Common Distribution Families - II
Continuous Distributions

Continuous Probability Models

There are other types of random variables besides the discrete ones you studied in the previous chapter. This chapter covers another major class: *continuous random variables*. These are central in statistics and are extensively used in applied probability. Calculus is essential to understand this topic.

Running Example: a Random Dart

Imagine that we throw a dart at random at the interval (0,1). Let D denote the spot we hit. By "at random", we mean that all subintervals of equal length are equally likely to get hit, such as (0.7,0.8) and (0.2,0.3).

Randomness

$$P(u \leq D \leq v) = v - u$$

For any $0 \leq u < v \leq 1$.

D is termed a **continuous** random variable because its support is a continuum of points in the interval (0,1).

Individual Values Now Have Probability Zero

The key point:

$$P(D = c) = 0 \tag{1}$$

For any individual point c.

How to make sense of it?

- Take $c = 0.3$ as an example:

  $$P(D = 0.3) \leq P(0.29 \leq D \leq 0.31) = 0.02$$

  But using smaller intervals, we can deduce $P(D = 0.3)$ must be smaller than any positive number, hence it's 0.
- Given infinite points, if all had some nonzero probability w, the probabilities would sum to infinity, not 1. Thus, they must have a probability of 0.

This observation holds true for any continuous random variable.

But Now We Have a Problem

Equation (1) presents a problem. In the case of discrete random variables M, their distribution was defined using the probability mass function, p_M. In the continuous case, however, all the probabilities of individual values are 0. We need an alternative approach.

Our Way Out of the Problem: Cumulative Distribution Functions:

For any random variable W, its **cumulative distribution function** (cdf), F_W, is defined by:

$$F_W(t) = P(W \leq t), -\infty < t < \infty \tag{2}$$

(Note: It is customary to use capital F for a cdf, with a subscript denoting the random variable's name.)

Example: Random Dart cdf

Using the "random dart" example:

$$F_D(0.23) = P(D \leq 0.23) = P(0 \leq D \leq 0.23) = 0.23$$

Additionally,

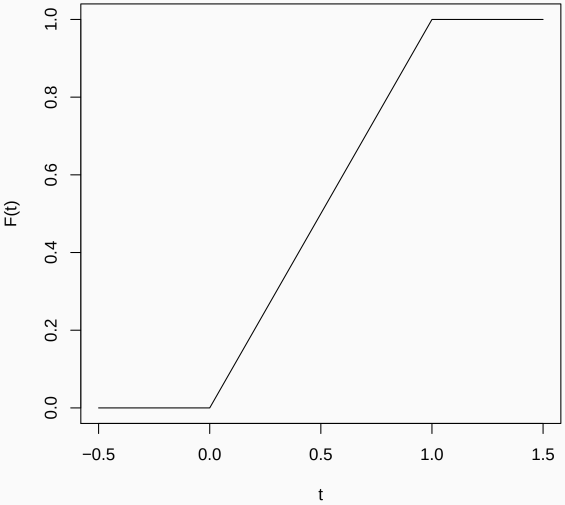
$$F_D(-10.23) = 0 \quad \text{and} \quad F_D(10.23) = 1$$

The general cdf for our dart is:

$$F_D(t) = \begin{cases} 0, & \text{if } t \leq 0 \\ t, & \text{if } 0 < t < 1 \\ 1, & \text{if } t \geq 1 \end{cases}$$

Visualizing the cdf

Here is the graph of F_D:



cdf of a Discrete Random Variable

For a discrete variable, say Z (number of heads from two coin tosses):

