Plolygenic traits and the central limit theorem

José Miguel Ponciano: josemi@ufl.edu

University of Florida, Biology Department

Darwin was once also frustrated with math...

Mathematics ...was repugnant to me ...[but] I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense. -Charles Darwin (Autobiography)

But this quote shows he **came to understand why math was important, even for him**.

Have you heard of Darwin Finches from the Galapagos Islands and how they are used to illustrate the interaction between evolution and ecology (beak length variation)?

Have you heard explanations of evolutionary changes that go all the way to the effect of multiple genes? Well, Ricklefs, my undergrad ecology class book author brings genes into the the explanation of evolutionary phenomena

An example: polygenic traits

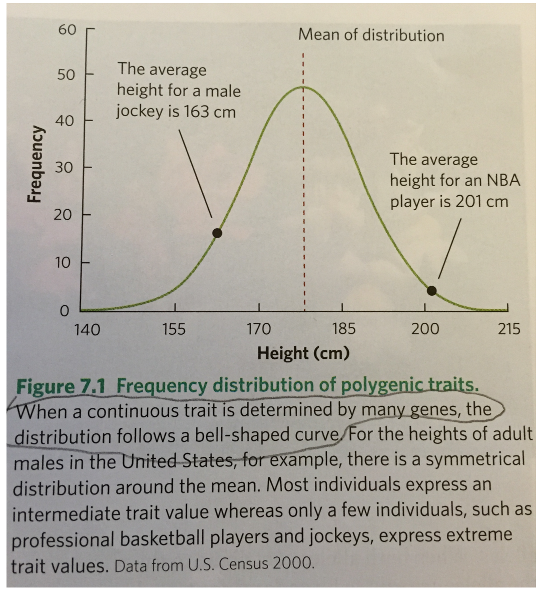


Figure 7.1 Frequency distribution of polygenic traits. When a continuous trait is determined by many genes, the distribution follows a bell-shaped curve. For the heights of adult males in the United States, for example, there is a symmetrical distribution around the mean. Most individuals express an intermediate trait value whereas only a few individuals, such as professional basketball players and jockeys, express extreme trait values. Data from U.S. Census 2000.

Why (to the first sentence of the caption)? What does the book author means here????

From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes.

From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes
- Suppose that the contribution of a single gene to the total beak size of a polygenic trait can be either Small, Medium or Large (say 1, 2 or 3 cms. of width respectively) depending on the physiology, other genes present, etc...

