

# Continuous Random Variables

## (Devore Chapter Four)

1016-345-01: Probability and Statistics for Engineers\*

Fall 2012

### Contents

<b>1</b>	<b>Continuous Random Variables</b>	<b>2</b>
1.1	Defining the Probability Density Function . . . . .	2
1.2	Expected Values . . . . .	6
1.3	Percentiles and Medians . . . . .	7
1.4	Worksheet . . . . .	8
<b>2</b>	<b>Normal Distribution</b>	<b>8</b>
2.1	The Standard Normal Distribution . . . . .	9
2.1.1	Cumulative Distribution Function . . . . .	11
2.2	$z_\alpha$ Values and Percentiles . . . . .	12
2.3	Approximating Distributions of Discrete Random Variables . . . . .	13
2.3.1	Example: Approximating the Binomial Distribution . . . . .	14
<b>3</b>	<b>Other Continuous Distributions</b>	<b>14</b>
3.1	Exponential Distribution . . . . .	15
3.1.1	Gamma Distribution . . . . .	17
3.1.2	The Gamma Function . . . . .	17
3.1.3	The Standard Gamma Distribution . . . . .	18
3.2	Chi-Squared Distribution . . . . .	19

---

\*Copyright 2012, John T. Whelan, and all that

Thursday 20 September 2012

# 1 Continuous Random Variables

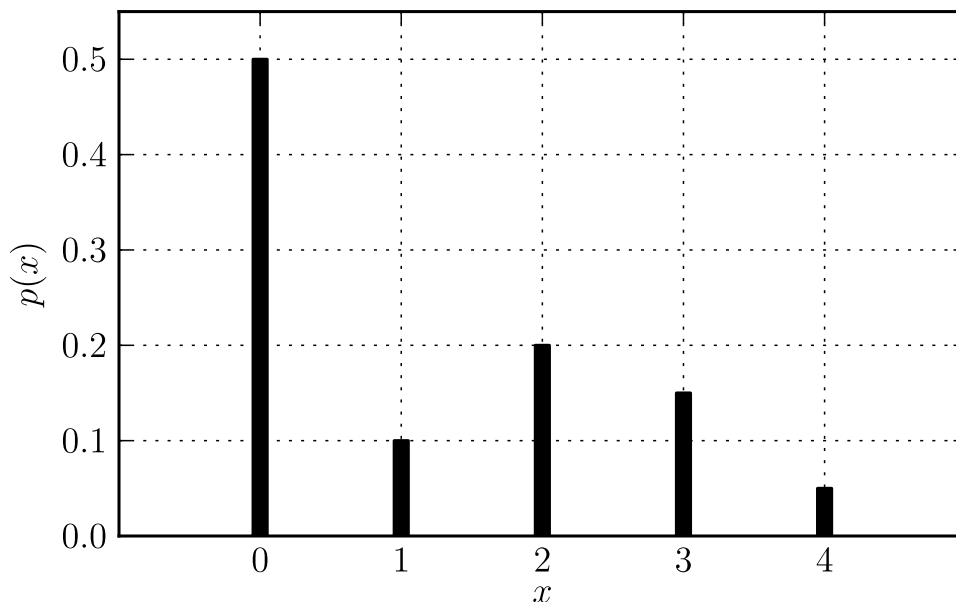
Recall that when we introduced the concept of random variables last chapter we focussed on discrete random variables, which could take on one of a set of specific values, either finite or countably infinite. Now we turn attention to **continuous random variables**, which can take on an value in one or more intervals, but for which there is zero probability for any single value.

## 1.1 Defining the Probability Density Function

Devore starts with the definition of the probability density function, and deduces the cumulative distribution function for a continuous random variable from that. We will follow a complementary presentation, starting by extending the cdf to a continuous rv, and then deriving the pdf from that.

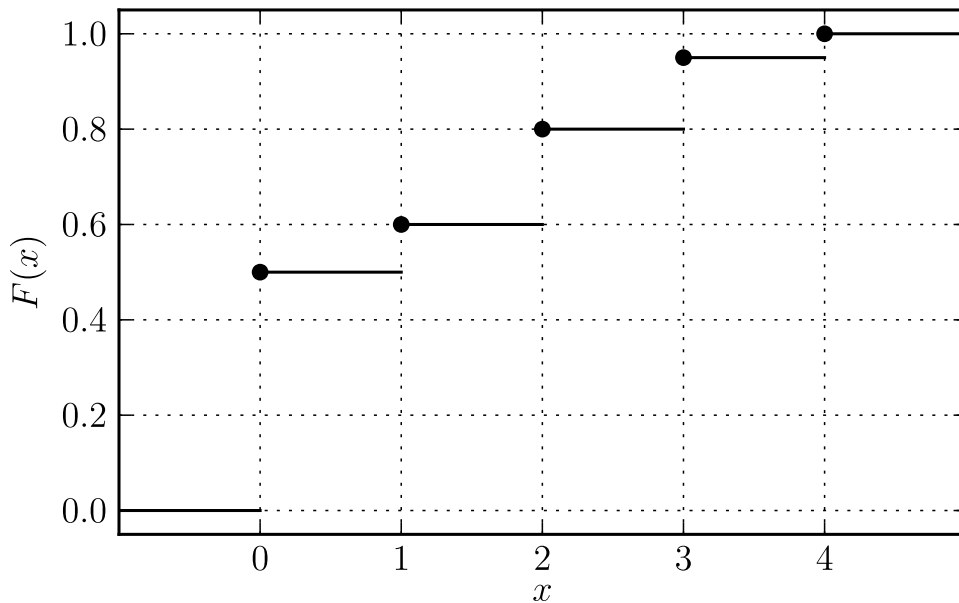
Recall that if  $X$  is a discrete random variable, we can define the probability mass function