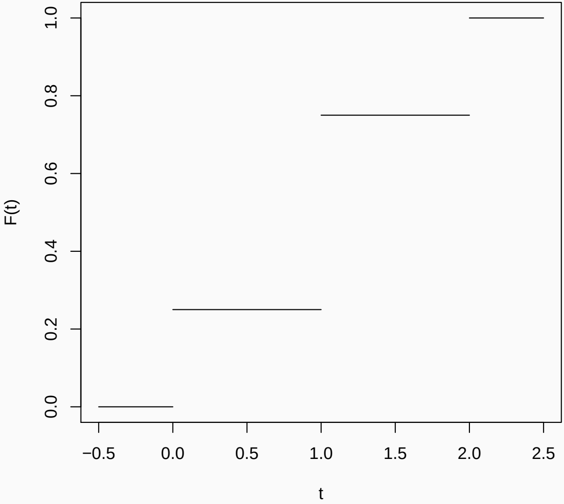
$$F_Z(t) = \begin{cases} 0, & \text{if } t < 0 \\ 0.25, & \text{if } 0 \le t < 1 \\ 0.75, & \text{if } 1 \le t < 2 \\ 1, & \text{if } t \ge 2 \end{cases}$$

For example:

$$F_Z(1.2) = P(Z = 0 \text{ or } Z = 1) = 0.25 + 0.50 = 0.75$$

Visualizing the cdf of Z

Here's the graph of F_Z:



What's Next?

In our study, random variables are either discrete or continuous. However, some exist that are neither (although rare).

With cdfs in hand, our goal is to find a counterpart for continuous random variables that mirrors the probability mass functions for discrete ones.

Intuitive Understanding

Intuition is key here. Make SURE you develop a good intuitive understanding of density functions, as it is vital in being able to apply probability well. We will use it a lot in our course.

From pmfs to Densities (pdfs)

- From Equation (2), for a discrete variable, its cdf is calculated by summing its pmf.
- In continuous cases, we integrate, not sum.
- Hence, the analog of pmf for continuous case should be something that integrates to the cdf.
- This is the derivative of the cdf, termed as **density**.

Density (pdf) Definition Consider a continuous random variable W. Define

$$f_W(t) = \frac{d}{dt} F_W(t), -\infty < t < \infty \tag{3}$$

wherever the derivative exists. The function f_W is called the **probability density function** (pdf), or just the **density** of W.

Properties of Densities

Equation (3) implies:

Property A

$$P(a < W \leq b) = F_W(b) - F_W(a) \tag{4}$$

$$= \int_a^b f_W(t)\, dt \tag{5}$$

- (4) arises from the difference in probabilities accumulated from $-\infty$ to b and a.
- (5) is from the Fundamental Theorem of Calculus.

More Properties

Property B

$$P(a < W \le b) = P(a \le W \le b) = P(a \le W < b) = P(a < W < b) = \int_a^b f_W(t)\,dt$$

Property C

$$\int_{-\infty}^{\infty} f_W(t)\,dt = 1 \tag{6}$$

- Note: $f_W(t)$ will be 0 for ranges of t where W cannot take on values.

Characteristics of a Density (pdf)

- Any nonnegative function integrating to 1 is a density.
- A density can be increasing, decreasing, or mixed.
- Densities can have values greater than 1 at points but must integrate to 1.

Considering a Continuous Random Variable X:

Suppose we have some continuous random variable X, with density f_X, graphed in Figure 1.

Consider probabilities of the form:

$$P(s - 0.1 < X < s + 0.1) \tag{7}$$

Visual Representation

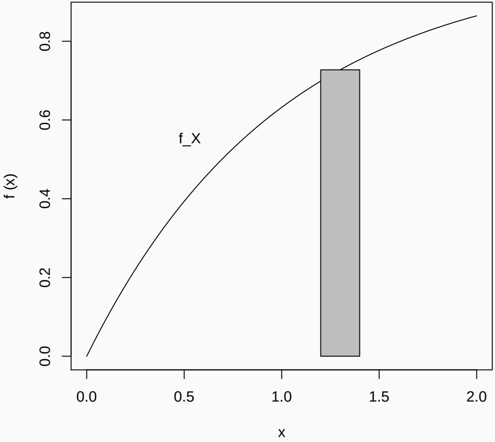


Figure 1: Approximation of Probability by a Rectangle

An Illustrative Case: s = 1.3

The rectangular strip in the figure reminds us of early calculus.

