Topic 10: The Central Limit Theorem

October 15 and 20, 2009

In the discussion leading to the law of large numbers, we saw that the standard deviation of an average has size inversely proportional to $\sqrt{n}$, the square root of the number of observations. So, in the graphs that show the law of large numbers, we see the running averages moving to its distributional mean. For the first simulation, based on 200 observations of independent random variables, uniformly distributed on $[0, 1]$, we see the running average converging to 1/2, the distributional mean, as anticipatted.

The result of magnifying the difference between the running average and the mean by a factor of $\sqrt{n}$. This suggests investigating the graph of

$$\sqrt{n}\left(\frac{1}{n}S_n - \mu\right)$$

versus n. We simulate this for the uniform random variables on the interval $[0, 1$. Thus, we center at $\mu = 1/2$. The results can be seen in Figure 1.

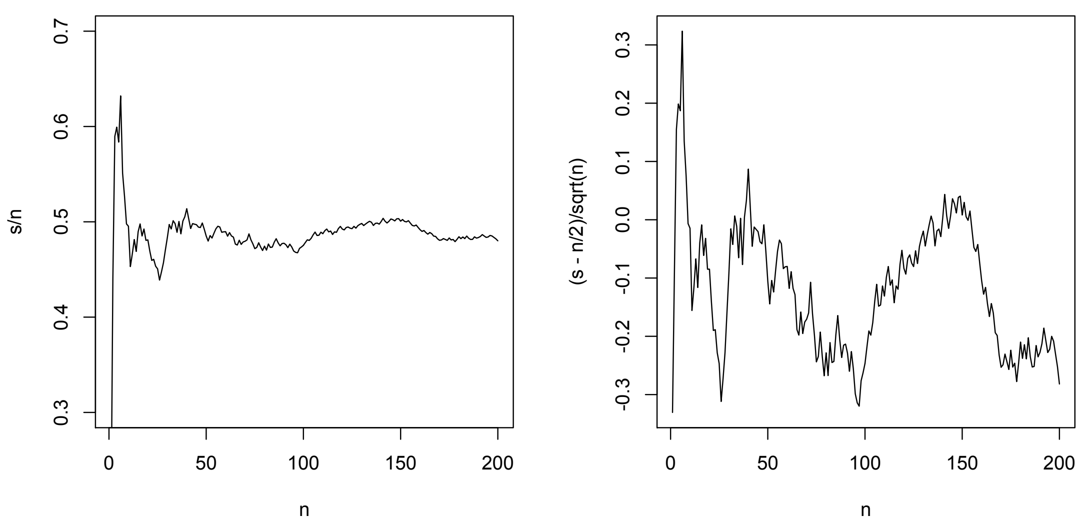

Figure 1: a. Running average of independent random variables, uniform on $[0, 1]$. b. Running average centered at the mean value of 1/2 and magnified by $\sqrt{n}$.

As we see in Figure 2, even if we continue this for larger values of n, we continue to see fluctuations about the center of roughly the same size and the nature of these fluctuations cannot be prodicted in advance.

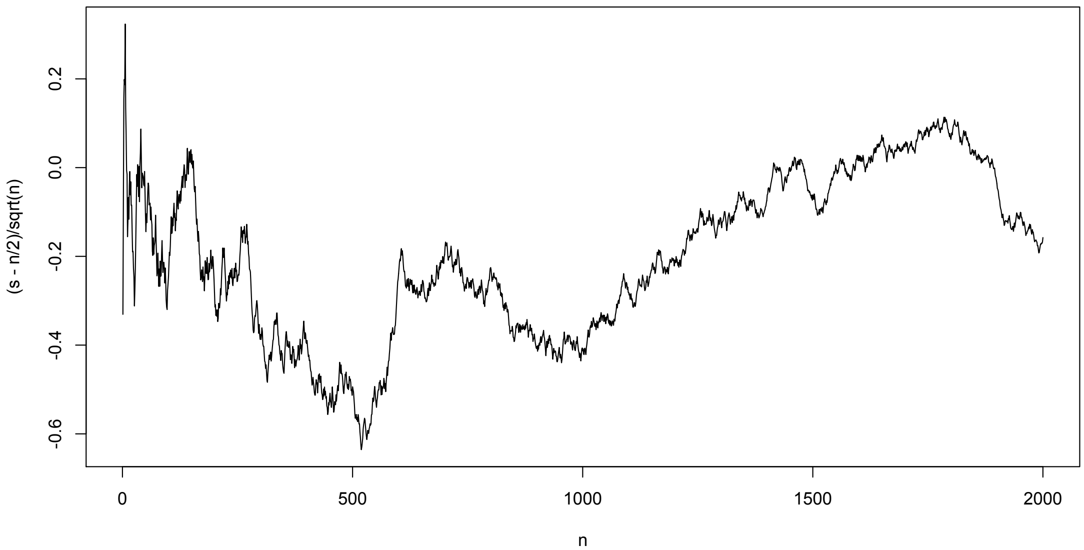

Figure 2: Running average centered at the mean value of 1/2 and magnified by $\sqrt{n}$ extended to 2000 steps.

