ELECTRONOTES
1016 Hanshaw Road
Ithaca, NY 14850

APPLICATION NOTE NO. 391

November 25, 2012

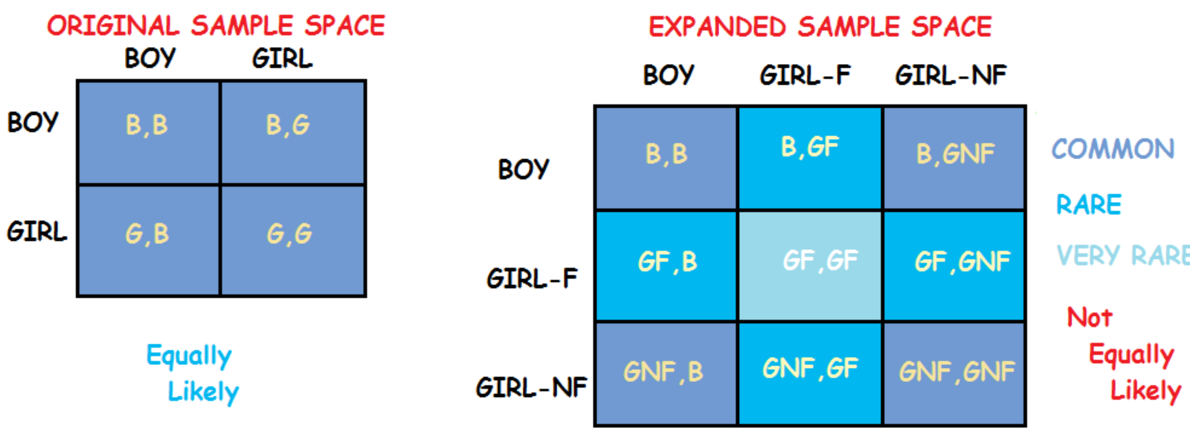
FUN WITH PROBABILITY PUZZLES 3 – THE GIRL NAMED FLORIDA

In the previous two ANs, we looked at a couple of probability puzzles. This one I am calling the "Girl Named Florida" puzzle because that is what Leonard Mlodinow in his book *The Drunkard's Walk* (2008) called it. It is also found in several places on the internet as such. We should begin at the beginning.

This puzzle begins discussing a family with two children. Given the assumption that the probabilities of a child being a boy (B) or a girl (G) are both 1/2 (perhaps not strictly true), what is the chance that both children are girls? Everyone reading this has answered correctly: 1/4. This is exactly the same math that we use for coin flips. The chance that the first child is a girl is 1/2, and that the second one is a girl is 1/2, so the product is 1/4. Equivalently, the "sample space" would be the possible combinations of (B,B), (B,G), (G,B), and (G,G) all equally likely (see left 2x2 figure in the illustration below). We mention this math here only because we will use similar but more complicated devices in just a bit.

Looking at the sample space, we also see that the chance of a boy and a girl is 1/2, and that the chance of at least one girl is 3/4. These calculations rely on the fact that we have no additional information. If there is more information, we need to take that into account.

For example, what is the probability of two girls if we are told the family has at least one girl? We might suppose it is 1/2 – just the probability that the second child is a girl. But this is not so. Looking at the sample space, we see that if there is at least one girl, the combination (B,B) is ruled out. This leave three equally likely possibilities, and only (G,G) meets the condition, so the probability is 1/3. What is the probability of two girls if we are told that the first child is a girl. Now the (B,B) and (B,G) are ruled out, leaving equally likely cases (G,B) and (G,G) or which only (G,G) is favorable, so that probability is 1/2. Okay, the first child is already known to be a girl, so we really did expect in this case that the probability of two girls was just the probability that the second child was a girl, or 1/2. But things are getting weird. It seems to matter which order the children were born in as to the resulting gender ratio. Impossible of course. What matters is the information we have and how we use it. Perhaps it is not much more that a word game. In this sense, it may be well to consider the probability calculations as being related to the **fair betting odds** we should be given if we are gambling on a result.



So next things will get really weird because we ask what the probability is that a family with two children has two girls given that one child is a girl named Florida? Can this possibly differ from the problem where we want the probability of two girls given that they have at least one girl (probability 1/3)? Could the name possibly matter? Yes, the probability changes to 1/2.

