

Chapter 5. Continuous Probability Distributions

5.1. Introduction

In this chapter, we introduce some of the common probability density functions (PDFs) for continuous sample spaces. The goal of this section is to become familiar with these PDFs and to be able to easily evaluate quantities associated with the PDFs. When given a word problem, we will also learn how to identify which PDF is appropriate.

This chapter is organized with each section describing a different PDF, including continuous, normal, Student's t, gamma, exponential, chi-squared and F distributions. To be clear this is simply a small set of continuous PDFs with common application in the sciences and in statistics. A final section discusses functions of random variables.

5.2. Continuous Uniform Distribution

The continuous uniform distribution describes a sample space where every point within the space is equally (uniformly) likely.

The density function of the continuous uniform random variable X on the interval $[A,B]$ is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & \text{for } A \leq x \leq B \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

For any of the continuous PDF, the population mean of any function of a random variable, $g(x)$, is given by the definition first provided in Chapter 3.

$$\mu_{g(x)} \equiv \int_{-\infty}^{\infty} g(x)f(x)dx \quad (3.X)$$

The application of this definition to the mean of x for the continuous uniform PDF leads to

$$\mu_x = \int_A^B x \frac{1}{B-A} dx = \frac{1}{B-A} \int_A^B x dx = \frac{1}{B-A} \left[\frac{x^2}{2} \right]_A^B = \frac{1}{B-A} \frac{B^2 - A^2}{2} \quad (5.2)$$

Further simplification leads to an intuitive formula for the mean of x for the continuous uniform PDF

$$\mu_x = \frac{B+A}{2} \quad (5.3)$$

For any of the continuous PDF, the population variance of any function of a random variable, $g(x)$, is given by the definition first provided in Chapter 3.

$$\sigma_{g(x)}^2 \equiv \int_{-\infty}^{\infty} [g(x) - \mu_{g(x)}]^2 f(x) dx \quad (3.Y)$$

The application of this definition to the variance of x for the continuous uniform PDF leads to

$$\sigma_x^2 = \int_A^B (x - \mu_x)^2 \frac{1}{B-A} dx = \frac{1}{B-A} \left[\frac{(x - \mu_x)^3}{3} \right]_A^B = \frac{1}{B-A} \left[\frac{(B - \mu_x)^3}{3} - \frac{(A - \mu_x)^3}{3} \right] \quad (5.4)$$

Further simplification leads to a formula for the variance of x for the continuous uniform PDF

$$\sigma_x^2 = \frac{(B-A)^2}{12} \quad (5.5)$$

The Continuous Uniform Probability Distribution in MATLAB

There are a variety of useful tools for dealing with continuous distributions in MATLAB. A few of them are reviewed here with respect to their application to the continuous uniform distribution.

In order to generate a random number from the continuous uniform PDF defined from A to B , one can use the `random` function in MATLAB,

```
>> x = random('Uniform',A,B)
```

For example, to generate a random number from the continuous uniform PDF defined from -1 to 1, one can type the command,

```
>> x = random('Uniform',-1,1)
```

which yields an output such as

```
x = -0.804919190001181
```

In order to evaluate a PDF at a given value of x , one can use the `pdf` function in MATLAB

```
>> f = pdf('Uniform',x,A,B)
```

For example, to generate the value of the PDF for the continuous uniform PDF defined from -1 to 1 at $x = 0.3$, one can type the command,

```
>> f = pdf('Uniform',0.3,-1,1)
```

which yields the following output

```
f = 0.5000000000000000
```

In order to evaluate a cumulative PDF at a given value of x , one can use the `cdf` function in MATLAB. If we are interested in the probability that $x \leq x_{hi}$, then the appropriate function is the cumulative distribution function.

$$F = p(x \leq x_{hi}) \equiv \int_{-\infty}^{x_{hi}} f(x) dx$$

In MATLAB, we can directly evaluate the cumulative distribution function for a number of common PDFs, including all of the continuous PDFs studies in this course.

```
>> F = cdf('Uniform',xhi,A,B)
```

For example, to generate the value of the cumulative PDF for the continuous uniform PDF defined from -1 to 1 at $x = 0.3$, one can type the command,

```
>> F = cdf('Uniform',0.3,-1,1)
```

which yields the following output

```
F = 0.6500000000000000
```

In other words,

$$F = p(x \leq 0.3) \equiv \int_{-\infty}^{0.3} f(x) dx = 0.65$$

Above we were given x_{hi} and asked to find the probability, p . Frequently, we are asked the inverse problem, where we are given, p , and asked to find the value of x_{hi} . In order to determine the value of a random variable x at which the cumulative PDF of the continuous uniform distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> xhi = icdf('Uniform',F,A,B)
```

For example, to determine the value of a random variable x at which the cumulative uniform PDF, defined from -1 to 1, has a value of 0.6, one can type the command,

```
>> xhi = icdf('Uniform',0.6,-1,1)
```

which yields the following output

```
xhi = 0.2000000000000000
```

In other words,

$$F = 0.6 = p(x \leq 0.2) \equiv \int_{-\infty}^{0.2} f(x) dx$$

Example 5.1: The default random number generators for most programming languages provides a uniform distribution on the interval $[0,1]$. What is the probability that a MATLAB random number generator yields (a) $X=0.25$ (b) $X<0.25$, (c) $x \leq 0.25$, (d) $X>0.25$, (e) $0.1 < X < 0.25$?

- (a) The probability of getting exactly any number in a continuous sample space is zero.
- (b) Here we use the definition of a continuous PDF.

$$P(X < 0.25) = P(0 < X < 0.25) = \int_0^{0.25} f(x) dx = \int_0^{0.25} \frac{1}{1-0} dx = 0.25$$

Alternatively, we can solve this problem in MATLAB.

```
>> F = cdf('Uniform',0.25,0,1)
```

```
F = 0.2500000000000000
```

(c) The answer for (c) is the same as the answer for (b).

$$P(X \leq 0.25) = P(X < 0.25) + P(X = 0.25) = P(X < 0.25) = 0.25$$

(d) The continuous uniform PDF is normalized. So, the answer to (d) is given by

$$P(X > 0.25) = 1 - P(X < 0.25) = 1 - 0.25 = 0.75$$

(e)

$$P(0.1 < X < 0.25) = \int_{0.1}^{0.25} f(x) dx = \int_{0.1}^{0.25} \frac{1}{1-0} dx = 0.15$$

Alternatively, we can solve this problem in MATLAB.