$$p(x) = P(X = x) \tag{1.1}$$



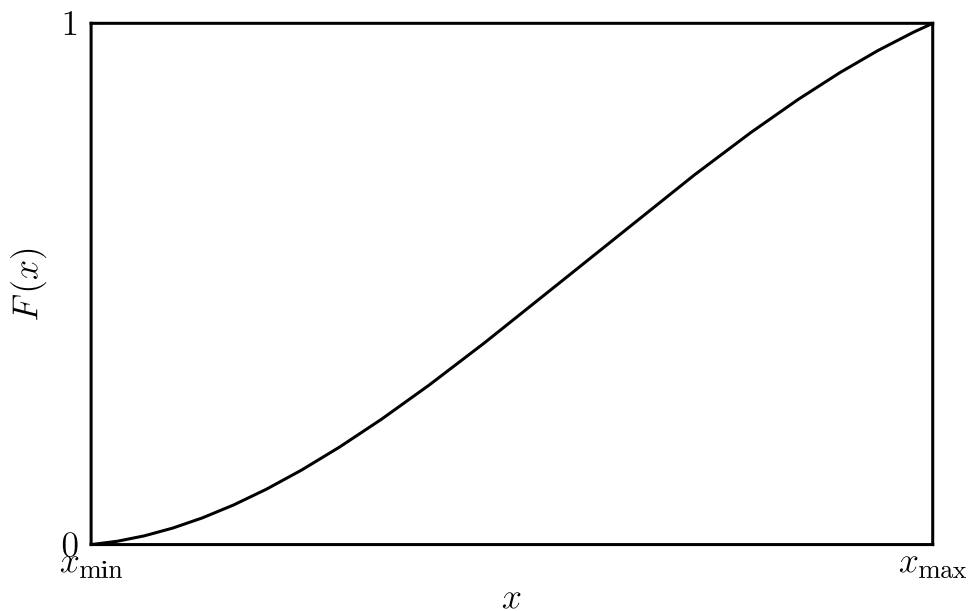
and the cumulative distribution function

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y) \tag{1.2}$$



Note that for a discrete random variable, the cdf is constant for a range of values, and then jumps when we reach one of the possible values for the rv.

For a continuous rv  $X$ , we know that the probability of  $X$  equalling exactly any specific  $x$  is zero, so clearly  $p(x) = P(X = x)$  doesn't make any sense. However,  $F(x) = P(X \leq x)$  is still a perfectly good definition. Now, though,  $F(x)$  should be a continuous function with no "jumps", since a jump corresponds to a finite probability for some specific value:



Note that in both cases, if  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum possible values of

$X$ , then  $F(x) = 0$  when  $x < x_{\min}$ , and  $F(x) = 1$  when  $x \geq x_{\max}$ . We can also write

$$F(-\infty) = 0 \tag{1.3a}$$

$$F(\infty) = 1 \tag{1.3b}$$

Just as before, we can calculate the probability for  $X$  to lie in some finite interval from  $a$  to  $b$ , using the fact that

$$(X \leq b) = (X \leq a) \cup (a < X \leq b) \tag{1.4}$$

(where  $b > a$ ):

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) . \tag{1.5}$$

But now, since  $P(X = a) = 0$  and  $P(X = b) = 0$ , we don't have to worry about whether or not the endpoints are included:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) \quad (X \text{ a cts rv}) \tag{1.6}$$

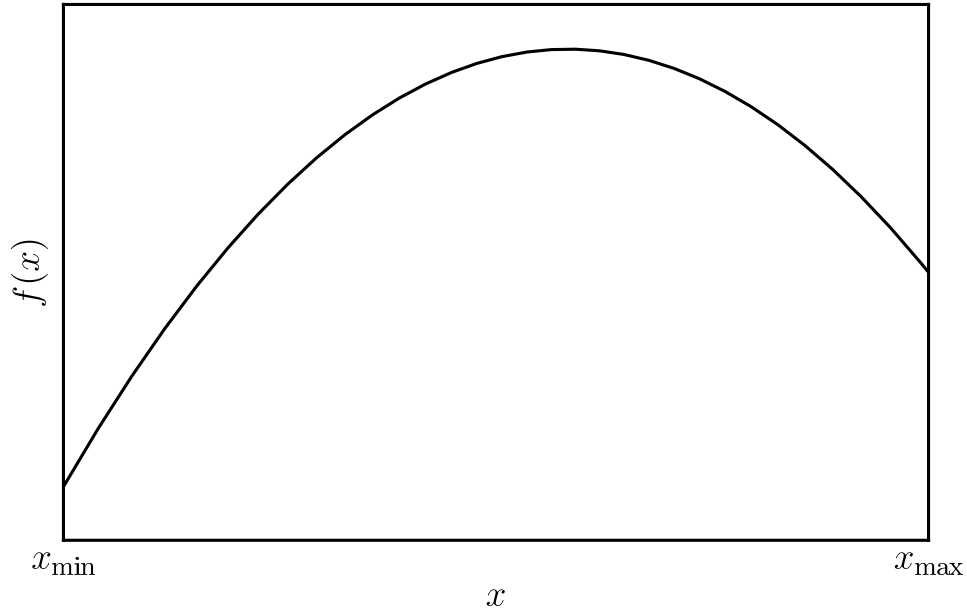
We can't talk about the probability for  $X$  to be exactly some value  $x$ , but we can consider the probability for  $X$  to lie in some little interval of width  $\Delta x$  centered at  $x$ :

$$P(x - \Delta x/2 < X < x + \Delta x/2) = F(x + \Delta x/2) - F(x - \Delta x/2) \tag{1.7}$$

We know that if we make  $\Delta x$  smaller and smaller, the difference between  $F(x + \Delta x/2)$  and  $F(x - \Delta x/2)$  has to get smaller and smaller. If we define

$$\begin{aligned} f(x) &= \lim_{\Delta x \rightarrow 0} \frac{P(x - \Delta x/2 < X < x + \Delta x/2)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x/2) - F(x - \Delta x/2)}{\Delta x} = \frac{dF}{dx} = F'(x) \end{aligned} \tag{1.8}$$

this is just the derivative of the cdf. This derivative  $f(x)$  is called the *probability density function*.



