Probability
An intro for calculus students

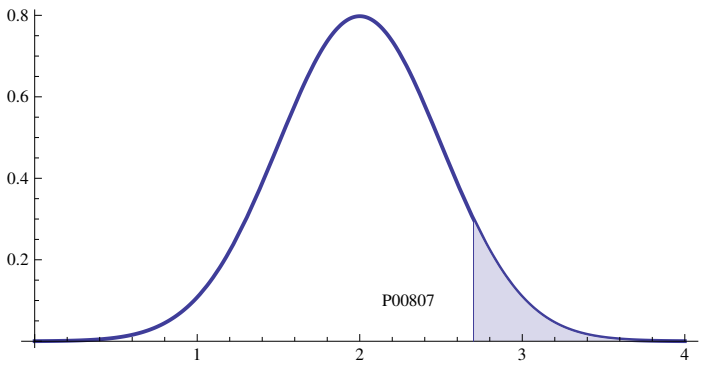

Figure 1: A normal integral

[fig:aNormalIntegral]

Suppose we flip a coin 20 times; what is the probability that we get more than 12 heads? Suppose we roll a six-sided die 9 times; what is the probability that our sum total exceeds 20? What is the probability that a college graduate will earn $50000/year, as compared to a high school graduate? These questions, and many like them, can be answered by integrating a probability distribution function.

Continuous and discrete distributions

The function shown in figure [fig:aNormalIntegral] is an example of a *continuous distribution*. To understand this and how it relates to probabilistic computations, we should first examine a few simpler distributions.

Uniform distributions

Suppose we pick a *real number* randomly from the interval $[0, 1]$. What does that even mean? What is the probability we pick 1 or 0.1234 or $1/\pi$? What is the probability that our pick lies in the left half of the interval? One way to make sense of this is to suppose the probability that our pick lies in any particular interval is proportional to the length of that interval. This might make sense if, for example, we choose the number by throwing a dart at a number line while blindfolded. Then, the answer to our second question should be $1/2$. The probability that our pick lies in the interval $[0, 0.3]$ should be $3/10$.

More generally, we can express such a probability via integration against a probability density function. A *probability density function* is simply a non-negative function whose total integral is 1; i.e.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

In our example involving $[0, 1]$ our probability density function would be

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{else.} \end{cases}$$

Then, the probability that a point chosen from $[0, 1]$ lies in the left half of the interval is

$$\int_0^{1/2} 1\,dx = \frac{1}{2}.$$

The probability that we pick a number from the interval $[0, 0.3]$ is the area of the darker, rectangular region shown in figure [fig:uniformDist].

[fig:uniformDist]

In some sense, this is a natural generalization of a discrete problem: Pick an integer between 1 and 10 uniformly and at random. In that case, it makes sense to suppose that each number has an equal probability $1/10$ of being chosen. The probability of choosing a 1, 2, or 3 would be $1/10 + 1/10 + 1/10$ or $3/10$; this is called a *uniform discrete distribution*. The sub-rectangles indicated by the dashed lines in figure [fig:uniformDist] are meant to emphasize the relationship, since they all have area $1/10$. A discrete visualization of this is shown in the top of figure [fig:aDiscretDistribution]. The bottom of figure [fig:aDiscretDistribution] illustrates the uniform discrete distribution on the numbers $\{1, 2, \ldots, 100\}$. Note how the continuous uniform distribution on $[0, 1]$ shown in figure [fig:uniformDist] appears to be a limit of these discrete distributions, after rescaling.

[fig:aDiscretDistribution]