1 The Classical Central Limit Theorem

Do the size of these fluctuations have any regular and predictable structure? In other words, if we look at distributions for the sum of independent and identically distributed random variables $X_i, 1 = 1, 2, \ldots$, and take their sum,

$$S_n = X_1 + X_2 + \cdots + X_n,$$

what disstribution do we see? Let's look first to the simplest case, X_i Bernoulli random variables.

The curves in Figure 3 are looking like bell curves. Their center and spread vary in ways that are predictable. In order to make comparisons, we should examine standardized versions of the random variables with mean μ and variance σ^2.

To accomplish this,

- we can either standardize using the **sum** S_n having mean $n\mu$ and standard deviation $\sigma\sqrt{n}$, *or*
- we can standardize using the **sample mean** $\bar{X}_n$ having mean μ and standard deviation $\sigma/\sqrt{n}$.

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu). \quad (1)$$

For our next example, we look at the density of the sum of standardized exponential random variables. The exponential density is strongly skewed and so we have have to wait for larger values of n before we see the bell curve emerge.

In Figure 4, we see the densities approaching that of the bell curve for a standard normal random variables. Even for the case of $n = 32$ we see a small amount of skewness that is a remnant of the skewness in the exponential density.

The theoretical result behind these numerical explorations is called the **classical central limit theorem**:

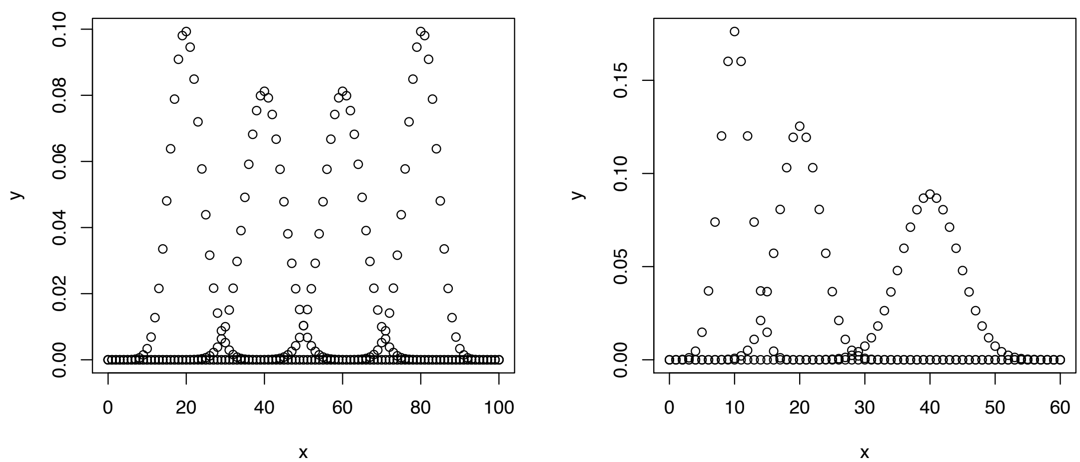

Figure 3: a. Successes in 100 Bernoulli trials with $p = 0.2, 0.4, 0.6$ and 0.8. b. Successes in Bernoulli trials with $p = 1/2$ and $n = 20, 40$ and 80.

Let $\{X_i; i \geq 1\}$ be independent random variables having a common distribution. Let μ be their mean and σ^2 be their variance. Then Z_n as defined by equation (1) converges **in distribution** to Z a standard normal random variable. More precisely, the distribution function for Z_n converges to Φ, the distribution function of the standard normal.

$$\lim_{n\to\infty} P\{Z_n \leq z\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx = \Phi(z).$$

We abbreviate convergence in distribution by writing

$$Z_n \to^{\mathcal{D}} Z.$$

This theorem is an enormously practical tool in providing good estimates for probabilities of events depending on either S_n or $\bar{X}_n$. We shall begin to show this in the following examples.