```
>> Fhi = cdf('Uniform', 0.25, 0, 1);
>> Flo = cdf('Uniform', 0.1, 0, 1);
>> p = Fhi - Flo
```

```
p = 0.15000000000000000
```

5.3. Normal Distribution

The normal distribution is one of the most important continuous PDF in statistics and in science. It is also known as the Gaussian distribution and many other application-specific names. For example, the Maxwell-Boltzmann distribution is the name given to the normal distribution when it is applied to the distribution of atomic velocities in a material. The shape of the normal distribution has given rise to the nickname “the bell curve”.

The probability density function of the normal random variable, x , with mean μ and variance σ^2 , is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (5.6)$$

You will notice immediately that this distribution is different than the PDF's we've studied previously. The normal PDF is defined by two parameters, the mean and variance, μ and σ^2 . In our previous cases, the PDF defined μ and σ^2 . The effect of the mean on the normal PDF is shown in Figure 5.1. Clearly the normal distribution is centered around the mean.

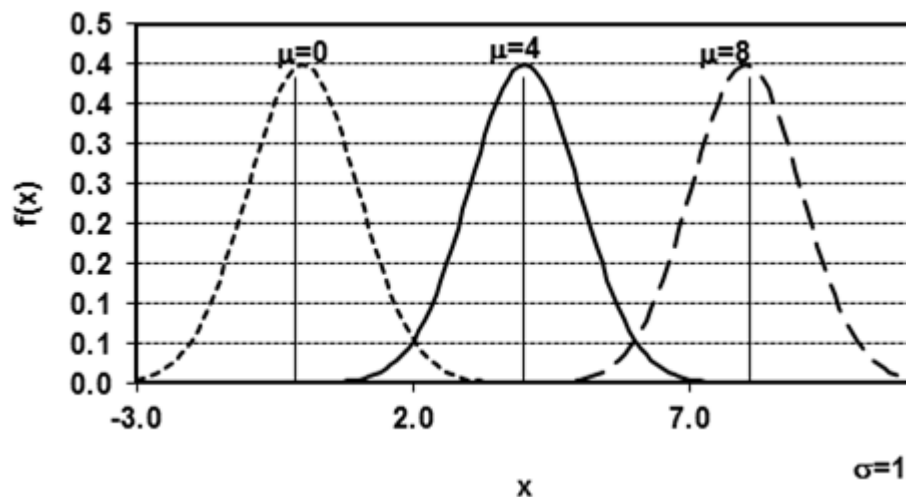


Figure 5. The effect of mean on the normal distribution with a variance of one.

The effect of the variance on the normal PDF is shown in Figure 5.2. Clearly, the normal distribution grows broader as the variance increases.

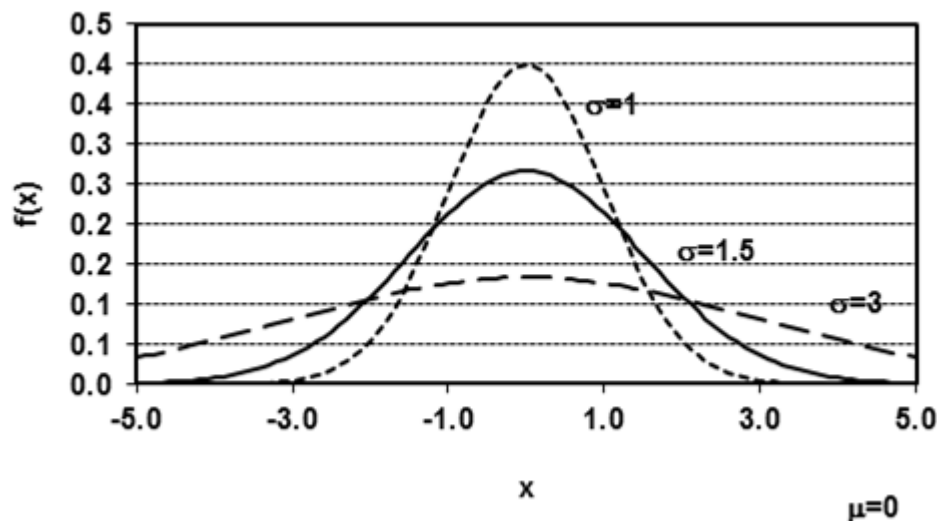


Figure 5.2. The effect of variance on the normal distribution with a mean of zero.

Some characteristics of the Normal PDF include

- The mode, (the most probable value occurs at the mean).
- The distribution is symmetric about the mean.
- The curve has inflection point at $X = \mu \pm \sigma$. (Remember: inflection points are where the second derivative is zero, where the curve changes from concave up to concave down.
- The normal curve approaches the x-axis as we move from the mean.
- The total area under the curve and above the x-axis is one (as it is for all PDF's.)

• The probability $P(a < X < b) = \int_a^b f(x)dx$ is the area under the normal curve between a and b .

• The Normal PDF with $\mu = 0$ and $\sigma = 1$ is called the **Standard Normal PDF**.

• Any Normal PDF, $f(x; \mu, \sigma)$, can be converted to a standard normal PDF, $f(z; 0, 1)$, with the change of variable $z = \frac{x - \mu}{\sigma}$.

• The integral of the normal PDF in equation (5.6) cannot be analytically evaluated.

• The normal PDF $f(x; \mu, \sigma)$ is equal to the binomial PDF $b(x; n, p) = \binom{n}{x} p^x q^{n-x}$ with $\mu = np$ and $\sigma = npq$ in the limit when $n \rightarrow \infty$.

Since the PDF cannot be analytically integrated, tables of integrals for this commonly used distribution are widely available. Typically, the tables are given for the cumulative PDF,

$$P(x \leq a) = F(a) = \int_{-\infty}^a f(x; \mu, \sigma) dx = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (5.7)$$

When the pioneers were crossing the plains in their covered wagons and they wanted to evaluate probabilities from the normal distribution, they used tables of the cumulative normal PDF, such as those provided in the back of the Walpole, Myers, Myers and Ye textbook. These tables are also available online. For example wikipedia has a table of cumulative standard normal PDFs at

http://en.wikipedia.org/wiki/Standard_normal_table

These tables provide values of the cumulative standard normal PDF, $f(z; 0, 1)$, for discretized values of the random variable, z .

The Normal Distribution in MATLAB

Alternatively, one can use tools in MATLAB to evaluate problems concerned with the normal distribution. In order to generate a random number from the normal PDF defined by mean, m , and standard deviation, s , one can use the `random` function in MATLAB,

