GENETIC DRIFT

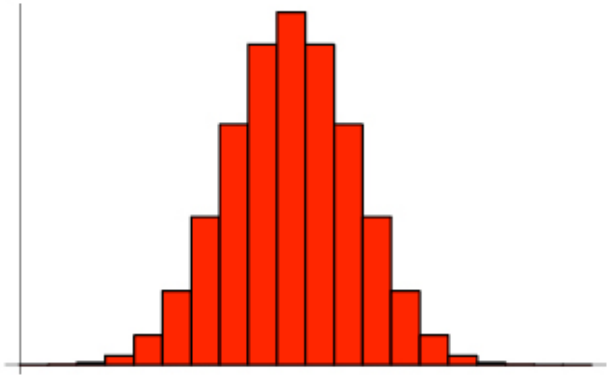
Carol Beuchat PhD

By now you should be very comfortable with the notion that for every gene location - a locus - an animal has two alleles, one that came from the sire and one from the dam. And which of the two alleles present in the parent gets passed on to each offspring is random. So the pair of alleles that the offspring inherits for each gene is determined only by chance.

The Binomial Situation

Take out a coin. Every coin has two sides - heads and tails - and for this reason when flip a coin we are talking about *binomial* probability. If it's a fair coin, there is a 50:50 chance of getting heads every time you flip it. You might get 5 heads in a row, but nevertheless at the next toss the chance of getting a heads is one out of two.

If you only flip it once and get heads, then the outcome of the trial is 100% heads. If you flip it 5 times and get 4 heads and 1 tails, then the outcome of the trial is 80% heads. If you flip it 100 times, or 1000 times, the probability of getting an extreme result goes down, and it should tend towards 50:50. This is the basis of the "bell curve" - most of the results will be 50:50, and deviations from this will become rarer as they become more extreme.



What does this have to do with dog breeding? Remember, which of the two possible alleles an offspring inherits from each parent is determined randomly - but when there are only a small number of "trials" (puppies, in this case), the results could be extreme just by chance. This is relevant to dogs because in terms of "large" vs "small", the typical size of a population of dogs is small - and here we could be talking about an entire breed, or a subset of a breed that is isolated by geography,

politics, or whatever. Because dog populations are statistically small, extreme results from a binomial sampling can occur.

Simulating Binomial Sampling With Beans

It's easy to demonstrate what we're talking about here. I'm sure you understand the example of the coin toss (or if you didn't, get out a coin and do some tossing). Let's do the same sort of thing, but now using beans to represent the alleles a dog could inherit at a particular locus.

This is where you get to play with the beans! Get out a small bowl and some cups. Start with two types (colors) of beans, and count out 50 of each into the bowl. Mix them all up with your hand. Now we're going to simulate inheritance.

On a piece of paper make three columns - label the first column LL (for light-light, or you could use RR for red-red, or whatever color beans you're using), the middle column LD, and the third column DD.

Before we do any bean breeding, let's do a bit of math. Based on what we've talked about already, answer these questions:

1) If you reach into the bowl and randomly (without peeking!) select one bean, what is the probablility it will be a "light" bean?

2) If you put that bean back in the bowl and mix it around, what is the probability of selecting another light bean?

The probability for each independent draw is 50% (0.5). So, what is the probability of drawing two light beans in a row (LL)? It's the product of their independent probabilities - (0.5) x (0.5) = 0.25, or 25%.

If we know this for the light beans, it must be equally true of the dark beans (or whatever color you're using) (DD).

Now, what is the probability of choosing a *light bean first* and a dark bean second (LD)? It's still (0.5) x (0.5) = 0.25, or 25%.

And what about drawing the *dark bean first* and the light bean second (DL)? So the probability of getting DL is also 25%.

Go back to your piece of paper with the columns, and above LL write 25%, and above DD write 25%. What about the middle column? If we consider all the beans of a particular color to be identical, then LD is the same as DL. *So*

the probability of getting two different colors is 25% + 25% = 50%, and you can write that above the middle column.

Okay, back to our beans. Mix them up, then reach into the bowl and pull out TWO beans at once, and put a tick mark in the appropriate column (if two light beans, count one for "LL"). Put those beans back (so there are always 100 beans in the bowl in a 50:50 ratio), draw another pair, and log the result. Do this a total 20 times.

Now tally up the number of occurrences of each combination (e.g., 8 LL, 4 LD, 8 DD). Then divide each of these numbers by the total number of draws (20) to determine the frequency of each outcome (e.g., 8/20 = 0.4, or 40%). How close did this come to the statistically "expected" outcome?

Draw a line under those data and repeat this exercise 2 more times, and calculating the fractional outcomes as before.

In all liklihood, none of the three trials produced the results you predicted at the beginning. But extreme deviations from expected can occur with small sample sizes. Since we did all three of these trials exactly the same way, we can pool the results and calculate the overall frequency of each outcome by adding the results in each column and dividing by 60. For example, if for LL you got 8, 5, and 3 for the three trials, the total is 16, and you divide that by 60; likewise for the other two outcomes.

What you should find is that the sum of the three draws (a total of 60) comes closer to the expected values you wrote at the top of the columns than the trials with only 20 draws. (Did you??? If you didn't , do another few draws of 20 just to convince yourself that eventually you will come close to the expected values.)

The bottom line here is that when you are working with a small sample (20), you are more likely to get frequencies that are different than expected. As the number of samples increases, the proportions should get closer and closer to the predictions.