Example 1. *For Bernoulli random variables, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. S_n is the number of successes in n Bernoulli trials. In this situation, the sample mean is generally denoted by $\hat{p}$ to indicate the fact that it is a* **sample proportion**

The normalized version of S_n is

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{\bar{X}_n - p}{\sqrt{p(1-p)/n}}.$$

Toss a fair coin 100 times, then $\mu = 1/2$ and $\sigma = \sqrt{1/2(1-1/2)} = 1/2$. Thus,

$$Z_{100} = \frac{S_{100} - 50}{5}.$$

So,

$$P\{S_{100} > 65\} = P\left\{\frac{S_{100} - 50}{5} > 3\right\} \approx P\{Z_{100} > 3\} \approx 0.0013.$$

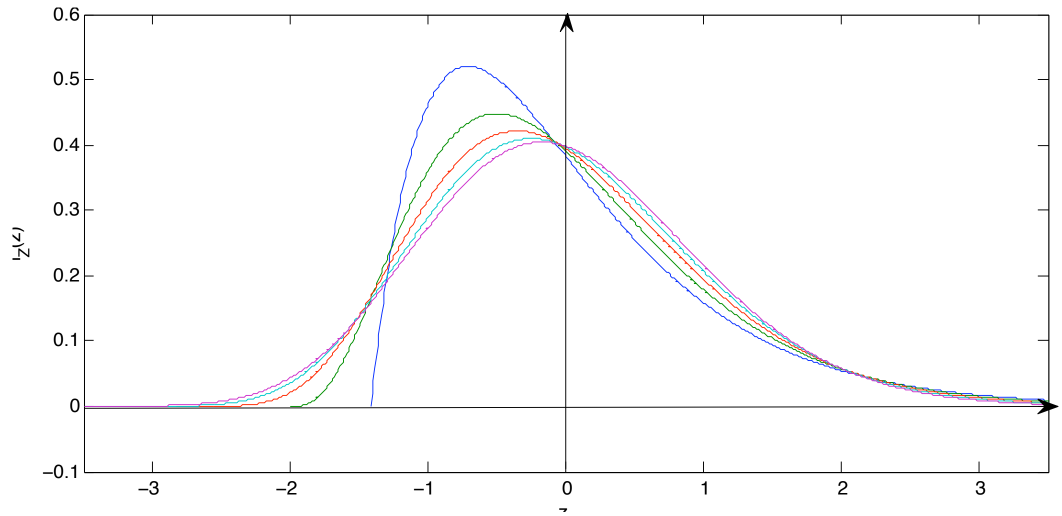

Figure 4: Density of the standardized version of the sum of n independent exponential random variables for $n = 2, 4, 8, 16$ and 32.

```
> 1-pnorm(3)
[1] 0.001349898
```

We could, alternatively write $\hat{p}$ for the empirical franction of coin tosses that result in heads. Then,

$$Z_{100} = \frac{\hat{p} - 1/2}{1/20}.$$

and

$$P\{\hat{p} \leq 0.40\} = P\{20(\hat{p} - 1/2) \leq 20(0.4 - 1/2)\} = P\{Z_n \leq -2\} \approx 0.023.$$

```
> pnorm(-2)
[1] 0.02275013
```

Example 2. *For exponential random variables $\mu = 1/\lambda$ and $\sigma = 1/\lambda$ and therefore*

$$Z_n = \frac{S_n - n/\lambda}{\sqrt{n}/\lambda} = \frac{\lambda S_n - n}{\sqrt{n}}.$$

Let T_{64} be the sum of 64 independent with parameter $\lambda = 1$. Then, $\mu = 1$ and $\sigma = 1$. So,

$$P\{T_{64} < 60\} = P\left\{\frac{T_{64} - 64}{8} < -\frac{1}{2}\right\} = P\{Z_{64} < -0.5\} \approx 0.309.$$

Example 3. *Video projector light bulbs are known to have a mean lifetime of $\mu = 100$ hours and standard deviation $\sigma = 5$. The university uses the projectors for 9000 hours per semester. How likely are the light bulbs to be sufficient for the semester?*

Let S_{100} be the total lifetime of the 100 bulbs. We are asking for the probability that $\{S_n > 9000\}$. Note that this event is equivalent to

$$Z_n = \frac{S_n - 10000}{75 \cdot \sqrt{100}} > \frac{9000 - 10000}{75 \cdot \sqrt{100}} = -\frac{4}{3}$$

and, by the central limit theorem, $P\{Z_n > 4/3\} \approx 0.909$.