If we want to calculate the probability for  $X$  to lie in some interval, we can use the pdf  $f(x)$  or the cdf  $F(x)$ . If the interval is small enough that the cdf  $F(x)$  can be approximated as a straight line, the change in  $F(x)$  over that interval is approximately the width of the interval times the derivative  $f(x) = F'(x)$ :

$$P(x - \Delta x/2 < X < x + \Delta x/2) = F(x + \Delta x/2) - F(x - \Delta x/2) \approx F'(x) \Delta x = f(x) \Delta x \quad (1.9)$$

If the interval is not small enough, we can always divide it up into pieces that are. If the  $i$ th piece is  $\Delta x$  wide and centered on  $x_i$ ,

$$P(a < X < b) = F(b) - F(a) \approx \sum_{i=1}^N f(x_i) \Delta x \quad (1.10)$$

where

$$x_{i+1} - x_i = \Delta x \quad (1.11)$$

and

$$x_1 - \Delta x/2 = a \quad (1.12a)$$

$$x_N + \Delta x/2 = b \quad (1.12b)$$

$$(1.12c)$$

In the limit that  $N \rightarrow \infty$ , so that  $\Delta x \rightarrow 0$ , the sum becomes an integral, so that

$$P(a < X < b) = F(b) - F(a) = \int_a^b f(x) dx \quad (1.13)$$

Of course, this is also a consequence of the Second Fundamental Theorem of Calculus:

$$\text{if } f(x) = F'(x) \text{ then } \int_a^b f(x) dx = F(b) - F(a) \quad (1.14)$$

The normalization condition, which for a discrete random variable is  $\sum_x p(x) = 1$ , becomes for a continuous random variable

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (1.15)$$

## 1.2 Expected Values

Recall that for a discrete random variable, the expected value is

$$E(X) = \sum_x x p(x) \quad (X \text{ a discrete rv}) \quad (1.16)$$

In the case of a continuous random variable, we can use the same construction as before to divide the whole range from  $x_{\min}$  to  $x_{\max}$  into little intervals of width  $\Delta x$ , centered at points  $x_i$ , we can define a discrete random variable  $X_{\Delta x}$  which is just  $X$  rounded off to middle of whatever interval it's in. The pmf for this discrete random variable is

$$p_{\Delta x}(x_i) = P(x_i - \Delta x/2 < X < x_i + \Delta x/2) \approx f(x_i) \Delta x \quad (1.17)$$

$X_{\Delta x}$  and  $X$  will differ by at most  $\Delta x/2$ , so we can write, approximately,

$$E(X) \approx E(X_{\Delta x}) = \sum_i x_i p_{\Delta x}(x_i) \approx \sum_i x_i f(x_i) \Delta x \quad (1.18)$$

In the limit that we divide up into more and more intervals and  $\Delta x \rightarrow 0$ , this becomes

$$E(X) = \int_{x_{\min}}^{x_{\max}} x f(x) dx = \int_{-\infty}^{\infty} x f(x) dx \quad (1.19)$$

The same argument works for the expected value of any function of  $X$ :

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx \quad (1.20)$$

In particular,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx \quad (1.21)$$

We can also define the variance, and as usual, the same shortcut formula applies:

$$V(X) = \sigma_X^2 = E((X - \mu_X)^2) = E(X^2) - (E(X))^2 \quad (1.22)$$

where

$$\mu_X = E(X) \quad (1.23)$$

### 1.3 Percentiles and Medians

A concept which is easier to define for continuous random variables than discrete ones is that of percentiles, and specifically the 50th percentile or median. The median  $\tilde{\mu}$  of a probability distribution is the value such that there is a 50% chance of the random variable lying above the median, and a 50% chance of lying below it. We didn't talk about it much for discrete random variables, since a careful definition would be something like this:

$$P(X \geq \tilde{\mu}) \geq .5 \quad \text{and} \quad P(X \leq \tilde{\mu}) \geq .5 \quad (1.24)$$

With continuous random variables, it's a little easier. Since the probability of  $X$  equalling exactly  $\tilde{\mu}$  is zero,  $P(X \geq \tilde{\mu}) = P(X > \tilde{\mu}) = 1 - P(X \leq \tilde{\mu})$  and therefore

$$F(\tilde{\mu}) = P(X \leq \tilde{\mu}) = .5 \quad (\text{for continuous random variables}) \quad (1.25)$$

I.e., the median of a probability distribution is the value at which the cumulative distribution function equals one-half.

Percentiles are defined similarly; the  $(100 \times p)$ th percentile, called  $\eta(p)$ , is the value at which the cdf equals  $p$ :

$$F(\eta(p)) = p \quad (1.26)$$

If we have the functional form of the cdf, we can just solve (1.26) for  $\eta(p)$ ; if not, we can still define the percentile indirectly using the definition of the cdf:

$$\int_{-\infty}^{\eta(p)} f(x) dx = p \quad (1.27)$$

### Practice Problems

4.1, 4.5, 4.11, 4.13, 4.17, 4.21

Tuesday 25 September 2012

## 1.4 Worksheet

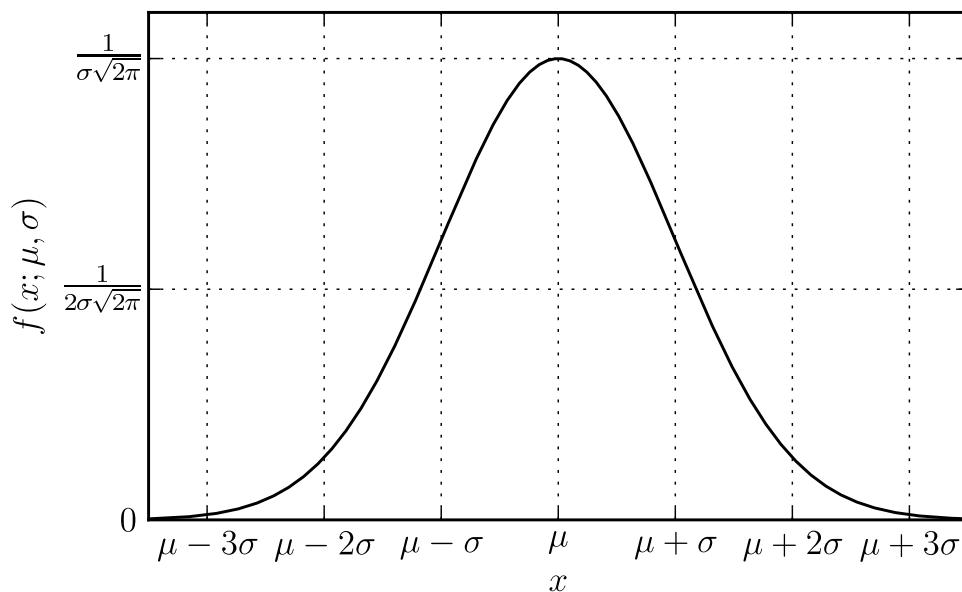
- worksheet
- solutions

## 2 Normal Distribution

On this worksheet, you considered one simple pdf for a continuous random variable, a uniform distribution. Now we're going to consider one which is more complicated, but very useful, the normal distribution, sometimes called the Gaussian distribution. The distribution depends on two parameters,  $\mu$  and  $\sigma$ ,<sup>1</sup> and if  $X$  follows a normal distribution with those parameters, we write  $X \sim N(\mu, \sigma^2)$ , and the pdf of  $X$  is

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (2.1)$$

The shape of the normal distribution is the classic “bell curve”:



Some things we can notice about the normal distribution:

- It is non-zero for all  $x$ , so  $x_{\min} = -\infty$  and  $x_{\max} = \infty$

---

<sup>1</sup> $\mu$  can be positive, negative or zero, but  $\sigma$  has to be positive



- The pdf is, in fact, normalized. This is kind of tricky to show, since it relies on the definite integral result

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi} \quad (2.2)$$

This integral can't really be done by the usual means, since there is no ordinary function whose derivative is  $e^{-z^2/2}$ , and actually only the definite integral from  $-\infty$  to  $\infty$  (or from 0 to  $\infty$ ) has a nice value. The proof of this identity is cute, but unfortunately it requires the use of a two-dimensional integral in polar coordinates, so it's beyond our scope at the moment.

- The parameter  $\mu$ , as you might guess from the name, is the mean of the distribution. This is not so hard to see from the graph: the pdf is clearly symmetric about  $x = \mu$ , so that point should be both the mean and median of the distribution. It's also not so hard to show from the equations that

$$\int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)} dx = \mu \int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)} dx \quad (2.3)$$

(and thus  $E(X) = \mu$ ) by using the change of variables  $y = x - \mu$ .

- The parameter  $\sigma$  turns out to be the standard deviation. You can't really tell this from the graph, but it is possible to show that

$$V(X) = E((X - \mu)^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/(2\sigma^2)} dx \quad (2.4)$$

This is harder, though, since the integral

$$\int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/(2\sigma^2)} dx = \sigma^2 \int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)} dx \quad (2.5)$$

has to be done using integration by parts.

- Although the pdf depends on two parameters, neither one of them fundamentally changes the shape. Changing  $\mu$  just slides the pdf back and forth. Changing  $\sigma$  just changes the scale of the horizontal axis. If we increase  $\sigma$ , the curve is stretched horizontally, and it has to be squeezed by the same factor vertically, in order to keep the area underneath it equal to 1 (since the pdf is normalized).

## 2.1 The Standard Normal Distribution

The fact that, for  $X \sim N(\mu, \sigma^2)$ , changing the parameters  $\mu$  and  $\sigma$  just slides and stretches the  $x$  axis is the motivation for constructing a new random variable  $Z$ :

$$Z = \frac{X - \mu}{\sigma} \quad (2.6)$$

Now, because the expected values for continuous random variables still have the same behavior under linear transformations as for discrete rvs, i.e.,

$$E(aX + b) = aE(X) + b \quad (2.7a)$$

$$V(aX + b) = a^2V(X) \quad (2.7b)$$

we can see that

$$E(Z) = \frac{E(X) - \mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0 \quad (2.8a)$$

$$V(Z) = \frac{V(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1 \quad (2.8b)$$

In fact,  $Z$  is itself normally distributed  $Z \sim N(0, 1)$ . We call this special case of the normal distribution the *standard normal distribution*.

The standard normal distribution is often useful in calculating the probability that a normally distributed random variable falls in a certain finite interval:

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < Z \leq \frac{b - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a - \mu}{\sigma}}^{\frac{b - \mu}{\sigma}} e^{-z^2/2} dz \quad (2.9)$$

The problem is that, as we noted before, there's no ordinary function whose derivative is  $e^{-z^2/2}$ . Still, the cumulative distribution function of a standard normal random variable is a perfectly sensible thing, and we could, for example, numerically integrate to find an approximate value for any  $z$ . So we define a function  $\Phi(z)$  to be equal to that integral:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du \quad (\text{definition}) \quad (2.10)$$

Note that this is defined so that

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} . \quad (2.11)$$

We can now write down the cdf for the original normal rv  $X$ :

$$F(x; \mu, \sigma) = P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (2.12)$$

and from the cdf we can find the probability for  $X$  to lie in any interval:

$$P(a < X \leq b) = F(b; \mu, \sigma) - F(a; \mu, \sigma) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \quad (2.13)$$

### 2.1.1 Cumulative Distribution Function

The function  $\Phi(z)$  has been defined so that it equals the cdf of  $Z$ :

$$F(z; 1, 0) = P(Z \leq z) = \int_{-\infty}^z f(u; 0, 1) du = \Phi(z) \quad (2.14)$$