```
>> x = random('normal', m, s)
```

For example, to generate a random number from the normal PDF defined by mean and standard deviation, 0 and 1, one can type the command,

```
>> x = random('normal',0,1)
```

which yields an output such as

```
x = -2.258846861003648
```

In order to evaluate a PDF at a given value of x , one can use the `pdf` function in MATLAB

```
>> f = pdf('normal',x,m,s)
```

For example, to generate the value of the PDF for the normal PDF defined by mean and standard deviation, 0 and 1 at $x = 0.3$, one can type the command,

```
>> f = pdf('normal',0.3,0,1)
```

which yields the following output

```
f = 0.381387815460524
```

In order to evaluate a cumulative PDF at a given value of x , one can use the `cdf` function in MATLAB

```
>> F = cdf('normal',xhi,m,s)
```

For example, to generate the value of the cumulative PDF for the normal PDF defined by mean and standard deviation, 0 and 1 at $x = 0.3$, one can type the command,

```
>> F = cdf('normal',0.3,0,1)
```

which yields the following output

```
F = 0.617911422188953
```

In other words,

$$F = p(x \leq 0.3) \equiv \int_{-\infty}^{0.3} f(x) dx = 0.61791$$

In order to determine the value of a random variable x at which the cumulative PDF of the normal distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> xhi = icdf('normal',F,m,s)
```


For example, to determine the value of the random variable x at which the cumulative normal PDF with mean and standard deviation, 0 and 1, has a value of 0.6, one can type the command,

```
>> xhi = icdf('normal',0.6,0,1)
```

which yields the following output

```
xhi =    0.253347103135800
```

In other words,

$$F = 0.6 = p(x \leq 0.25335) = \int_{-\infty}^{0.25335} f(x) dx$$

Example 5.2: A type of battery lasts on the average 3.0 years with a standard deviation of 0.5 years. What is the probability that the battery will last (a) less than 2.3 years? (b) between 2.5 and 3.5 years? (c) more than four years?

(a) If we solve this problem manually, the first step is to convert the non-standard normal PDF to the standard normal PDF via the transformation

$$z_{2.3} = \frac{x - \mu}{\sigma} = \frac{2.3 - 3.0}{.5} = -1.4$$

One can then go visit a standard normal table and look up the value corresponding to $z = -1.4$.

$$P(x \leq 2.3) = P(z \leq -1.4) = 0.0808$$

Alternatively, one can solve the problem in MATLAB.

```
>> F = cdf('normal',2.3,3,0.5)
```

```
F =    0.080756659233771
```

(b) Manually, we transform both limits.

$$z_{2.5} = \frac{x - \mu}{\sigma} = \frac{2.5 - 3.0}{.5} = -1.0, \quad z_{3.5} = \frac{x - \mu}{\sigma} = \frac{3.5 - 3.0}{.5} = 1.0$$

We use the cumulative table twice and take the difference.

$$P(2.5 \leq X \leq 3.5) = P(-1.0 \leq Z \leq 1.0) = P(Z \leq 1.0) - P(Z \leq -1.0) \\ = 0.8413 - 0.1587 = 0.6826$$

Alternatively, one can solve the problem in MATLAB.

```
>> Fhi = cdf('normal', 3.5, 3, 0.5);
>> Flo = cdf('normal', 2.5, 3, 0.5);
>> p = Fhi - Flo
```

```
p = 0.682689492137086
```

In this problem, Fhi provides the probability less than 3.5 and Flo provides the probability less than 2.5. The difference of these two probabilities is the probability that x lies between 2.5 and 3.5.

(c) In order to find the probability that a battery lasts more than four years,

$$z_{4.0} = \frac{x - \mu}{\sigma} = \frac{4.0 - 3.0}{.5} = 2.0$$

$$P(4.0 \geq X) = P(2.0 \geq Z) = 1 - P(2.0 \leq Z) = 1 - 0.9772 = 0.0228$$

Alternatively, one can solve the problem in MATLAB.

```
>> p = 1 - cdf('normal', 4.0, 3, 0.5)
```

```
p = 0.022750131948179
```

In other words, the probability that a battery lasts more than 4.0 years, is 2.28% when the lifetime of this battery is given by a normal distribution with mean 3 years and variance 0.5 years.

5.4. Student's t Distribution

The student's t distribution or simply the t distribution is a generalization of the normal distribution. It is defined as

$$f(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} \quad \text{for} \quad -\infty < t < \infty \quad (5.8)$$

where ν is a parameter, called the degrees of freedom, and the gamma function is a standard mathematical function defined as .

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \text{for } \alpha > 0 \quad (5.9)$$

This definite integral cannot be analytically evaluated. The Gamma function has some special values: $\Gamma(1) = 1$, $\Gamma(n) = (n-1)!$ where n is a positive integer, and $\Gamma(1/2) = \pi$. The Gamma function can be evaluated in MATLAB via

```
>> g = gamma(0.5)
```

```
g =  
1.772453850905516
```

In the limit that the degrees of freedom approaches infinity, the t distribution approaches the normal distribution. One may wonder how this is possible given that the t distribution does not explicitly have the mean and a variance as parameters as does the normal distribution, but we shall save that discussion until the next chapter on sampling.

The t distribution is shown in Figure 5.3. Like the normal distribution it is symmetric about a mean. The random variable t ranges from negative to positive infinity.

When the cavemen were evaluating probabilities from the t distribution with charcoal on the walls of their caves, they used tables of the cumulative t distribution PDF, such as those provided in the back of the statistics textbooks. These tables are also available online. Here we will rely on a modern, computational approach to evaluating the t distribution.

The t Distribution in MATLAB

One can use tools in MATLAB to evaluate problems concerned with the t distribution. In order to generate a random number from the t PDF defined by degrees of freedom, ν , one can use the random function in MATLAB,

```
>> x = random('t',v)
```

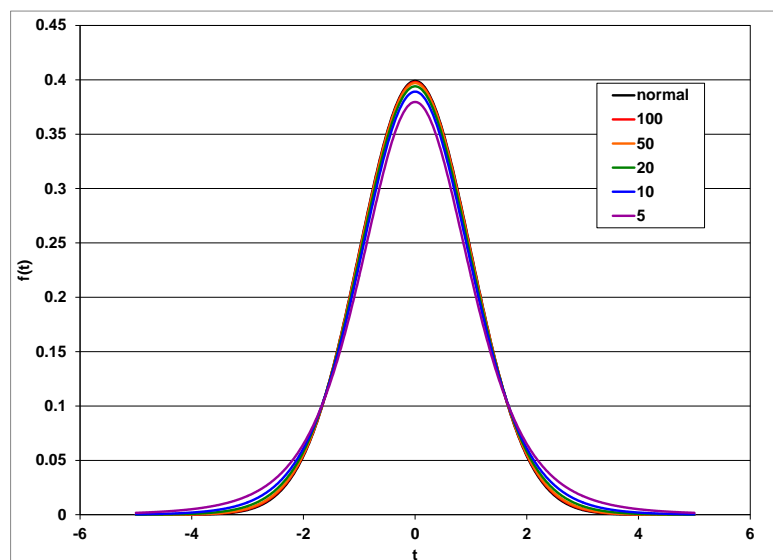


Figure 5.3. The t -distribution as a function of the degrees of freedom and the normal distribution.

