Topic 11: The Central Limit Theorem*

October 11 and 18, 2011

1 Introduction

In the discussion leading to the law of large numbers, we saw visually that the sample means converges to the distributional mean. In symbols,

$$\bar{X}_n \to \mu \text{ as } n \to \infty.$$

Using the Pythagorean theorem for independent random variables, we obtained the more precise statement that the standard deviation of $\bar{X}_n$ is inversely proportional to $\sqrt{n}$, the square root of the number of observations. For example, for simulations based on observations of independent random variables, uniformly distributed on $[0, 1]$, we see, as anticipated, the running averages converging to

$$\mu = \int_0^1 x f_X(x)\,dx = \int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2},$$

the distributional mean.

Now, we zoom around the mean value of $\mu = 1/2$. Because the standard deviation $\sigma_{\bar{X}_n} \propto 1/\sqrt{n}$, we magnify the difference between the running average and the mean by a factor of $\sqrt{n}$ and investigate the graph of

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

versus n. The results of a simulation are displayed in Figure 1.

As we see in Figure 2, even if we extend this simulation for larger values of n, we continue to see fluctuations about the center of roughly the same size and the size of the fluctuations for a single realization of a simulation cannot be predicted in advance.

Thus, we focus on addressing a broader question: Does the distribution of the size of these fluctuations have any regular and predictable structure? This question and the investigation that led to led to its answer, the **central limit theorem**, constitute one of the most important episodes in mathematics.

2 The Classical Central Limit Theorem

Let's begin by examining the distribution for the sum of $X_1, X_2 \ldots X_n$, independent and identically distributed random variables

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

what distribution do we see? Let's look first to the simplest case, X_i Bernoulli random variables. In this case, the sum S_n is a binomial random variable. We look at two cases - in the first we keep the number of trials the same at $n = 100$ and vary the success probability p. In the second case, we keep the success probability the same at $p = 1/2$, but vary the number of trials.

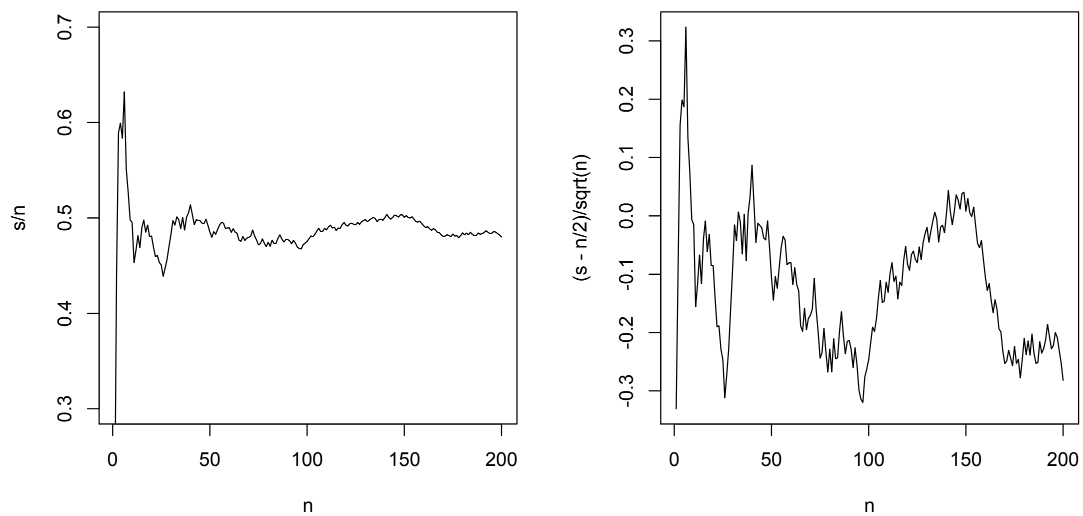

Figure 1: a. Running average of independent random variables, uniform on $[0, 1]$. b. Running average centered at the mean value of 1/2 and magnified by $\sqrt{n}$.