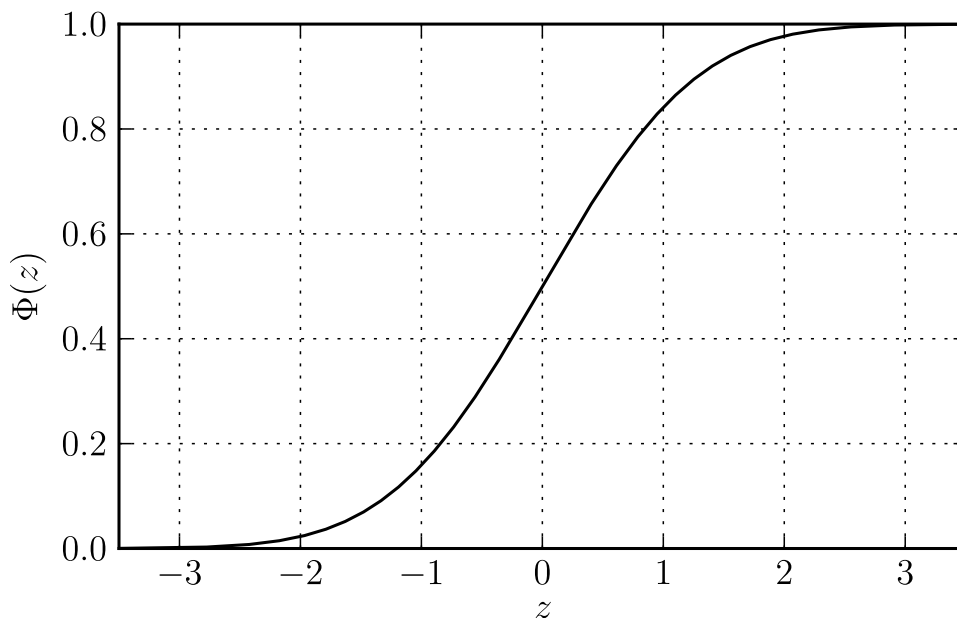
This identification of  $\Phi(z)$  as a cdf tells us a few values exactly:

$$\Phi(-\infty) = 0 \quad (2.15a)$$

$$\Phi(0) = \frac{1}{2} \quad (2.15b)$$

$$\Phi(\infty) = 1 \quad (2.15c)$$

We know  $\Phi(\infty)$  from the fact that the normal distribution is normalized. We can deduce  $\Phi(0)$  from the fact that it's symmetric about its mean value, and therefore the mean value is also the median. For any other value of  $z$ ,  $\Phi(z)$  can in principle be calculated numerically. Here's what a plot of  $\Phi(z)$  looks like:



Of course, since  $\Phi(z)$  is an important function—like, for example,  $\sin \theta$  or  $e^x$ —its value has been tabulated. Table A.3, in the back of the book, collects some of those values. Now, in the old days trigonometry books also had tables of values of trig functions, but now we don't need those because our calculators have `sin` and `cos` and  $e^x$  buttons. Unfortunately, our calculators don't have  $\Phi$  buttons on them yet, so we have the table. Note that if you're using a mathematical computing environment like `scipy` or `Mathematica` or `matlab` which provides the “error function”  $\text{erf}(x)$ , you can evaluate

$$\Phi(z) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{z}{\sqrt{2}} \right) \right] \quad (2.16)$$

(That's how I made the plot of  $\Phi(z)$ .)

Some useful sample values of  $\Phi(z)$  are

$$P(X \leq \mu + \sigma) = \Phi(1) \approx .8413 \quad (2.17a)$$

$$P(X \leq \mu + 2\sigma) = \Phi(2) \approx .9772 \quad (2.17b)$$

$$P(X \leq \mu + 3\sigma) = \Phi(3) \approx .9987 \quad (2.17c)$$

$$P(\mu - \sigma < X \leq \mu + \sigma) = \Phi(1) - \Phi(-1) \approx .6827 \quad (2.17d)$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx .9545 \quad (2.17e)$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) = \Phi(3) - \Phi(-3) \approx .9973 \quad (2.17f)$$

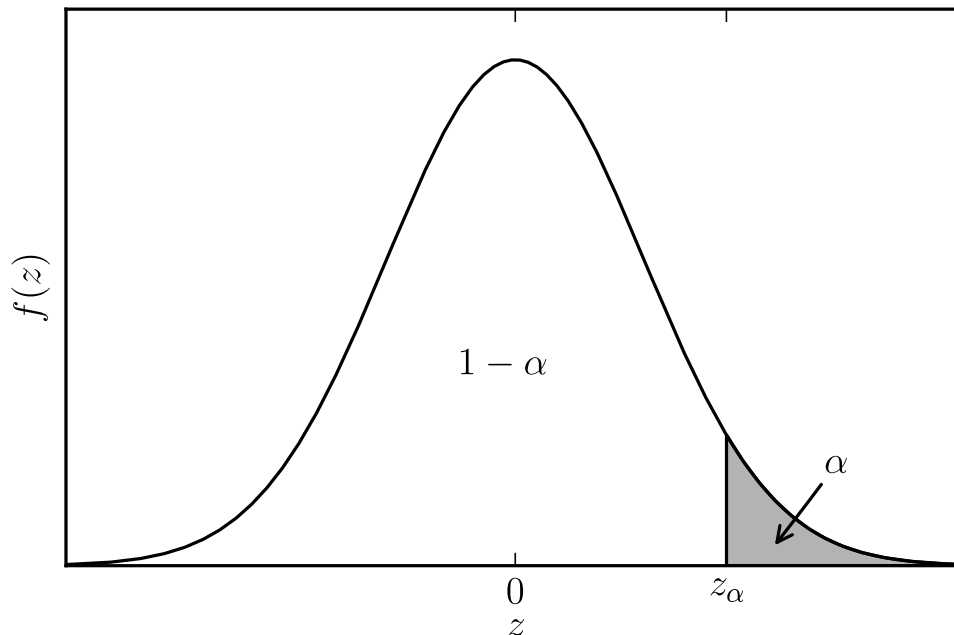
The last few in particular are useful to keep in mind: A given value drawn from a standard normal distribution has

- about a 2/3 (actually 68%) chance of being within one sigma of the mean
- about a 95% chance of being within two sigma of the mean
- about a 99.7% chance of being within three sigma of the mean

## 2.2 $z_\alpha$ Values and Percentiles

Often you want to go the other way, and ask, for example, what's the value that a normal random variable will only exceed 10% of the time. To get this, you basically need the inverse of the function  $\Phi(z)$ . The notation is to define  $z_\alpha$  such that  $\alpha$  (which can be between 0 and 1) is the probability that  $Z > z_\alpha$ :

$$P(Z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha \quad (2.18)$$



We know a few exact values:

$$z_0 = \infty \quad z_{.5} = 0 \quad z_1 = -\infty \quad (2.19)$$

The rest can be deduced from the table of  $\Phi(z)$  values; a few of them are in Table 4.1 of Devore (on page 156 of the 8th edition). Note that his argument about  $z_{.05}$  actually only shows that it's between 1.64 and 1.65. A more accurate value turns out to be  $z_{.05} \approx 1.64485$  which means that, somewhat coincidentally,  $z_{.05} \approx 1.645$  is accurate to four significant figures. (But to three significant figures  $z_{.05} \approx 1.64$ .)

