Numerical Comparative Dynamics: Ball Python Breeding

Donald DiJacklin

July 30, 2018

Introduction

Selective breeding is performed on many species, whether it is done to increase the size of the offspring, to increase the speed of the offspring, or simply to make healthier offspring. All of the preceding reasons are implemented in an effort to make the offspring worth more in some sense. In the case of Ball Pythons, the selective breeding is done (most of the time) in an effort to increase the number of visually expressed genes, and among those to express rare genes.

Ball Pythons (*python regius*) are a species of snake indigenous to Africa that constrict their prey, but in recent years they have been imported to other continents and seen some success as exotic pets. As with any pet, some people prefer different traits to be expressed in a Ball Python, like some people prefer small dogs to big dogs. Ball Pythons can have traits that affect colors or patterns, or both. As an example, shown below on the left is a Ball Python that looks like most of the Ball Pythons in Africa, referred to as a Normal by snake breeders. On the right is a Ball Python that expresses the trait known as Pastel.
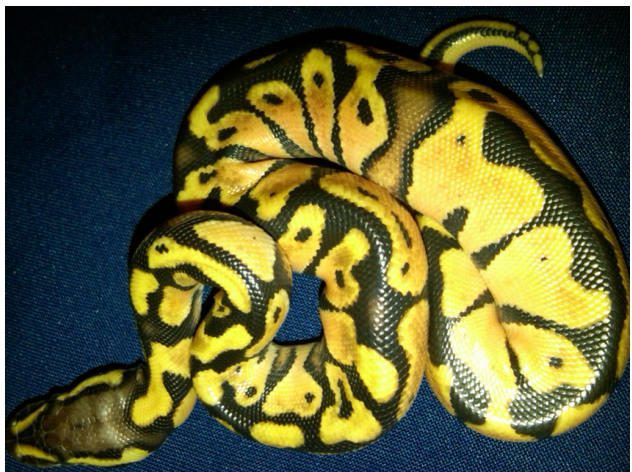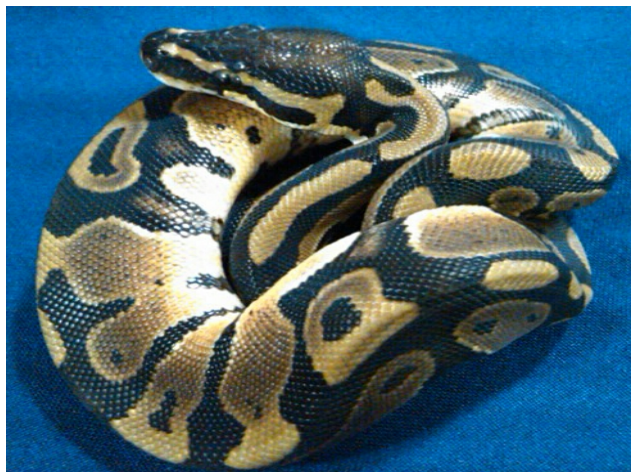
Note the marked difference that one trait can make in the appearance of the snake. Note too, that one pairing can result in an extremely variable set of snakes, as shown below.

As with dogs, certain traits are valued much more than others. A Normal costs about $20, whereas a Pastel costs about $45, and Ball Pythons with another trait, called Banana, cost about $200. As one might imagine the 'cooler' looking the snake the more it will cost, so breeders seek to make more money by breeding offspring that express more genes and, therefore, look 'cooler.'

I will not go over the details of snake reproduction here, but there are certain facts that will be used in this paper and the related programs, without which the reader will undoubtedly be lost. First, a successful (and sometimes unsuccessful) pairing of a male and female snake, produce a set of eggs, called a clutch. Second, female snakes can breed up to once per breeding season, whereas a male can breed up to five times per breeding season. Finally, a male reaches sexual maturity in about a year, whereas a female reaches sexual maturity in about two years. For more information on snakes' genetic traits **?** has written an excellent paper, which unfortunately is interested in every trait except the ones that I am concerned with in this paper.

Theory

Initially, one can think of this problem as a portfolio management problem in which the Ball Python breeders are portfolio managers and the snakes are capital goods. One can imagine that the breeders would like to maximize the present discounted value of profits from their stock of snakes. The preceding method was pioneered in Jarvis (1974). Comparative statics for this interpretation have been derived in the preceding

paper and Paarsch (1985).

