Continuous Probability Distributions

These two probability models have sample spaces with only a finite number of points:

Bernoulli(θ)	has S = {0,1}	(two points)
Binomial(n,θ)	has S = {0,1,2,...,n}	(n+1 points)

The Poisson distribution, however, has an <u>infinite</u> number of sample points, S={0,1,2,3,...}.

All three are similar in that

(1) there is a set of points, each of which has a certain probability, and

(2) the sum of the probabilities of all the points is 1.

Probability distributions like these are "discrete." They have discrete sample spaces, so they are <u>discrete</u> probability distributions.

Discrete probability distributions are represented by simply listing the probabilities of all their possible values:

Binomial(n,θ)

$$P(X = k) = \binom{n}{k} \theta^{k} (1-\theta)^{n-k} \quad k = 0,1,2,..., n$$

Poisson(λ)

$$P(X = k) = \lambda^{k} e^{-\lambda} / k! \qquad k = 0,1,2,...$$

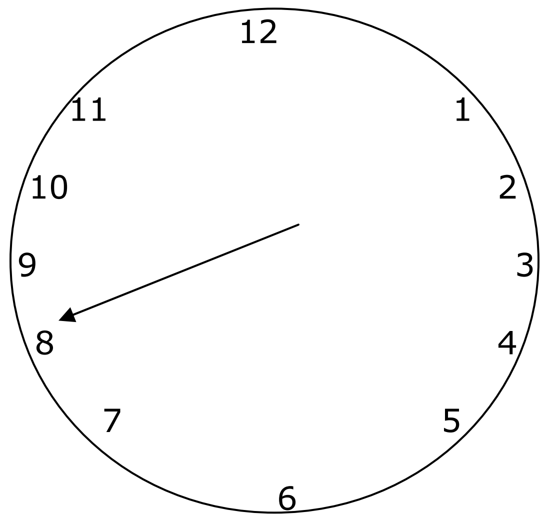
This representation of the probability distribution (i.e., listing all possible values and their probabilities) won't work for other types of variables such as

... the time until something happens
... the temperature tomorrow at noon
... your systolic blood pressure

For continuous variables like these, we must use a probability density function to represent the distribution.

Example:

I have a clock with a sweeping second hand that moves continuously (not in little jumps).
Let X be the exact position of the second hand at the instant the telephone rings.



What is the probability that the second hand will land between

6 and 12 ?	1/2	
1 and 7 ?	1/2	
3 and 6 ?	1/4	
5 and 6 ?	1/12	
4.5 and 5.5 ?	1/12	
5 and 5.5 ?	(0.5)(1/12)	= 0.04167
5 and 5.21 ?	(0.21)(1/12)	= 0.01750
5 and 5.01 ?	(0.01)(1/12)	= 0.00083

In general,

$$P(\ a \leq X \leq b\ ) = (b\text{-}a)/12 \qquad 0 \leq a \leq b \leq 12$$

What is the probability that the hand will point exactly at 3?

<u>Zero</u>. Look:

$$\begin{aligned}
P(X=3) \quad &< P(\ 2 \leq X \leq 4\ ) = 1/6 &&= 0.1667 \\
&< P(\ 2.5 \leq X \leq 3.5\ ) = 1/12 &&= 0.0833 \\
&< P(\ 2.9 \leq X \leq 3.1\ ) = (0.2)/12 &&= 0.0167 \\
&< P(\ 2.99 \leq X \leq 3.01) = (0.02)/12 &&= 0.0017 \\
\\
&< P(\ 2.99999 \leq X \leq 3.00001) &&= 0.00000167
\end{aligned}$$

...etc.

Clearly, $P(\ X = 3\ ) = 0$!