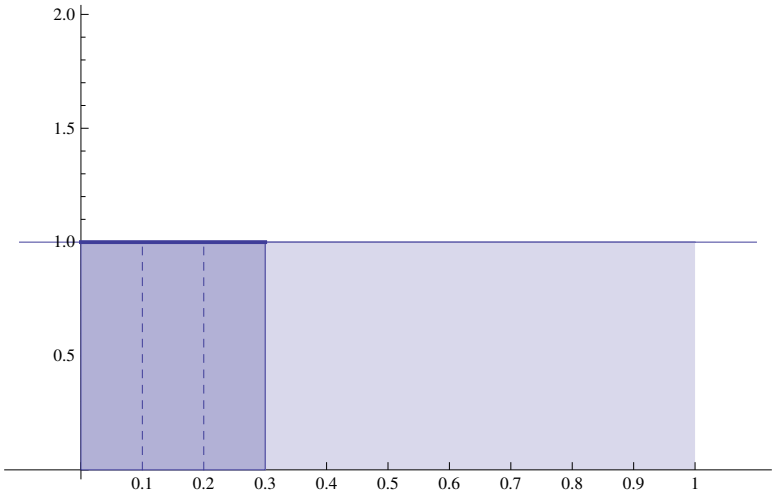

Figure 2: The uniform distribution on $[0, 1]$

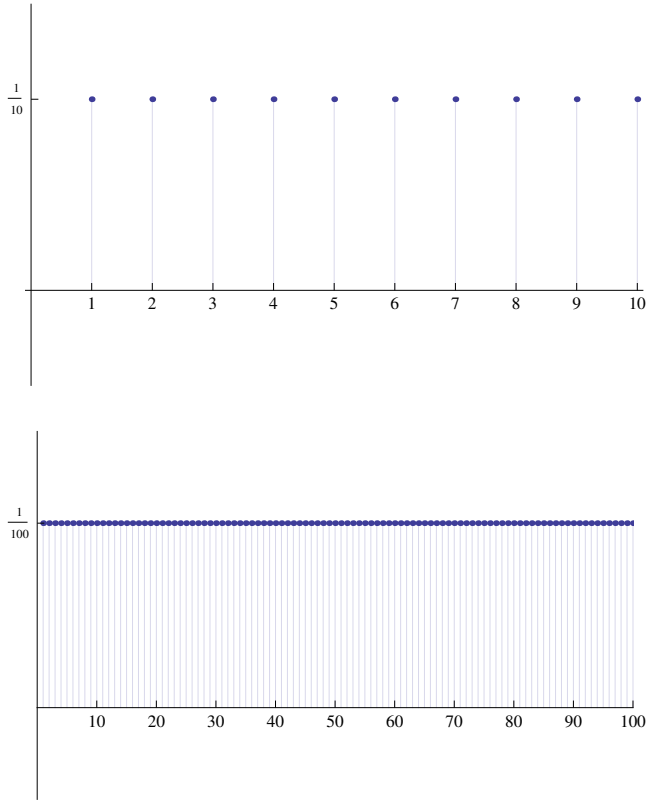

Figure 3: Uniform discrete distributions

Now suppose we pick an integer between 1 and 1000, all with equal probability 1/1000. Then the probability of generating a number between 1 and 314 would be

$$\sum_{1}^{314} \frac{1}{1000} = \frac{314}{1000} = \int_{0}^{0.314} 1dx.$$

I've included the integral here to emphasize the relationship with the continuous distribution. In a real sense, the continuous, uniform distribution on $[0, 1]$ is a limit of discrete distributions.

A bell shaped distribution

Now suppose we generate an integer between 0 and 10 by flipping a coin 10 times and counting the number of heads. There are 11 possible outcomes, but they are not all equally likely. The probability of generating a zero is $1/2^{10} = 1/1024$, which is much smaller than 1/11. This is because we must throw a tail on each throw and the throws are independent of one another. Since the probability of getting a tail on a single throw is 1/2, the probability of getting 10 straight heads is $1/2^{10}$. The probability of generating a 1 is $10/2^{10}$, since the single head could occur on any of 10 possible throws; this probability is ten times bigger than the probability of a zero, yet still much smaller than 1/11.

In a discrete math class or introductory statistics class, we would talk carefully about the binomial coefficients:

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}.$$

This is read n choose k and represents the number of ways to choose k objects from n given objects. Thus, if we flip a coin n times and want exactly k heads, there are n choose k possible ways to be successful. If, for example, we flip the coin five times and want exactly two heads, there are

$$\left(\begin{array}{c} 5 \\ 2 \end{array} \right) = \frac{5!}{2!(5-2)!} = 10$$

ways to make this happen. These are all illustrated in figure [fig:headsTails]. Note that each particular sequence of heads and tails has equal probability $1/2^5$ of occurring. Thus, the probability of getting exactly 2 heads in five flips is 10/32.