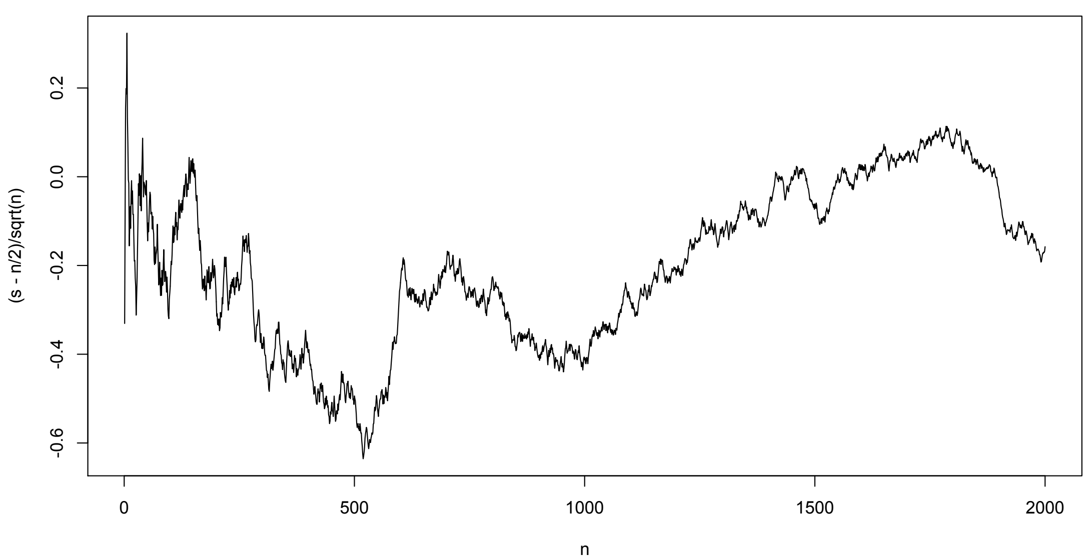

Figure 2: Running average centered at the mean value of 1/2 and magnified by $\sqrt{n}$ extended to 2000 steps.

The curves in Figure 3 are looking like bell curves. Their center and spread vary in ways that are predictable. The binomial random variable S_n has

$$\text{mean } np \text{ and standard deviation } \sqrt{np(1-p)}.$$

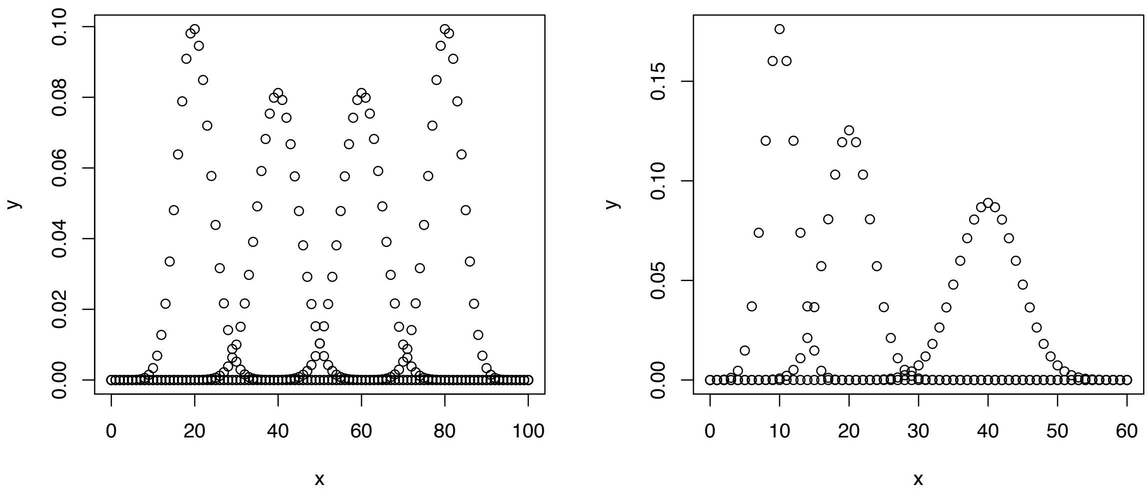

Figure 3: a. Successes in 100 Bernoulli trials with $p = 0.2, 0.4, 0.6$ and 0.8. b. Successes in Bernoulli trials with $p = 1/2$ and $n = 20, 40$ and 80.

Thus, if we take the standarized version of these sums of Bernoulli random variables

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}},$$

then these bell curve graphs would lie on top of each other.

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 order to make comparisons, we 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 $\sigma/\sqrt{n}$.

This yields two equivalent versions of the z-score.

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

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**:

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, the standardized scores defined by equation (1), converges **in distribution** to Z a standard normal random variable. More precisely, the distribution function F_{Z_n} converges to Φ, the distribution function of the standard normal.