2 Propagation of Error

For any random variable Y with mean μ_Y and standard deviation σ_Y, we will be looking at linear functions $aY + b$ for Y. recall that

$$E[aY + b] = a\mu_Y + b, \qquad \text{Var}(aY + b) = a^2\text{Var}(Y).$$

We will apply this to the linear approximation of $g(Y)$ about the point μ_Y.

$$g(Y) \approx g(\mu_Y) + g'(\mu_Y)(Y - \mu_Y). \tag{2}$$

If we take expected values, then

$$Eg(Y) \approx E[g(\mu_Y) + g'(\mu_Y)(Y - \mu_Y)] = g(\mu_Y) + g'(\mu_Y)E]Y - \mu_Y] = g(\mu_Y) + 0 = g(\mu_Y).$$

The variance

$$\text{Var}(g(Y)) \approx \text{Var}(g'(\mu_Y)(Y - \mu_Y)) = g'(\mu_Y)^2\text{Var}(Y - \mu_Y) = g'(\mu_Y)^2\sigma_Y^2.$$

Thus, the standard deviation

$$\sigma_{g(Y)} \approx |g'(\mu_Y)|\sigma_Y \tag{3}$$

gives what is known as the **propagation of error**.

If Y is meant to be some measurement of a quantity q with a measurement subject to error. Then, saying that

$$q = \mu_Y = EY$$

is stating that Y is an **unbiased estimator** of q. In other words, Y does not systematically overestimate or underestimate q. The standard deviation σ_Y gives a sense of the variability in the measurement apparatus. However, if we measure Y but want to give not an estimate for q, but an estimate for a function of q, namely $g(q)$. Its standard deviation is approximation by the formula (3).

Example 4. *If Y is the measurement of a side of a cube with length ℓ. Then Y^3 is an estimate of the volume of the cube. If the measurement error has standard deviation σ_Y, then, taking $g(y) = y^3$, we see that the standard deviation of the error in the measurement of the volume*

$$\sigma_{Y^3} \approx 3q^2\sigma_Y.$$

If we estimate q with Y, then

$$\sigma_{Y^3} \approx 3Y^2\sigma_Y.$$

To estimate the coefficient volume expansion α_3 of a material, we begin with a material of known length ℓ_0 at temperature T_0 and measure the length ℓ_1 at temperature T_1. Then, the coefficient of linear expansion

$$\alpha_1 = \frac{\ell_1 - \ell_0}{\ell_0(T_1 - T_0)}.$$

If the measure length of ℓ_1 ia Y. We estimate this by

$$\hat{\alpha}_1 = \frac{Y - \ell_0}{\ell_0(T_1 - T_0)}.$$

Then, if measurements Y has variance σ_Y^2, then

$$\text{Var}(\hat{\alpha}_1) = \frac{\sigma_Y^2}{\ell_0^2(T_1 - T_0)^2} \quad \sigma_{\hat{\alpha}_1} = \frac{\sigma_Y}{\ell_0|T_1 - T_0|}.$$

Now, we estimate

$$\alpha_3 = \frac{\ell_1^3 - \ell_0^3}{\ell_0^3(T_1 - T_0)} \quad \text{by} \quad \hat{\alpha}_3 = \frac{Y^3 - \ell_0^3}{\ell_0^3(T_1 - T_0)}$$

and

$$\sigma_{\hat{\alpha}_3} \approx 3\ell_1^2\frac{\sigma_Y}{\ell_0^3|T_1 - T_0|}..$$

Often, the function g is a function of several variables. We will show the multivariate propagation of error in the two dimensional case noting that extension to higher the higher dimensional case is straightforward. Now, for **independent** random variables Y_1 and Y_2 with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. The linear approximation about the point (μ_1, μ_2) is

$$g(Y_1, Y_2) \approx g(\mu_1, \mu_2) + \frac{\partial g}{\partial y_1}(\mu_1, \mu_2)(Y_1 - \mu_1) + \frac{\partial g}{\partial y_2}(\mu_1, \mu_2)(Y_2 - \mu_2).$$