Note also that the definition of  $z$ -critical values is “backwards” from that of percentiles. E.g.,  $z_{.05}$  is the 95th percentile of the standard normal distribution.

## 2.3 Approximating Distributions of Discrete Random Variables

We described some aspects of a continuous random variable as a limiting case of a discrete random variable as the spacing between possible values went to zero. It turns out that continuous random variables, and normally-distributed ones in particular, can be good approximations of discrete random variables when the numbers involved are large. In this case,  $\Delta x$  doesn't go to zero, but it's small compared to the other numbers in the problem.

Suppose we have a discrete random variable  $X$  with pmf  $p(x)$  which can take on values which are  $\Delta x$  apart.<sup>2</sup> The continuous random variable  $\mathcal{X}$  with pdf  $f(x)$  is a good approximation of  $X$  if

$$p(x) = P(X = x) \approx P(x - \Delta x/2 < \mathcal{X} < x + \Delta x/2) \approx f(x) \Delta x \quad (2.20)$$

One thing to watch out for is the discreteness of the original random variable  $X$ . For instance, if we want to estimate its cdf,  $P(X \leq x)$ , we should consider that the value  $X = x$  corresponds to the range  $x - \Delta x/2 < \mathcal{X} < x + \Delta x/2$ , so

$$P(X \leq x) \approx P(\mathcal{X} < x + \Delta x/2) = \int_{-\infty}^{x+\Delta x/2} f(y) dy \quad (2.21)$$

Often the continuous random variable used in the approximation is a normally-distributed rv with the same mean and standard deviation as the original discrete rv. I.e., if  $E(X) = \mu$  and  $V(X) = \sigma^2$ , it is often a good approximation to take

$$F(x) = P(X \leq x) \approx \int_{-\infty}^{x+\Delta x/2} f(y; \mu, \sigma) dy = \Phi\left(\frac{x + .5(\Delta x) - \mu}{\sigma}\right) \quad (2.22)$$

---

<sup>2</sup>This is made even simpler if the discrete rv can take on integer values, so that  $\Delta x = 1$ , but it also works for other examples, like SAT scores, for which  $\Delta x = 10$ .

### 2.3.1 Example: Approximating the Binomial Distribution

As an example, consider a binomial random variable  $X \sim \text{Bin}(n, p)$ . Remember that for large values of  $n$  it was kind of a pain to find the cdf

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^x \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=0}^x \binom{n}{y} p^y q^{n-y} \quad (2.23)$$

(Did anyone forget what  $q$  is? I did.  $q = 1 - p$ .) But if we can approximate  $X$  as a normally-distributed random variable with  $\mu = E(X) = np$  and  $\sigma^2 = V(X) = np(1-p) = npq$ , life is pretty simple (here  $\Delta x = 1$  because  $x$  can take on integer values):

$$P(X \leq x) = B(x; n, p) \approx \Phi\left(\frac{x + .5 - np}{\sqrt{npq}}\right) \quad (2.24)$$

At the very least, we only have to look in one table, rather than in a two-parameter family of tables!

## Practice Problems

4.29, 4.31, 4.35, 4.47, 4.53, 4.55

Thursday 27 September 2012

## 3 Other Continuous Distributions

Sections 4.4 and 4.5 unleash upon us a dizzying array of other distribution functions. (I count six or seven.) In this course, we'll skip section 4.5, and therefore only have to contend with the details of a few of them. Still, it's handy to put them into some context so we can understand how they fit together. First, a qualitative overview:

- The *exponential distribution* occurs in the context of a Poisson process. It is the pdf of the time we have to wait for a Poisson event to occur.
  - The *gamma distribution* is a variant of the exponential distribution with a modified shape. It describes a physical situation where the expected waiting time changes depending on how long we've been waiting. (The *Weibull distribution*, which is in section 4.5, is a different variation on the exponential distribution.)
- The *chi-squared distribution* is the pdf of the sum of the squares of one or more independent standard normal random variables. This pdf turns out to be a special case of the gamma distribution.

Part of the complexity of the formulas for these various random variables arises from several things which are not essential to the fundamental shape of the distribution. Consider a normal distribution

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (3.1)$$

First, notice that the factor of  $\frac{1}{\sigma\sqrt{2\pi}}$  does not depend on the value  $x$  of the random variable  $X$ . It's a constant (which happens to depend on the parameter  $\sigma$ ) which is just needed to normalize the distribution. So we can write

$$f(x; \mu, \sigma) = \mathcal{N}(\sigma) e^{-(x-\mu)^2/(2\sigma^2)} \quad (3.2)$$

or

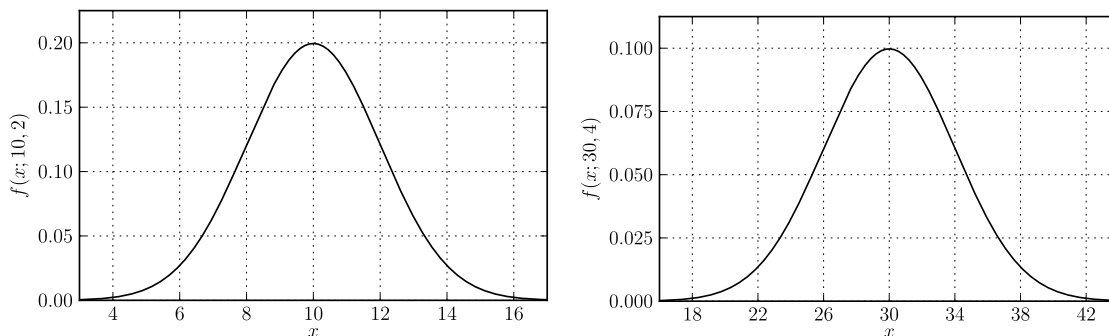
$$f(x; \mu, \sigma) \propto e^{-(x-\mu)^2/(2\sigma^2)} \quad (3.3)$$

We will try to separate the often complicated forms of the normalization constants from the often simpler forms of the  $x$ -dependent parts of the distributions.