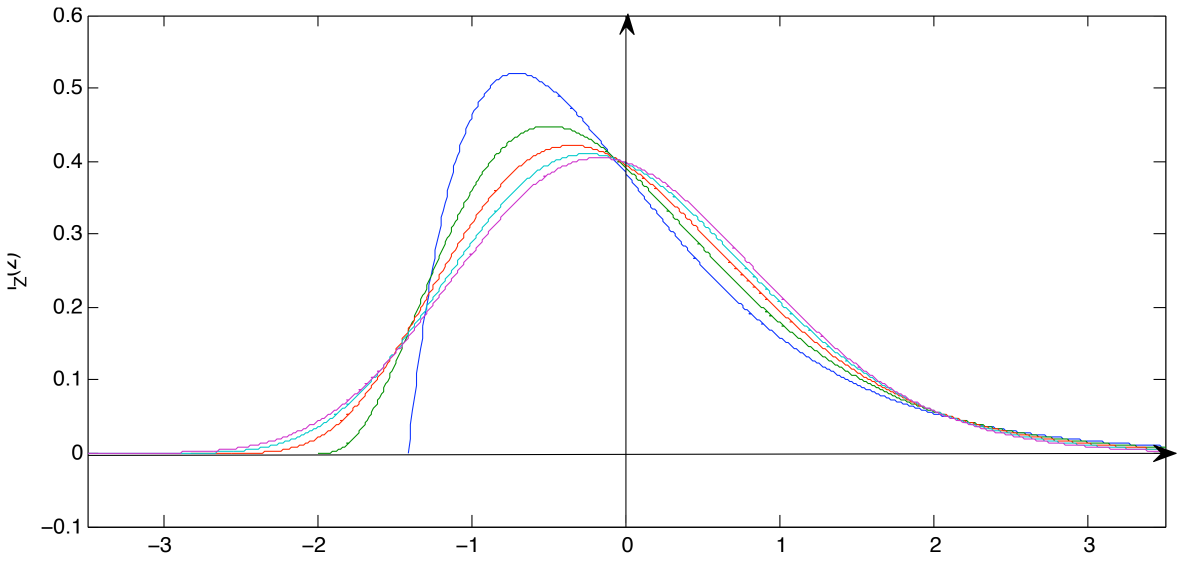

Figure 4: Displaying the central limit theorem graphically. Density of the standardized version of the sum of n independent exponential random variables for $n = 2$ (dark blue), 4 (green), 8 (red), 16 (light blue), and 32 (magenta). Note how the skewness of the exponential distribution slowly gives way to the bell curve shape of the normal distribution.

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

We abbreviate convergence in distribution by writing

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

In practical terms the central limit theorem states that $P\{a < Z_n \le b\} \approx P\{a < Z \le b\} = \Phi(b) - \Phi(a)$.

This theorem is an enormously useful 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 the fraction of trials that result in a success. This is generally denoted by $\hat{p}$ to indicate the fact that it is a* **sample proportion***.*

The normalized versions of S_n and $\hat{p}$ are equal to

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{\hat{p}_n - p}{\sqrt{p(1-p)/n}},$$

For example, in 100 tosses of a fair coin, $\mu = 1/2$ and $\sigma = \sqrt{1/2(1-1/2)} = 1/2$. Thus,

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

So,

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

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

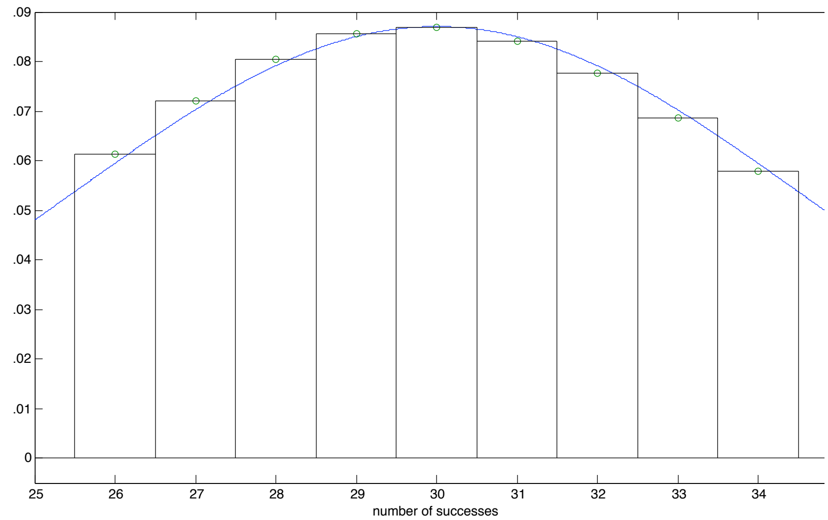

Figure 5: Mass function for a $Bin(100, 0.3)$ random variable (black) and approximating normal density $N(100 \cdot 0.3, \sqrt{100 \cdot \cdot 0.3 \cdot 0.7})$.

We could also write,

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

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
```

Remark 2. *We can improve the normal approximation to the binomial random variable by employing the* **continuity correction***. For a binomial random variable X, the distribution function*

$$P\{X \leq x\} = P\{X < x + 1\} = \sum_{y=0}^{x} P\{X = y\}$$

can be realized as the area of $x+1$ rectangles, height $P\{X = y\}$, $y = 0, 1, \ldots, x$ and width 1. These rectangles look like a Riemann sum for the integral up to the value $x + 1/2$. For the example in Figure 5, $P\{X \leq 32\} = P\{X < 33\}$ is the area of 33 rectangles. This right side of rectangles is at the value 32.5. Thus, for the approximating normal random variable Y, this suggests computing $P\{Y \leq 32.5\}$. In this example the exact value

```
> pbinom(32,100,0.3)
[1] 0.7107186
```

Comparing this to possible choices for the normal approximations

```
> x<-c(32,32.5,33)
> data.frame(x,prob)
     x      prob
1 32.0 0.6687397
2 32.5 0.7073105
3 33.0 0.7436546
```

This shows a difference of 0.0034.

Example 3. *Opinion polls are generally designed to be modeled as Bernoulli trials. The number of trials n is set to give a prescribed value m of two times the standard deviation of $\hat{p}$. This value of m is an example of a* **margin of error**. *The standard deviation*

$$\sqrt{p(1-p)/n}$$

takes on its maximum value for $p = 1/2$. For this case,

$$m = 2\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)/n} = \frac{1}{\sqrt{n}}$$

Thus,

$$n = \frac{1}{m^2}$$

We display the results in R *for typical values of m.*