For example, to generate a random number from the t PDF defined by degrees of freedom, 15, one can type the command,

```
>> x = random('t',15)
```

which yields an output such as

```
x = 0.902666218739684
```

In order to evaluate a PDF at a given value of t , one can use the `pdf` function in MATLAB

```
>> f = pdf('t',t,v)
```

For example, to generate the value of the PDF for the t PDF defined by degree of freedom, 15, at $t = 0.3$, one can type the command,

```
>> f = pdf('t',0.3,15)
```

which yields the following output

```
f = 0.374018696345112
```

In order to evaluate a cumulative PDF at a given value of t , one can use the `cdf` function in MATLAB

```
>> F = cdf('t',thi,v)
```

For example, to generate the value of the cumulative PDF for the t PDF defined by degree of freedom, 15, at $t = 0.3$, one can type the command,

```
>> F = cdf('t',0.3,15)
```

which yields the following output

```
F = 0.615852732732224
```

In other words,

$$F = p(x \leq 0.3) = \int_{-\infty}^{0.3} f(x) dx = 0.61583$$

In order to determine the value of a random variable t at which the cumulative PDF of the t distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> thi = icdf('t',F,v)
```

For example, to determine the value of a random variable t at which the cumulative t PDF, with degrees of freedom of 15, has a value of 0.6, one can type the command,

```
>> thi = icdf('t',0.6,15)
```

which yields the following output

```
thi = 0.257885300937261
```

In other words,

$$F = 0.6 = p(x \leq 0.25789) = \int_{-\infty}^{0.25789} f(x) dx$$

We shall postpone an example of applying the t distribution until the next chapter, where we demonstrate its use in statistical sampling.

5.5. Gamma Distribution

The Gamma distribution commonly appears in the sciences. The Gamma distribution for the continuous random variable x with parameters α and β is given by

$$f_{\Gamma}(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (5.10)$$

where the gamma function, defined in equation (5.9) again appears in the distribution.

The Gamma distribution is shown in Figure 5.4. The gamma distribution is only defined for values of x greater than zero. Increasing α while holding β constant causes the peak in the distribution to move to larger values of x and causes the magnitude of the peak to decrease.

Increasing β while holding α constant also causes the peak in the distribution to move to larger values of x and causes the magnitude of the peak to decrease.

One can use the definitions to explicitly evaluate the mean variance of the random variable x in the gamma distribution. Only the results are given here.

$$\mu = \alpha\beta \quad (5.11)$$

$$\sigma^2 = \alpha\beta^2 \quad (5.12)$$

The Gamma Distribution in MATLAB

One can use tools in MATLAB to evaluate problems concerned with the gamma distribution. In order to generate a random number from the gamma PDF defined by parameters, α and β , one can use the random function in MATLAB,

```
>> x =
random('gamma', a, b)
```

For example, to generate a random number from the gamma PDF defined by parameters, $\alpha = 2$ and $\beta = 3$, one can type the command,

```
>> x = random('gamma', 2, 3)
```

which yields an output such as

```
x = 8.310470316364883
```

In order to evaluate a PDF at a given value of x , one can use the pdf function in MATLAB

```
>> f = pdf('gamma', x, a, b)
```

For example, to generate the value of the PDF for the gamma PDF defined by parameters, $\alpha = 2$ and $\beta = 3$, at $x = 1$, one can type the command,

```
>> f = pdf('gamma', 1, 2, 3)
```

which yields the following output

```
f = 0.079614590063754
```

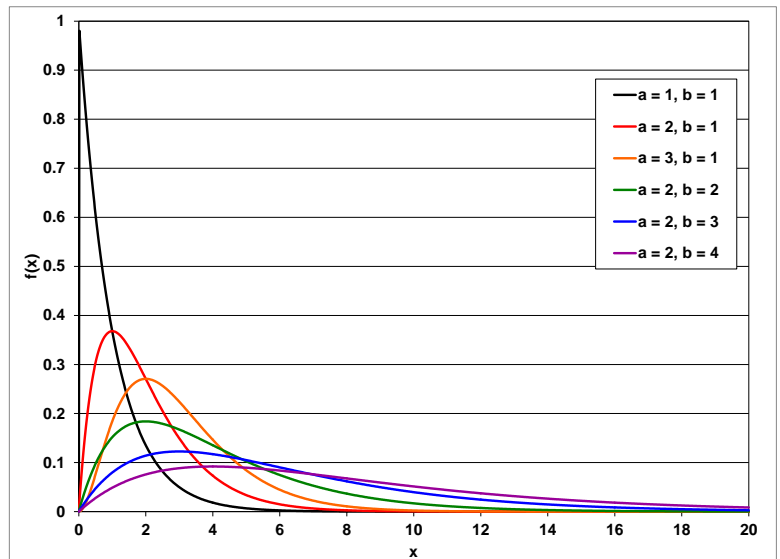


Figure 5.4. The gamma distribution for various values of α and β .

In order to evaluate a cumulative PDF at a given value of x , one can use the `cdf` function in MATLAB

```
>> F = cdf('gamma', xhi, a, b)
```

For example, to generate the value of the cumulative gamma PDF defined by parameters, $\alpha = 2$ and $\beta = 3$, at $x = 5.3$, one can type the command,

```
>> F = cdf('gamma', 5.3, 2, 3)
```

which yields the following output

```
F = 0.527171927915781
```

In other words,

$$F = p(x \leq 5.3) = \int_{-\infty}^{5.3} f(x) dx = 0.52717$$

In order to determine the value of a random variable x at which the cumulative PDF of the gamma distribution has a value of F , one can use the `icdf` (the inverse `cdf`) function in MATLAB

```
>> xhi = icdf('gamma', F, a, b)
```

For example, to determine the value of a random variable x at which the cumulative gamma PDF, with parameters, $\alpha = 2$ and $\beta = 3$, has a value of 0.6, one can type the command,