As before,

$$Eg(Y_1, Y_2) \approx g(\mu_1, \mu_2).$$

Because Y_1 and Y_2 are independent, then so are

$$\frac{\partial g}{\partial y_1}(\mu_1, \mu_2)(Y_1 - \mu_1) \quad \text{and} \quad \frac{\partial g}{\partial y_2}(\mu_1, \mu_2)(Y_2 - \mu_2)$$

and

$$\begin{aligned}
\text{Var}(g(Y_1, Y_2)) &\approx \text{Var}(\frac{\partial g}{\partial y_1}(\mu_1, \mu_2)(Y_1 - \mu_1)) + \text{Var}(\frac{\partial g}{\partial y_2}(\mu_1, \mu_2)(Y_2 - \mu_2)) \\
&= \left(\frac{\partial g}{\partial y_1}(\mu_1, \mu_2)\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial y_2}(\mu_1, \mu_2)\right)^2 \sigma_2^2.
\end{aligned}$$

$$\sigma_{g(Y_1, Y_2)} = \sqrt{\left(\frac{\partial g}{\partial y_1}(\mu_1, \mu_2)\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial y_2}(\mu_1, \mu_2)\right)^2 \sigma_2^2}.$$

Example 5. *In the previous example, we now estimate the volume of an*

$$\ell_0 \times w_0 \times h_0$$

rectangular solid with the measurements Y_1, Y_2, and Y_3 for, respectively, the length, width, and height with respective variances σ_1, σ_2, and σ_3. Here, we take $g(\ell, w, h) = \ell w h$, then

$$\frac{\partial g}{\partial \ell}(\ell, w, h) = wh, \quad \frac{\partial g}{\partial w}(\ell, w, h) = \ell h, \quad \frac{\partial g}{\partial h}(\ell, w, h) = \ell w$$

and

$$\sigma_{g(Y_1, Y_2, Y_3)} = \sqrt{(wh)^2\sigma_1^2 + (\ell h)^2\sigma_2^2 + +(\ell w)^2\sigma_3^2}.$$

3 Delta Method

Let's use repeated measurements, $Y_1, Y_2, \ldots Y_n$ to estimate a quantity q by $\bar{Y}$. If each measurement has mean μ_Y and variance σ_Y^2, then $\bar{Y}$ has mean μ_Y and variance σ_Y^2/n. We can apply the propagation of error analysis based on a linear approximation of $g(\bar{Y})$ to obtain

$$g(\bar{Y}) \approx g(\mu_Y), \quad \text{and} \quad \text{Var}(g(\bar{Y}) \approx g'(\mu_Y)^2 \frac{\sigma_Y^2}{n}.$$

Thus, the reduction in the variance in the estimate of q "propagates" a reduction in variance in the estimate of $g(q)$.

However, the central limit theorem gives us some additional information. Returning to the linear approximation (2)

$$g(\bar{Y}) \approx g(\mu_Y) + g'(\mu_Y)(\bar{Y} - \mu_Y).$$

The central limit theorem tells us that $\bar{Y}$ has a nearly normal distribution. Consequently, the linear approximation to $g(\bar{Y})$ also has nearly a normal distribution. In addition, the linear approximation is a better approximation in this case because the difference $\bar{Y} - \mu_Y$ is more likely to be small.

The **delta method** combines the central limit theorem and the propagation of error. To see this write,

$$Z_n = \frac{g(\bar{Y}) - g(\mu_y)}{|g'(\mu_Y)|\sigma_Y/\sqrt{n}}.$$

The, Z_n converges in distribution to a standard normal random variable. The delta method greatly extends the applicability of the central limit theorem.

Let's return to our previous example on thermal expansion.

Example 6. *Let $Y_1, Y_2, \ldots, Y_n$ be repeated unbiased measurement of a side of a cube with length ℓ_1 and temperature T_1. We use $\bar{Y}$ to estimate the length at temperature T_1 to estimate for the coefficient of linear expansion.*

$$\hat{\alpha}_1 = \frac{\bar{Y} - \ell_0}{\ell_0(T_1 - T_0)}.$$

Then, if each measurement Y_i has variance σ_Y^2, then

$$\text{Var}(\hat{\alpha}_1) = \frac{\sigma_Y^2}{\ell^2(T_1 - T_0)^2 n} \quad \sigma_{\hat{\alpha}_1} = \frac{\sigma_Y}{\ell|T_1 - T_0|\sqrt{n}}.$$

Now, we estimate the coefficient of volume expansion by

$$\hat{\alpha}_3 = \frac{\bar{Y}^3 - \ell_0^3}{\ell_0^3(T_1 - T_0)}$$

and

$$\sigma_{\hat{\alpha}_3} \approx \frac{3\ell_1^2\sigma_Y}{\ell_0^3|T_1 - T_0|n}.$$

By the delta method,

$$Z_n = \frac{\hat{\alpha}_3 - \alpha_3}{\sigma_{\alpha_3}}$$

has a distribution that can be well approximated by a standard normal random variable.