```
> m<-seq(0.01,0.05,0.01)
> n<-1/m^2
> data.frame(m,n)
     m         n
1 0.01 10000.000
2 0.02  2500.000
3 0.03  1111.111
4 0.04   625.000
5 0.05   400.000
```

Exercise 4. *We have two approximation methods for a large number n of Bernoulli trials - Poisson, which applies then p is small and their product $\lambda = np$ is moderate and normal when the mean number of successes np or the mean number of failures $n(1-p)$ is sufficiently large. Investigate the approximation of the distribution, X, a Poisson random variable, by the distribution of a normal random variable, Y, for the case $\lambda = 16$. Use the continuity correction to compare*

$$P\{X \le x\} \quad \text{to} \quad P\{Y \le x + \frac{1}{2}\}.$$

Example 5. *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,

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

Example 6. *Video projector light bulbs are known to have a mean lifetime of $\mu = 100$ hours and standard deviation $\sigma = 75$. The university uses the projectors for 9000 hours per semester. How likely are 100 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 $\{S_{100} > 9000\}$. Note that this event is equivalent to

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

and, by the central limit theorem, $P\{Z_{100} > 4/3\} \approx 0.909$.

Exercise 7. *Use the central limit theorem to estimate the number of light bulbs necessary to have a 1% chance of running out of light bulbs before the semester ends.*

Exercise 8. *Simulate 1000 times, $\bar{x}$, the sample mean of 100 random variables, uniformly distributed on $[0,1]$. Show a histogram for these simulations to see the approximation to a normal distribution. Find the mean and standard deviations for the simulations and compare them to their distributional values. Use both the simulation and the central limit theorem to estimate the 35th percentile of $\bar{X}$.*

3 Propagation of Error

Propagation of error or **propagation of uncertainty** is a strategy to estimate the impact on the standard deviation of the consequences of a nonlinear transformation of a measured quantity whose measurement is subject to some uncertainty.

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 formula (3).

Example 9. *Let Y be 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 is Y. We estimate this by

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

Then, if a measurement Y of ℓ_1 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

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

Exercise 10. *In an effort to estimate the angle θ of the sun, the length ℓ of a shadow from a 10 meter flag pole is measured. If $\sigma_{\hat{\ell}}$ is the standard deviation for the length measurement, use propagation of error to estimate $\sigma_{\hat{\theta}}$, the standard deviation in the estimate of the angle.*

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 the higher dimensional case is straightforward. Now, for 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,

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

For Y_1 and Y_2 **independent**, we also have that the random variables

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

are independent. Because the variance of the sum of independent random variables is the sum of their variances, we have the approximation

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

and consequently, the standard deviation,

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

Exercise 11. *Repeat the exercise in the case that the height h if the flag poll is also unknown and is measured independently of the shadow length with standard deviation $\sigma_{\hat{h}}$. Comment on the case in which the two standard deviations are equal.*

Exercise 12. *Generalize the formula for the variance to the case of $g(Y_1, Y_2, \ldots, Y_d)$ for independent random variables $Y_1, Y_2, \ldots, Y_d$.*

Example 13. *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 ℓ_0, width w_0, and height h_0 with respective standard deviations σ_ℓ, σ_w, and σ_h. Here, we take $g(\ell, w, h) = \ell wh$, 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

$$\sigma_{g(Y_1,Y_2,Y_3)} \approx \sqrt{\left(\frac{\partial g}{\partial \ell}(\ell_0,w_0,h_0)\right)^2\sigma_\ell^2 + \left(\frac{\partial g}{\partial w}(\ell_0,w_0,h_0)\right)^2\sigma_w^2 + \left(\frac{\partial g}{\partial h}(\ell_0,w_0,h_0)\right)^2\sigma_h^2}$$