The area under f_X from 1.2 to 1.4 is approximately:

$$2(0.1)f_X(1.3) \approx \int_{1.2}^{1.4} f_X(t)\ dt \tag{8}$$

From the properties we've discussed:

$$P(1.2 < X < 1.4) \approx 2(0.1)f_X(1.3) \tag{9}$$

Another Case: s = 0.4

Similarly, for s = 0.4,

$$P(0.3 < X < 0.5) \approx 2(0.1)f_X(0.4) \tag{10}$$

And, in general:

$$P(s - 0.1 < X < s + 0.1) \approx 2(0.1)f_X(s) \tag{11}$$

A Key Insight

> *Regions in the number line (X-axis in the picture) with low density have low probabilities, while regions with high density have high probabilities.*

Although densities are not probabilities, they indicate which regions are likely or unlikely.

Expected Values - Understanding the Nature of Expected Values in Continuous Cases

Expected Value for Continuous RVs

Recall expectation for Discrete rv W:

For a discrete variable W:

$$E(W) = \sum_{c} c\, p_W(c) \tag{12}$$

For instance, if W is the number of dots in rolling two dice, c ranges over values 2, 3, ..., 12.

Continuous Analog: Expected Value

The analog for a continuous W: **Property D:**

$$E(W) = \int_{t} t\, f_W(t)\ dt \tag{13}$$

Where t ranges over possible values of W. Alternatively:

$$E(W) = \int_{-\infty}^{\infty} t\, f_W(t)\ dt \tag{14}$$

Calculating $E(W^2)$ and More

For the square of W:

$$E(W^2) = \int_t t^2 \, f_W(t) \; dt \tag{15}$$

In general: **Property E:**

$$E[g(W)] = \int_t g(t) \, f_W(t) \; dt \tag{16}$$

Properties of Expected Value and Variance

Many properties of expected value and variance for discrete random variables also apply to continuous ones:

Property F: The following properties of discrete variables remain valid in the continuous scenario.

- $E(U+V) = E(U) + E(V)$
- $E(aX + bY) = aEX + bEY$
- If U and V *are* independent, then $E(UV) = EU \cdot EV$
- $Var(U) = E(U^2) - (EU)^2$
- $Var(cU) = c^2 Var(U)$

A First Example: Density Function: $2t/15$

Consider the density function:

$$f(t) = \begin{cases} 2t/15 & \text{for } t \in (1,4) \\ 0 & \text{elsewhere} \end{cases}$$

Say X has this density.

Computations

$$E(X) = \int_1^4 t \cdot \frac{2t}{15}\, dt = 2.8 \tag{17}$$

$$P(X > 2.5) = \int_{2.5}^4 \frac{2t}{15}\, dt = 0.65 \tag{18}$$

$$F_X(s) = \int_1^s \frac{2t}{15}\, dt = \frac{s^2 - 1}{15} \quad \text{for } s \in (1,4) \tag{19}$$

$$Var(X) = E(X^2) - (E(X))^2 = \int_1^4 t^2 \cdot \frac{2t}{15}\, dt - 2.8^2 = 0.66$$

Lifetime of a Light Bulb

Suppose L is the lifetime of a light bulb (say in years), with the above density. Let's find some quantities in that context:

Proportion of bulbs with lifetime less than the mean lifetime:

$$P(L < 2.8) = \int_{1}^{2.8} 2t/15 \; dt = (2.8^2 - 1)/15 = 0.456 \tag{20}$$

Mean of 1/L:

$$E(1/L) = \int_{1}^{4} \frac{1}{t} \cdot 2t/15 \; dt = \frac{2}{5} \tag{21}$$

Concept of Support

The *support* of a discrete distribution is its domain, i.e., where pmf is positive.

It is similar for a continuous random variable: The support represents the range where the density is non-zero.

For the density in the slides before, the support is (1,4).

Common Parametric Families of Continuous Distributions

The Uniform Distribution

Uniform Distribution

Consider the example of throwing a dart at an interval (a,b).

For a uniform distribution with all points being "equally likely", the density must be constant in this interval and integrate to 1.