The previous papers assume the animals (cattle) are all homogeneous except in age and there was no improvement in breed over time, which are simply not valid assumptions in the case of snake breeding. The value of each snake in an inventory depends on the genes it expresses, which depends on what its parents were, and there is an overall improvement in the stock over time because of the selective breeding exercise, which can't be handled by the aforementioned model.

One could expect that after a certain point the system will settle into a steady state, and it might be a useful exercise to characterize what the steady state would look like. If the breeder's set of snakes is not homozygous (having two copies of a certain allele) for all the different genes contained in their snakes, then the system will not be in a steady state, since they will try to breed so that their snakes are homozygous for all traits since that would make them more valuable. If their whole stock of snakes is homozygous for all of the same genes except for one snake, then they will either breed the snake that's different to at least one other snake, or they will sell it. If they breed the snake that's different then there will be some snakes in stock that are not homozygous (not homozygous is heterozygous) for all genes, which would continue as a selective breeding exercise until all snakes were homozygous for all genes in the stock of snakes. The steady state can't occur unless all snakes have the same genes and are homozygous for them all, at which point the problem changes to be akin to the previous models. Supposing that all snakes are, however, homozygous for all genes except for a gene which all snakes are heterozygous for, there is some positive probability that in any finite number of periods none of the snakes will be born homozygous for that gene, so a steady state is not guaranteed to be achievable in a finite number of periods.

Suppose instead, that one is interested in how the system behaves before a steady state, if one would even occur (as this paper is). As was argued in the previous paragraph, this would be where at least one snake is different from the others, or the snakes are all the same, but not homozygous for all traits. In this world it matters which snake one breeds to which, so this becomes a matching problem: deciding which male to breed to which females. Thinking about a single period yields the problem courtesy of Becker (1973) with a slight tweak:

$$\max_{\Pi} \left\{ \sum_{i \in I} g\left[m_i, f_{\Pi(i)}\right] \right\}$$

where I is the number of male snakes, m signifies a male snake, f signifies a female snake, g is a profit function, and Π is a mapping from the set of males to subsets of 5 or less snakes out of the set of females. In this setting, the matching problem will result in neither PAM nor NAM (Positive/Negative Assortative

Matching) as shown in the simple example below.

Suppose that one has a homozygous Fire male ($350), a heterozygous Fire female ($90), and a homozygous Pastel female ($75). Suppose further that one can only breed one male to one female. The offspring of the Fire Ball Pythons paired together would have an expected value of $220, whereas with the Fire and Pastel paired together the offspring have an expected value of $250. This may seem to imply NAM, however if one takes the same females but replaces the male with a heterozygous Pastel ($45), the expected value of the offspring of Pastel and Fire is $103.75 and Pastel with Pastel has an expected value of $60, which contradicts NAM. Now there's a serious problem, if it was PAM or NAM it would be a simple process to choose which snakes to breed to which.

The previous two paragraphs were about a single period decision, which is not really what a snake breeder should be doing. The snake breeder should, ideally, decide what to do this period with an eye to future time periods. Unfortunately, that turns an already difficult problem into an intractable problem, maximizing time discounted profit over an infinite time horizon, i.e.,

$$\max_{\Pi_t} \left\{ \sum_{t=1}^{\infty} \sum_{i \in I_t} g\left[m_i, f_{\Pi_t(i)}, t\right] \delta^t \right\}.$$

So what does one do since true optimality is intractable? The approach this paper uses is to find a reasonable rule that does well, where "well" is defined as doing better than the rule snake breeders actually follow, and to simulate in order to find the changes instilled by a change in the parameters of the system.

Data

I retrieved the data used for my program from the breeders at a show called Repticon in Tampa. I took pictures of snake breeders' displays, from which I gathered the traits of the snakes, the sex of the snakes, and the price of the snakes, and typed the data into the file called `snakedat.txt`. I processed the data with a program called `datacleaner.py` into a more usable format of a large csv with a column for every trait, as well as a column for sex and price. In each column a snake got a 1 if it had a trait and a zero if it did not. I loaded the output of `datacleaner.py` into a script called `modeler.R` in which I ran a linear regression of the traits and sex of the snake against the log of the price of the snake for snakes that had only genes represented in at least 10 snakes. This was accomplished by eliminating genes that had less than 10 snakes and eliminating the snakes with those genes. The resulting genes (first column) and their prices (second column) are found in `newprices.csv`.