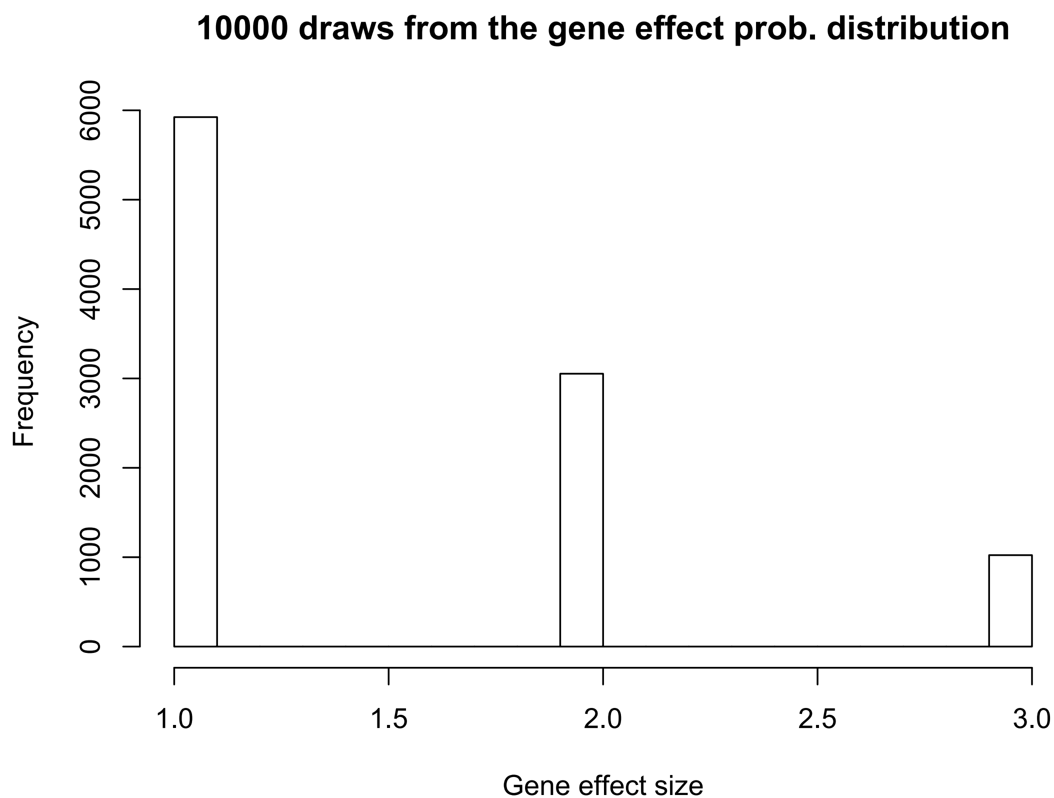
From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes
- Suppose that the contribution of a single gene to the total beak size of a polygenic trait can be either Small, Medium or Large (say 1, 2 or 3 cms. of width respectively) depending on the physiology, other genes present, etc...
- Then, suppose that for a single gene involved in the beak size, the probability that the conditions are such that its effect is small is 0.6, medium 0.3, large 0.1

From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes
- Suppose that the contribution of a single gene to the total beak size of a polygenic trait can be either Small, Medium or Large (say 1, 2 or 3 cms. of width respectively) depending on the physiology, other genes present, etc...
- Then, suppose that for a single gene involved in the beak size, the probability that the conditions are such that its effect is small is 0.6, medium 0.3, large 0.1
- This probabilistic model for the trait effect is what in probability is known as a "discrete probability distribution"

The probability distribution of the effect of a single gene



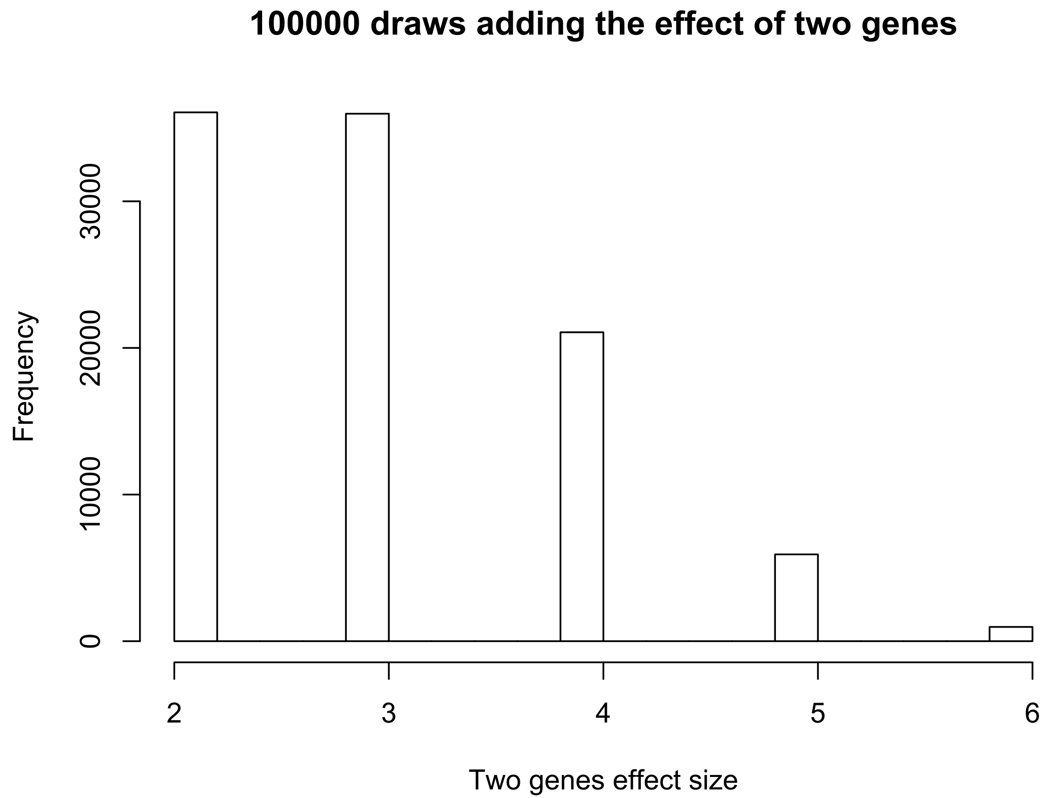
From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes
- Suppose that the contribution of a single gene to the total beak size of a polygenic trait can be either Small, Medium or Large (say 1, 2 or 3 cms. of width respectively) depending on the physiology, other genes present, etc...
- Then, suppose that for a single gene involved in the beak size, the probability that the conditions are such that its effect is small is 0.6, medium 0.3, large 0.1
- This probabilistic model for the trait effect is what in probability is known as a "discrete probability distribution"
- Calculation and definition of the mean effect size and the variance in the effect size for a single gene.