```
>> xhi = icdf('gamma', 0.6, 2, 3)
```

which yields the following output

```
xhi = 6.066939735973971
```

In other words,

$$F = 0.6 = p(x \leq 6.06694) = \int_{-\infty}^{6.06694} f(x) dx$$

In an example problem, you more or less have to be told to apply the gamma distribution.

5.6. Exponential Distribution

The exponential distribution is the gamma distribution when α is set to one.. The exponential distribution for the continuous random variable x with parameters β is given by

$$f_e(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (5.13)$$

The exponential distribution is shown in Figure 5.5. The exponential distribution is only defined for values of x greater than zero. Increasing β causes decay of this exponential curve to lengthen.

One can use the definitions to explicitly evaluate the mean variance of the random variable x in the exponential distribution. Only the results are given here.

$$\mu = \beta \quad (5.14)$$

$$\sigma^2 = \beta^2 \quad (5.15)$$

The Exponential Distribution in MATLAB

One can use tools in MATLAB to evaluate problems concerned with the exponential distribution. In order to generate a random number from the exponential PDF defined by parameter, β , one can use the random function in MATLAB,

```
>> x =
random('exp', b)
```

For example, to generate a random number from the exponential PDF defined by parameter, $\beta = 3$, one can type the command,

```
>> x = random('exp', 3)
```

which yields an output such as

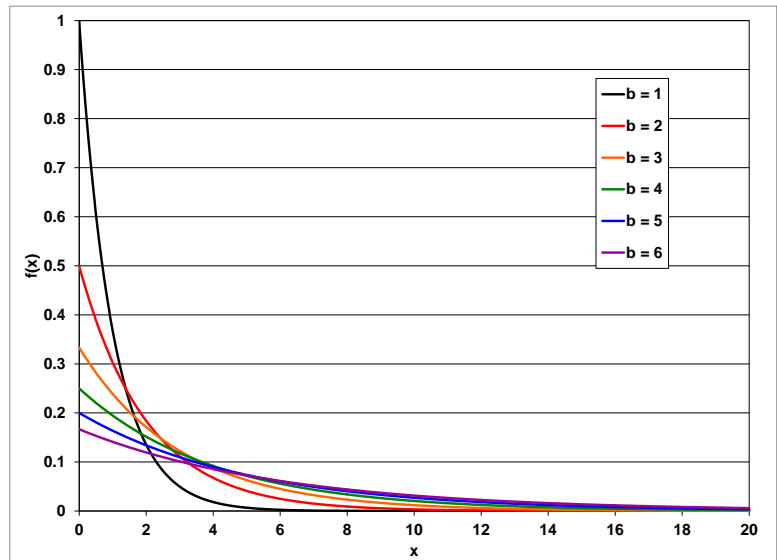


Figure 5.5. The exponential distribution for various values of β .


```
x = 5.858186922301461
```

In order to evaluate a PDF at a given value of x , one can use the `pdf` function in MATLAB

```
>> f = pdf('exp',x,b)
```

For example, to generate the value of the PDF for the exponential PDF defined by parameter, $\beta=3$, at $x=1$, one can type the command,

```
>> f = pdf('exp',1,3)
```

which yields the following output

```
f = 0.238843770191263
```

In order to evaluate a cumulative PDF at a given value of x , one can use the `cdf` function in MATLAB

```
>> F = cdf('exp',xhi,b)
```

For example, to generate the value of the cumulative exponential PDF defined by parameter, $\beta=3$, at $x=5.3$, one can type the command,

```
>> F = cdf('exp',5.3,3)
```

which yields the following output

```
F = 0.829098287198475
```

In other words,

$$F = p(x \leq 5.3) = \int_{-\infty}^{5.3} f(x) dx = 0.82910$$

In order to determine the value of a random variable x at which the cumulative PDF of the exponential distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> xhi = icdf('exp',F,b)
```

For example, to determine the value of a random variable x at which the cumulative exponential PDF, with parameter, $\beta = 3$, has a value of 0.6, one can type the command,

```
>> xhi = icdf('exp',0.6,3)
```

which yields the following output

```
xhi = 3.105378597263349
```

In other words,

$$F = 0.6 = p(x \leq 3.10538) = \int_{-\infty}^{3.10538} f(x)dx$$

Example 5.3: The lifetime of sparkplugs is measure in time, t , and is modelled by the exponential distribution with an average time to failure of 5 years, $\beta = 5$. If new sparkplugs are installed in an 8-cylinder engine and never replaced, what is the probability that m sparkplugs are still alive at the end of t_i years where $0 \leq m \leq 8$ and $t_i = 1, 2.5, 5, 7.5, 10$. The sparkplugs are independent of each other.

The probability that a single independent spark plug is still alive at the end of t_i years is given by:

$$P(t_i < t) = \int_{t_i}^{\infty} f_e(t; \beta) dt$$

$$= \int_{t_i}^{\infty} \frac{1}{\beta} e^{-t/\beta} dt = e^{-t_i/\beta}$$

We can compute this for any desired value of t_i . Now, we need that probability that m of $n=8$ sparkplugs are still functioning at t_i , given the probability above. This is precisely the function of the binomial distribution, $b(x;n,p)$ where $x = m$, the number of functioning sparkplugs, $n=8$, the number of total sparkplugs; and $p = P(t_i < t)$, the probability that a single sparkplug makes it to

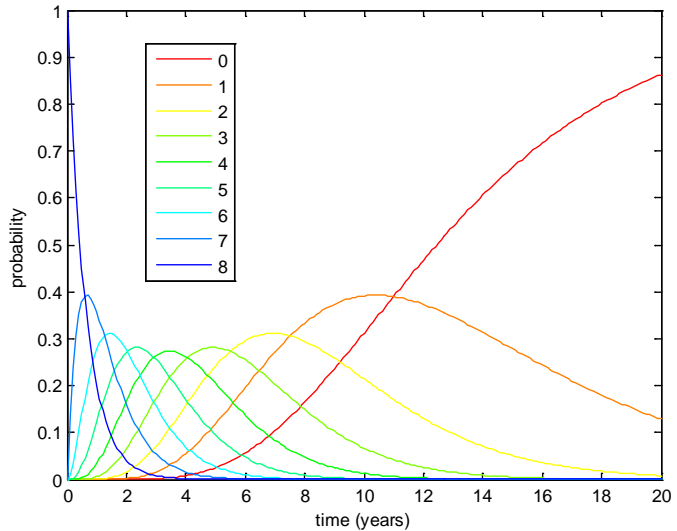


Figure 5.6. The probability that m sparkplugs still work as a function of time, given that the mean sparkplug lifetime is 5 years.

time t_i , as given by the exponential distribution. We can calculate $b(m, 8, P(t_i < t))$ for all values of m , (namely $0 \leq m \leq 8$) and for several values of t_i .