Program

Since most of the work in this paper has gone into developing the simulation program, `snakes.py`, it seems reasonable to talk a little about what is going on in the program. The first major part of the program is that a snake is an instance of a class aptly called Snake which contains the relevant information about a snake: name, age, sex, price, traits, forebears, age it can breed, and number of times it has bred this period. The second major part of the program is the function tick which simulates what a breeder with a given decision rule will do in a year. The function tick itself calls three separate functions: breed, hypobreed, and hypobreedg.

The function breed does the following: checks whether the snakes can breed at all with regards to sex, age, and inbreeding. If they were able to it has a 60 percent chance to generate offspring, and if it does, it generates a clutch size from the discrete triangular distribution where the minimum observable value is 3, maximum observable value is 14, and the value with the highest probability is 6, and generates that number of new snakes. The functions hypobreed and hypobreedg are used to calculate the expected profit and expected number of genes of offspring respectively, for use with the decision rules.

The two decision rules that I implement are: maximize expected profit in the next time period, and maximize the expected number of genes in the next period. The first does the following:

$$\max_{\mathbf{X},\mathbf{y},\mathbf{z}} \left\{ \sum_{i \in I} \mathbf{r}_i \mathbf{x}_i^T - 80\mathbf{y}^T \mathbf{1}_I - 80\mathbf{z}^T \mathbf{1}_J \right\}$$

$$s.t. \quad \mathbf{X}\mathbf{1}_J \leq 51_I$$

$$\mathbf{X}^T \mathbf{1}_I \leq \mathbf{1}_J$$

$$\mathbf{X}\mathbf{1}_J \leq M\mathbf{y}$$

$$\mathbf{X}^T \mathbf{1}_I \leq M\mathbf{z}$$

$$\mathbf{y}^T \mathbf{1}_I + \mathbf{z}^T \mathbf{1}_J \leq C$$

$$X_{ij}, y_i, z_j \in \{0,1\}$$

where:

- I- the set of male snakes in stock.

- J- the set of female snakes in stock.

- $\mathbf{R}$- (not shown) a matrix, the ijth element of which is the expected revenue of breeding male $i \in I$ and female $j \in J$.

- $\mathbf{r}_i$- the ith row of $\mathbf{R}$.

- $\mathbf{X}$- a decision matrix, the ijth element of which has a 1 if one should breed male i with female j.

- $\mathbf{x}_i$- the ith row of $\mathbf{X}$.

- $\mathbf{y}$- a decision vector, the ith entry of which should contain 1 if one should keep male i.

- $\mathbf{z}$- a decision vector, the jth entry of which should contain 1 if one should keep female j.

- M- a large integer.

- C- the capacity constraint.

Taken together the above maximizes expected profit in the next period subject to a capacity constraint and constrained by the fact that males can breed up to 5 times per period and females up to once per period. One might also wonder why the number 80 appears in the objective, and the answer is that the average yearly cost to keep a ball python is \$80.

The second decision rule does the following:

$$\max_{\mathbf{X},\mathbf{y},\mathbf{z}} \left\{ \sum_{i \in I} \mathbf{g}_i \mathbf{x}_i^T \right\}$$

$$s.t. \quad \mathbf{X}\mathbf{1}_J \leq 5\mathbf{1}_I$$

$$\mathbf{X}^T\mathbf{1}_I \leq \mathbf{1}_J$$

$$\mathbf{X}\mathbf{1}_J \leq M\mathbf{y}$$

$$\mathbf{X}^T\mathbf{1}_I \leq M\mathbf{z}$$

$$\mathbf{y}^T\mathbf{1}_I + \mathbf{z}^T\mathbf{1}_J \leq C$$

$$X_{ij}, y_i, z_j \in \{0,1\}$$

where the variables are the same as the last decision rule except:

- $\mathbf{G}$- (not shown) a matrix, the ijth element of which is the expected number of genes in the offspring of male i and female j.

- $\mathbf{g}_i$- the ith row of $\mathbf{G}$.