[fig:headsTails]

More generally, the probability of getting exactly k heads in n flips is

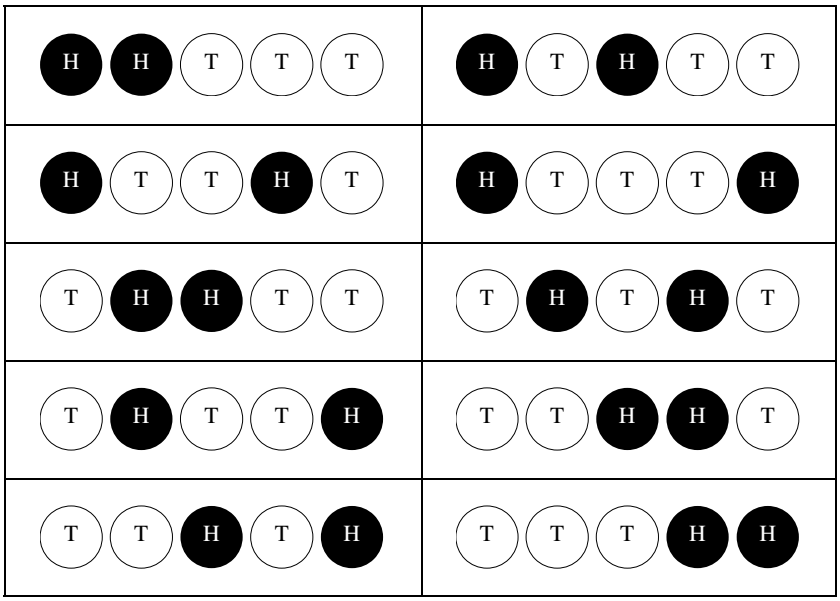

Figure 4: Ways to get two heads in five flips

$$\left(\begin{array}{c} n \\ k \end{array} \right) \frac{1}{2^n}.$$

We can plot these numbers in a manner that is analogous to the uniform discrete distributions shown in figure [fig:aDiscretDistribution]; the result is shown in figure [fig:binomialDistributions]. Note that each discrete plot is accompanied by a continuous curve that approximates the points very closely. There is a particular formula for this curve that defines a continuous distribution, called a *normal distribution*. This continuous distribution is, in a natural sense, the limit of the discrete distributions when properly scaled. A basic understanding of the normal distribution is our primary objective here. We've got a bit more notation we'll have to slog through first, however.

[fig:binomialDistributions]

Formalities

Let's try to write down some careful definitions for all this. The outcome of a random experiment (tossing a coin, throwing a dart at a number line, etc.) will be denoted by X. Probabilists would call X a *random variable*. We can feel that we thoroughly understand X if we know its *distribution*. The two broad classes

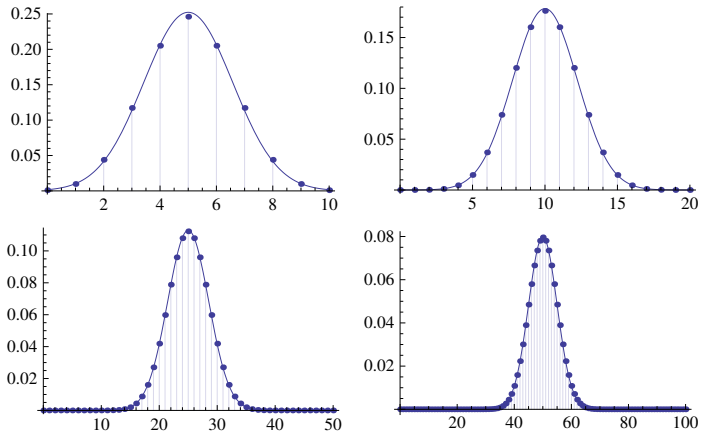

Figure 5: Binomial distributions together with their normal approximations.

of distributions we've seen to this point are *discrete* and *continuous* leading to discrete or continuous random variables.