$$= \sqrt{(wh)^2\sigma_1^2 + (\ell h)^2\sigma_2^2 + (\ell w)^2\sigma_3^2}.$$

4 Delta Method

Let's use repeated independent measurements, $Y_1, Y_2, \ldots Y_n$ to estimate a quantity q by its sample mean $\bar{Y}$. If each measurement has mean μ_Y and variance σ_Y^2, then $\bar{Y}$ has mean $q = \mu_Y$ 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" to 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. Thus, the linear approximation to $g(\bar{Y})$ also has nearly a normal distribution. Moreover, with repeated measurements, the variance of $\bar{Y}$ is the variance of a single measurement divided by n. As a consequence, the linear approximation under repeated measurements yields a better approximation because the reduction in variance implies that 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,

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

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

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

Example 14. *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 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,

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

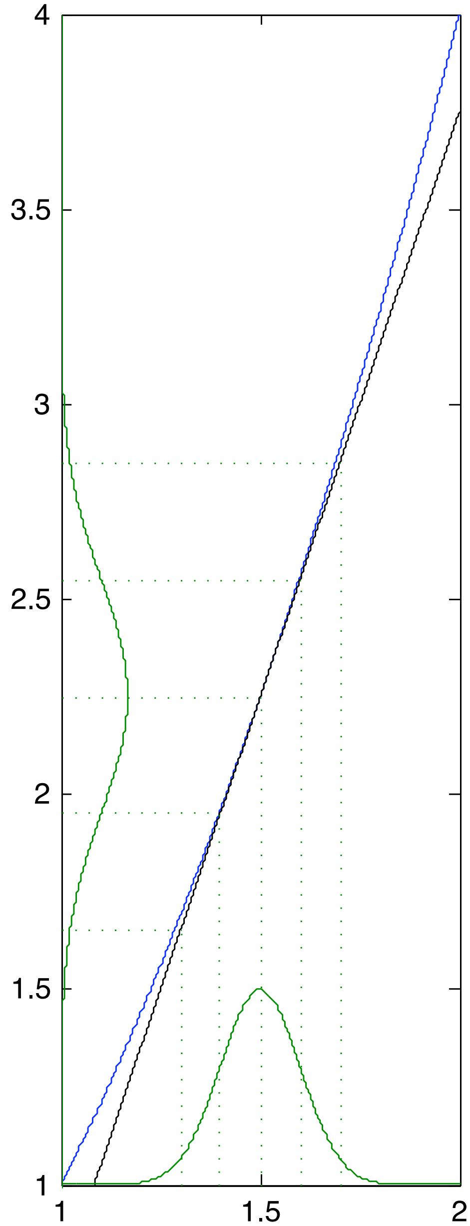

Figure 6: Illustrating the delta method. Here $\mu = E\bar{X} = 1.5$ and the blue curve $g(x) = x^2$. Thus, $g(\bar{X})$ is approximately normal with approximate mean 2.25 and $\sigma_{g(\bar{X})} \approx 3\sigma_{\bar{X}}$. The bell curve on the y-axis is the reflection of the bell curve on the x-axis about the (black) tangent line $y = g(\mu) + g'(\mu)(x - \mu)$.

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

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

By the delta method,

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

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

$$\sigma^2_{g(\bar{Y}_1,\bar{Y}_2,\bar{Y}_3)} = \text{Var}(g(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3)) \approx \frac{\partial g}{\partial y_1}(\mu_1, \mu_2, \mu_3)^2\frac{\sigma_1^2}{n_1} + \frac{\partial g}{\partial y_2}(\mu_1, \mu_2, \mu_3)^2\frac{\sigma_2^2}{n_2} + \frac{\partial g}{\partial y_3}(\mu_1, \mu_2, \mu_3)^2\frac{\sigma_3^2}{n_3}. \quad (4)$$

To obtain the normal approximation associated with 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 15. *In avian biology, the fecundity B is defined as the number of female fledglings per female per year. B is a product of three random variables,*

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

where μ_F equals mean number of female fledglings per successful nest, p equals nest survival probability, and μ_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, using (4), to $B = g(F, p.N) = F \cdot p \cdot N$

$$\sigma^2_{B,n} = \frac{1}{n_1}(p\mu_N\sigma_F)^2 + \frac{1}{n_2}(\mu_F\mu_F N\sigma_p)^2 + \frac{1}{n_3}(p\mu_F\sigma_N)^2.$$

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

$$\left(\frac{\sigma^2_{B,n}}{\mu_B}\right)^2 = \frac{1}{n_1}\left(\frac{\sigma_F}{\mu_F}\right)^2 + \frac{1}{n_2}\left(\frac{\sigma_p}{p}\right)^2 + \frac{1}{n_3}\left(\frac{\sigma_N}{\mu_N}\right)^2 = \frac{1}{n_1}\left(\frac{\sigma_F}{\mu_F}\right)^2 + \frac{1}{n_2}\left(\frac{1-p}{p}\right) + \frac{1}{n_3}\left(\frac{\sigma_N}{\mu_N}\right)^2.$$

This gives the individual contributions to the variance of B from each of the three data collecting activities. 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 in parentheses..

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

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

Exercise 16. *Give the formula for $\sigma^2_{B,n}$ in the case that measurements for F, p, and N are not independent.*

This topic leads us naturally to the next topic - a more general discussion of estimation of parameters. We finish the discussion on the central limit theorem with a summary of some of its applications.

5 Summary of Normal Approximations

The z-score of some random quantity is

$$Z = \frac{\text{random quantity} - \text{mean}}{\text{standard deviation}}.$$

The **central limit theorem** and extensions like the **delta method** tell us when the z-score has an approximately standard normal distribution. Thus, using R, we can find good approximations for the probabilities of $P\{Z < z\}$, `pnorm(z)` and $P\{Z > z\}$ using `1-pnorm(z)` or $P\{z_1 < Z < z_2\}$ using the difference `pnorm(z2) - pnorm(z1)`