Results

Before I can make conclusions about the differences brought about by changing the parameters of the system, I need to show that my decision rule performs well. Having setup both my decision rule and a snake breeder's decision rule I ran both through 8 time periods 124 times apiece and calculated the total profit of each simulation with a capacity constraint of 15 snakes. The empirical cumulative distribution functions are shown below.
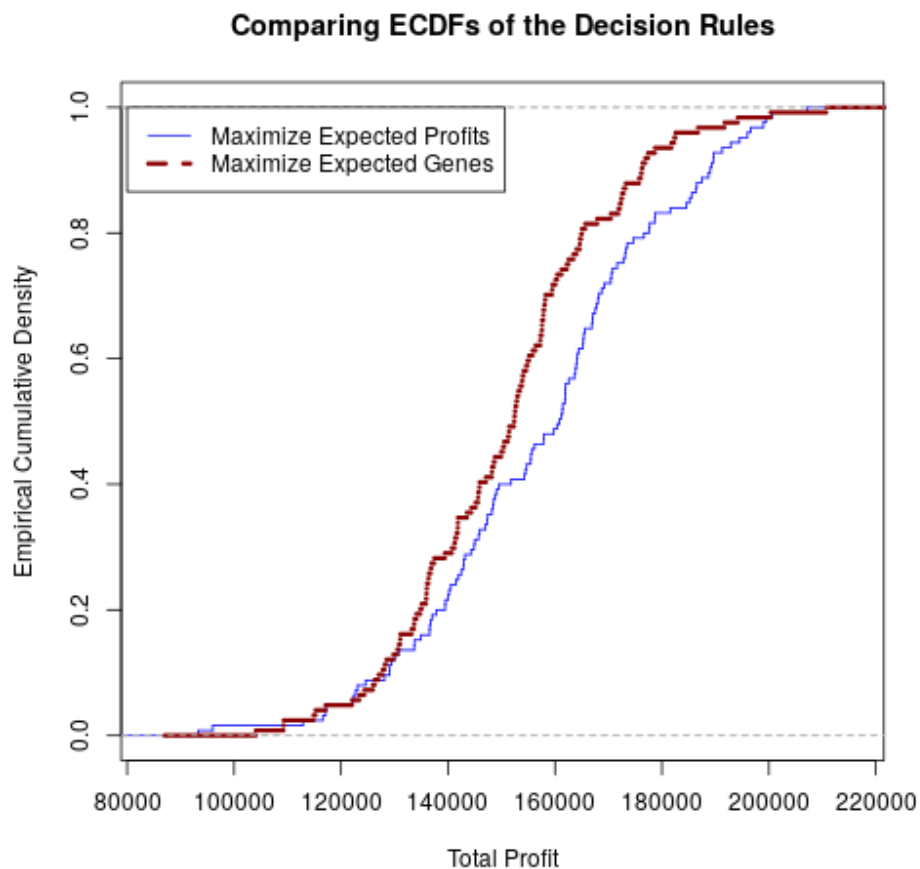
The first thing to note about the above is that according to a two-sided two-sample Kolmogorov-Smirnov test we can be 95 percent (p-value .001) confident that these two ecdfs are drawn from different distributions. The second thing to note is that all risk-neutral and most risk-averse snake breeders should prefer to use my decision rule (the thinner line).

If one changes the snake breeder's decision rule slightly to have a relative weighting system where the most weight is placed on the best gene, second most is placed on the second best gene, and so on and so forth, the graph of the ecdfs are shown on the next page.