The next natural step is to take the approach used for the propagation of error in a multidimensional setting and extend the delta method. Focusing on the three dimensional case, we have three **independent** sequences $(Y_{1,1}, \ldots, Y_{1,n_1})$, $(Y_{2,1}, \ldots, Y_{2,n_2})$ and $(Y_{3,1}, \ldots, Y_{3,n_3})$ of independent random variables. The observations in the i-th sequence have mean μ_i and variance σ_i^2 for $i = 1, 2$ and 3. We shall use $\bar{Y}_1$, $\bar{Y}_2$ and $\bar{Y}_3$ to denote the sample means for the three sets of observations. Then, $\bar{Y}_i$ has

$$\text{mean} \quad \mu_i \quad \text{and} \quad \text{variance} \quad \frac{\sigma_i^2}{n_i} \quad \text{for } i = 1, 2, 3.$$

From the propagation of error linear approximation, we obtain

$$Eg(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3) \approx g(\mu_1, \mu_2, \mu_3)$$

and

$$\sigma^2_{g(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3)} = \text{Var}(g(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3) \approx \frac{\partial g}{\partial y_1}(\mu_1, \mu_2, \mu_3)^2 \frac{\sigma_1^2}{n_1} + \frac{\partial g}{\partial y_2}(\mu_1, \mu_2, \mu_3)^2 \frac{\sigma_2^2}{n_2} + \frac{\partial g}{\partial y_3}(\mu_1, \mu_2, \mu_3)^2 \frac{\sigma_3^2}{n_3}.$$

To obtain the normal approximation associated to the delta method, we need to have the additional fact that **the sum of independent normal random variables is also a normal random variable**. Thus, we have that, for n large,

$$Z_n = \frac{g(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3) - g(\mu_1, \mu_2, \mu_3)}{\sigma_{g(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3)}}$$

is approximately a standard normal random variable.

Example 7. *In avian biology, the fecundity B is defined as the number of female fledglings per year. B is a product of three random variables,*

$$B = F \cdot p \cdot N,$$

where F equals mean number of female fledglings per successful nest, p equals nest survival probability, and N equals the mean number of nests built per female per year. Let's collect measurement on n_1 nests to count female fledglings in a successful nest, check n_2 nests for survival probability, and follow n_3 females to count the number of successful nests per year. Our experimental design is structured so that measurements are independent. Then,

$$\sigma^2_{B,n} = \frac{1}{n_1}(pN\sigma_F)^2 + \frac{1}{n_2}(FN\sigma_p)^2 + \frac{1}{n_3}(pF\sigma_N)^2.$$

Using the fact that $\sigma_p^2 = p(1-p)$ for a Bernoulli random variable, we can write the expression above as

$$\left(\frac{\sigma_{B,n}}{B}\right)^2 = \frac{1}{n_1}\left(\frac{\sigma_F}{F}\right)^2 + \frac{1}{n_2}\left(\frac{\sigma_p}{p}\right)^2 + \frac{1}{n_3}\left(\frac{\sigma_N}{N}\right)^2 = \frac{1}{n_1}\left(\frac{\sigma_F}{F}\right)^2 + \frac{1}{n_2}\left(\frac{1-p}{p}\right) + \frac{1}{n_3}\left(\frac{\sigma_N}{N}\right)^2.$$

This gives the individual contributions to the variance of B. The values of n_1, n_2, and n_3 can be adjusted in the collection of data to minimize the variance of B under a variety of experimental designs. In general, the largest choice of the n_i should be chosen to reduce the highest of the three ratios.

Estimates for $\sigma^2_{B,n}$ can be found from the field data. Compute sample means

$$\bar{F}, \quad \hat{p}, \quad \text{and} \quad \bar{N},$$

and sample variance

$$s_F^2, \quad \hat{p}(1-\hat{p}) \quad \text{and} \quad s_N^2.$$

So we estimate the variance in fecundity

$$s^2_{B,n} = \frac{1}{n_1}(\hat{p}\bar{N}s_F)^2 + \frac{1}{n_2}(\bar{F}\bar{N})^2\hat{p}(1-\hat{p}) + \frac{1}{n_3}(\hat{p}\bar{F}s_N)^2$$

This topic leads us naturally into a more general discussion of estimation of parameters..