1. Results of sampling distributions from $N(\mu, \sigma^2)$

Recall two basic facts about the normal distribution:

(1) If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

(2) If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ and they are independent, then $X+Y \sim N(\mu+\nu, \sigma^2+\tau^2)$.

You can prove these results using the MGF technique.

Result 1: Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Then, we have the following:

(1) $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, where $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$

(2) $\frac{nW_n}{\sigma^2} \sim \chi^2_n$, where $W_n = \frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2$

(3) $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$, where $s_n^2 = \frac{1}{n-1}\sum_{k=1}^n \left(X_k - \bar{X}_n\right)^2$

(4) $\bar{X}_n$ and s_n^2 are independent

(5) $$\frac{\bar{X} - \mu}{s_n/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{s_n} \sim t_{n-1}.$$

The *first result* is easy.

The *second result* is also familiar, since $Z_i = (X_i - \mu)/\sigma$ for $1 \le i \le n$ are i.i.d. $N(0,1)$ variables and we have seen that sum of squares of n i.i.d. $N(0,1)$ variables has χ^2_n distribution.

Exercise 1. Let $Z_1, \ldots, Z_n$ be i.i.d. $N(0,1)$ random variables. Then, $Z_1^2 + \cdots + Z_n^2 \sim \text{Gamma}(n/2, 1/2)$. In statistics, the distribution $\text{Gamma}(n/2, 1/2)$ is usually called the *chi-squared distribution with n degrees of freedom*.
Also, recall 'Sampling Distributions' from Module 1 of Note 9.

It remains to show the *next three* facts. They can be done as follows. Firstly, we use the standardized variables $Z_i = (X_i - \mu)/\sigma$ for $1 \le i \le n$ which are i.i.d. $N(0,1)$. The goal is to show that $\bar{Z}_n$ and $\sum_{k=1}^n \left(Z_k - \bar{Z}_n\right)^2$ are independent, and the latter has χ^2_{n-1} distribution. Define

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{h,n} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix},$$

where the matrix $A = (a_{i,j})_{i,j \le n}$ is chosen as follows: Let the entries in the first row be $1/\sqrt{n}$. Then, the first row be orthogonal to the second, third, ... rows of the matrix. There is a lot of choice, but it does not matter how we pick the orthonormal basis. With this, the matrix A becomes an

orthogonal matrix, i.e., $AA' = A'A = I_n$. Because of this, $Y_1, \ldots, Y_n$ are also i.i.d. $N(0,1)$ random variables (recall properties of multivariate normal distribution to verify that $Y = (Y_1, \ldots, Y_n) \sim N_n(0, I_n)$).

Further, $Y_1 = \sqrt{n}\bar{Z}_n$. From the orthogonality of A, it follows that $Y_1^2 + \cdots + Y_n^2 = Z_1^2 + \cdots + Z_n^2$ (because $Y'Y = Z'A'AZ = Z'Z$). Consequently, we get

$$\begin{aligned} Y_2^2 + \cdots + Y_n^2 &= Z_1^2 + \cdots + Z_n^2 - n\bar{Z}_n^2 \\ &= \left(Z_1 - \bar{Z}_n\right)^2 + \cdots + \left(Z_n - \bar{Z}_n\right)^2. \end{aligned}$$

This shows that $\left(Z_1 - \bar{Z}_n\right)^2 + \cdots + \left(Z_n - \bar{Z}_n\right)^2$ depends only on $Y_2, \ldots, Y_n$ and hence is independent of $\bar{Z}_n$ which depends on Y_1 alone. Further, $Y_2^2 + \cdots + Y_n^2$ has χ^2_{n-1} distribution, being a sum of squares of $(n-1)$ i.i.d. $N(0,1)$ variables.