Also, notice that the parameters  $\mu$  and  $\sigma$  affect the location and scale of the pdf, but not the shape, and if we define the rv  $Z = \frac{X-\mu}{\sigma}$ , the pdf is

$$f(z) \propto e^{-z^2/2} \quad (3.4)$$

BTW, what I mean by “doesn't change the shape” is that if I plot a normal pdf with different parameters, I can always change the ranges of the axes to make it look like any other one. For example:



### 3.1 Exponential Distribution

Suppose we have a Poisson process with rate  $\lambda$ , so that the probability mass function of the number of events occurring in any period of time of duration  $t$  is

$$P(n \text{ events in } t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (3.5)$$

In particular, evaluating the pmf at  $n = 0$  we get

$$P(\text{no events in } t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \quad (3.6)$$

Now pick some arbitrary moment and let  $X$  be the random variable describing how long we have to wait for the next event. If  $X \leq t$  that means there are one or more events occurring in the interval of length  $t$  starting at that moment, so the probability of this is

$$P(X \leq t) = 1 - P(\text{no events in } t) = 1 - e^{-\lambda t} \quad (3.7)$$

But that is the definition of the cumulative distribution function, so

$$F(x; \lambda) = P(X \leq x) = 1 - e^{-\lambda x} \quad (3.8)$$

Note that it doesn't make sense for  $X$  to take on negative values, which is good, since  $F(0) = 0$ . This means that technically,

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \quad (3.9)$$

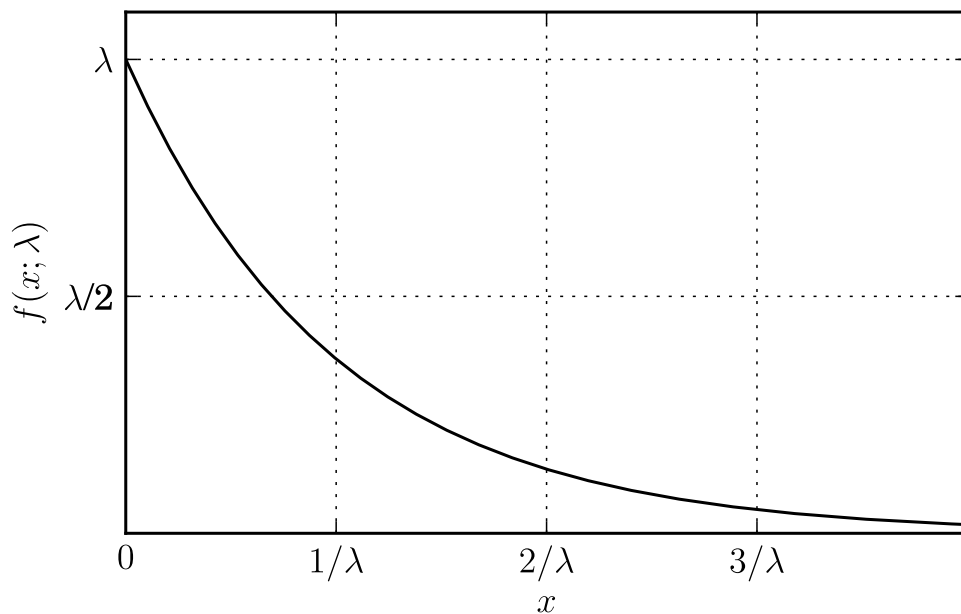
We can differentiate (with respect to  $x$ ) to get the pdf

$$f(x; \lambda) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \quad (3.10)$$

If we isolate the  $x$  dependence, we can write the slightly simpler

$$f(x; \lambda) \propto e^{-\lambda x} \quad x \geq 0 \quad (3.11)$$

In any event, we see that changing  $\lambda$  just changes the scale of the  $x$  variable, so we can draw one shape:





You can show, by using integration by parts to do the integrals, that

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \quad (3.12)$$

and

$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2} \quad (3.13)$$

so

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2} \quad (3.14)$$

### 3.1.1 Gamma Distribution

Sometimes you also see the parameter in an exponential distribution written as  $\beta = 1/\lambda$  rather than  $\lambda$ , so that

$$f(x; \beta) \propto e^{-x/\beta} \quad x \geq 0 \quad (3.15)$$

The *gamma distribution* is a generalization which adds an additional parameter  $\alpha > 0$ , so that

$$f(x; \alpha, \beta) \propto \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \quad x \geq 0 \quad \text{gamma distribution} \quad (3.16)$$

Note that  $\alpha$  (which need not be an integer) really does influence the shape of the distribution, and  $\alpha = 1$  reduces to the exponential distribution.

### 3.1.2 The Gamma Function

Most of the complication associated with the gamma distribution in particular is associated with finding the normalization constant. If we write

$$f(x; \alpha, \beta) = \mathcal{N}(\alpha, \beta) \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \quad (3.17)$$

then

$$\frac{1}{\mathcal{N}(\alpha, \beta)} = \int_0^{\infty} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} dx = \beta \int_0^{\infty} u^{\alpha-1} e^{-u} du = \frac{\beta}{\mathcal{N}(\alpha, 1)} \quad (3.18)$$

This integral

$$\frac{1}{\mathcal{N}(\alpha, 1)} = \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \Gamma(\alpha) \quad (3.19)$$

is defined as the ‘‘Gamma function’’, which has lots of nifty properties:

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$  if  $\alpha > 1$  (which can be shown by integration by parts)
- $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$  from which it follows that  $\Gamma(n) = (n - 1)!$  if  $n$  is a positive integer
- $\Gamma(1/2) = \sqrt{\pi}$  which actually follows from  $\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}$

- Note that since  $1 = \Gamma(1) = 0 \cdot \Gamma(0)$ ,  $\Gamma(0)$  must be infinite, which is what I meant by  $\frac{1}{(-1)!} = 0$ .