Comparing ECDFs of the Decision Rules
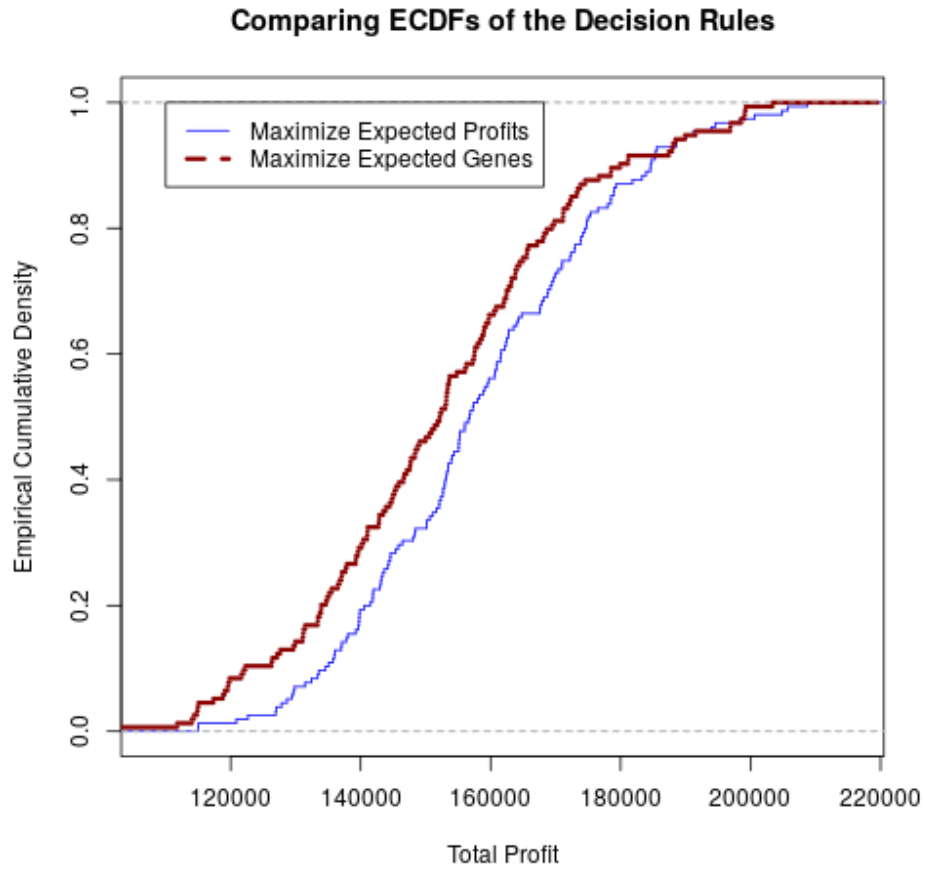
We can be 90 percent (p-value .078) confident that the two are drawn from different distributions using the same test as before. Once again all risk-neutral and most risk-averse snake breeders would prefer my decision rule (in fact all risk averse snake breeders if these distributions are close enough to the truth, since my decision rule looks like it Second Order Stochastically Dominates the snake breeder's decision rule). One might wonder what would happen if one used a weighting system that wasn't naive, and the answer is that it would turn into maximizing expected revenue, and should do slightly worse than my decision rule, in general, since the goal is profit not revenue.

Having shown that my decision rule does well, we can move on to the goal of this paper: finding how much the system changes when one alters the parameters of the system. Before we do that, let me first establish what parameters will be changed and what one might expect the result of that parameter change might be.

Parameters and Expectations

Probability of Clutch

The current probability that a pairing will result in a clutch is 0.6, and since increasing that probability would increase the average number of offspring in each period it's fairly obvious to expect that the profit in each period would increase on average. Thinking a little more about it yields the following observation: there will be an effect that accumulates each period due to the fact that, since there are more snakes on average, there is a higher probability of getting better snakes each period, and since one should be choosing to keep better snakes, this will have an additional positive effect on profits.

Distribution of Clutch Size

There are three natural ways to modify the distribution I specified. One can change the lower bound, the number with the highest probability, or the upper bound. Modifying one of these numbers upwards should work in a similar way to increasing the probability of a clutch, in that it will, on average, generate more snakes. This may once again have the same benefit of getting better snakes more often which will have a cumulative additional effect. On a different note, moving the lower bound up, or the upper bound down should decrease the variance of the profits, since the variance of the triangular distribution will decrease.

Male Virility

Since the base assumption is that a male can breed with up to five females, one could think of increasing that ratio as something that could be plausibly attained, and therefore something to be investigated. Since a male could still breed with five females if that would do better with regards to profit, I posit that profits could only stay the same or increase, on average, and I would place my bet on increase. Another affect that one can imagine, is that the ratio of males to females would tend to be lower, since one needs fewer males to inseminate as many females, and the number of clutches can be at most the same as the number of females.

Cost

As with any producer, increasing their costs should decrease their profits. This shouldn't have any strange cumulative effects until the increased cost makes the program start discarding snakes it would have otherwise kept, which shouldn't happen for small increases since most of the snakes in the initial stock are worth more than $80.

Capacity

This particular parameter is simultaneously one of the more interesting ones and the most vexing to think about. One can imagine that loosening this constraint can only have a zero or positive effect on profits, and at first blush one might guess that it would be positive. Think about the following scenario: moving from a capacity of 18 to a capacity of 19. At 18 snakes it's likely that one would keep 3 males and 15 females. Suppose one has the same choices, but a capacity of 19. It's unclear whether one should use that additional space. My working hypothesis is that moving upwards will normally increase profits, but moving upwards from multiples of 6 may have no effect.

