

Probability and statistics

We will be using several of Maple's probability and statistics functions in Math 151. Most of these functions are used the same way that standard mathematical functions (like sin, cos, ln etc.) are used -- you just need to be sure to know what the inputs and outputs of the functions mean. Some of Maple's statistical functions are in the special "stats" library and are accessed in a way that is a little different than most of Maple's functions. But all this is illustrated below.

Combinatorial functions:

1. **Permutations:** there are two special functions for permutations that are useful for counting problems. Of course, the number of permutations of the elements of a set of n distinct elements is $n!$, and the standard factorial notation is used in Maple:

```
> 15!;  
1307674368000
```

The two special Maple functions are found in the "combinat" library (and must be "with"ed -- they are `numbperm` and `permute`):

```
> with(combinat,numbperm,permute);  
[numbperm, permute]
```

The `numbperm` function tells how many permutations there are of a list, which must be enclosed in square brackets [] -- the list may have duplicate elements:

```
> numbperm([a,b,c]);  
6  
> numbperm([a,a,b]);  
3
```

It is also possible to ask for the number of different ordered subsets having a specified number of elements of a list. For example, for the number of different 3-element lists taken from the list [a,a,b,c], we would say:

```
> numbperm([a,a,b,c],3);
```

12

Next, the `permute` function acts like the `numbperm` function, except instead of saying how many permutations there are, the `permute` function simply lists them all. For instance:

```
> permute([a,b,c]);
```

```
[[a,b,c],[a,c,b],[b,a,c],[b,c,a],[c,a,b],[c,b,a]]
```

```
> permute([a,a,b]);
```

```
[[a,a,b],[a,b,a],[b,a,a]]
```

```
> permute([a,a,b,c],3);
```

```
[[a,a,b],[a,a,c],[a,b,a],[a,b,c],[a,c,a],[a,c,b],[b,a,a],[b,a,c],[b,c,a],  
 [c,a,a],[c,a,b],[c,b,a]]
```

You get the idea -- these three examples are consistent with the previous three. The most important thing to remember when using `permute` and `numbperm` is that the list being permuted must be enclosed in square brackets.

2. Combinations: There are two (or three) functions which do the same thing as those above, except for combinations (unordered lists). They are "binomial " (which is always available), "numbcomb " and "choose " (the latter two must be "with"ed from the `combinat` package):

```
> with(combinat,choose, numbcomb);
```

```
[choose, numbcomb]
```

First, `binomial` is used to calculate binomial coefficients -- `binomial(n,k)` is the number of ways to choose `k` things out of `n`:

```
> binomial(6,2);
```

15

Next, `numbcomb` does the same thing except the first argument of `numbcomb` may be a list (enclosed in square brackets, just like `numbperm`) instead of a number:

```
> numbcomb([a,b,c,d,e,f],2);
```

15

Finally, choose produces the list of all ways to choose the subsets whose number is reported by numbcomb :

```
> choose([a,b,c,d,e,f],2);
```

```
[[a, b], [a, c], [a, d], [a, e], [a, f], [b, c], [b, d], [b, e], [b, f], [c, d], [c, e], [c, f], [d, e],  
 [d, f], [e, f]]
```

STATISTICAL FUNCTIONS :

Two kinds of Maple's statistical functions will be useful in Math 151. They are the functions that calculate "descriptive statistics" for a set of data -- i.e., numbers like the mean, median, variance and standard deviation. The other kind of useful functions are those that give values of probability distributions or their related cumulative distribution functions.

Descriptive statistical functions :

As indicated above, these are the functions that calculate means, medians and such of sets of data. Although Maple allows you to input the data in a variety of ways, we will use only one of them. You might find it useful later to explore some of the other descriptive statistical functions Maple can compute, and the other ways to enter data (or read it in from external files).

The statistical functions, like the combinatorial functions, are stored in libraries and must be loaded from the disk before they can be used. To get the descriptive statistical functions, one uses both of the following commands:

```
> with(stats,describe);
```

```
[describe]
```

```
> with(describe);
```

```
[coefficientofvariation, count, countmissing, covariance, decile, geometricmean,  
harmonicmean, kurtosis, linearcorrelation, mean, meandeviation, median, mode, moment,  
percentile, quadraticmean, quantile, quartile, range, skewness, standarddeviation, sumdata,  
variance]
```

All of the descriptive statistical functions can do their computations on a data list. This is simply a list of numbers enclosed in square brackets. The functions for mean, median, variance, and standard deviation are called "mean", "median", "variance" and "standarddeviation", respectively.