A random variable X has the *uniform* distribution on (a, b) with $a < b$, denoted $X \sim \text{Uniform}(a, b)$, if its pdf has the form

$$f_X(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{for } x \in (a,b), \\ 0 & \text{otherwise.} \end{cases}$$

That is, the (continuous) uniform distribution is defined by spreading density uniformly over the interval (a, b).

Uniform Distribution (with parameters (a, b))

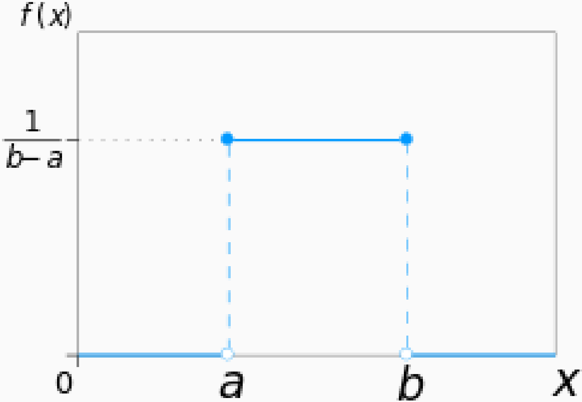


Figure 2: Plot of the Uniform(a, b) pdf.

Uniform Distribution (with parameters (a, b)

Descriptive Measures:

Mean: $E[X] = \dfrac{a+b}{2}$,

Variance: $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$,

cdf: $F_X(x) = \dfrac{x-a}{b-a}$ for $a < x < b$,

MGF: $M_X(t) = \dfrac{e^{tb} - e^{ta}}{t(b-a)}$.

Application (Probability Integral Transform) If X is a continuous random variable and has cdf $F_X(x)$. Then $Y = F_X(X)$ has the uniform distribution on $(0, 1)$. It is very important in random number generation and hence simulation studies. Many other applications including order statistics will be illustrated later.

R Functions for Uniform Distribution

For a uniformly distributed random variable X on (a, b):

- **dunif(x,a,b)**: to find $f_X(x)$
- **punif(q,a,b)**: to find $P(X \leq q)$
- **qunif(q,a,b)**: to find c such that $P(X \leq c) = q$
- **runif(n,a,b)**: to generate n independent values of X

For many of these functions in R, both **x** and **q** can be vectors.

Modeling of Denial-of-Service Attack

A uniform distribution can indicate a possible **denial-of-service attack**. In this scenario, an attacker aims to monopolize resources by overwhelming them with requests.

Research by David Marchette suggests that attackers choose uniformly distributed false IP addresses, a pattern not typically observed at servers.[1]

[1] *Statistical Methods for Network and Computer Security*, David J. Marchette, Naval Surface Warfare Center, `rion.math.iastate.edu/IA/2003/foils/marchette.pdf`.

The Normal (Gaussian) Family of Continuous Distributions

Normal (Gaussian) Distribution

Bell-shaped Curves These are known as the famous "bell-shaped curves" due to the shape of their densities.

"All that glitters is not gold" — Shakespeare

While other families like the Cauchy also have bell shapes, the Central Limit Theorem makes the normal family especially significant.

Normal Distribution: Density and Properties

Density and Parameters:

The density for a normal distribution is given by:

$$f_W(t) = \frac{1}{\sqrt{2\pi}\sigma}\ e^{-0.5\left(\frac{t-\mu}{\sigma}\right)^2}, -\infty < t < \infty$$

This is a two-parameter family, indexed by:

- μ (mean or expected value)
- σ (standard deviation)

It's denoted as $N(\mu, \sigma^2)$, with the custom being to specify the variance σ^2 instead of the standard deviation.