In Figure 5.6, we show the probability that m sparkplugs still work as a function of time. Let's examine the plot. At any time t , the sum of the probabilities is 1. At time near 0, it is most probable that all 8 sparkplugs still function. At six years, it is most probable that only 2 sparkplugs still function, followed by 3, 1, 4, 0, 5, 6, 7, 8. At twelve years, it is most probable that no sparkplugs function anymore.

5.7. Chi-Squared Distribution

A second special case of the gamma distribution occurs when $\alpha = v/2$ and $\beta = 2$. This distribution is called the chi-squared distribution and v is called the degrees of freedom. The chi-squared distribution for the continuous random variable x with parameter v is given by

$$f_{\chi^2}(x; v) = \begin{cases} \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (5.16)$$

The chi-squared distribution is shown in Figure 5.7. The chi-squared distribution is only defined for values of x greater than zero. Increasing v causes the peak to move out to larger values of the random variable and to decrease in magnitude.

One can use the definitions to explicitly evaluate the mean variance of the random variable x in the chi-squared distribution. Only the results are given here.

$$\mu = v \quad (5.17)$$

$$\sigma^2 = 2v \quad (5.18)$$

The Chi-Squared Distribution in MATLAB

One can use tools in MATLAB to evaluate problems concerned with the chi-squared distribution. In order to generate a random number from the chi-squared PDF defined by parameter, β , one can use the random function in MATLAB,

```
>> x =
random('chi2', v)
```

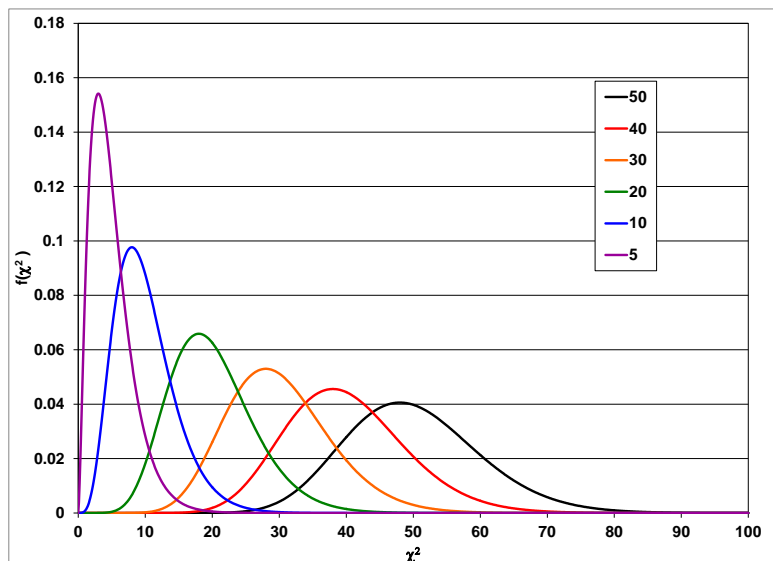


Figure 5.7. The chi-squared distribution for various values of v .

For example, to generate a random number from the chi-squared PDF defined by parameter, $v=13$, one can type the command,

```
>> x = random('chi2',13)
```

which yields an output such as

```
x = 21.310077394704155
```

In order to evaluate a PDF at a given value of x , one can use the `pdf` function in MATLAB

```
>> f = pdf('chi2',x,v)
```

For example, to generate the value of the PDF for the chi-squared PDF defined by parameter, $v=13$, at $x = 10$, one can type the command,

```
>> f = pdf('chi2',10,13)
```

which yields the following output

```
f = 0.081773608489854
```

In order to evaluate a cumulative PDF at a given value of x , one can use the `cdf` function in MATLAB

```
>> F = cdf('chi2',xhi,v)
```

For example, to generate the value of the cumulative chi-squared PDF defined by parameter, $v = 13$, at $x = 10$, one can type the command,

```
>> F = cdf('chi2',10,13)
```

which yields the following output

```
F = 0.306065632019251
```

In other words,

$$F = p(x \leq 10) = \int_{-\infty}^{10} f(x) dx = 0.30607$$

In order to determine the value of a random variable x at which the cumulative PDF of the chi-squared distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> xhi = icdf('chi2',F,v)
```

For example, to determine the value of a random variable x at which the cumulative chi-squared PDF, with parameter, $v = 13$, has a value of 0.6, one can type the command,

```
>> xhi = icdf('chi2',0.6,13)
```

which yields the following output

```
xhi = 13.635570993661943
```

In other words,

$$F = 0.6 = p(x \leq 13.63557) = \int_{-\infty}^{13.63557} f(x) dx$$

We shall postpone an example of applying the chi-squared distribution until the next chapter, where we demonstrate its use in statistical sampling.

5.8. F distribution

The last distribution of this chapter is the F distribution, which is used in statistical analysis. The F distribution for the continuous random variable x with parameters v_1 and v_2 is given by

$$h_f(f; v_1, v_2) = \begin{cases} \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)} \frac{f^{\frac{v_1}{2}-1}}{\left(1 + \frac{v_1}{v_2} f\right)^{\frac{v_1+v_2}{2}}} & \text{for } f > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (5.19)$$

The F distribution is shown in Figure 5.8. The F distribution is only defined for values of x greater than zero. Increasing v_1 while holding v_2 constant causes the peak to move out to larger values of the random variable and to increase in magnitude. Increasing v_2 while holding v_1 constant also causes the peak in the distribution to move out to larger values of the random variable and to increase in magnitude.

The F Distribution in MATLAB

One can use tools in MATLAB to evaluate problems concerned with the F distribution. In order to generate a random number from the F PDF defined by parameter, β , one can use the random function in MATLAB,

```
>> x = random('f',v1,v2)
```

For example, to generate a random number from the F PDF defined by parameters, $v_1 = 13$ and $v_2 = 17$, one can type the command,

```
>> x =
random('f',13,17)
```

which yields an output such as

```
x =
2.625364477079602
```

In order to evaluate a PDF at a given value of x , one can use the pdf function in MATLAB

```
>> f =
pdf('f',x,v1,v2)
```

For example, to generate the value of the PDF for the F PDF defined by parameters, $v_1 = 13$ and $v_2 = 17$, at $x = 1$, one can type the command,