1. The mean function calculates the mean of a list of numbers -- the list must be enclosed in square brackets:

```
> mean([3,6,4.2,7,7,2,3]);  
4.600000001
```

You can also name the list ahead of time (so you can calculate mean and variance without typing the list twice, for example):

```
> data:=[3,6,4.2,7,7,2,3];  
data := [3, 6, 4.2, 7, 7, 2, 3]  
> mean(data);  
4.600000001
```

2. The variance function has the same syntax as the "mean" function, except it computes the variance of the list:

```
> variance([3,6,4.2,7,7,2,3]);  
3.645714288  
> variance(data);  
3.645714288
```

3. The standard deviation is just the square root of the variance, but there is also the

Maple function "standarddeviation " for this in the "describe " subset of the "stats " package:

```
> standarddeviation(data);  
1.909375366
```

Now that you see the pattern, you can figure out how to use the Maple function median to compute the median of a data list.

Statistical Distribution Functions

There is another sub-package of the stats package that deals with probability distributions -- it is called statevalf , and it must be loaded into computer memory using both of the commands:

```
> with(stats,statevalf);  
[statevalf]  
> with(statevalf);  
[cdf, dcdf, icdf, idcdf, pdf, pf]
```

The commands within statevalf correspond to the operations one wishes to perform on either discrete probability distributions (like the binomial distribution) or continuous probability distributions (like the normal distribution). The operations are

1. Evaluate the probability density function at a given value for a given random variable.
2. Evaluate the cumulative distribution function at a given value of a random variable (to find the probability that a random sample will yield a value less than or equal to the given value). This operation answers questions of the form "What is the probability that a sample from this distribution will be less than or equal to x?".

3. Evaluate the "inverse cumulative distribution function" of a random variable -- this is like looking up a probability in the body of the normal distribution table in the back of the book. This operation answers questions of the form "What value is 95% of the population less than?".

1. DISCRETE DISTRIBUTIONS:

Finite discrete random variables assume only finitely many values, like the sum of what comes up on two dice, or the number of pennies that come up heads when ten are flipped. Maple understands three discrete distributions that will be useful in Math 151: "empirical", uniform and binomial.

The discrete uniform distribution is denoted "`discreteuniform[a,b]`" in Maple. In this distribution, a and b are whole numbers, and the distribution assigns equal probabilities to the integers from a to b (inclusive). For example, the distribution `discreteuniform[1..6]` assigns the probability $1/6$ to each of the whole numbers from 1 through 6 (it is the distribution of the outcomes of rolling one of a pair of dice).

An empirical distribution is one that is completely specified by the user's input. For example, the distribution of the sum of two (fair) dice is given in the Finite Math text. To communicate this distribution to Maple, the proper notation is:

```
empirical[op(evalf(0, 1/36, 2/36, 3/36, 4/36, 5/36, 6/36,
5/36, 4/36, 3/36, 2/36, 1/36))]
```

This indicates that the probability of $x=1$ is 0, the probability of $x=2$ is $1/36$, the probability of $x=3$ is $2/36$ and so on up to the probability of $x=12$ is $1/36$. In other words, the twelve numbers in the list represent the probabilities of rolling a 1, 2, 3,...,12 respectively. For some reason, "empirical" only works when you use floating point numbers -- it gives error messages when you try to put in the actual

fractions. That's why there is an `op(evalf(...))` in the statement. One doesn't type this alone as a Maple statement -- we illustrate below how to use it.

Finally, binomial distributions are denoted `binomiald[n,p]` -- this notation has the obvious meaning. (Notice the "d" in the spelling -- leaving this out will result in a "Requested distribution does not exist" error message.)

Now we come to the uses of `statevalf` for each of these distributions. First:

```
> pf[binomiald[5,0.3]](2);  
                                .30870
```

This indicates that the probability of getting 2 out of 5 successes when the probability of success on each trial is 0.3 is 0.30870. The "pf" in the statement indicates that what is desired is the probability that the random variable is *exactly equal* to the number in parentheses.