Exercise 2. Let $U \sim N(0,1)$ and $V^2 \sim \chi^2_n$ be independent. Then, the density of $\frac{U}{V/\sqrt{n}}$ is given by

$$\frac{1}{\sqrt{n-1}\text{Beta}(\frac{1}{2}, \frac{n-1}{2})} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}}}$$

for all $t \in \mathbb{R}$. This is known as *Student's t-distribution*.
This can be derived from the change of variable formula, and its CDF (check R!) can be tabulated by numerical integration. Again, recall 'Sampling Distributions' from Module 1 of Note 9.

How does this result help us? We know that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$, $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$, and the two are independent. Take these random variables in the above exercise to conclude that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n}$ has t_{n-1} distribution.

Now, recall 'Student's Theorem' from Module 2 of Note 9.

2. CONFIDENCE INTERVALS

So far, in estimating of an unknown parameter, we give a *single number* as our guess for the known parameter. It would be better to give an interval, and say with what confidence we expect the true parameter to lie within it. As a very simple example, suppose we have one random variable X with $N(\mu, 1)$ distribution. How do we estimate μ? Suppose the observed value of X is 2.7. Going by any method, the guess for μ would be 2.7 itself. But of course μ is not equal to X, so we would like to give an interval in which μ lies. How about $[X-1, X+1]$? Or $[X-2, X+2]$? Using normal tables, we see that $\mathbf{P}(X-1 < \mu < X+1) = \mathbf{P}(-1 < (X-\mu) < 1) = \mathbf{P}(-1 < Z < 1) \approx 0.68$ and similarly $\mathbf{P}(X-2 < \mu < X+2) \approx 0.95$. Thus, by making the interval longer we can be more confident that the true parameter lies within. But, the accuracy of our statement goes down (if you want to know the average height of students in MSO201A, and the answer you give is "between 100cm and 200cm", it is very probably correct, but of little use!). The probability with which our confidence interval (CI) contains the unknown parameter is called the level of confidence. Usually we fix the level of confidence (say, 0.90) and find an interval *as short as possible* but subject to the condition that it should have a confidence level of 0.90.

In this section, we consider the problem of confidence intervals for the **normal** population.

The setting: Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. We consider four situations.

(1) Confidence interval for μ (when σ^2 is known).

(2) Confidence interval for σ^2 (when μ is known).

(3) Confidence interval for μ (when σ^2 is unknown).

(4) Confidence interval for σ^2 (when μ is unknown).

A starting point in finding a confidence interval for a parameter is to first start with an estimate for the parameter. For example, in finding a CI for μ, we may start with $\bar{X}_n$ and enlarge it to an interval $[\bar{X}_n - a, \bar{X}_n + a]$. Similarly, in finding a CI for σ^2, we use the estimate $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ if μ is unknown and $W_n = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ if the value of μ is known.

2.1. **Estimating μ when σ^2 is known.** We look for a confidence interval of the form $I_n = [\bar{X}_n - a, \bar{X}_n + a]$. Then,

$$\mathbf{P}\left(I_n \ni \mu\right) = \mathbf{P}\left(-a \le \bar{X}_n - \mu \le a\right) = \mathbf{P}\left(-\frac{a\sqrt{n}}{\sigma} \le \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le \frac{a\sqrt{n}}{\sigma}\right).$$

Recall that $\bar{X}_n \sim N(0, \sigma^2/n)$ and $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$. Therefore,

$$\mathbf{P}\left(I_n \ni \mu\right) = \mathbf{P}\left(-\frac{a\sqrt{n}}{\sigma} \le Z \le \frac{a\sqrt{n}}{\sigma}\right),$$

where $Z \sim N(0,1)$. Fix $0 < \alpha < 1$ and denote by z_α the number such that $\mathbf{P}(Z > z_\alpha) = \alpha$ (in other words, z_α is the $(1-\alpha)$-quantile of the standard normal distribution). For example, from normal tables (use R!) we find that $z_{0.05} \approx 1.65$ and $z_{0.005} \approx 2.58$, etc.

If we set $a = z_{\alpha/2}\sigma/\sqrt{n}$, then we get

$$\mathbf{P}\left(\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right] \ni \mu\right) = 1 - \alpha.$$

This is a $(1-\alpha)$-confidence interval for μ.