A discrete random variable X takes values on a discrete set, like $\{0, 1, 2, \ldots, n\}$ and a discrete distribution is simply a list of non-negative probabilities, like $\{p_0, p_1, p_2, \ldots, p_n\}$ associated with these that add up to one. The uniform discrete distribution, for example, takes all these probabilities to be the same. The binomial distribution weights the middle terms much more heavily. In either case, the probability that X takes on some particular value i is simply p_i. To compute the probability that X takes on one of a set S of values, we simply sum the corresponding p_is, i.e. we compute

$$\sum_{i \in S} p_i.$$

A continuous random variable X takes its values in an interval or even the whole real line $\mathbb{R}$. The distribution of X is a non-negative function $f(x)$. To compute the probability that X lies in some interval $[a, b]$, we compute the integral

$$\int_a^b f(x)dx.$$

Of course, a real valued random variable must take on *some* value. That is the probability of choosing some number must be one. Thus, we require that

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

Measures of distributions

There are two very general and important descriptive properties defined for distributions, namely the mean μ and standard deviation σ. We must understand these to understand how the normal distributions are related to the binomial distributions.

Mean and standard deviation for discrete random variables

As we've just described, if X is a random variable taking on values $\{0, 1, \ldots, n\}$, it's distribution is simply the list $\{p_0, p_1, \ldots, p_n\}$ where p_k indicates the probability that $X = k$. The *mean* μ of a distribution simply represents the weighted average of its possible values. We express this concretely as

$$\mu = \sum_k k p_k.$$

For example, if we choose a number $\{0, 1, 2, 3, 4\}$ uniformly (so each term has probability $p = 1/5$), then the mean is

$$\mu = \frac{(0 + 1 + 2 + 3 + 4)}{5} = 2,$$

exactly as we'd expect. The mean of the binomial distribution is also "near the middle" but distributions can certainly be weighted otherwise.

The binomial distribution is particularly useful for us, since we ultimately want to understand the normal distribution. Recall that a binomially distributed random variable is constructed by flipping a coin n times and counting the number of heads. If we flip a coin once, we generate either a zero or a one with probability $1/2$ each. Thus, the mean of one coin flip is $1/2$. If we add random variables, then their means add. Thus, the mean of the binomial distribution with n flips is $n/2$. This reflects the fact that we expect to get a head about half the time.

Standard deviation σ, and its square the *variance* σ^2, both measure the dispersion of the data; the larger the value of σ, the more spread out the data. They're quite similar conceptually but sometimes one is more easy to work with than the other. The variance of a random variable with mean μ is defined by

$$\sigma^2 = \sum_k (k - \mu)^2 p_k.$$

Note that the expression $k - \mu$ is the (signed) difference between the particular value and the average value. We want to measure how large this is on average

so we take the weighted average. It makes sense to square first, since we don't want the signs to cancel.

The variance of our coin flip example is

$$\sigma^2 = \left(0 - \frac{1}{2}\right)^2 \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \frac{1}{2} = \frac{1}{4}.$$

It follows that the standard deviation is $\sigma = 1/2$. If we add random variables, then their variances add. Thus, the variance of the binomial distribution with n flips is $n/4$ and its standard deviation is $\sqrt{n}/\,2$.

Mean and standard deviation for continuous random variables

The mean, standard deviation, and variance of continuous probability distributions can be defined an a way that is analogous to discrete distributions. In particular, the mean μ and variance σ^2 are defined by

$$\mu = \int_{-\infty}^{\infty} xp(x)dx$$

and

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx.$$

As with discrete distributions, the standard deviation is the square root of the variance.

Suppose, for example that X is uniformly distributed on the interval $[a, b]$. Thus, X has distribution

$$p(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{else.} \end{cases}$$

Thus, we can compute the mean as follows:

$$\frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\frac{1}{2}x^2\Big|_a^b = \frac{1}{2(b-a)}\left(b^2-a^2\right) = \frac{a+b}{2}.$$

This is, of course, exactly what we'd expect. In your homework, you'll show that $\sigma^2 = (b-a)^2/\,12$. Note that the larger the interval, the larger the variance.

An example continuous distribution

Here's an example continuous distribution which is complicated enough to be interesting yet simple enough to do some computations. We'll take our distribution function to be

$$p(x) = \begin{cases} \frac{1}{(1+x)^2} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

Note that

$$\int_0^\infty \frac{1}{(1+x)^2}\, dx = \lim_{b\to\infty} -\frac{1}{1+x}\Big|_0^b = 1.$$

Thus, p is a good probability density function. The graph of $p(x)$ is shown in figure [fig:paretoPDF].