```
>> f = pdf('f',1,13,17)
```

which yields the following output

```
f = 0.752702954780597
```

In order to evaluate a cumulative PDF at a given value of x , one can use the cdf function in MATLAB

```
>> F = cdf('f',xhi,v1,v2)
```

For example, to generate the value of the cumulative F PDF defined by parameters, $v_1 = 13$ and $v_2 = 17$, at $x = 1$, one can type the command,

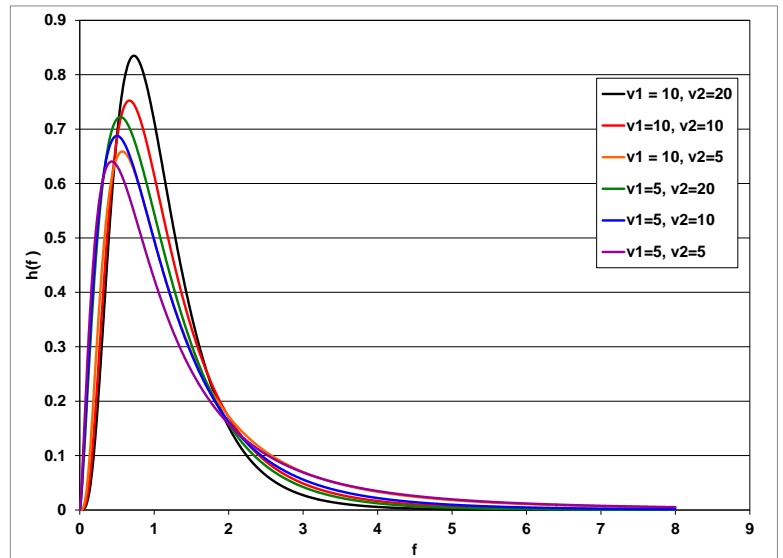


Figure 5.8. The F distribution for various values of v_1 and v_2 .

```
>> F = cdf('f',1,13,17)
```

which yields the following output

```
F = 0.509323073450741
```

In other words,

$$F = p(x \leq 1) = \int_{-\infty}^{10} f(x) dx = 0.50932$$

In order to determine the value of a random variable x at which the cumulative PDF of the F distribution has a value of F , one can use the `icdf` (the inverse cdf) function in MATLAB

```
>> xhi = icdf('f',F,v1,v2)
```

For example, to determine the value of a random variable x at which the cumulative F PDF, with parameters, $v_1 = 13$ and $v_2 = 17$, has a value of 0.6, one can type the command,