Normal Densities

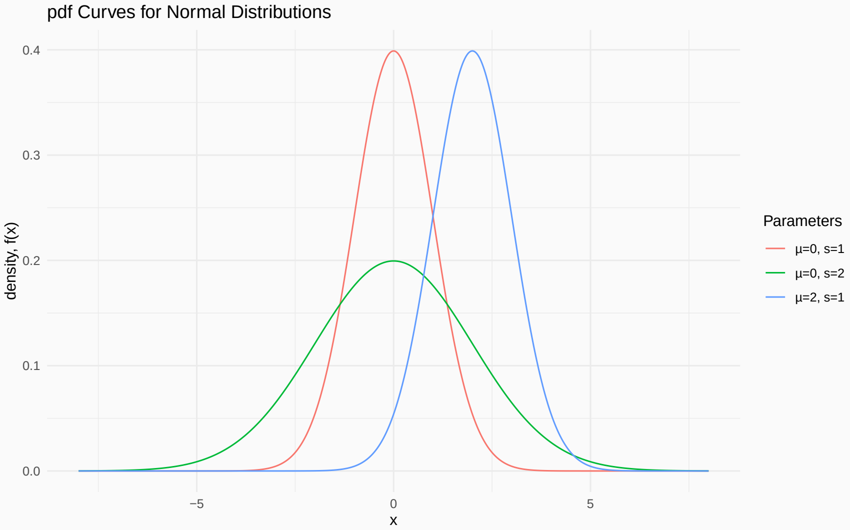


Figure 3: Various Normal Densities

Importance of the Normal Family

The significance of the normal family cannot be overstated (see also Chapter 9).

1. It is analytically simple; it has a bell shape.
2. Many real life phenomena appear to be approximately normally distributed.
3. A large portion of statistical theory is built on the normal distribution.
4. To compute probabilities associated with the normal distribution, we can use standard normal tables (or software).
5. Normal distribution is often used to *approximate other distributions* (based on CLT). For instance, when n is large enough,
   - 5.1 Binomial(n, p) distribution may be approximated by a normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$,
   - 5.2 Poisson(λ) distribution may be approximated by a normal distribution with $\mu = \sigma^2 = \lambda$.

Normal Distribution

Because of CLT, most scientists in the late 19th and early 20th centuries believed that almost all data sets were normal.

For example, the famous French mathematician Henri Poincaré said "*Everyone believes it: experimentalists believe it is a mathematical theorem, and mathematicians believe that it is an empirical fact*".

Descriptive Measures:
Mean: $E[X] = \mu$,
Variance: $\text{Var}(X) = \sigma^2$,
MGF: $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$.

The Exponential Distributions

Clarification on the Term "Exponential Family":
We have been discussing parametric families of distributions. In this context, we introduce the family of exponential distributions.

Important: This should not be confused with the term *exponential family* in mathematical statistics, which encompasses exponential distributions but is more extensive.

Exponential Distribution: Density and Properties

The densities in this family are given by:

$$f_W(t) = \lambda e^{-\lambda t}, \quad 0 < t < \infty$$

After integration, we find:

$$E(W) = \frac{1}{\lambda}$$

$$Var(W) = \frac{1}{\lambda^2}$$

It's intriguing why λ is preferred over $1/\lambda$ (mean) as the indexing parameter.

