Gallery of Continuous Random Variables
Class 5, 18.05, Spring 2014
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.

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.

Although it comes towards the end, we call your attention to the normal distribution. It is easily the most important distribution defined here.

3   Uniform distribution

1. Parameters: a, b.

2. Range: $[a, b]$.

3. Notation: uniform(a, b) or U(a, b).

4. Density: $f(x) = \dfrac{1}{b-a}$ for $a \leq x \leq b$.

5. Distribution: $F(x) = (x - a)/(b - a)$ for $a \leq x \leq b$.

6. Models: All outcomes in the range have equal probability (more precisely all outcomes have the same probability density).

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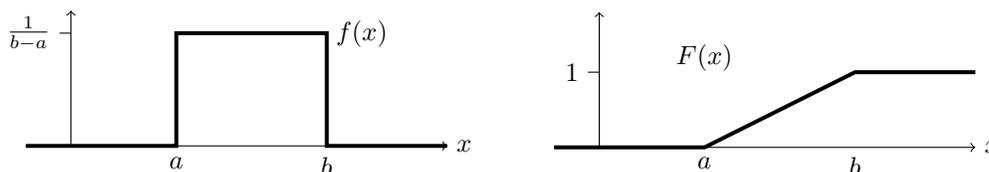
pdf and cdf for uniform(a,b) distribution.

Examples. 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 uniformly distributed between -0.5 and 0.5 millimeters.

2. Many boardgames 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. Density: $f(x) = \lambda e^{-\lambda x}$ for $0 \le x$.

5. Distribution: (easy integral)

$$F(x) = 1 - e^{-\lambda x} \text{ for } x \ge 0$$

6. *Right tail distribution*: $P(X > x) = 1 - F(x) = e^{-\lambda x}$.

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

Examples. 1. 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 one over the average number of taxis that pass per minute (on weekday afternoons).

2. 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 1. Suppose that the probability that a taxi arrives within the first five minutes is p. If I wait five minutes and in fact no taxi arrives, then the probability that a taxi arrives within the next five minutes is still p.

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 the 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 unaffected by having already waited s minutes without incident. In symbols, $P(X > s + t \,|\, X > s) = P(X > t)$.

Proof of memorylessness: Since $(X > s + t) \cap (X > s) = (X > s + t)$ we have

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

Graphs:
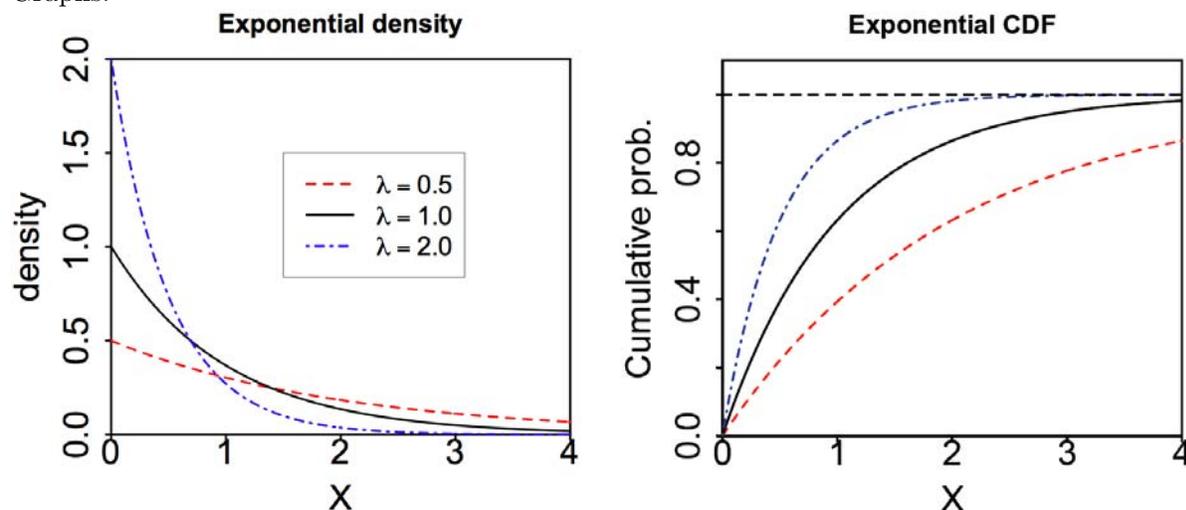
5 Normal distribution

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 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. Density: $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$.

