Exercise

Problem: At a certain T-intersection, 25% of cars typically turn left and 75% turn right. On a particular day, you counted 283 cars that turned left and 752 cars that turned right.

a) What uncertainty should you quote on these measurements to reflect the expected variability in traffic?
b) Was it a typical day? With what probability can you say that? Of course, you have to define "typical".

Solution:

1) Each event in your experiment can have only one of the two possible outcomes: turning left or right. Hence, the statistics is described by Binomial distribution. The expected variance on the number of cars is then

   $$\sigma^2 = n*p*(1-p),$$

   where n=283+752=1035 is total number of events, and p is the probability of the outcome. For p=0.25 (or 0.75), you get σ=14.
   Alternatively, you could estimate p from your data as p=283/n, getting a fairly similar estimate for σ. The σ you obtained gives you a measure of the expected variability in traffic.

2) By looking at the formula for the Binomial pdf you can see that knowing n and p completely defines the distribution, i.e. the probability of finding a certain outcome x, such as the number (x) of cars turning left. Plot the distribution to get the idea of its width and see where your measurements fall:

```
n= 1035; p=0.25;
x= 1:n;        % this defines a set of points 1,2,…,1035 at which you
               % evaluate the values
pdf=binopdf(x,n,p);     % define an array variable "pdf"
plot(x,pdf);
```

The function binopdf, as you can guess, returns the values of pdf (not probabilities!) at each point in x. You can zoom in on the plot to investigate it closer. You can see that your measured value of 283 is fairly far away from the mean.

Side note: What is the mean of this distribution? Can you calculate it from the pdf using the formula for the mean? If you try something like "mean(pdf)", you get nonsense, of course; see for yourself why. Remember, calculating the mean of x involves the integral of x*p(x)*dx, where p(x) is the probability distribution function. Now, in this case, your dx=1 (only integer number of cars allowed), and the integral becomes a sum:

```
sum(x.*pdf)
```

(Note the "." in the multiplication of two arrays.) Verify your result for the mean by a direct calculation using the formula for the mean of the Binomial distribution with given n and p.
Now, for a trick question, calculate the variance σ from the pdf, and compare your result to the value that you obtained above.

3) So how should you define what a "typical" or "not typical" day is? There are several ways of doing this, and it's to a large degree your choice. Let's say, for the sake of concreteness, that we are looking for a probability with which one can expect a fluctuation (deviation from the mean) of the observed magnitude or larger.

The simplest way to go to probability from pdf is through the cumulative distribution function (c.d.f.).

```
cdf = binocdf(x,n,p);   % define the value of cdf at any point in set x
plot(x(200:300),cdf(200:300))      % plot only for x from 200 to 300, to zoom.
```

Remember, the c.d.f. at a point *x* is an integral of the pdf from minus infinity to *x*. (Note that it has to reach 1 for large x). Thus, the value of cdf at x=283 represents the probability (that's what the integral of pdf is) that, on a typical day, one would count 283 cars *or less*. (I.e. you integrate that bell curve of pdf up until x=283)

```
binocdf(283,n,p)        % or simply cdf(283), since you have cdf as variable
```

gives you P(<283)=0.9612 = 96%. That is, there is about 4% chance of finding more than 283 cars turning left on a typical day.
(Of course, you could have executed this command right away, without plotting pdf and cdf, but what would be the point of doing the exercise?)

Not finished yet. We defined "typical" through deviation from the mean, without specifying the sign. Meaning, that to the 4% chance of finding *more* that 283 cars you have to add the chance of finding *less* than 258.75-(283-258.75) = 234.5 cars.

```
binocdf(234,n,p)
```

gives you 0.0397, that is also approximately 4%, as could be expected. The total probability of finding a fluctuation of this magnitude is then about 8%.