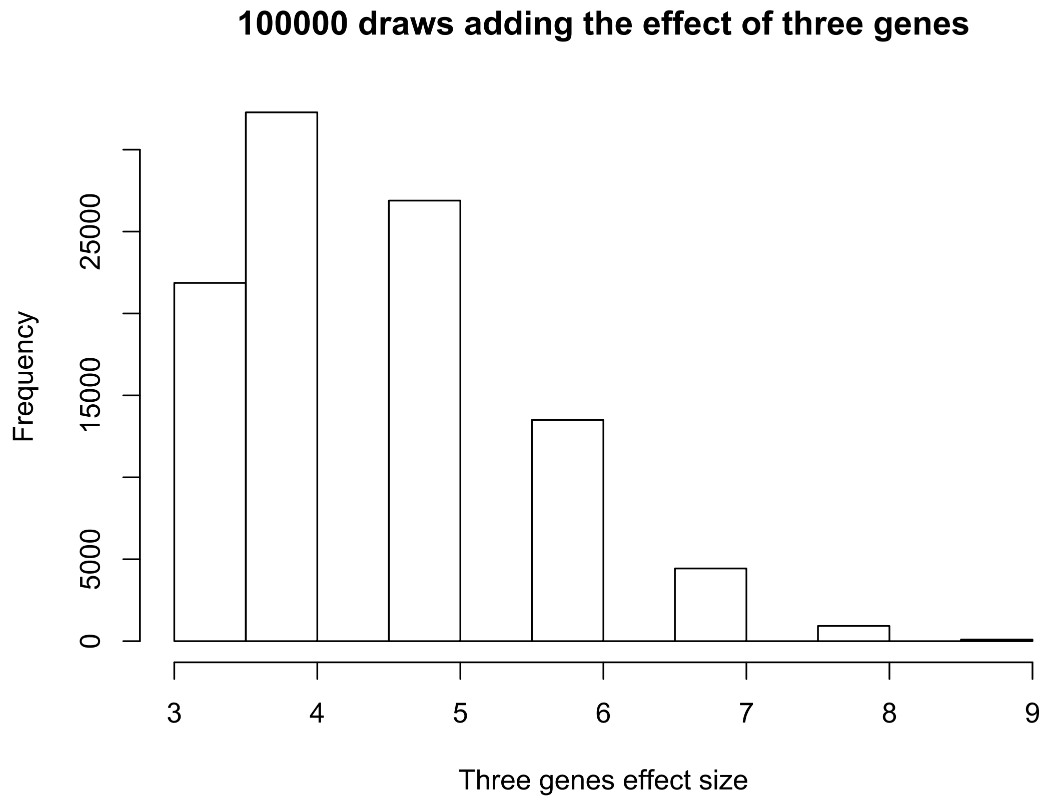
From the effect of one gene to the effect of many genes