Physical Simulation of Genetic Drift

Get out a cup, and put in it 100 beans in the proportions you got from your first trial above. (Just multiply the numbers in each column by 5.) If by some fluke you got the exact expected proportions (25% LL + 50% LD + 25% DD), pick the second trial (or third). We want to do a new simulation with a population where the frequencies of D and L are not equal.

Do the same thing you did before - record the results of 20 draws of a pair of beans, this time just once. Compute the proportions, then mix up another bowl of 100 beans in these new proportions. Do one more set of 20 draws and record your data.

Let's look at your data. You know you started with L and D beans in equal proportion at the very beginning. You then used the data from the first round of draws to create a the next generation of our bean population, which just by chance alone has a different proportion of L and D. And you repeated this again, creating another new generation that probably had again a different ratio of L:D. With every subsenquent generation, the frequency of alleles in the population will vary, just by chance.

This change in the frequency of alleles in our population with each generation is called *genetic drift*. If you continued doing these trials, say 100 or 1000 of them, you would see that the effect of genetic drift on the genetics of a population can be profound. But instead of playing with beans for a few more hours, we can do the same kind of simulation very quickly using a computer program that will create a virtual population of alleles, then randomly select, replace, and select again, in the same way you just have, for as many generations as you want. Using this, we can do a bunch of experiments very quickly.

Computer Simulation of Genetic Drift

> <If you skipped going the bean experiment because you thought you would be fine just reading it instead, or because you thought it was hokey and a waste of time, please reconsider. Without doing this yourself, you won't understand how the computer simulation works that we're about to do. Get out your beans and just do it. Trust me, it will make what follows much easier to understand.>

You can put the beans away for now, and go to the Red Lynx Population Genetics Simulator (http://scit.us/redlynx/).

To see how it works, run some simulations using the default settings - 2000 generations, population size of 800, and initial frequency of each of our alleles (A_1 and A_2) of 50%.

Each time you click on "Run Simulation", it will do the same thing you just did for beans - starting with a 50:50 mix of alleles, it will create 800 new individuals with alleles drawn at random. Then starting over again with alleles at their new frequencies, it will repeat again for 2000 generations. It will draw a line for each run showing how the frequency of the A_1 allele changed over time. The total number of alleles in the population stays the same over time, so if A_1 goes up, A_2 must go down. If A_1 goes all the way to 100%, that means A_2 has - just by chance - been lost from the population. Likewise, if A_1 goes to zero, then all of the alleles in the population are A_2. Each time you click on run, it does another simulation of 2000 generations and plots a new line.

Okay, let's do some experiments. From the bean counting experiment we did above, we decided that if a population is very large, the proportions of alleles drawn randomly should be close to what is predicted. When the population is small, just by chance you can get a result that is extreme.

We can do that experiment now with hundreds of generations in a few seconds. Try this:

1) Run 10 simulations with the default settings (population size of 800) except change the number of generations to 200, which is more reasonable for purebred dogs. How many times was the A_1 allele lost from the population (frequency went to zero)? How many times did the A_1 allele go to fixation (100%) - i.e, A_2 was lost, and all individuals were therefore homozygous for A_1?

2) Clear the graph, change the population size to 400, run 10 simulations, and note as above the number of times A_1 was eliminated or became fixed in the population.

3) Do the same thing with population sizes of 100, 50, and 25.

You should be getting the picture. What is the effect of population size on the genetic stability of our virtual population?

Now, let's simulate something more interesting. Let's pretend A_1 is the gene for PRA or some other genetic disorder, and we'll make it rare in the population - say 10%. Change initial frequency to 10%, put the population size back to 800, and run 10 trials, followed by population sizes of 400, 100, 50, and 25, as before.

As before, you will notice that population size has a large influence on the stability of the allele in the population, with the results getting more and more unpredictable as the population gets smaller. In these simulations, you probably found that many times your rare PRA allele was completely eliminated from the population, but occasionally (and more frequently at small population sizes), the frequency of this allele increased, perhaps substantially.

What is the size of the reproductive population in your breed? Think about some genetic disease that occurs in your breed that is caused by an autosomal recessive allele - e.g., PRA, or von Wildebrand disease. This disease gene could start out being rare in your breed, but in just a few generations - *by chance alone* - it could be lost entirely, or it could become very common and even fixed in the breed. Of course, as this allele becomes more common, the frequency of affected animals will go up (because the number of homozygous offspring will increase), and suddenly a genetic disorder shows up in your breed. This isn't a spontaneous mutation - it is an allele that has been there all along, and just by chance has become more common by genetic drift.

The frequencies of all alleles can vary each generation because of genetic drift, not just disease alleles. Just by chance, dogs might get larger, or bolder, or a rare color could become more common, or they might become more sensitive to a particular disease or have more allergies. The point to remember is these changes are occurring because of changes in allele frequencies of the *population*.

The Power of Genetic Drift

You now might be worrying about your own breed, wondering how large your breeding population is, and what nasty gene might be lurking in your gene pool waiting for the chance - just by chance - to become a serious problem. This is definitely something breeders should be thinking about. In most breeds, only a small percentage of puppies born each year are bred, and those are not selected randomly from each litter. Under these conditions, as you have seen, some dramatic shifts in allele frequencies can be occurring by chance without breeders even being aware. Population size is far more influential on the genetic status of a breed than most breeders realize.

Next time, we'll look at two additional things that can have a profound effect on the genetic composition of a population - the founder effect and population bottlenecks.