12. Let $X \sim \text{uniform}(0, 1)$. Find $E(X)$.

<u>**answer:**</u> X has range $[0, 1]$ and density $f(x) = 1$. Therefore,

$$E(X) = \int_0^1 x\, dx = \left. \frac{x^2}{2} \right|_0^1 = \boxed{\frac{1}{2}}.$$

Not surprisingly the mean is at the midpoint of the range.

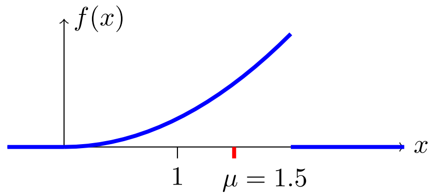
Example 13. Let X have range $[0, 2]$ and density $\frac{3}{8}x^2$. Find $E(X)$.

<u>**answer:**</u>

$$E(X) = \int_0^2 x f(x)\, dx = \int_0^2 \frac{3}{8} x^3\, dx = \left. \frac{3x^4}{32} \right|_0^2 = \boxed{\frac{3}{2}}.$$

Does it make sense that this X has mean is in the right half of its range?

answer: Yes. Since the probability density increases as x increases over the range, the average value of x should be in the right half of the range.

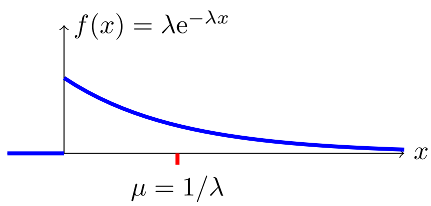


μ is "pulled" to the right of the midpoint 1 because there is more mass to the right.

Example 14. Let $X \sim \exp(\lambda)$. Find $E(X)$.

answer: The range of X is $[0, \infty)$ and its pdf is $f(x) = \lambda \mathrm{e}^{-\lambda x}$. So (details in appendix)

$$E(X) = \int_0^\infty \lambda \mathrm{e}^{-\lambda x}\, dx = -\lambda \mathrm{e}^{-\lambda x} - \left.\frac{\mathrm{e}^{-\lambda x}}{\lambda}\right|_0^\infty = \boxed{\frac{1}{\lambda}}.$$

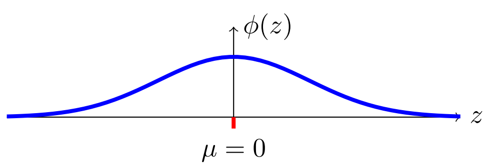


Mean of an exponential random variable

Example 15. Let $Z \sim \mathrm{N}(0, 1)$. Find $E(Z)$.

answer: The range of Z is $(-\infty, \infty)$ and its pdf is $\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}$. So (details in appendix)

$$E(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z \mathrm{e}^{-z^2/2}\, dz = -\left.\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\right|_{-\infty}^{\infty} = \boxed{0}.$$



The standard normal distribution is symmetric and has mean 0.

9.2 Properties of $E(X)$

The properties of $E(X)$ for continuous random variables are the same as for discrete ones:

1. If X and Y are random variables on a sample space Ω then

$$E(X + Y) = E(X) + E(Y). \qquad \text{(linearity I)}$$

2. If a and b are constants then

$$E(aX + b) = aE(X) + b. \qquad \text{(linearity II)}$$

Example 16. In this example we verify that for $X \sim \mathrm{N}(\mu, \sigma)$ we have $E(X) = \mu$.

answer: Example (15) showed that for standard normal Z, $E(Z) = 0$. We could mimic the calculation there to show that $E(X) = \mu$. Instead we will use the linearity properties of $E(X)$. In the class 5 notes on manipulating random variables we showed that if $X \sim \mathrm{N}(\mu, \sigma^2)$ is a normal random variable we can standardize it:

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

Inverting this formula we have $X = \sigma\, Z + \mu$. The linearity of expected value now gives

$$E(X) = E(\sigma\, Z + \mu) = \sigma\, E(Z) + \mu = \mu$$

9.3 Expectation of Functions of X

This works exactly the same as the discrete case. if $h(x)$ is a function then $Y = h(X)$ is a random variable and

$$E(Y) = E(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx.$$

Example 17. Let $X \sim \exp(\lambda)$. Find $E(X^2)$.

answer: Using integration by parts we have

$$E(X^2) = \int_0^{\infty} x^2 \lambda \mathrm{e}^{-\lambda x}\, dx = \left[-x^2 \mathrm{e}^{-\lambda x} - \frac{2x}{\lambda}\mathrm{e}^{-\lambda x} - \frac{2}{\lambda^2}\mathrm{e}^{-\lambda x}\right]_0^{\infty} = \boxed{\frac{2}{\lambda^2}}.$$

10 Variance

Now that we've defined expectation for continuous random variables, the definition of variance is identical to that of discrete random variables.

Definition: Let X be a continuous random variable with mean μ. The variance of X is

$$\mathrm{Var}(X) = E((X - \mu)^2).$$

10.1 Properties of Variance

These are exactly the same as in the discrete case.

1. If X and Y are independent then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

2. For constants a and b, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

3. **Theorem**: $\mathrm{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2$.

For Property 1, note carefully the requirement that X and Y are independent.

Property 3 gives a formula for $\text{Var}(X)$ that is often easier to use in hand calculations. The proofs of properties 2 and 3 are essentially identical to those in the discrete case. We will not give them here.

Example 18. Let $X \sim \text{uniform}(0,1)$. Find $\text{Var}(X)$ and σ_X.

answer: In Example 12 we found $\mu = 1/2$. Next we compute

$$\text{Var}(X) = E((X-\mu)^2) = \int_0^1 (x-1/2)^2\,dx = \boxed{\frac{1}{12}}.$$

Example 19. Let $X \sim \exp(\lambda)$. Find $\text{Var}(X)$ and σ_X.

answer: In Examples 14 and 17 we computed

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda} \quad \text{and} \quad E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2}.$$

So by Property 3,

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \text{and} \quad \sigma_X = \frac{1}{\lambda}.$$

We could have skipped Property 3 and computed this directly from $\text{Var}(X) = \int_0^\infty (x-1/\lambda)^2\lambda e^{-\lambda x}\,dx$.

Example 20. Let $Z \sim \text{N}(0,1)$. Show $\text{Var}(Z) = 1$.

Note: The notation for normal variables is $X \sim \text{N}(\mu, \sigma^2)$. This is certainly suggestive, but as mathematicians we need to prove that $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Above we showed $E(X) = \mu$. This example shows that $\text{Var}(Z) = 1$, just as the notation suggests. In the next example we'll show $\text{Var}(X) = \sigma^2$.

answer: Since $E(Z) = 0$, we have

$$\text{Var}(Z) = E(Z^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\,dz.$$

(using integration by parts with $u = z$, $v' = ze^{-z^2/2} \Rightarrow u' = 1$, $v = -e^{-z^2/2}$)

$$= \frac{1}{\sqrt{2\pi}}\left(-ze^{-z^2/2}\Big|_{-\infty}^{\infty}\right) + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2}\,dz.$$

The first term equals 0 because the exponential goes to zero much faster than z grows at both $\pm\infty$. The second term equals 1 because it is exactly the total probability integral of the pdf $\varphi(z)$ for $\text{N}(0,1)$. So $\text{Var}(X) = 1$.

Example 21. Let $X \sim \text{N}(\mu, \sigma^2)$. Show $\text{Var}(X) = \sigma^2$.

answer: This is an exercise in change of variables. Letting $z = (x-\mu)/\sigma$, we have

$$\begin{aligned}\text{Var}(X) = E((X-\mu)^2) &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\,dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\,dz = \sigma^2.\end{aligned}$$

The integral in the last line is the same one we computed for $\text{Var}(Z)$.