- The beak size of Darwin's finches is known to be controlled by many genes
- Suppose that the contribution of a single gene to the total beak size of a polygenic trait can be either Small, Medium or Large (say 1, 2 or 3 cms. of width respectively) depending on the physiology, other genes present, etc...
- Then, suppose that for a single gene involved in the beak size, the probability that the conditions are such that its effect is small is 0.6, medium 0.3, large 0.1
- This probabilistic model for the trait effect is what in probability is known as a "discrete probability distribution"
- Calculation and definition of the mean effect size and the variance in the effect size for a single gene.
- What would the distribution of the sum of two identical genes look like? And three?

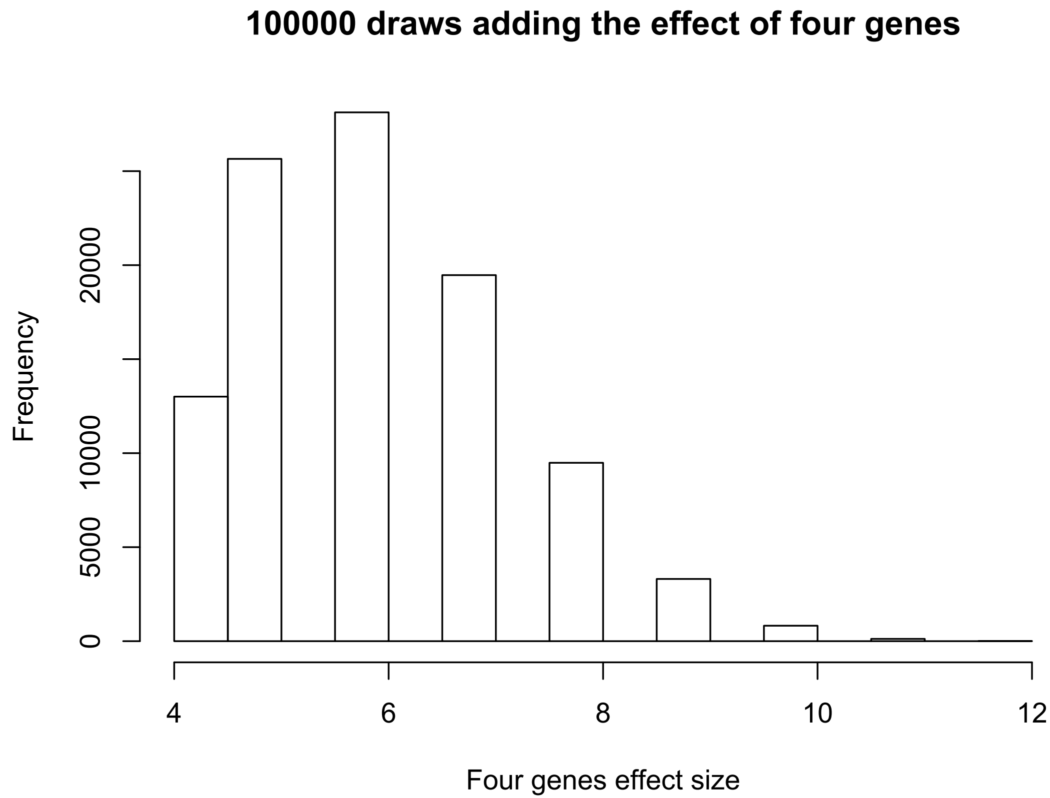
The probability distribution of the effect of 2 genes