This being counterintuitive to the extent of a suspected swindle, we will look at the explanation in two ways – the second being a computer simulation. The illustration shows on the left the original 2x2 sample space we have been using above, while the space expands to 3x3 for the Florida problem. The possibilities include, as before, a boy. But now there are two choices for a girl – her name might be Florida, denoted GF, (unlikely) or it might be anything else (denoted GNF). Thus we have a 3x3 grid of the 9 possible combinations. Unlike the 2x2 grid, the possibilities are not all equally probable. Note that except for the four corner squares, the others require at least one girl named Florida. These are rare because girls named Florida are rare. [Indeed, we could choose almost any other name, and even the most common of these would likely be no more than a few percent.]

The four corners of the 3x3 on the other hand are common. In fact, they are little changed from what the corners were in the 2x2. All that changes is that we use GNF instead of G. We expect that most girls are <u>not</u> named Florida. So these corners are shown in the same shade of blue as the 2x2 case. At the same time, we note that the center square requires two girls with the name Florida. It is of course the case that most families don't name children with the same name (!) but the point is that if one with the name Florida is rare, two with the name Florida would seem exceedingly rare. So we can if we wish include the center square, recognizing it may be insignificant. But it is convenient to ignore it, and concentrate on the squares in the middle of the sides. These are weak compared to the corners. In fact, they are a subspace of the sample space that are lightly populated, but mostly we can think of these four as equally probable. [The corners are much more probable, and the center much less probable.]

So now we have reduced the sample space to include only families with two children, one being a girl named Florida. Most families have been excluded at this point, so the absolute numbers of families being studied is greatly reduced. GF is seen in all four squares. So envision Florida and her sibling. The sibling is equally likely to be a boy or a girl (a girl not named Florida). The probabilities are 1/2 for either.

We note that this probability is 1/2, and we saw this above for the case where we were told that the <u>first child</u> was a girl. In this case, a particular girl is not interchangeable with her sister. The family might not remember who was born first, but perhaps they named a particular girl "First". Or they might have named her Florida.

In the next AN we will look at this problem again using Bayes theorem. Here we close this out by presenting a Matlab simulation program named ***florida.m***. An example output is here:

epsilon =	0.0100	probability of name Florida, ε, (1%)
N =	10000000	number of simulated families (10,000,000)
TGFam =	2500516	number of two-girl families (about 1/4 of them)
NF =	100203	number named Florida (about 100,000, 20,000,000 children, 1/2 girls times 1%, in 10 million)
FamFl =	99933	number of families with girl Florida (note that 100,203 – this 99,933 = 270 families with two girls named Florida)
TGFamFl =	49945	two-girl families with name Florida
TwoFl =	270	two girls named Florida in same family
Theory1 =	0.0050	$(1/2 - \varepsilon/4)*\varepsilon$
Sim =	0.0050	
Theory2 =	0.4987	$(1/2 - \varepsilon/4)/(1 - \varepsilon/4)$
Ratio1 =	0.4998	

```
% florida.m

epsilon=0.01
FamFl=0;       % families with girl named Florida
TGFam=0;       % two girl families
TwoFl=0;       % two girls both Florida
TGFamFl=0;     % two girl families with girl named Florida
NF=0;   N=10000000;

for k=1:N
  % set up family
  % C1 and C2 children 0=boy 1=girl
  C1=0;
  if rand>.5; C1=1; end
  C2=0;
  if rand>.5; C2=1; end
  % assign name Florida CN1 and/or CN2 ---> 1
  C1N=0;
  F=0;
  if C1==1; if rand>(1-epsilon); C1N=1;F=1;NF=NF+1; end; end
  C2N=0;
  if C2==1; if rand>(1-epsilon); C2N=1;F=1;NF=NF+1; end; end
      % These two tries, for C1N and for C2N, would seem to offer 1/100 + 1/100
      %   probability only add up to 1/100 because C1 and C2 only have probability
      %   of 1/2 of passing to the rand>.5 condition.

  % add up families with one/both girls Florida
  if F==1;FamFl=FamFl+1; end

  % add up number of two-girl families
  if C1==1; if C2==1; TGFam=TGFam+1; end; end

  % add up families with two girls + Florida
  if F==1; if C1==1; if C2==1; TGFamFl=TGFamFl+1; end; end; end

  % check families with two girls Florida
  if C1N==1;if C2N==1; TwoFl=TwoFl+1;end;end

end

N
TGFam      % two girl families
NF         % number of Floridas
FamFl      % families with girl named Florida
TGFamFl    % two girl families with two girls and Florida
TwoFl      % two girls both Florida
Theory1=(1/2 - epsilon/4)*epsilon
Sim=TGFamFl/N
Theory2=(1/2 - epsilon/4)/(1 - epsilon/4)
Ratio1=TGFamFl/(FamFl)
```