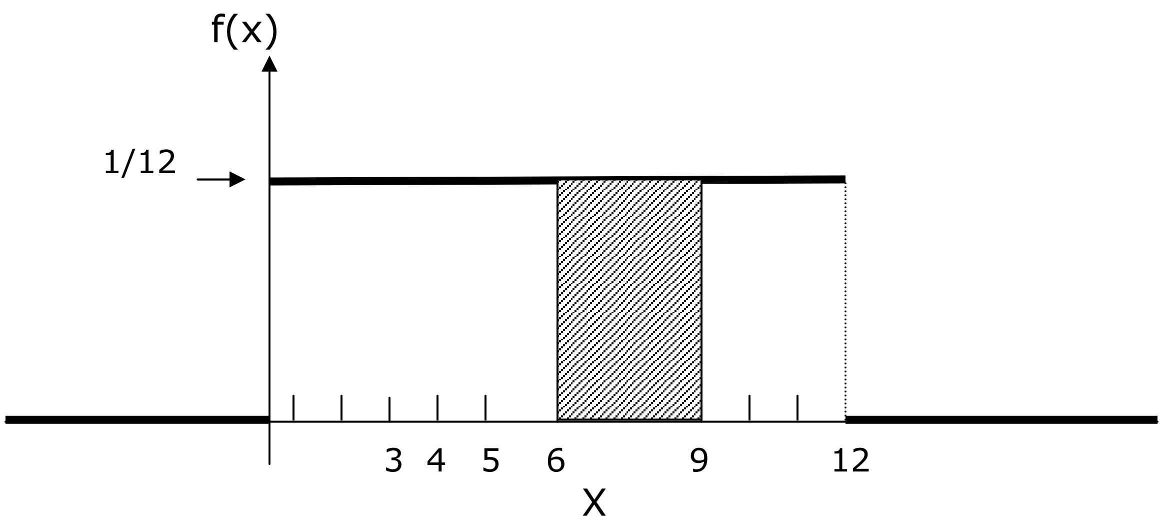
And the same is true of <u>every</u> value k:

$$P(X = k) = 0$$

As a result we cannot represent the probability distribution by listing all of the possible values and their probabilities, because the probabilities <u>are</u> <u>all</u> <u>zero</u>.

But we can represent it with a <u>probability density function</u>:

$$f(x) = \begin{cases} 0 & x < 0 \\ 1/12 & 0 \leq x \leq 12 \\ 0 & 12 < x \end{cases}$$



For any two points, like 6 and 9, the **<u>area</u>** under the probability density function represents the probability that *X* will fall between those points.

$$P(6 \leq X \leq 9) = 3(1/12) = 1/4$$
$$P(15 \leq X \leq 20) = 0$$
$$P(X < -2) = 0$$
$$P(5 \leq X \leq 5.5) = (0.5)/12 = 0.04167$$

This random variable, *X*, has a

<u>Uniform Probability Distribution</u>
<u>on the interval (0,12)</u>

The probability that *X* will fall in any interval (a,b) is represented by the area under the probability density function between a and b.

The value of the probability density function at a point, say 5, (which is 1/12 in this example) does not represent the probability that X will equal 5. That probability is zero!

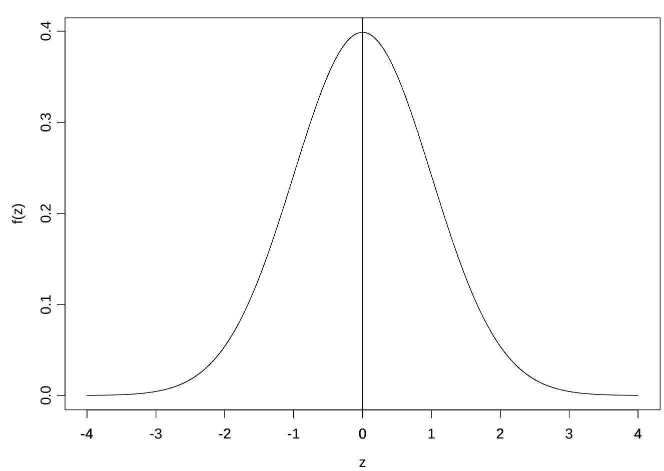
Standard Normal Distribution

The simplest continuous probability distribution is the uniform. The most important one is the standard normal distribution, whose probability density function is

(1) A random variable X is continuous
(2) S={-∞,∞}
(3) The probability distribution function (PDF) is

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$$

If *Z* is a random variable with this probability distribution, then *Z* is a "standard normal" variable.