Results

Probability of Clutch
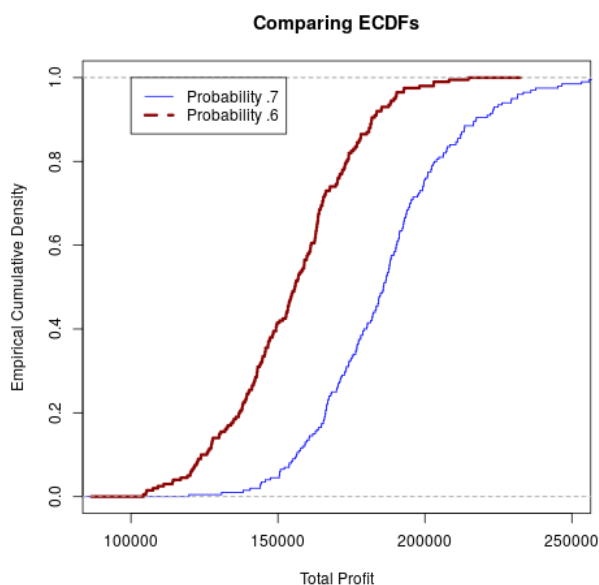
As one can see from the above, it is clear that increasing the probability increases the profits, as expected. The part that requires some thought is whether or not there was the hypothesized benefit from the cumulative improvement of the genetic base. Imagine with me what would happen if there wasn't a cumulative improvement. If the snakes were the same average value as they were previously, and there were just more of them than the revenue should be, on average, 16.7 percent higher since 0.7 is that percent more than 0.6. The costs should be roughly the same on average, which, coupled with the reasonable assumption that they should be keeping exactly 15 snakes in almost every scenario, leads me to posit that the total cost for these

8 years is $80 \times 15 \times 8 = 9,600$. The average total profit from a probability of 0.6 was $154,768.60, so adding the $9,600 of cost yields an average total revenue of $164,368.60, which should increase to $191,763.40 if my hypothesis is wrong. It instead increased to $195,813. That value rejects the null hypothesis that there is no cumulative effect at an α of 0.05 using a two-sample test of means (unless otherwise specified this is the test that will be used in the following results). In total increasing the probability from 0.6 to 0.7 increased average profits by $31,444.40 after 8 years, of which $4,049.60 were from the cumulative effect.

Distribution of Clutch Size

11
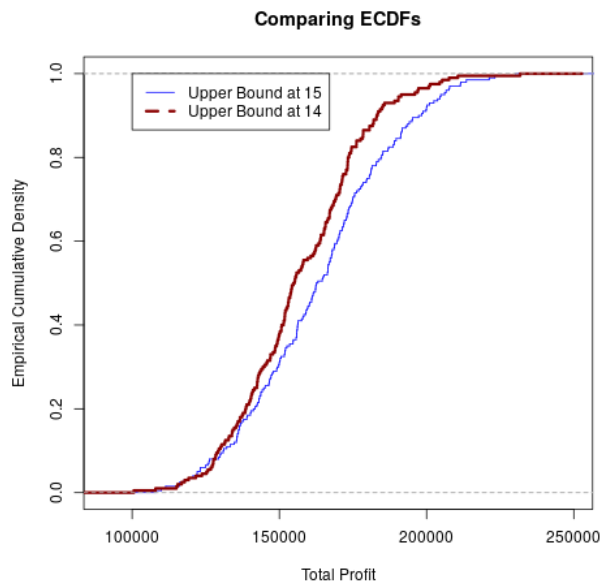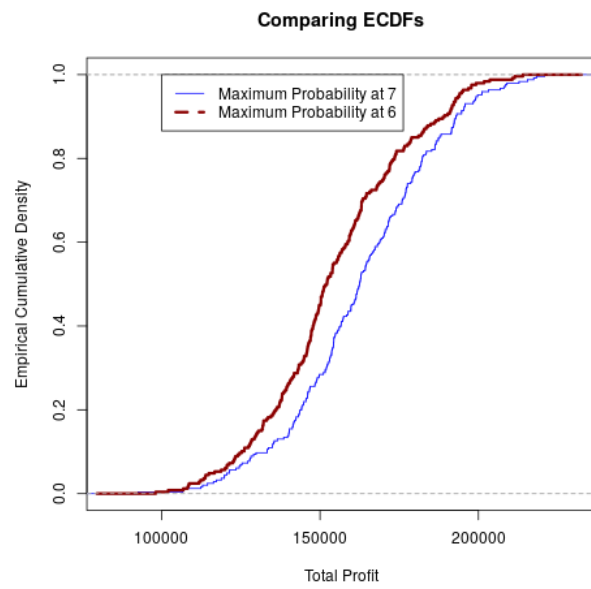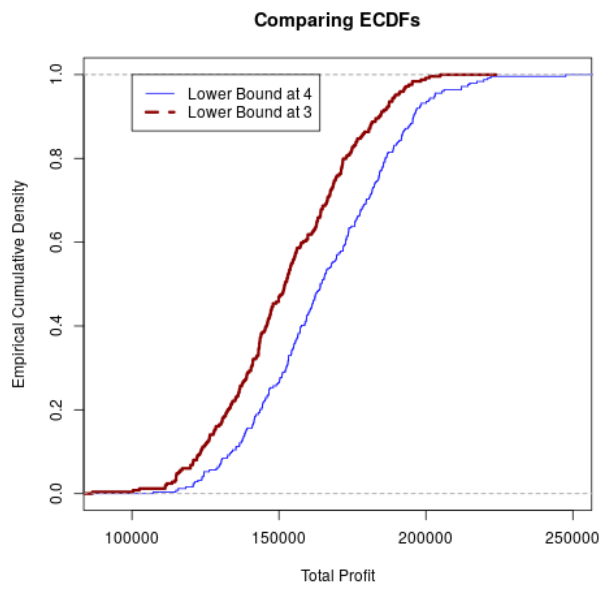
As is easily seen in all of the above, increasing the distributional parameters all increase total profits. Increasing the lower bound of the distribution from 3 to 4 yielded an increase of \$12,696.40, which is statistically significant. Increasing the highest probability parameter from 6 to 7 resulted in an increase of \$8,783.40 in average profits, which is statistically significant. Increasing the upper bound from 14 to 15 increased average profits by \$6,316.20, which is statistically significant. As far as the statement that variance should go down, there is insufficient evidence to show that it did with an F-test (p-value = 0.102) in the case of the lower bound, and no real evidence (p-value = 0.4564) in the case of decreasing the upper bound. The variance results are probably attributable to the fact that, although the variance of the triangular distribution goes down, increasing the lower bound in particular can raise the variance of the kept snakes. As for the upper bound, I am at a loss.