Exponential Densities

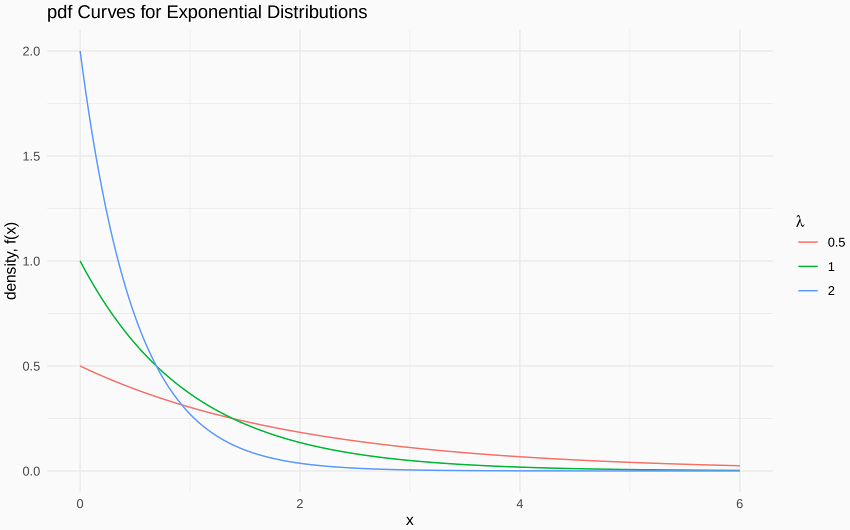


Figure 4: Various Exponential Densities

Exponential Distribution

For $X \sim \text{Exponential}(\lambda)$, **Descriptive Measures:**

Mean: $E[X] = 1/\lambda$,

Variance: $\text{Var}(X) = 1/\lambda^2$,

cdf: $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$,

MGF: $M_X(t) = \dfrac{\lambda}{\lambda - t}$ for $t < \lambda$.

Memoryless Property of the Exponential Distribution: If $X \sim \text{Exponential}(\lambda)$, for $s, t > 0$

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

Why? If $X \sim \text{Exponential}(\lambda)$, then for $t > 0$, we have $P(X > t) = \int_t^\infty \lambda\, e^{-\lambda x} dx = e^{-\lambda t}$. Then

$$\begin{aligned} P(X > t + s | X > t) &= P(X > t + s)/P(X > t) \\ &= e^{-\lambda (t+s)}/e^{-\lambda t} = e^{-\lambda s} = P(X > s). \quad \square \end{aligned}$$

Exponential Distribution

Ex 1: Suppose that the lifetime of a certain type of electric component has an exponential distribution with a mean life of 500 hours (i.e., $1/\lambda = 500$). If X denotes the lifetime of the component, then $X \sim$ Exponential(1/500). Then

$$P(X > x) = \int_x^\infty \frac{1}{500} e^{-t/500} dt = e^{-x/500}.$$

Ex 2: Under the setting of Ex 1 above, suppose that given that the component has been operating for 300 hours, find that probability that the component will last for another 600 hours.

Solution:

The "component has been operating for 300 hours" means that "the lifetime of the component is larger than 300", i.e., "$X > 300$" is given.

Solution (continued)

And "the component will last for another 600 hours (in addition to 300 hours" implies it will "last for 600+300=900 hours", i.e., "the lifetime of the component will be larger than 900", i.e., "$X > 900$". Thus,

$$P(X > 900 \mid X > 300) = P(X > 600) = e^{-600/500} = e^{-6/5} \approx 0.30$$

This means that, for such components, an old component is as good as new. $\square$

The exponential distribution is the only continuous distribution with support in $(0, \infty)$ that is memoryless (i.e. "If X is a positive continuous random variable with the memoryless property, then $X \sim \text{Exponential}(\lambda)$ for some $\lambda > 0$").

R Functions for Exponential Distributions

Relevant R functions for a uniformly distributed random variable X with parameter λ include:

- **dexp(x,lambda)**: Determine $f_X(x)$
- **pexp(q,lambda)**: Calculate $P(X \leq q)$
- **qexp(q,lambda)**: Find c such that $P(X \leq c) = q$
- **rexp(n,lambda)**: Generate n independent values of X

Example: Refunds on Failed Components

Scenario: A manufacturer's electronic component has an exponentially distributed lifetime with mean 10000 hours. Refunds are given for failures before 500 hours. Let L denote the lifetime of a component and M denote the number of sold items up to the first refund. We aim to find EM and $Var(M)$.

Insight: M has a geometric distribution with success probability:

$$p = P(L < 500) = \int_0^{500} 0.0001e^{-0.0001t}\, dt = 0.05$$

Using formulas for $Geo(p = .05)$, we can determine EM and $Var(M)$.

Importance in Modeling

The exponential family is crucial in modeling due to its unique properties, making it applicable to various scenarios:

- Air conditioner lifetimes on airplanes
- Interarrival times (e.g., bank customers, network messages)
- Software reliability studies

The Gamma Family of Distributions

Gamma Distribution

Motivation Ex:

Suppose at time 0 we install a light bulb in a lamp, which burns X_1 amount of time. We immediately install a new bulb, which burns for time X_2, and so on. The X_i are independent random variables having an exponential distribution with parameter λ.