The probability that *Z* will fall in any specified interval is represented by the area under this curve in that interval.

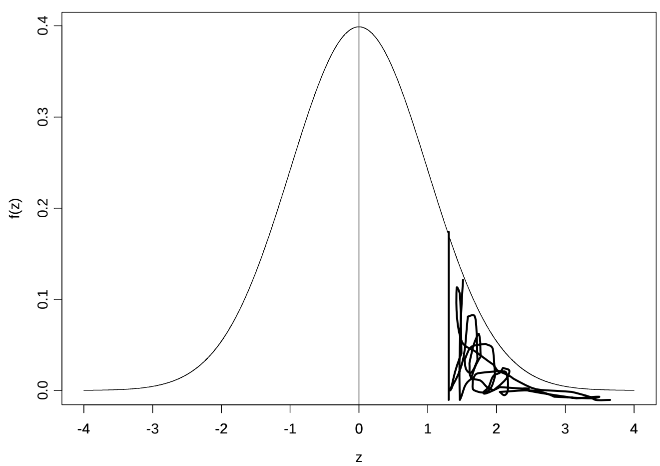
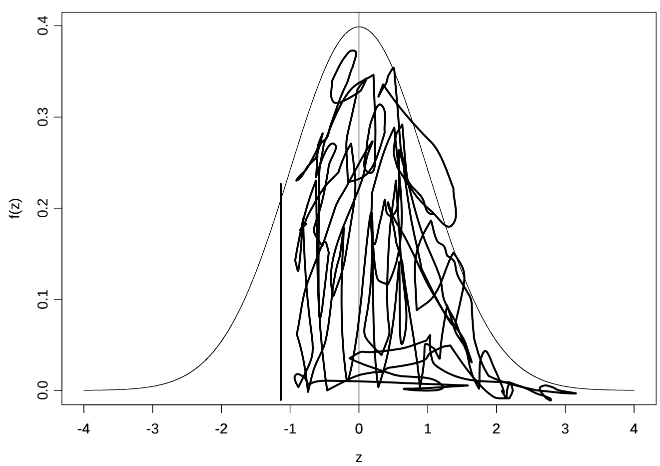


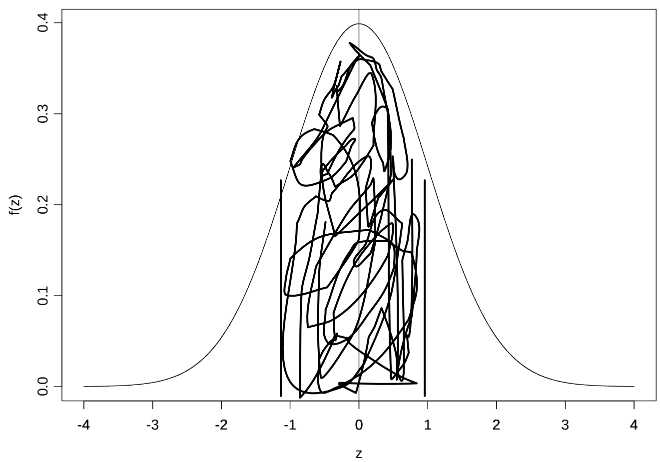
Table A.3 on page A-9 gives P(*Z* > z) for values of z>0. The probability that *Z* will fall above 1.06 is given by the shaded area. We can find this probability in Table A.3 (p.A-9). It is 0.145.

The probability that *Z* will exceed 1 is P(*Z* > 1) = 0.159. What is the probability that *Z* will exceed -1?