5.1 Sample Sum

If we have a sum S_n of n independent random variables, $X_1, X_2, \ldots X_n$ whose common distribution has mean μ and variance σ^2, then

- the mean $ES_n = n\mu$,
- the variance $\text{Var}(S_n) = n\sigma^2$,
- the standard deviation is $\sigma\sqrt{n}$.

Thus, S_n is approximately normal with mean $n\mu$ and variance $n\sigma^2$. The z-score in this case is

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

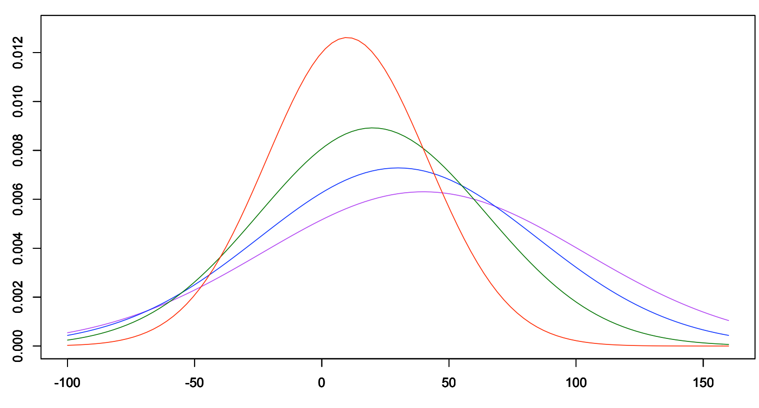

Figure 7: The density function for S_n for a random sample of size $n = 10$ (red), 20 (green), 30 (blue), and 40 (purple). In this example, the observations are normally distributed with mean $\mu = 1$ and standard deviation $\sigma = 10$.

We can approximate $P\{S_n < x\}$ by noting that this is the same as computing

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} < \frac{x - n\mu}{\sigma\sqrt{n}} = z$$

and finding $P\{Z_n < z\}$ using the standard normal distribution.

For the special case of Bernoulli trials with success probability p, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. In this case, normal approximations often use a continuity correction.

5.2 Sample Mean

For a sample mean $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$,

- the mean $E\bar{X} = \mu$,
- the variance $\text{Var}(\bar{X}) = \sigma^2/n$,
- the standard deviation is $\sigma/\sqrt{n}$.

Thus, $\bar{X}$ is approximately normal with mean μ and variance σ^2/n. The z-score in this case is

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Thus,

$$\bar{X} < x \quad \text{is equivalent to} \quad Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{x - \mu}{\sigma/\sqrt{n}}.$$

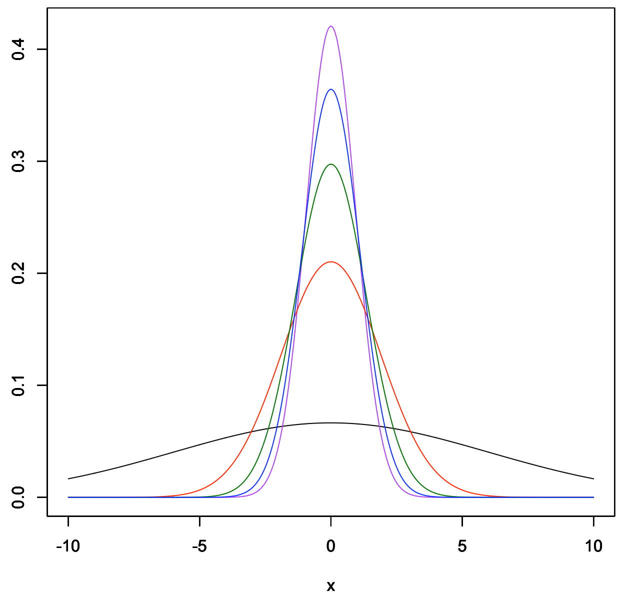

Figure 8: The density function for $\bar{X} - \mu$ for a random sample of size $n = 1$ (black), 10 (red), 20 (green), 30 (blue), and 40 (purple). In this example, the observations are normally distributed with standard deviation $\sigma = 10$.

5.3 Sample Proportion

For Bernoulli trials $X_1, X_2, \ldots X_n$ with success probability p, Let $\hat{p} = (X_1 + X_2 + \cdots + X_n)/n$ be the sample proportion. Then

- the mean $E\hat{p} = p$,
- the variance $\text{Var}(\hat{p}) = p(1-p)/n$,
- the standard deviation is $\sqrt{p(1-p)/n}$.

Thus, $\hat{p}$ is approximately normal with mean p and variance $p(1-p)/n$. The z-score in this case is

$$Z_n = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}.$$

5.4 Delta Method

For the delta method in one variable using $\bar{X}$ and a function g, for a sample mean $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$, we have

- the mean $Eg(\bar{X}) \approx g(\mu)$,
- the variance $\text{Var}(g(\bar{X})) \approx g'(\mu)^2\sigma^2/n$,
- the standard deviation is $|g'(\mu)|\sigma/\sqrt{n}$.

Thus, $g(\bar{X})$ is approximately normal with mean $g(\mu)$ and variance $g'(\mu)^2\sigma^2/n$. The z-score is

$$Z_n = \frac{g(\bar{X}) - g(\mu)}{|g'(\mu)|\sigma/\sqrt{n}}.$$

For the two variable delta method, we now have two independent sequences of independent random variables, $X_1, X_2, \ldots X_{n_1}$ whose common distribution has mean μ_1 and variance σ_1^2 and $Y_1, Y_2, \ldots Y_{n_2}$ whose common distribution has mean μ_2 and variance σ_2^2. For a function g of the sample means, we have that

- the mean $Eg(\bar{X}, \bar{Y}) \approx g(\mu_1, \mu_2)$,
- the variance
$$\text{Var}(g(\bar{X}, \bar{Y})) = \sigma_{g,n}^2 \approx \left(\frac{\partial}{\partial x} g(\mu_1, \mu_2)\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{\partial}{\partial y} g(\mu_1, \mu_2)\right)^2 \frac{\sigma_2^2}{n_2},$$
- the standard deviation is $\sigma_{g,n}$.

Thus, $g(\bar{X}, \bar{Y})$ is approximately normal with mean $g(\mu_1, \mu_2)$ and variance $\sigma_{g,n}^2$. The z-score is

$$Z_n = \frac{g(\bar{X}, \bar{Y}) - g(\mu_1, \mu_2)}{\sigma_{g,n}}.$$

The generalization of the delta method to higher dimensional data will add terms to the variance formula.

6 Answers to Selected Exercises

3. A Poisson random variable with parameter $\lambda = 16$ has mean 16 and standard deviation $4 = \sqrt{16}$. Thus, we first look at the maximum difference in the distribution function of a $Pois(4)$ random variable, X, and a $N(16, 4)$ random variable, Y, by comparing $P\{X \leq x\}$ to $P\{Y \leq x + \frac{1}{2}\}$ in a range around the mean value of 16.