- $T_r = \sum_{i=1}^{r} X_i = X_1 + ... + X_r, \quad r = 1, 2, 3, ...$
- T_r is the time of the r^{th} light bulb replacement.
- Its density is:

$$f_{T_r}(t) = \frac{1}{(r-1)!} \lambda^r t^{r-1} e^{-\lambda t}, \ t > 0 \tag{22}$$

Note: This is Gamma distribution with parameters r and λ. If r is an integer (as it is the case here), it is also called *Erlang Distribution*.

Gamma Distribution

Generalization: The function defined as

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\ dx \tag{23}$$

is called the *gamma function*. The gamma function leads to the Gamma family of distributions with pdf:

$$f_W(t) = \frac{1}{\Gamma(r)} \lambda^r t^{r-1} e^{-\lambda t},\ t > 0 \tag{24}$$

Properties of the Gamma Distribution

- Mean: r/λ
- Variance: r/λ^2
- *MGF:* $M_X(t) = \left(\frac{1}{1 - t/\lambda}\right)^\alpha$ for $t < \lambda$.
- Reduces to exponential distribution when $r = 1$.

Gamma Densities

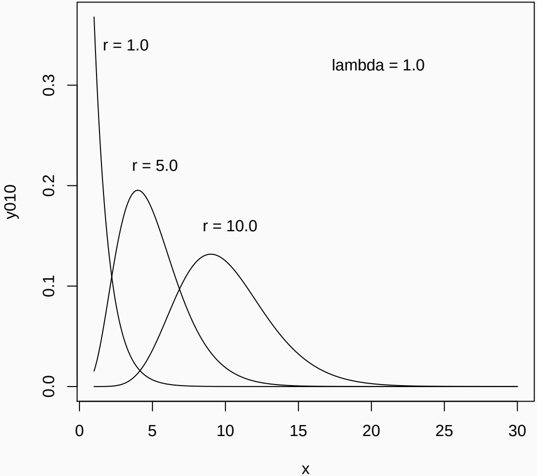


Figure 5: Various Gamma Densities

The figures are generated by running:

```
> curve(dbeta(x,0.2,0.2))
> curve(dbeta(x,2,3)
```

Example: Network Buffer

Suppose in a network context (not the ALOHA example), a node does not transmit until it has accumulated five messages in its buffer. Suppose the times between message arrivals are independent and exponentially distributed with mean 100 milliseconds. Let's find the probability that more than 552 ms will pass before a transmission is made, starting with an empty buffer.

Solution:

Let X_1 be the time until the first message arrives, X_2 the time from then to the arrival of the second message, and so on. Then the time until we accumulate five messages is $Y = X_1 + ... + X_5$.

Example: Network Buffer (continued)

Then from the definition of the gamma family, we see that Y has a gamma distribution with r = 5 and $\lambda = 0.01$. Then

$$P(Y > 552) = \int_{552}^{\infty} \frac{1}{4!} 0.01^5 t^4 e^{-0.01t}\, dt \approx 0.3544101375 \quad (25)$$

This integral could be evaluated via repeated integration by parts, but let's use R instead:

```
> 1 - pgamma(552,5,0.01)
[1] 0.3544101
```

Note that the parameter r is called **shape** in R, and λ is **rate**.

R Functions for Gamma Distribution

```
Usage:
dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE)
pgamma(q, shape, rate = 1, scale = 1/rate,
  lower.tail = TRUE, log.p = FALSE)
qgamma(p, shape, rate = 1, scale = 1/rate,
  lower.tail = TRUE, log.p = FALSE)
rgamma(n, shape, rate = 1, scale = 1/rate)
```

Importance in Modeling

- Has *applications associated with intervals between events*: Sums of exponentially distributed random variables often arise in applications and such sums have gamma distributions.
- Gamma distributions useful for approximating certain data sets.
- Specific applications include *queuing models, the flow of items through manufacturing and distribution processes, and the load on web servers and the many and varied forms of telephone exchanges*.
- Also, owing to its moderately skewed pdf, it can be used as a probability model in a range of disciplines, including *climatology, where it is a workable model for rainfall, and financial services, where it has been used for modeling insurance claims and the size of loan defaults*.