As an example, consider the following problem from the Finite Math book: "In a certain congressional district, it is known that 40 percent of the registered voters classify themselves as conservatives. If ten registered voters are selected at random from this district, what is the probability that four of them will be conservatives?"

Since ten voters are chosen, and the probability of choosing a conservative is 0.4, the relevant distribution is the binomial distribution with $n=10$ and $p=0.4$, in Maple this is `binomiald[10,0.4]`. We want the probability that four of the choices are conservatives -- so 4 goes in the parentheses. The answer to the problem is obtained via the Maple statement:

```
> pf[binomiald[10,0.4]](4);  
                                .2508226560
```

The second way to use `statevalf` with discrete distributions is to calculate the

cumulative distribution: this is the probability that a random variable is *less than or equal to* a given value. To illustrate, another problem from the Finite Math texts asks what is the probability that *at most* 8 of a random sample of 20 photocells are defective if it is known that 5% of all cells produced are defective. The relevant probability distribution is the binomial distribution with $n=20$ and $p=0.05$, and we want the probability that the random variable (number of defectives) is less than or equal to 8. The answer is obtained via the Maple statement:

```
> dcdf[binomiald[20,0.05]](8);
.9999998021
```

Note that to get "*less than or equal to*" we use "dcdf" (which stands for discrete cumulative distribution function).

To illustrate the use of an empirical distribution, consider the probability of rolling a number less than or equal to 6 with a pair of dice. We define:

```
> dice:=empirical[op(evalf([0,1/36,2/36,3/36,4/36,5/36,6/36,5/36,
4/36,3/36,2/36,1/36]))];
dice := empirical[0,.0277777778,.0555555556,.0833333333,.1111111111,.1388888889,.1666666667,.1388888889,.1111111111,
.0833333333,.0555555556,.0277777778
```

Then the probability of getting 6 or less is:

```
> dcdf[dice](6);
.4166666667
```

2. CONTINUOUS DISTRIBUTIONS

Continuous random variables may take on all real values between the endpoints of some interval. As with discrete distributions, Maple knows many continuous distributions. The three that we will use most often are:

The (continuous) *uniform distribution*, denoted "uniform[a,b]" in Maple -- its distribution function is equal to the constant $1/(b-a)$ for values of x between a and b

(and zero otherwise).

The *Normal Distribution* (see section 8.5 of the Finite Math book) with mean μ and standard deviation σ . It is denoted "normald[μ , σ]" in Maple. (The *standard normal distribution* is normald[0,1] .

The *exponential distribution* (used for waiting times, etc..) with parameter α . It is denoted exponential[α ,0] in Maple.

It is rare that the value of the probability density function is required for the solution to a problem involving continuous distributions. There are two typical kinds of problems, however. The first involves the *cumulative distribution function* (cdf) -- which gives the probability that the value of a continuously distributed random variable is *less than or equal to* a given number. For example, a problem from the Finite Math book reports that IQ scores are found to have a mean of 100 and a standard deviation of 15. To find the probability that a random person's IQ is 90 or less, use the Maple statement:

```
> cdf[normald[100,15]](90);  
                                .2524925376
```

Another part of the problem asks for the probability that a random person's IQ is between 100 and 120. This is solved via:

```
> cdf[normald[100,15]](120)-cdf[normald[100,15]](100);  
                                .4087887802
```

To simplify the typing, it is convenient to give the distribution a name. For example:

```
> nd:=normald[100,15];  
                                nd := normald100,15
```

Then the previous result can be computed via:

```
> cdf[nd](120)-cdf[nd](100);  
                                .4087887802
```

The other kind of problem occurs when a probability value p is given and you want

to know for what x it is true that the probability that a random variable is less than x is p . For example, one might ask for the IQ level such that 75 percent of the population has IQ less than the level. This calls for the "*inverse cumulative distribution function*". In Maple, the answer to this problem is

```
> icdf[nd](0.75);
```

```
110.1173463
```

This indicates that 75% of this particular population has IQ less than about 110.

A final note -- for a complete list of the probability distributions (discrete and continuous) known to Maple, type

```
> ?stats[distributions];
```

A help screen will appear that lists all distributions known to Maple.