```
> x<-c(4:28)
> max(abs(pnorm(x+0.5,16,4)-ppois(x,16)))
[1] 0.01648312
```

The maximum difference between the distribution function is approximately 1.6%. To compare the density functions, we have

```
> poismass<-dpois(x,16)
> plot(x,poismass,ylim=c(0,0.1),
  ylab="probability")
> par(new=TRUE)
> x<-seq(4,28,0.01)
> normd<-dnorm(x,16,4)
> plot(x,normd,ylim=c(0,0.1),
  ylab="probability",type="l",col="red")
```

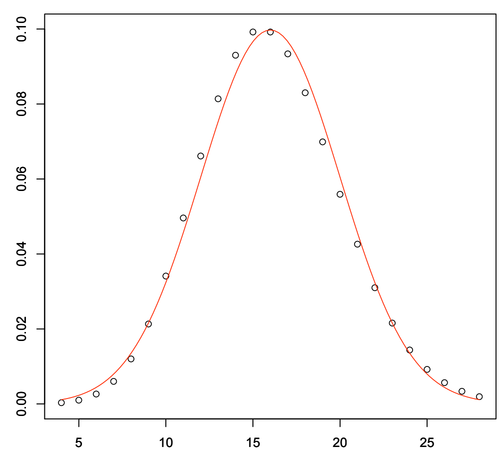

Figure 9: Circles indicate the mass function for a $Pois(16)$ random variable. The red curve is the density function of a $N(16, 4)$ random variable. The plots show that the Poisson random variable is slightly more skewed to the right that the normal.

6. For Z a standard normal random variable to determine $z_{0.01}$ that satisfies $P\{Z > z_{0.01}\} = 0.01$, we use the R command