In terms of the Gamma function,

$$\mathcal{N}(\alpha, \beta) = \frac{\mathcal{N}(\alpha, 1)}{\beta} = \frac{1}{\beta \Gamma(\alpha)} \quad (3.20)$$

which makes the pdf for the Gamma distribution

$$f(x; \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad (3.21)$$

which is the form in which the pdf appears in Devore.

It's straightforward to show, using integration by parts, that

$$E(X) = \int_0^\infty x f(x; \alpha, \beta) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x/\beta} dx = \alpha\beta \quad (3.22)$$

and

$$E(X^2) = \int_0^\infty x^2 f(x; \alpha, \beta) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x/\beta} dx = \alpha(\alpha+1)\beta^2 \quad (3.23)$$

so

$$V(X) = E(X^2) - [E(X)]^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2. \quad (3.24)$$

Note that this makes sense from a dimensional analysis point of view, since  $\beta$  has the same units as  $X$ , and  $\alpha$  is dimensionless.

### 3.1.3 The Standard Gamma Distribution

Just as we could take any normally-distributed random variable  $N(\mu, \sigma^2)$  and construct from it a standard normal random variable  $N(0, 1) \equiv (N(\mu, \sigma^2) - \mu)/\sigma$ , we can convert a gamma random variable into a so-called standard gamma random variable. If  $X$  obeys a gamma distribution with parameters  $\alpha$  and  $\beta$ , then  $X/\beta$  obeys a gamma distribution with parameters  $\alpha$  and 1.<sup>3</sup> This means that the cdf for the gamma random variable  $X$  can be found from the cdf for the standard gamma random variable  $X/\beta$ :

$$P(X \leq x) = P\left(\frac{X}{\beta} \leq \frac{x}{\beta}\right) = \int_0^{x/\beta} f(y; \alpha, 1) dy = \frac{1}{\Gamma(\alpha)} \int_0^{x/\beta} y^{\alpha-1} e^{-y} dy \quad (3.25)$$

The standard gamma cdf

$$F(x; \alpha) = \int_0^x f(y; \alpha, 1) dy = \frac{1}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y} dy \quad (3.26)$$

---

<sup>3</sup>Note that the standard gamma distribution still has an  $\alpha$  parameter; only  $\beta$  has been transformed away by the scaling of the random variable. We can say that  $\alpha$  is an *intrinsic parameter* while  $\beta$  is an *extrinsic parameter*.

is referred to by Devore as the *incomplete gamma function*. (Be careful, though, since some sources and software packages omit the  $1/\Gamma(\alpha)$  from the definition and/or reverse the order of the arguments.) Its value is collected for some choices of  $\alpha$  and  $x$  in Table A.4 of Devore. The cdf for a gamma distribution with parameters  $\alpha$  and  $\beta$  is written in terms of this as

$$P(X \leq x) = \int_0^x f(y; \alpha, \beta) dy = F\left(\frac{x}{\beta}; \alpha\right) \quad (3.27)$$

## 3.2 Chi-Squared Distribution

We are often interested in the sum of the squares of a number of independent standard normal random variables,  $\sum_{k=1}^{\nu} (Z_k)^2$ . To see why this would be, consider a model that predicts values  $\{\mu_1, \mu_2, \dots, \mu_{\nu}\}$  for a bunch of measured quantities, where  $\nu$  is the number of things that can be measured. Suppose it also associates normally-distributed random errors with each of them, with standard deviations  $\{\sigma_1, \sigma_2, \dots, \sigma_{\nu}\}$ , so that the results of a set of  $\nu$  measurements are  $\{X_1, X_2, \dots, X_{\nu}\}$ . We'd like to have a single measure of how far off the measurements  $\{X_k | k = 1, \dots, n\}$  are from the predicted values  $\{\mu_k | k = 1, \dots, n\}$ . Keeping in mind the pythagorean theorem, we might want to find the distance between the predicted and measured points in space by adding  $(X_k - \mu_k)^2$  for  $k = 1, 2, \dots, n$ . But that wouldn't really be fair, since if e.g.,  $\sigma_1$  is much greater than  $\sigma_2$ , the model tells us that the offset in the 1st direction,  $|X_1 - \mu_1|$ , will be bigger than the offset in the 2nd direction,  $|X_2 - \mu_2|$ . Even worse,  $X_1$  might be a distance, and  $X_2$  an energy or something, and it would make no sense to add them. So instead, we want to scale the offsets by the sizes of the errors predicted by the model, and take the distance in that scaled space, which is

$$\sqrt{\left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2 - \mu_2}{\sigma_2}\right)^2 + \dots + \left(\frac{X_{\nu} - \mu_{\nu}}{\sigma_{\nu}}\right)^2} \quad (3.28)$$

but if we construct a standard normal random variable  $Z_k = \frac{X_k - \mu_k}{\sigma_k}$  from each of the measurements, the distance is just the square root of the sum of their squares:

$$\sqrt{Z_1^2 + Z_2^2 + \dots + Z_{\nu}^2} \quad (3.29)$$

It's convenient and slightly more conventional to talk about the square of this scaled distance; if we call it a random variable  $X$  (no relation to the  $\{X_1, \dots, X_{\nu}\}$  we started with), it is indeed the sum of the squares of  $\nu$  independent standard normal random variables:

$$X = Z_1^2 + Z_2^2 + \dots + Z_{\nu}^2 = \sum_{k=1}^{\nu} Z_k^2 \quad (3.30)$$

This distribution obeyed by this random variable is called a *chi-squared distribution with  $\nu$  degrees of freedom*. It can be shown that the pdf of this random variable is

$$f(x) \propto x^{(\nu/2)-1} e^{-x/2} \quad (3.31)$$

But if we compare this to equation (3.16) we see that the chi-squared distribution with  $\nu$  degrees of freedom is just a Gamma distribution with  $\alpha = \nu/2$  and  $\beta = 2$ . This means its mean and variance are

$$E(X) = \nu \quad \text{and} \quad V(X) = 2\nu \quad (3.32)$$

and its cdf is

$$P(X \leq x) = F\left(\frac{x}{2}; \frac{\nu}{2}\right) \quad (3.33)$$

## Practice Problems

4.59, 4.61, 4.63, 4.65, 4.67, 4.69, 4.71