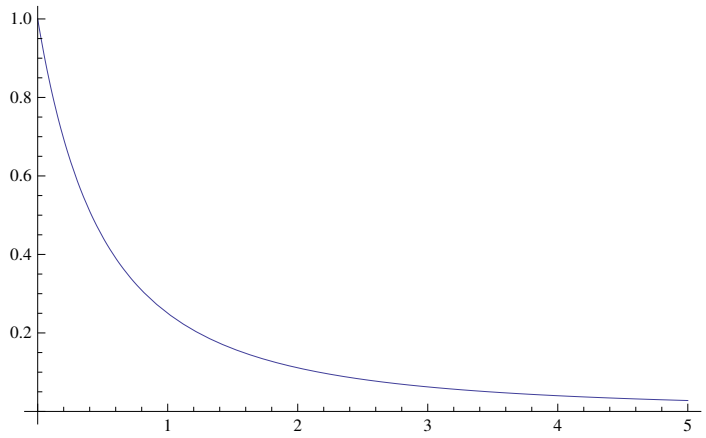

Figure 6: The graph of our simple distribution

[fig:paretoPDF]

The shape of the graph of $p(x)$ indicates that this density function is more likely to generate a number close to zero, than far away. More precisely, we can compute the probability that we generate a number between zero and one as follows:

$$\int_0^1 \frac{1}{(1+x)^2}\, dx = -\frac{1}{1+x}\Big|_0^1 = \frac{1}{2}.$$

The probability that we generate a number between two and four, on the other hand is only 2/15. We could use a computer to generate thousands of numbers

with this distribution and plot the corresponding histogram. The result is shown in figure [fig:paretoHistogram], together with a plot of the distribution function.

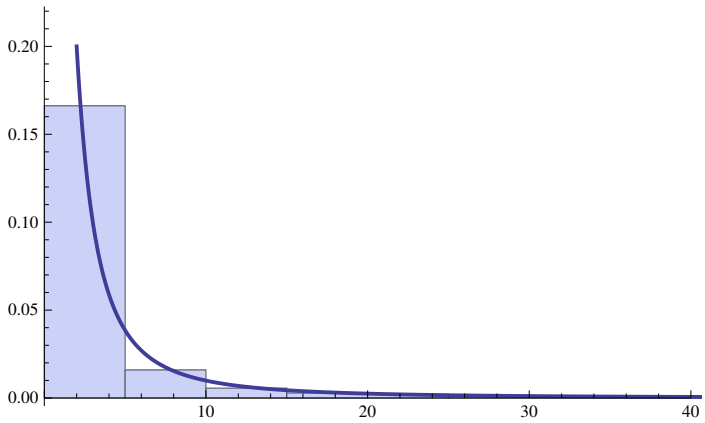

Figure 7: A histogram generated by our simple probability density function

[fig:paretoHistogram]

This distribution is an example of a *Pareto distribution*, which has been used to model distribution of wealth among other things. The general form of a Pareto distribution is

$$p(x) = \begin{cases} \frac{\alpha}{k}\left(\frac{k}{k+x-\mu}\right)^{\alpha+1} & x \geq \mu \\ 0 & x < \mu. \end{cases}$$

In the example above, $\mu = 0$ and $\alpha = k = 1$. In your homework, you'll play with Pareto distributions that might reasonably be used to model distribution of income.

The normal distribution

One of the most important, perhaps *the* most important, continuous distributions is the *normal distribution*.

Definition

The formula for the normal distribution with mean μ and standard deviation σ is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\left(2\sigma^2\right)}.$$

The graphs of several normal distributions are shown in figure [fig:normalDistributions]. When $\mu = 0$ and $\sigma = 1$ in equation [eq:normalDistribution], we get the *standard normal*. Thus, the probability distribution of the standard normal is

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

The standard normal is symmetric about the vertical axis in figure [fig:normalDistributions].