Male Virility
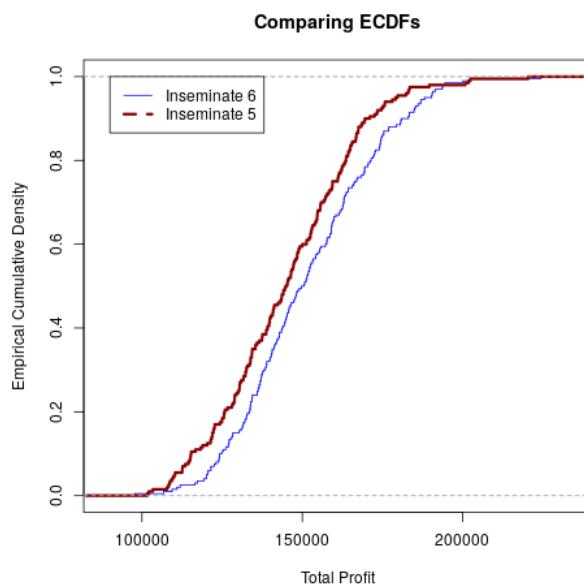
As one can see above, each male having the ability to inseminate 6 females instead of 5 results in an increase in profits, as I hypothesized previously. In point of fact, the average total profit rose by \$6,603, which rejected the null hypothesis of no change. The average male to female ratio went from 3:11 to 2.025:11.96591. Note that for these calculations a capacity of 14 was used instead of 15.

12
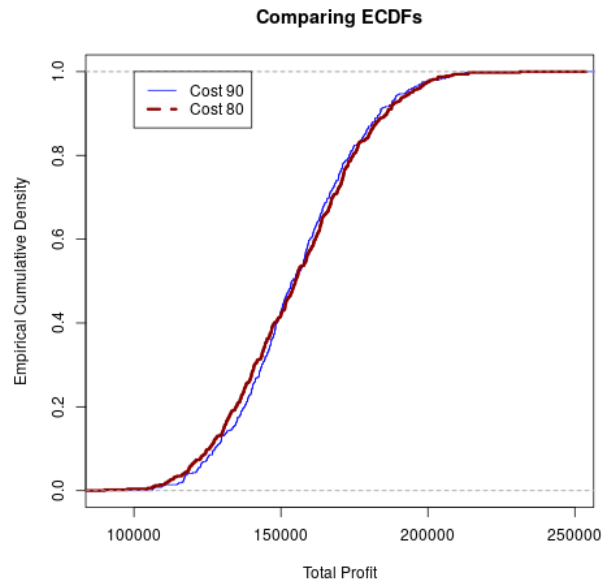
Cost

Clearly, no extreme difference exists in the distribution of profit when the cost changes from \$80 to \$90 to maintain a snake since the ecdfs are almost indistinguishable from each other. There is no statistically significant difference between the two. Even at a much larger number of simulations than the other parameters the difference that one would expect is simply not detectable. I simply cannot explain this.