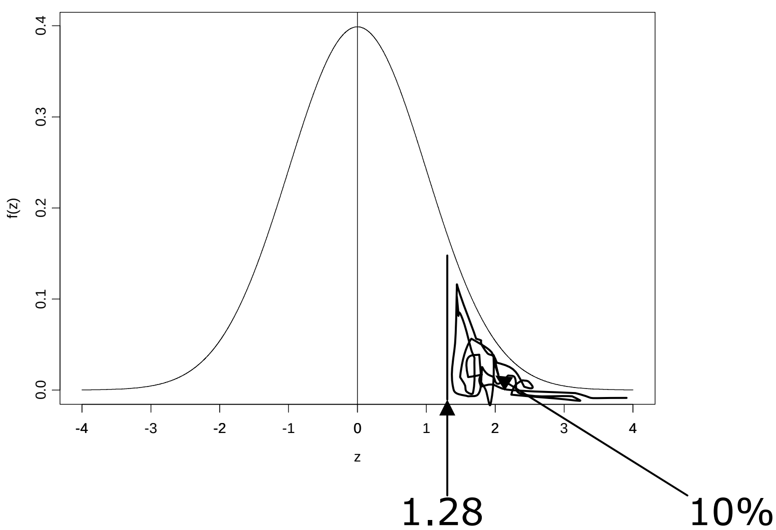
$$
\begin{aligned}
P(Z > -1) &= P(Z < 1) \\
&= 1 - P(Z > 1) \\
&= 1 - 0.159 \\
&= 0.841
\end{aligned}
$$

What is the probability that Z will fall <u>between</u> -1 and 1?



$$
\begin{aligned}
P(-1 < Z < 1) &= 1 - P(Z > 1) - P(Z < -1) \\
&= 1 - 2P(Z > 1) \\
&= 1 - 2(0.159) \\
&= 1 - 0.318 \\
&= 0.682
\end{aligned}
$$

Find the value of z that cuts off the upper 10% of the probability. That is, find the number, z, that satisfies P(Z > z) = 0.10.



Look inside the Table (A.3). Find 0.100 at z = 1.28.

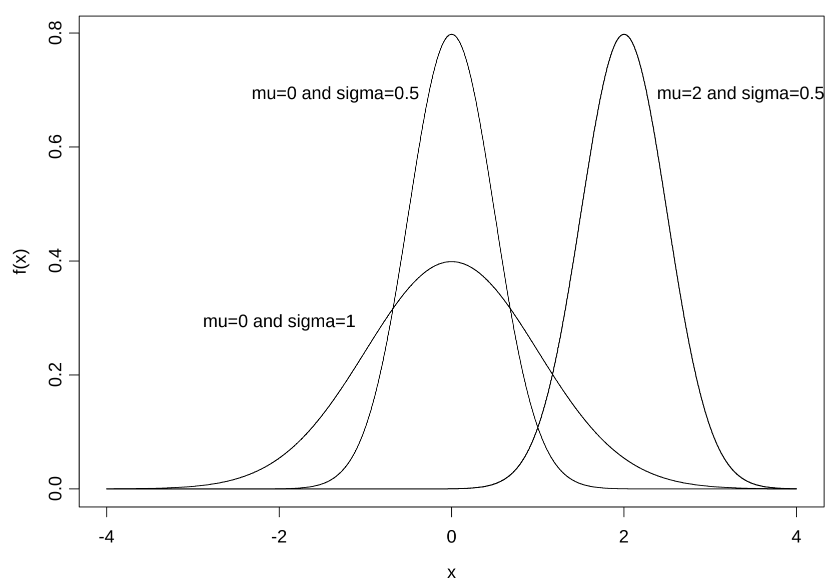
General Normal Distribution

The probability density function of the general normal probability distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

This probability distribution depends on two parameters, the variables μ ("mu") and σ ("sigma").

Varying these parameters changes the shape of the normal distribution:



The distribution is centered at μ, and σ measures how spread out it is:

μ is the mean and
σ is the standard deviation

μ determines the location of the distribution
σ determines the dispersion

Small σ means that the probability is concentrated near the mean μ.