Beta Distribution

As seen in Figure 5, the gamma family is a good choice to consider if our data are nonnegative, with the density having a peak near 0 and then gradually tapering off to the right. What about data in the range (0,1)?

For instance, say a trucking company transports many things, including furniture. Let X be the proportion of a truckload that consists of furniture. For instance, if 15% of a given truckload is furniture, then X $=$ 0.15. So here we have a distribution with support in (0,1). The beta family provides a very flexible model for this kind of setting, allowing us to model many different concave up or concave down curves.

Density Etc.

The densities of the family have the following form:

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1} \text{ for } 0 < t < 1 \tag{26}$$

There are two parameters, α and β.

For $X \sim \text{Beta}(\alpha, \beta)$,

Descriptive Measures:

Mean: $E[X] = \frac{\alpha}{\alpha+\beta}$,

Variance: $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$,

MGF: $M_X(t)$ is not in a nice closed form.

Beta Densities - I

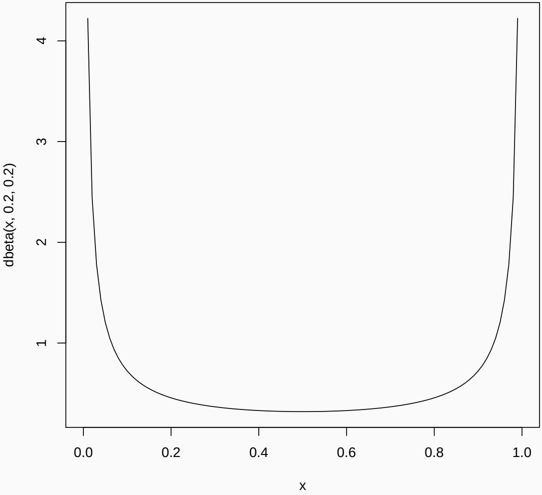


Figure 6: Beta Density, $\alpha = 0.2, \beta = 0.2$

Beta Densities - II

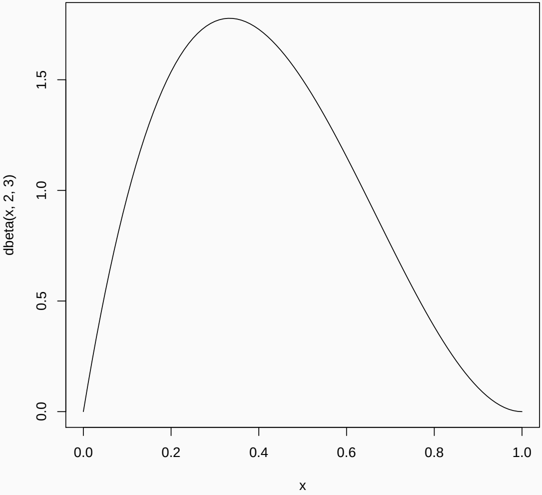


Figure 7: Beta Density, $\alpha = 2.0, \beta = 3.0$

Functions in R

Again, there are also **dbeta()**, **qbeta()** and **rbeta()**. From the R man page:

```
Usage:
dbeta(x, shape1, shape2, ncp = 0, log = FALSE)
pbeta(q, shape1, shape2, ncp = 0,
      lower.tail = TRUE, log.p = FALSE)
qbeta(p, shape1, shape2, ncp = 0,
      lower.tail = TRUE, log.p = FALSE)
rbeta(n, shape1, shape2, ncp = 0)
```

Importance in Modeling

As mentioned, the beta family is a natural candidate for modeling a variable having range the interval (0,1). This family is also popular among **Bayesian** statisticians.

Applications:

1. It has been used as a statistical description of allele frequencies in population genetics;
2. Time allocation in project management/control systems;
3. Variability of soil properties;
4. Proportions of the minerals in rocks in stratigraphy.