5. Distribution: $F(x)$ 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, $\phi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$ 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 they are shaped like a *bell curve*. Note also that as σ increases they become more spread out.
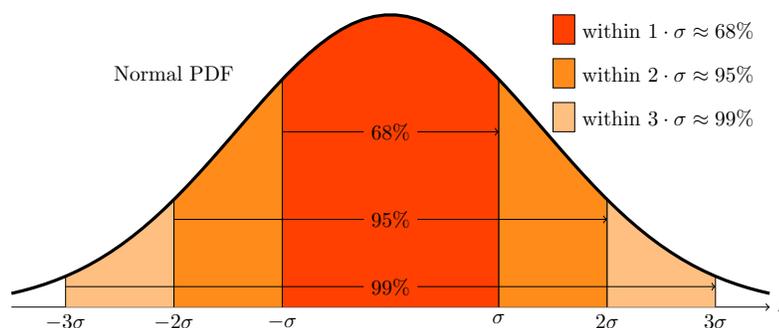
Graphs: (the *bell curve*):
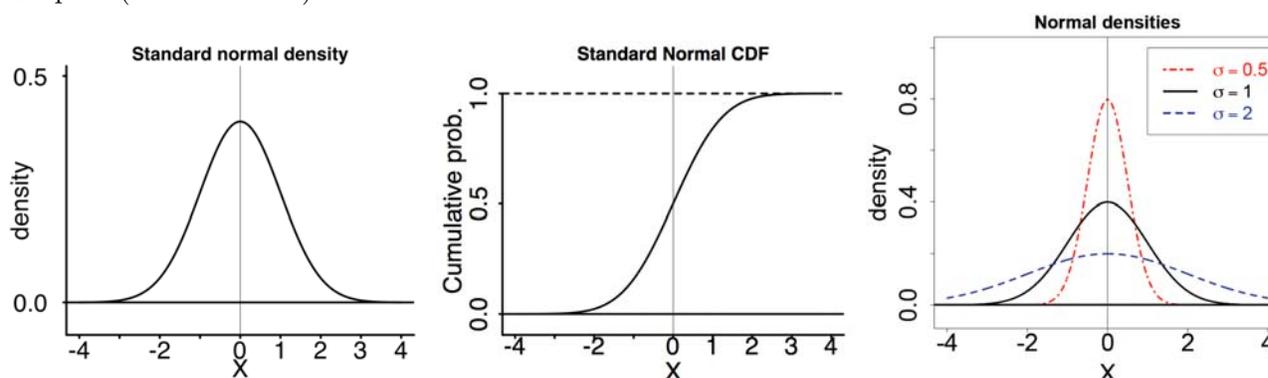
5.1 Normal probabilities

To make approximations it is useful to remember the following rule of thumb for three approximate probabilities
$$P(-1 \le Z \le 1) \approx .68, \quad P(-2 \le Z \le 2) \approx .95, \quad P(-3 \le Z \le 3) \approx .99$$

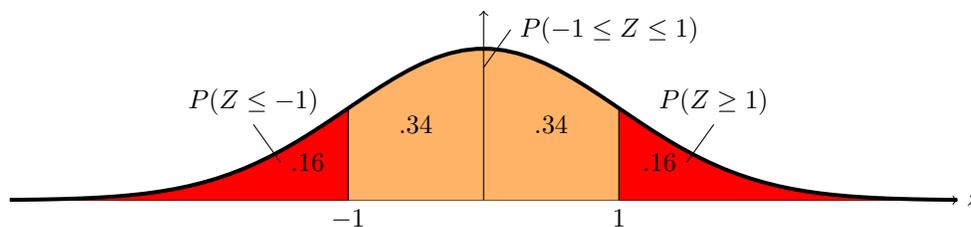**Symmetry calculations**

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

Example 1. The rule of thumb says $P(-1 \le Z \le 1) \approx .68$. Use this to estimate $\Phi(1)$.

answer: $\Phi(1) = P(Z \le 1)$. In the figure, the two tails (in red) have combined area 1-.68 = .32. By symmetry the left tail has area .16 (half of .32), so $P(Z \le 1) \approx .68 + .16 = .84$.

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

```
# Use the R function pnorm(x, μ, σ) to compute F(x) for N(μ, σ²)
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 N(μ, σ^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 2. Use R to compute $P(-1.5 \leq Z \leq 2)$.

<u>answer:</u> 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)

$$F(x) = 1 - \frac{m^\alpha}{x^\alpha}, \ \text{ for } x \geq m$$

6. Tail distribution: $P(X > x) = m^\alpha / x^\alpha$, for $x \geq m$.

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

18.05 Introduction to Probability and Statistic
Spring 2014