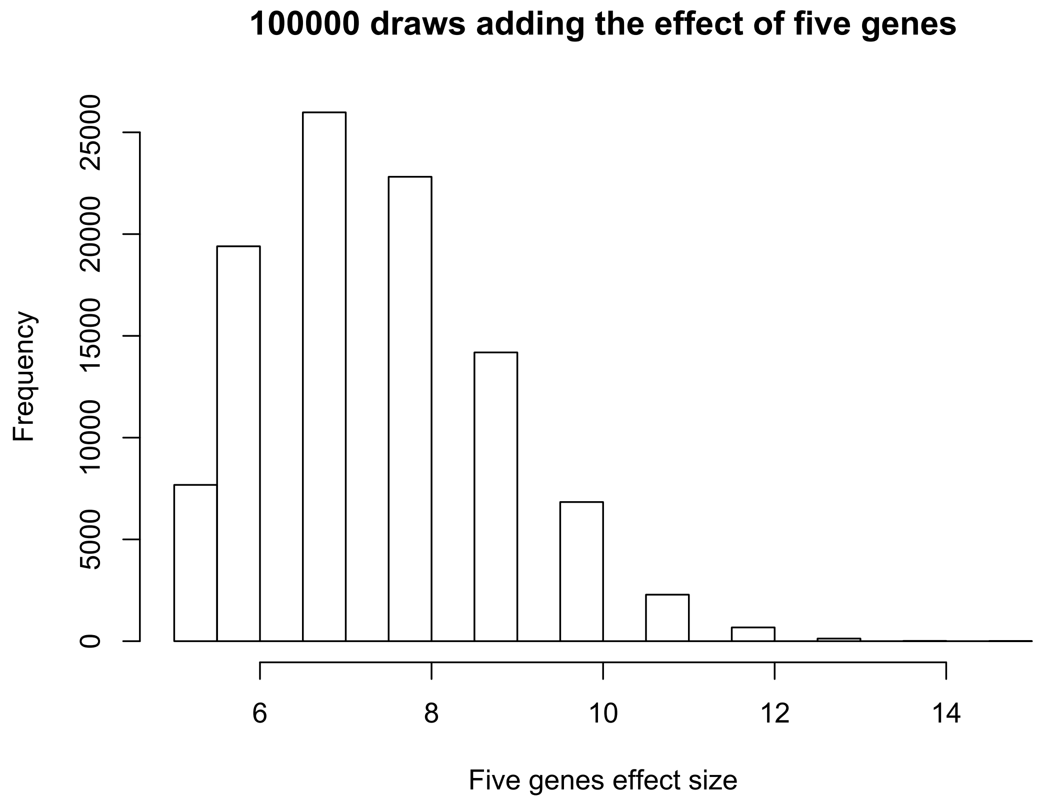
The probability distribution of the effect of 3 genes