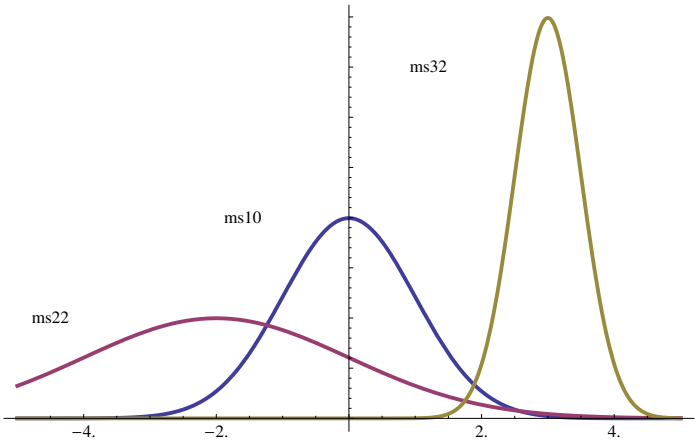

Figure 8: Several normal distributions

[fig:normalDistributions]

Relating normal distributions

Any normal distribution is related to the standard normal distribution because changing μ or σ in equation [eq:normalDistribution] changes the graph in predictable ways. A change of μ simply shifts the graph to the left of right; this changes the mean of the distribution, which is located where the maximum occurs. Reducing the size of σ increases the maximum value and concentrates the graph about that maximum value.

A major difficulty surrounding the normal distribution is that we *it has no elementary anti-derivative*! Elementary statistics courses get around this by providing a table of numerically computed values of

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_0^b e^{-x^2/2} dx.$$

From that information, one can immediately compute all sorts of integrals involving the standard normal. For example,

$$\frac{1}{\sqrt{2\pi}} \int_{-1}^2 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-x^2/2}\, dx + \frac{1}{\sqrt{2\pi}} \int_0^2 e^{-x^2/2} dx$$

and both of the integrals on the right can be computed from the table. Furthermore, integrals involving *any* normal distribution can be computed in terms of the standard normal. While the trick is described in an elementary statistics class, it ultimately boils down to the following formula:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\alpha}^{(b-\mu)/\sigma} e^{-x^2/2} dx.$$

One can use the substitution $u = (x - \mu)/\sigma$ to verify this.

The central limit theorem

There are two big theorems in probability theory - the law of large numbers and the central limit theorem; it is the second of these that explains the importance of the normal distribution. Both deal with a sequence of independent random variables $X_1, X_2, \ldots$ that all have the same distribution. The law of law large numbers simply states that, if each X_i has mean μ, then

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is almost certainly close to μ. That is, flip a coin a bunch of times and it will come up heads around half the time.

The central limit theorem states more precise information about the distribution of $\bar{X}_n$. Technically, the central limit theorem states that if each X_i has mean μ and standard deviation σ, then the random variable $\sqrt{n}\left(\bar{X}_n - \mu\right)$ converges to the normal distribution with mean 0 and standard deviation σ. In practice this means that we can approximate $S_n = X_1 + X_2 + \cdots + X_n$ using a normal distribution. Now the mean of S_n will be $n\mu$ and its standard deviation will be $\sqrt{n}\sigma$. Thus, we must approximate using the normal distribution with this same mean and standard deviation. That is

$$p(x) = \frac{1}{\sqrt{2n\pi}\sigma} e^{-(x-n\mu)^2/\left(2n\sigma^2\right)}.$$

It is important to understand that the distributions of the X_i play no role here; all that is important is that they be independent and have the same distributions. Thus, no matter what the distribution of the original X_is, their average will be approximately normal!

Examples

Coin flipping

Suppose we flip a coin 99 times. What is the probability that we get fewer than 47 heads?

Solution: As we've seen, the mean and standard deviation of a single coin flip are both 1/2. By the central limit theorem, the sum of n coin flips is approximately normally distributed with mean and standard deviation $n/2$ and $\sqrt{n}/\,2$ respectively. Taking $n = 99$ in formula [eq:centralLimitNormalIntegral], we find that we should evaluate the following integral.

$$\int_{-\infty}^{46.5} \frac{2}{\sqrt{2\ 99\pi}} e^{\,-(x-99/2)^2/(299/4)} dx$$

The upper bound of 46.5, rather than 47 arises as an adjustment to relate the discrete and continuous distributions. This integral *must* be evaluated numerically; we can do so with Sage as follows:

```
f(x) = 2*exp(-((x-99/2)**2)/(2*99/4))/sqrt(2*99*pi)
numerical_integral(f,-oo,46.5)

#Out: (0.27324679770329097, 8.621934595567032e-11)
```

This particular example can also be done using the binomial distribution. In fact, the answer computed by Sage is *exactly*

```
k = var('k')
sum(factorial(99)/(factorial(k)*factorial(99-k)*2**99), k,0,46)

#Out: 135359702272832325591553024 7/4951760157141521099596496896
```

The normal integral is an approximation, but it is a very good one. The difference between the previous two computations is about 0.000109944.