```
> qnorm(0.99)
[1] 2.326348
```

Thus, we look for the value n that gives a standardized score of $z_{0.01}$.

$$\begin{aligned} 2.326348 = z_{0.01} = \frac{100n - 9000}{75 \cdot \sqrt{n}} &= \frac{4n-360}{3 \cdot \sqrt{n}} \\ 6.979044\sqrt{n} &= 4n - 360 \\ 48.70705n &= 129600 - 2880n + 16n^2 \\ 0 &= 129600 - 2928.707n + 16n^2 \end{aligned}$$

By the quadratic formula, we solve for n.

$$n = \frac{2928.707 + \sqrt{(2928.707)^2 - 4 \cdot 16 \cdot 129600}}{2 \cdot 16} = 108.1442$$

So, take $n = 109$.

7. The R code for the simulations is

```
> xbar<-rep(0,1000)
> for (i in 1:1000)
   {x<-runif(100);xbar[i]<-mean(x)}
> hist(xbar)
> mean(xbar)
[1] 0.498483
> sd(xbar)
[1] 0.02901234
> quantile(xbar,0.35)
      35%
0.488918
> qnorm(0.35)
[1] -0.3853205
```

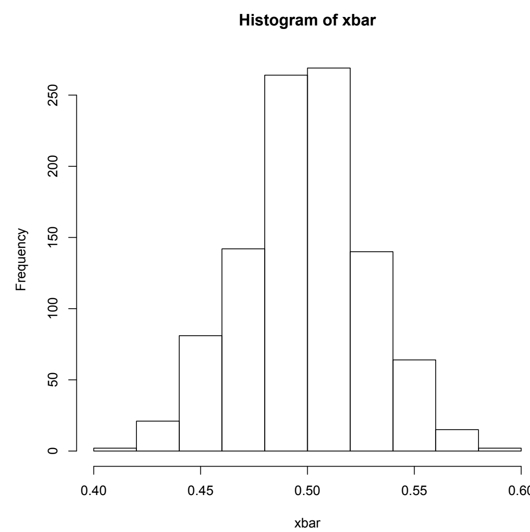

Figure 10: Histogram of the sample means of 100 random variables, uniformly distrobuted on $[0, 1]$.

The mean of a $U[0, 1]$ random variable is $\mu = 1/2$ and its variance is $\sigma^2 = 1/12$. Thus the mean of $\bar{X}$ is 1/2, its standard deviation is $\sqrt{1/(12 \cdot 100)} = 0.0289$, close to the simulated values.

The 35th percentile corresponds to a z-score of -0.3853205. Thus,

$$-0.3853 = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 0.5}{0.0289}$$

$$\bar{X} = -0.3853 \cdot 0.0289 + 0.5 = 0.4889,$$

agreeing to four decimal places the value given by the simulations of `xbar`.

10. Using right triangle trigonometry, we have that

$$\theta = g(\ell) = \tan^1 \left(\frac{\ell}{10} \right). \quad \text{Thus,} \quad g'(\ell) = \frac{1/10}{1 + (\ell/10)^2} = \frac{10}{100 + \ell^2}.$$

So, $\sigma_{\hat{\theta}} \approx 10/(100 + \ell^2) \cdot \sigma_\ell$. For example, set $\sigma_\ell = 0.1$ meter and $\ell = 5$. Then, $\sigma_{\hat{\theta}} \approx 10/125 \cdot 0.1 = 1/125$ radians = 0.49°.

11. In this case,

$$\theta = g(\ell, h) = \tan^1\left(\frac{\ell}{h}\right).$$

For the partial derivatives, we use the chain rule

$$\frac{\partial g}{\partial \ell}(\ell, h) = \frac{1}{1+(\ell/h)^2}\left(\frac{1}{h}\right) = \frac{h}{h^2+\ell^2} \quad \frac{\partial g}{\partial h}(\ell, h) = \frac{1}{1+(\ell/h)^2}\left(\frac{-\ell}{h^2}\right) = -\frac{\ell}{h^2+\ell^2}$$

Thus,

$$\sigma_{\hat\theta} \approx \sqrt{\left(\frac{h}{h^2+\ell^2}\right)^2\sigma_\ell^2 + \left(\frac{\ell}{h^2+\ell^2}\right)^2\sigma_h^2} = \frac{1}{h^2+\ell^2}\sqrt{h^2\sigma_\ell^2+\ell^2\sigma_h^2}.$$

If $\sigma_h = \sigma_\ell$, let σ denote their common value. Then

$$\sigma_{\hat\theta} \approx \frac{1}{h^2+\ell^2}\sqrt{h^2\sigma^2+\ell^2\sigma^2} = \frac{\sigma}{\sqrt{h^2+\ell^2}}.$$

In other words, $\sigma_{\hat\theta}$ is inversely proportional to the length of the hypotenuse.

12. Let μ_i be the mean of the i-th measurement. Then

$$\sigma_{g(Y_1,Y_2,\cdot,Y_dt)} \approx \sqrt{\left(\frac{\partial g}{\partial y_1}(\mu_1,\ldots,\mu_d)\right)^2\sigma_1^2 + \left(\frac{\partial g}{\partial y_2}(\mu_1,\ldots,\mu_d)\right)^2\sigma_2^2 + \cdots + \left(\frac{\partial g}{\partial y_d}(\mu_1,\ldots,\mu_d)\right)^2\sigma_d^2}.$$

14. Recall that for random variables X_1, X_2, X_3 and constants c_1, c_2, c_3,

$$\text{Var}(c_0 + c_1X_1 + c_2X_2 + c_3X_3) = \sum_{i=1}^3\sum_{j=3}^3 c_ic_j\text{Cov}(X_i, X_j) = \sum_{i=1}^3\sum_{j=3}^3 c_ic_j\rho_{i,j}\sigma_i\sigma_j.$$

where $\rho_{i,j}$ is the correlation of X_i and X_j. Note that the correlation of a random variable with itself, $\rho_{i,i} = 1$ Let F_0, p_0, N_0 be the actual values of the variables under consideration. Then we have the linear approximation,

$$\begin{aligned} g(\hat F, \hat p, \hat N) &\approx g(F_0,p_0,N_0) + \frac{\partial g}{\partial F}(F_0,p_0,N_0)(\hat F - F_0) + \frac{\partial g}{\partial p}(F_0,p_0,N_0)(\hat p - p_0) + \frac{\partial g}{\partial N}(F_0,p_0,N_0)(\hat N - N_0). \\ &= g(F_0,p_0,N_0) + p_0N_0(\hat F - F_0) + F_0N_0(\hat p - p_0) + F_0p_0(\hat N - N_0) \end{aligned}$$

Matching this to the covariance formula, we have

$$c_0 = g(F_0,p_0,N_0), \quad c_1 = p_0N_0, \quad c_2 = F_0N_0, \quad c_3 = F_0p_0,$$

$$X_1 = \hat F, \quad X_2 = \hat p, \quad X_3 = \hat N.$$

Thus,

$$\sigma_{B,n}^2 = \frac{1}{n_1}(pN\sigma_F)^2 + \frac{1}{n_2}(FN\sigma_p)^2 + \frac{1}{n_3}(pF\sigma_N)^2 + 2p_0F_0N_0^2\rho_{1,2}\frac{\sigma_1\sigma_2}{\sqrt{n_1n_2}} + 2p_0^2F_0N_0\rho_{1,3}\frac{\sigma_1\sigma_3}{\sqrt{n_1n_3}} + 2F_0^2p_0N_0\rho_{2,3}\frac{\sigma_2\sigma_3}{\sqrt{n_2n_3}}.$$