※ **The standard normal distribution has mean $\mu = 0$ and standard deviation $\sigma = 1$.**

The variance of a probability distribution is the square of its standard deviation (SD). So a distribution with SD σ has variance σ^2.

The most common way to specify the normal distribution is in terms of its mean, μ, and variance σ^2.

$$X \sim N(\mu, \sigma^2)$$

Thus N(3, 0.5) represents the normal distribution with mean 3 and variance 0.5 (SD = $\sqrt{0.5}$).

Standardizing any Normal Distribution

If X has a normal(μ, σ^2) distribution, then

$$Z = \frac{X - \mu}{\sigma}$$

has a <u>standard</u> normal distribution, N(0, 1).

⇨ This crucial fact enables us to use Table A.3 (which gives probabilities for the <u>standard</u> normal distribution) to find probabilities for <u>any</u> normal distribution.

Example:
Let *X* be a random variable that represents systolic blood pressure. For the population of 18-74-year old males in the United States, systolic blood pressure is approximately normally distributed with mean 129 millimeters of mercury and standard deviation 19.8 millimeters of mercury.

Determine the proportion of men in the population who have a systolic blood pressure greater than 150 millimeters of mercury.

We are told that $X \sim N(\ 129,\ (19.8)^2\ )$, i.e., that the mean is 129 and the SD is 19.8.

The probability that X is greater than 150 is

$$
\begin{aligned}
P(\ X > 150\ ) &= P(\ X\text{-}129 > 150\text{-}129\ ) \\
&= P(\ (X\text{-}129) / 19.8 > (150\text{-}129) / 19.8\ ) \\
&= P(\ Z > 1.06\ )
\end{aligned}
$$

From Table A.3 we see that this equals 0.145.

Try another:

For the population of 18- to 74-year-old females in the United States, height is normally distributed with mean μ = 63.9 inches and standard deviation σ = 2.6 inches. If we randomly select a woman between the ages of 18 and 74 from this population, what is the probability that she is between 60 and 68 inches tall?

Select a random female and observe her height, letting X represent the result. We use a <u>normal distribution</u> <u>model</u> to approximate the distribution of X : $X \sim N(\ 63.9, (2.6)^2\ )$. Then

$$
\begin{aligned}
P(\ 60 < X < 68\ ) &= P(60-63.9 < X-63.9 < 68-63.9\ ) \\
&= P(\ -3.9 < X-63.9 < 4.1\ ) \\
&= P\left(\frac{-3.9}{2.6} < \frac{X-63.9}{2.6} < \frac{4.1}{2.6} \right) \\
&= P(-1.50 < Z < 1.58\ ) \\
&= 1 - P(Z > 1.58) - P(Z > 1.50) \\
&= 1 - 0.057 - 0.067 \\
&= 0.876
\end{aligned}
$$

We might also want to know what value of height cuts off the upper 5% of the distribution.

Now we must find the number, x, that satisfies

$$P(\, X > x \,) = 0.05.$$

First we standardize, so that we can use the table of the standard normal distribution:

$$\begin{aligned} P(\, X > x \,) &= P(\quad X\text{-}63.9 \quad > \quad x\text{-}63.9 \quad) \\ &= P(\, (X - 63.9)/2.6 > (x - 63.9)/2.6 \,) \\ &= P(\qquad Z \qquad > (x - 63.9)/2.6 \,). \end{aligned}$$

Now from a better table than Table A.3 (or taking the value half way between 1.64 and 1.65 in Table A.3) we find that

$$P(\, Z > 1.645 \,) \;=\; 0.05,$$

so we see that $(x - 63.9)/2.6$ must equal 1.645.

$$(x - 63.9)/2.6 \;=1.645$$

$$x=(2.6)(1.645) + 63.9= 68.2$$

Thus only 5% of the women are taller than 68.2 inches, (just a little over 5 feet, 8 inches).