The real power arises when we have a very large number of trials - as might happen in a problem in statistical mechanics. For example, what's the probability of getting fewer than 500001000 heads in 1000000000 tosses? The binomial approach has half a billion terms in the sum but the normal integration approach is no harder. We still need to compute the integral with a numerical integrator, like Sage.

```
n = 10**9
b = 500001000.5
f(x) = 2*exp(-(x-n/2)**2/(2*n/4))/sqrt(2*n*pi)
integral_numerical(f,-oo,b)

#Out: (0.5253288961960733, 0.16413653376826715)
```

Pretty cool, eh?

Dice

Suppose we want to compute the mean and standard deviation of the roll of a standard six sided die. If we roll 100 six sided dice, what are the odds that our sum total is at least 400?

The distribution is simply $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$. Thus, we can compute μ and σ as follows.

$$\mu = \sum_{k=1}^{6} \frac{k}{6} = \frac{7}{2}$$
$$\sigma^2 = \sum_{k=1}^{6} (k - 7/2)^2/6 = \frac{35}{12}$$

If we roll 100 such dice, then the outcome is approximately normal with mean 100μ and standard deviation 10σ. Thus, the density function is

$$\frac{1}{\sqrt{2\pi}10\sigma} e^{-(x-100\mu)^2/\left(200\sigma^2\right)}$$

Where μ and σ are already defined. Thus the probability that our sum is at least 400 is

$$\frac{1}{\sqrt{2\pi}10\sigma} \int_{399.5}^{\infty} e^{-(x-100\mu)^2/\left(200\sigma^2\right)} dx \approx 0.00187522.$$

Again, that integral must be estimated numerically, here's how to do it with Sage.

Income

Let us suppose that average income in the US is \$44460/year with a standard deviation of \$48690. More precisely, we'll suppose that the distribution function $p(x)$ is given by

$$p(x) = \frac{3.5 \times 10^{55}}{(380000 + x)^{10.775}},$$

for $x > \$1037$.

(a) What is the probability that a randomly chosen individual earns more than \$50000?

(b) Suppose we pick 20 people at random. What is the probability that their collective income exceeds \$1000000?

Comments: The two parts are very different. For the first part, we'll use the given distribution $p(x)$, since the question is about one randomly chosen individual. The second part asks about the *sum of incomes of randomly chosen people*. As a result, we'll answer the question using a normal distribution with the proper mean and standard deviation.

The function $p(x)$ is based on data I obtained from the American Community Survey. It is an example of a Pareto distribution with $\mu = 1397$, $k = 381307$, and $\alpha = 9.77$. While over-simplified to be sure, it does a reasonable job for the purposes here. The lower bound \$1307 might be though of as a "minimum amount earned". Mathematically, there must be some lower bound because the integral of the function over all of $\mathbb{R}$ diverges.

Solution: To solve part (a), we simply use the given distribution function $p(x)$.

$$\int_{50000.5}^{\infty} \frac{3.5 \times 10^{55}}{(380000 + x)^{10.775}}\, dx = -\lim_{b\to\infty} \frac{3.5 \times 10^{55}}{(380000 + x)^{9.775} 9.775}\Bigg|_{50000.5}^{b} = 0.306755.$$

To solve part (b), we use a normal distribution with mean 20×44460 and standard deviation $\sqrt{20} \times 48690$. Thus we get

$$\frac{1}{\sqrt{2\pi \times 20} 48690} \int_{1000000.5}^{\infty} e^{-(x-20\times 44460)^2/(2\times 20\times 48690^2)} dx \approx 0.30543.$$

Remarkably close, but a bit smaller.