Capacity
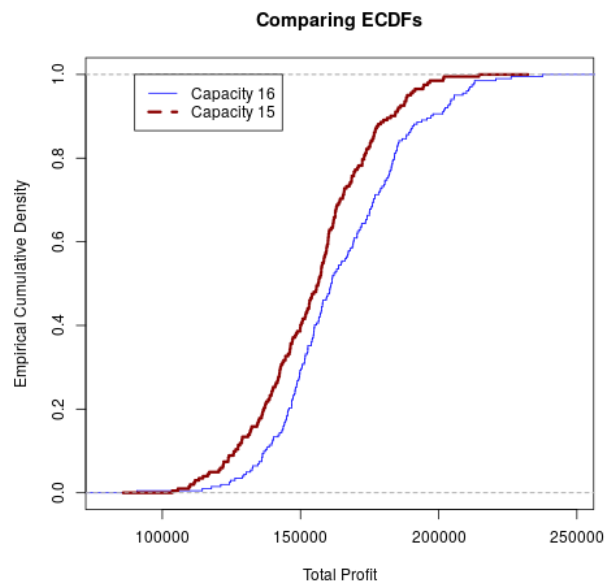
As one can see above, increasing the capacity from 15 to 16 increased average total profits, as was hypothesized. The change was $10098.3 which rejects the null hypothesis of no change.

Comparing ECDFs
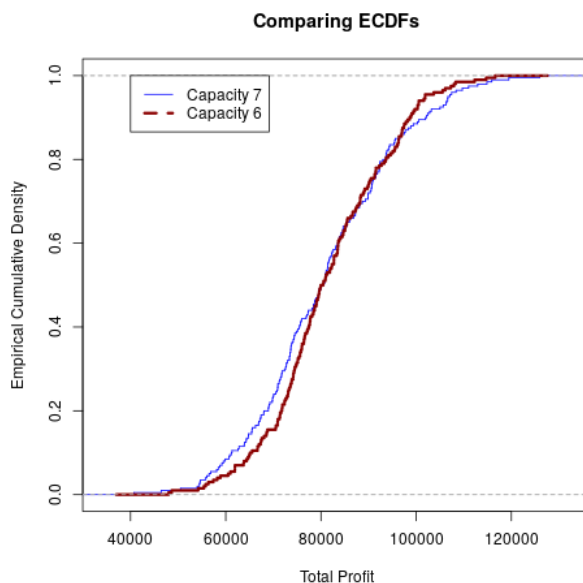
As was hypothesized the change when moving from a multiple of 6 to a non-multiple of 6 is neither a clear increase nor a clear decrease in profits. There is no statistically significant difference in the means of the two distributions.

Discussion

This paper is not without some limitations. First, is the assumption that snake breeders can immediately sell any snake they want. Oftentimes they might not find any buyers for a snake for years. The market for Normals in particular is almost nonexistent; one could be stuck with a Normal for a long time. Second, is that there was a dearth of data to identify how much difference homozygous versus heterozygous makes in snake prices. These problems will have opposite effects on total profits, however it's not quantifiable at the moment by how much the results would be off.

On the bright side, the program I developed for this project can have an extremely potent effect on any selective breeding program, not just Ball Pythons. The program is easily extensible to other animal specifications. Even though my results may be inaccurate, they can be updated upon the acquisition of more data to ameliorate the problems stated above. Even supposing that the problem of selling frictions cannot be rectified, this endeavor has still been worthwhile if only for the decision rules that were implemented. Even if breeders do not believe that my breeding rule would be accurate for them, I implemented their rule

as well, which could save them hours of thought.

References

Becker, G. S. (1973). A theory of marriage: Part 1. *Journal of Political Economy 81*, 813–846.

Jarvis, L. S. (1974). Cattle as capital goods and ranchers as portfolio managers: an application to the argentine cattle sector. *Journal of Political Economy 82*, 489–520.

Morrill, B. H. (2011). Quantitative genetic analysis of reproduction traits in ball pythons.

Paarsch, H. J. (1985). Micro-economic models of beef supply. *Canadian Journal of Economics*, 636–651.