11 Quantiles

Definition: The median of X is the value x for which $P(X \le x) = 0.5$, i.e. the value of x such that $P(X \le X) = P(X \ge x)$. In other words, X has equal probability of being above or below the median, and each probability is therefore $1/2$. In terms of the cdf $F(x) = P(X \le x)$, we can equivalently define the median as the value x satisfying $F(x) = 0.5$.

Think: What is the median of Z?

answer: By symmetry, the median is 0.

Example 22. Find the median of $X \sim \exp(\lambda)$.

answer: The cdf of X is $F(x) = 1 - \mathrm{e}^{-\lambda x}$. So the median is the value of x for which $F(x) = 1 - \mathrm{e}^{-\lambda x} = 0.5.$. Solving for x we find: $\boxed{x = (\ln 2)/\lambda}$.

Think: In this case the median does not equal the mean of $\mu = 1/\lambda$. Based on the graph of the pdf of X can you argue why the median is to the left of the mean.

Definition: The $\mathrm{p^{th}}$ quantile of X is the value q_p such that $P(X \le q_p) = p$.

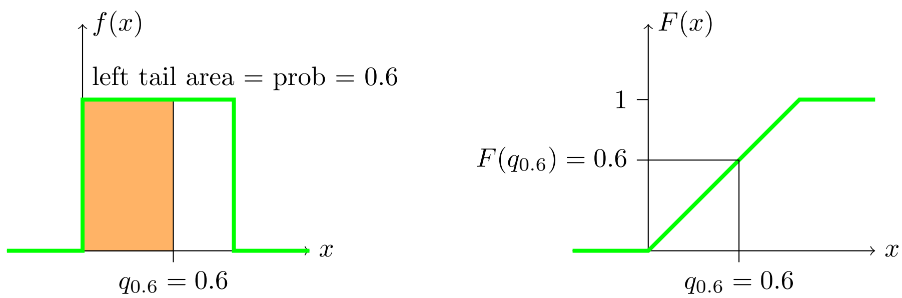
Notes. 1. In this notation the median is $q_{0.5}$.

2. We will usually write this in terms of the cdf: $F(q_p) = p$.

With respect to the pdf $f(x)$, the quantile q_p is the value such that there is an area of p to the left of q_p and an area of $1 - p$ to the right of q_p. In the examples below, note how we can represent the quantile graphically using either area of the pdf or height of the cdf.

Example 23. Find the 0.6 quantile for $X \sim U(0, 1)$.

answer: The cdf for X is $F(x) = x$ on the range [0,1]. So $q_{0.6} = 0.6$.

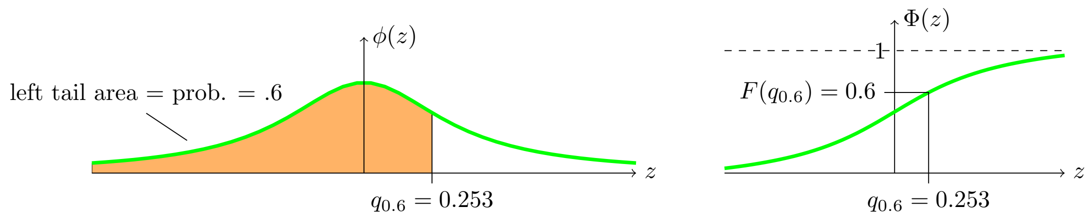


$q_{0.6}$: left tail area $= 0.6 \;\Leftrightarrow\; F(q_{0.6}) = 0.6$

Example 24. Find the 0.6 quantile of the standard normal distribution.

answer: We don't have a formula for the cdf, so we use the R 'quantile function' `qnorm`.

$$q_{0.6} = \texttt{qnorm(0.6, 0, 1)} = 0.25335$$



$q_{0.6}$: left tail area $= 0.6 \;\Leftrightarrow\; F(q_{.6}) = 0.6$

Quantiles give a useful measure of location for a random variable. We will use them more in coming lectures.