Problems

1. Referring to the table of standard normal integrals on the last page, compute the following.

   (a) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{1.3} e^{-x^2/2}\, dx$

(b) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-0.4}^{1.3} e^{-x^2/2}\, dx$

(c) $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{0.4}^{1.3} e^{-x^2/2}\, dx$

2. Using u-substitution, convert the following normal integrals into standard normal integrals. Then evaluate the integral using the table on the last page or your favorite numerical integrator.

(a) $\dfrac{1}{\sqrt{2\pi}2} \displaystyle\int_{0}^{1} e^{-(x-1)^2/8}\, dx$

(b) $\dfrac{1}{\sqrt{2\pi}4} \displaystyle\int_{12}^{18} e^{-(x-10)^2/32}\, dx$

3. Given that

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-x^2/2}\, dx = \frac{1}{2},$$

show that

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mu}^{\infty} e^{-(x-\mu)^2/\left(2\sigma^2\right)}\, dx = \frac{1}{2},$$

for all $\mu \in \mathbb{R}$ and $\sigma > 0$.

4. Below we see three probability distributions. I used each of these to generate 100 points and plotted the results in figure [fig:scatter]. Match the distribution functions with the point plots.

(a) $\dfrac{1}{\sqrt{2\pi}0.3} e^{-\frac{(x-1)^2}{2\cdot 0.3^2}}$ over $(-\infty, \infty)$

(b) $\dfrac{1}{\sqrt{2\pi}0.7} e^{-\frac{(x-1)^2}{2\cdot 0.7^2}}$ over $(-\infty, \infty)$

(c) $\dfrac{\log(5)}{24} 5^{2-x}$ over $[0, 2]$

5. For each of the following functions, find the constant c that makes the function a probabiltiy distribution over the specified interval.

(a) $cx(x-1)$ over $[0, 1]$

(b) $c2^x$ over $[0, \infty]$

(c) $c\sqrt{1-(x-1)^2}$ over $[0, 2]$

6. Compute the mean μ and standard deviation σ of the following distributions.

(a) The uniform distribtion over $[a, b]$

(b) The exponential distribution $p(x) = e^x$ over $[0, \infty]$

(c) The standard normal distribution

7. Suppose we flip a coin 1000 times. Use a normal integral to find the probability that you get more than 666 heads.

8. Suppose we roll a standard six sided die 12 times. Use a normal integral to find the probability that your rolls total more than 50.

9. Suppose we roll a fair 10 sided die 10 times. Use a normal integral to find the probability that your rolls total more than 60.

10. Compute the probability that a college graduate earns at least \$50000 and the probability that a high school graduate earns that same amount. For the purposes of this problem, suppose that the distribution functions p_h and p_c that describe the distribution of income for high school and college graduates respectively are

$$p_h(x) = \frac{4.64 \times 10^{1013}}{(4405254 + x)^{153.24}}$$

and

$$p_c(x) = \frac{1.415 \times 10^{229}}{(2113747 + x)^{36.983}}.$$

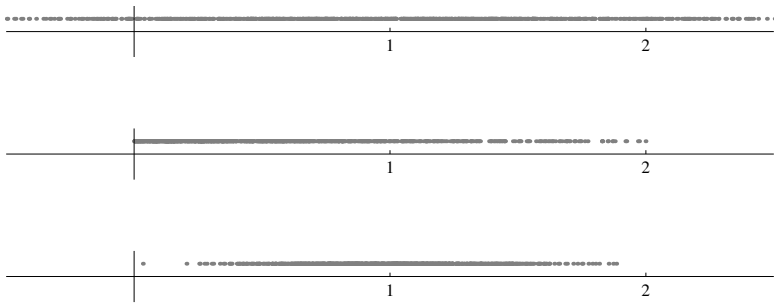

Figure 9: Three sets of randomly generated points

[fig:scatter]

Table of standard normal integrals

b	$\frac{1}{\sqrt{2\pi}} \int_0^b e^{-x^2/2}\, dx$
0	0.
0.1	0.0398278
0.2	0.0792597
0.3	0.117911
0.4	0.155422
0.5	0.191462
0.6	0.225747
0.7	0.258036
0.8	0.288145
0.9	0.31594
1	0.341345
1.1	0.364334
1.2	0.38493
1.3	0.4032
1.4	0.419243
1.5	0.433193
1.6	0.445201
1.7	0.455435
1.8	0.46407
1.9	0.471283
2	0.47725