```
>> xhi = icdf('f',0.6,13,17)
```

which yields the following output

```
xhi = 1.129264291748036
```

In other words,

$$F = 0.6 = p(x \leq 1.12926) = \int_{-\infty}^{1.12926} f(x) dx$$

We shall postpone an example of applying the F distribution until the next chapter, where we demonstrate its use in statistical sampling.

5.9. Functions of Random Variables

Any PDF of a variable can be transformed into a PDF of a function of that variable. In general, $f(x)$ can be transformed to $g(y(x))$, where $f(x)$ is the PDF describing the distribution of the random variable x and $g(y(x))$ is the PDF describing the distribution of $y(x)$.

Example 5.4: When we changed the variable of a normal distribution, $f(x; \mu, \sigma)$, to the standard normal distribution, $f(z; \mu = 0, \sigma = 1)$, we used the transformation,

$$z(x) = \frac{x - \mu}{\sigma}$$

Thus $f(x; \mu, \sigma)$ is the PDF describing the distribution of the random variable x and $f(z; \mu = 0, \sigma = 1)$ is the PDF describing the distribution of the random variable $z(x)$.

Example 5.5: Let x be a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{x}{12} & \text{for } 1 < x < 5 \\ 0 & \text{elsewhere} \end{cases}$$

then the probability, $P(x < a)$ is given by

$$P(x < a) = \int_1^a f(x') dx' = \int_1^a \frac{x'}{12} dx' = \frac{a^2 - 1}{24}$$

We can also find the probability distribution for y , where $y(x) = 2x + 2$.

Rearranging for x yields $x = \frac{y-2}{2}$ and $dx = \frac{dy}{2}$

$$g(y) = f((y-2)/2) = \frac{(y-2)/2}{12} = \frac{y-2}{24}$$

with limits $y(x=1) = 4$ and $y(x=5) = 12$. Therefore,

$$P(x < a) = P(y(x) < y(a)) = \int_1^a f(x') dx' = \int_4^{y(a)} g(y') \frac{dy'}{2}$$

$$P(y(x) < y(a)) = \int_4^{y(a)} \frac{y' - 2}{24} \frac{dy'}{2} = \left(\frac{y'^2}{96} - \frac{4y'}{96} \right) \Big|_4^{y(a)} = \frac{y(a)^2 - 4y(a)}{96}$$

In Figure 5.9, we show $f(x)$ and $F(x)=P(x<X)$, the cumulative PDF, as a function of x , and $g(y)$ and $G(y)=P(y<Y)$, the cumulative PDF, as a function of y .

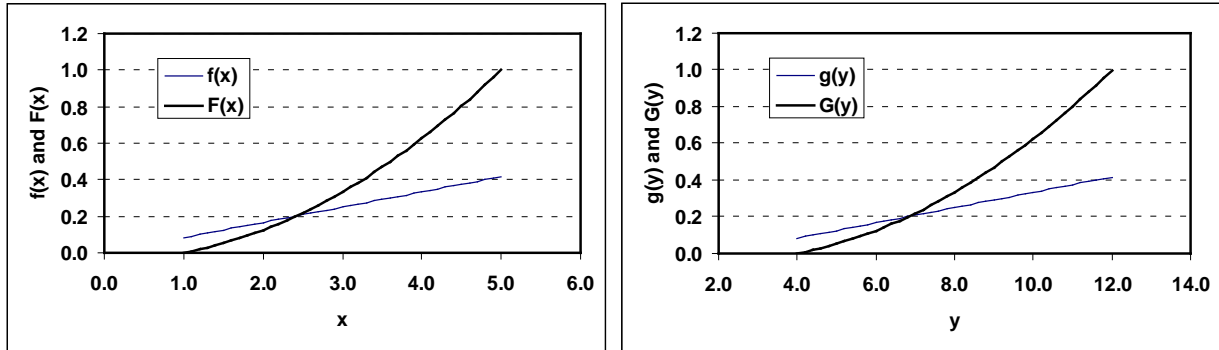


Figure 5.9. The PDF and cumulative PDF for a random variable, x , and a function of the random variable, $y(x) = 2x + 2$

Example 5.6: Consider the continuous random variable describing the temperature in a furnace, T , which is distributed between 500 K and 1000 K according to a PDF is given by $f(T) = \frac{c}{T^2}$, where $c = 1000$. The diffusivity is given as a function of temperature, $D(T) = D_o \exp\left(-\frac{E_a}{RT}\right)$ cm^2/s , where $D_o = 5 \text{ cm}^2/\text{s}$ and $E_a = 800R$. Construct a PDF in terms of the diffusivity rather than the temperature. Plot the PDF and CDF for both temperature and diffusivity.

We can invert the relationship between D and T to solve for the temperature.

$$T = \frac{E_a}{R} \frac{1}{\ln\left(\frac{D_o}{D}\right)}$$

We can substitute this into the PDF for T to obtain the functional form of the PDF for D .

$$f(D) = \frac{c}{T^2} = c \left[\frac{R}{E_a} \ln\left(\frac{D_o}{D}\right) \right]^2$$

The PDF in T was non-zero in the range $500 < T < 1000$. The PDF in D is non-zero in the range $D(500) < D < D(1000)$ or approximately, $1.009 < D < 2.247$.

Plots of both the PDF in temperature and the PDF in diffusivity are shown in Figure 5.10. The CDF is also plotted.

The cumulative PDF in T is given by

$$F(T) = \int_{-\infty}^T f(T) dT = \int_{-\infty}^T \frac{c}{T^2} dT = -c \left[\frac{1}{T} \right]_{T_{low}}^T = \frac{c}{T_{low}} - \frac{c}{T}$$

The cumulative PDF in D is obtained through substitution and is given by

$$F(D) = c \frac{R}{E_a} \ln \left(\frac{D(T)}{D(T_{low})} \right)$$

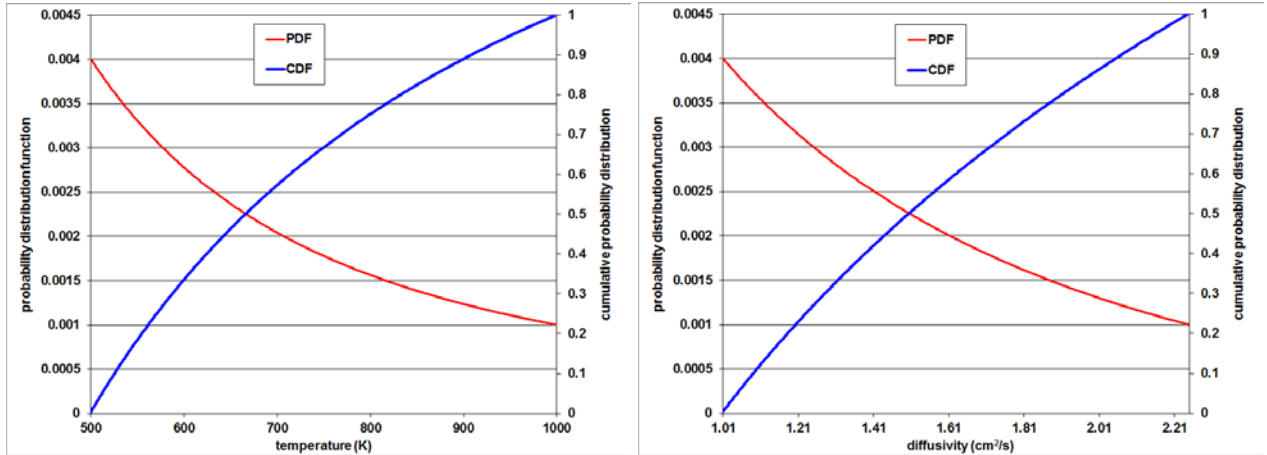


Figure 5.10. The PDF and CDF for the temperature (left) and the diffusivity (right). The curves are not exactly the same shape because the relationship was between temperature and diffusivity is nonlinear.

5.10. Problems

Problem 5.1.

Given a standard normal distribution, find the area under the curve which lies

- (a) to the left of $z = 1.43$
- (b) to the right of $z = -0.89$
- (c) between $z = -2.16$ and $z = -0.65$

Problem 5.2.

The daily amount of coffee, in liters, dispensed by a machine located in an airport lobby is a random variable X having a continuous uniform distribution with $A = 7$ and $B = 10$. Find the probability that on a given day the amount of coffee dispensed by this machine will be

- (a) at most 8.8 liters
- (b) more than 7.4 liters but less than 9.5 liters
- (c) at least 8.5 liters.

Problem 5.3.

In the November 1990 issue of Chemical Engineering Progress a study discussed the percent purity of oxygen from a certain supplier. Assume that the mean was 99.61 with a standard deviation of 0.08. Assume that the distribution of percent purity was approximately normal.

- (a) What percentage of the purity values would you expect to be between 99.5 and 99.7?
- (b) What purity value would you expect to exceed exactly 5% of the population?

Problem 5.4.

The average life of a certain type of small motor is 10 years with a standard deviation of two years. The manufacturer replaces all free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should he offer? Assume that the lifetime of a motor follows the normal distribution.

Problem 5.5.

You flip a coin 100 times.

- (a) What is the probability that you obtain at least 40 heads? Use the binomial distribution.
- (b) What is the probability that you obtain at least 40 heads? Use the normal approximation to the binomial distribution.

Problem 5.6.

The life of a certain type of device has an advertised failure rate of 0.01 per hour. The failure rate is constant and the exponential distribution applies.

- (a) What is the mean time to failure?
- (b) What is the probability that 200 hours will pass before a failure is observed?
- (c) What is the probability that exactly 3 of 5 devices are functioning after 200 hours?

Problem 5.7.

Using the gamma distribution with $\alpha = 4$ and $\beta = 1/3$, find

- (a) $P(x < 1/2)$
- (b) C where $P(x < C) = 0.05$

Problem 5.8.

Using the Chi-Squared distribution with $v=10$, find

- (a) $P(\chi^2 < 4)$
- (b) C where $P(\chi^2 < C) = 0.05$

Problem 5.9.

Using the t-distribution with $v=10$, find

- (a) $P(t < -1)$
- (b) C where $P(t < C) = 0.05$

Problem 5.10.

Using the F-distribution with $v_1=8$ and $v_2 = 12$, find

(a) $P(F < 1)$

(b) C where $P(F < C) = 0.05$