11.1 Percentiles, deciles, quartiles

For convenience, quantiles are often described in terms of percentiles, deciles or quartiles. The 60^{th} percentile is the same as the 0.6 quantile. For example you are in the 60^{th} percentile for height if you are taller than 60 percent of the population, i.e. the probability that you are taller than a randomly chosen person is 60 percent.

Likewise, deciles represent steps of 1/10. The third decile is the 0.3 quantile. Quartiles are in steps of 1/4. The third quartile is the 0.75 quantile and the 75^{th} percentile.

12 Appendix: Integral Computation Details

Example 14: Let $X \sim \exp(\lambda)$. Find $E(X)$.

The range of X is $[0, \infty)$ and its pdf is $f(x) = \lambda e^{-\lambda x}$. Therefore

$$E(X) = \int_0^\infty x f(x)\, dx = \int_0^\infty \lambda x e^{-\lambda x}\, dx$$

(using integration by parts with $u = x$, $v' = \lambda e^{-\lambda x} \;\Rightarrow\; u' = 1$, $v = -e^{-\lambda x}$)

$$= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx$$

$$= 0 - \left.\frac{e^{-\lambda x}}{\lambda}\right|_0^\infty = \frac{1}{\lambda}.$$

We used the fact that $x e^{-\lambda x}$ and $e^{-\lambda x}$ go to 0 as $x \to \infty$.

Example 15: Let $Z \sim \mathrm{N}(0, 1)$. Find $E(Z)$.

The range of Z is $(-\infty, \infty)$ and its pdf is $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. By symmetry the mean must be 0. The only mathematically tricky part is to show that the integral converges, i.e. that the mean exists at all (some random variable do not have means, but we will not encounter this very often.) For completeness we include the argument, though this is not something we will ask you to do. We first compute the integral from 0 to ∞:

$$\int_0^\infty z\phi(z)\, dz = \frac{1}{\sqrt{2\pi}} \int_0^\infty z e^{-z^2/2}\, dz.$$

The u-substitution $u = z^2/2$ gives $du = z\,dz$. So the integral becomes

$$\frac{1}{\sqrt{2\pi}}\int_0^\infty z\mathrm{e}^{-z^2/2}\,dz. = \frac{1}{\sqrt{2\pi}}\int_0^\infty \mathrm{e}^{-u}\,du = -\mathrm{e}^{-u}\Big|_0^\infty = 1$$

Similarly, $\displaystyle\int_{-\infty}^0 z\phi(z)\,dz = -1$. Adding the two pieces together gives $E(Z) = 0$.

Example 17: Let $X \sim \exp(\lambda)$. Find $E(X^2)$.

$$E(X^2) = \int_0^\infty x^2 f(x)\,dx = \int_0^\infty \lambda x^2 \mathrm{e}^{-\lambda x}\,dx$$

(using integration by parts with $u = x^2$, $v' = \lambda\mathrm{e}^{-\lambda x} \Rightarrow u' = 2x$, $v = -\mathrm{e}^{-\lambda x}$)

$$= -x^2\mathrm{e}^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x\mathrm{e}^{-\lambda x}\,dx$$

(the first term is 0, for the second term use integration by parts: $u = 2x$, $v' = \mathrm{e}^{-\lambda x} \Rightarrow u' = 2$, $v = -\frac{\mathrm{e}^{-\lambda x}}{\lambda}$)

$$\begin{aligned} &= -2x\frac{\mathrm{e}^{-\lambda x}}{\lambda}\Big|_0^\infty + \int_0^\infty \frac{\mathrm{e}^{-\lambda x}}{\lambda}\,dx \\ &= 0 - 2\frac{\mathrm{e}^{-\lambda x}}{\lambda^2}\Big|_0^\infty = \frac{2}{\lambda^2}. \end{aligned}$$