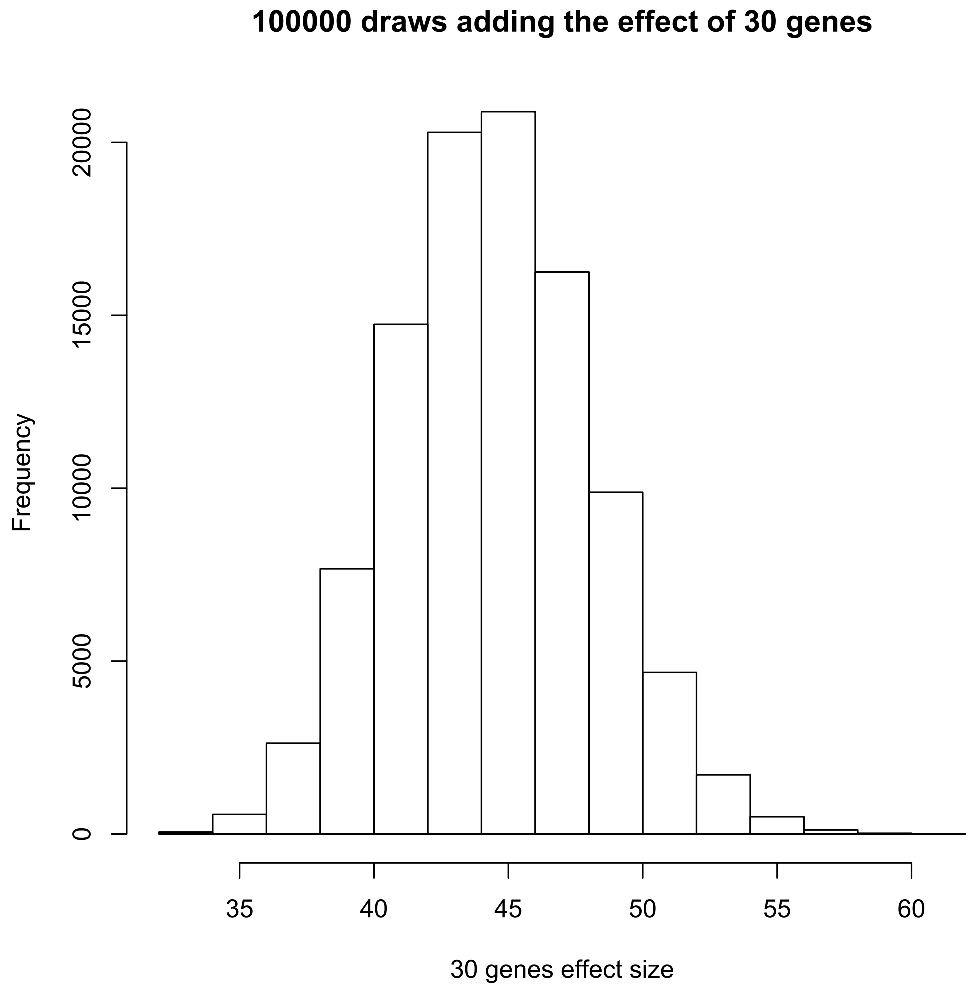
The probability distribution of the effect of 4 genes



The probability distribution of the effect of 5 genes



The probability distribution of the effect of many genes



And that's the explanation of the first sentence of the book figure's caption

- Suppose one starts with a single random variable (like the distribution of one gene's effect on the overall size of the beak), with a given mean (say m) and variance (say v). Then, the Central Limit Theorem (CLT) that you learn in your most basic stats class tells us that if we add the outcome of a large number (say n) of random variables that each have the same distribution and are independent from each other, then the resulting sum will be normally distributed. The mean of that normal distribution will be nm and its variance will be nv.

Now, in a dry year, mostly finches with larger beaks survive, and the population distribution shifts to the right. Did that shift (evolution) really happen?

We saw that the frequency distribution of the effect of 30 genes affecting beak size can be well described with a Normal distribution with mean $\mu = 45$ and variance $\sigma^2 = 13.5$ (standard deviation $= \sqrt(\sigma^2) \approx 3.674235$

Suppose that to gather evidence regarding whether evolution happened after drought we capture $n = 30$ birds and measure the beak, obtaining a sample mean of $\bar{x} = 46.5$. Is that evidence to show that evolution occurred?

Amazingly, the very same piece of math (the Central Limit Theorem) that allows us to describe a polygenic trait with a normal distribution, allows us to answer this scientific question

The central limit theorem and the sample mean

- If we want to answer our biological question of interest (did evolution occurred after a dry year), we'll need the following result.
- Suppose one starts with a single random variable (like the distribution of one gene's effect on the overall size of the beak), with a given mean (say m) and variance (say v). Then, the Central Limit Theorem (CLT) that you learn in your most basic stats class tells us that if we add the outcome of a large number (say n) of random variables that each have the same distribution and are independent from each other, then the resulting sum will be normally distributed. The mean of that normal distribution will be nm and its variance will be nv.
- Now, before answering the question of interest, we need to learn about sampling distributions, and the **distribution** of the sample mean!