2.2. **Estimating σ^2 when μ is known.** Since μ is known, we use $W_n = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ to estimate σ^2.

Let $\chi^2_n(\alpha)$ denote the $1-\alpha$ quantile of this distribution for $0 < \alpha < 1$. Similarly, $\chi^2_n(1-\alpha)$ is the α quantile (i.e., the probability for the chi-squared random variable to fall below $\chi^2_n(1-\alpha)$ is exactly α). When X_i for $1 \le i \le n$ are i.i.d. $N(\mu, \sigma^2)$, we know that $(X_i - \mu)/\sigma$ for $1 \le i \le n$ are i.i.d. $N(0,1)$. By the above fact, we see that

$$\frac{nW_n}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n.$$

Hence, we obtain

$$\mathbf{P}\left\{\frac{nW_n}{\chi^2_n\left(\frac{\alpha}{2}\right)} \le \sigma^2 \le \frac{nW_n}{\chi^2_n\left(1-\frac{\alpha}{2}\right)}\right\} = \mathbf{P}\left\{\chi^2_n\left(1-\frac{\alpha}{2}\right) \le \frac{nW_n}{\sigma^2} \le \chi^2_n\left(\frac{\alpha}{2}\right)\right\} = 1 - \alpha.$$

Thus, $\left[\frac{nW_n}{\chi^2_n\left(\frac{\alpha}{2}\right)}, \frac{nW_n}{\chi^2_n\left(1-\frac{\alpha}{2}\right)}\right]$ is a $(1-\alpha)$-confidence interval for σ^2.

An important result: Before going to the next two confidence interval problems, let us try to understand the two examples already covered. In both cases, we came up with a random variable ($\sqrt{n}(\bar{X}_n-\mu)/\sigma$ and W_n/σ^2, respectively) which involved data and the unknown parameter whose distributions we knew ($N(0,1)$ and χ^2_n, respectively) and these distributions *do not depend on any parameters*. This is generally the key step in any confidence interval problem. For the next two problems, we cannot use the same two random variables as above as they depend on the other unknown parameter too (i.e., $\sqrt{n}(\bar{X}_n-\mu)/\sigma$ uses σ which will be unknown and W_n/σ^2 uses μ which will be unknown).

Theorem 3. *Let $Z_1,\ldots,Z_n$ be i.i.d. $N(\mu,\sigma^2)$ random variables. Let $\bar{Z}_n$ and s_n^2 be the sample mean and the sample variance, respectively. Then,*

$$\bar{Z}_n \sim N(\mu,\frac{\sigma^2}{n}) \text{ and } \frac{(n-1)s_n^2}{\sigma^2}\sim\chi^2_{n-1},$$

and the two are independent.

Note two important features. First, the surprising independence of the sample mean and the sample variance. Second, the sample variance (appropriately scaled) has χ^2 distribution, just like W_n in the previous example, but the degrees of freedom is reduced by 1. Now, we use this theorem in computing confidence intervals.

2.3. **Estimating σ^2 when μ is unknown.** The estimate s_n^2 must be used (as W_n depends on μ which is unknown). Theorem 3 tells us that $\frac{(n-1)s_n^2}{\sigma^2}\sim\chi^2_{n-1}$. Fix $0<\alpha<1$. Hence, by the same logic as before, we get

$$\begin{aligned}\mathbf{P}\left\{\frac{(n-1)s_n^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}\le\sigma^2\le\frac{(n-1)s_n^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}\right\}&=\mathbf{P}\left\{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)\le\frac{(n-1)s_n^2}{\sigma^2}\le\chi^2_{n-1}\left(\frac{\alpha}{2}\right)\right\}\\&=1-\alpha.\end{aligned}$$

Thus, $\left[\frac{(n-1)s_n^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)},\frac{(n-1)s_n^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}\right]$ is a $(1-\alpha)$-confidence interval for σ^2.

If μ is known, we could use the earlier confidence interval using W_n, or simply ignore the knowledge of μ and use the above confidence interval using s_n^2. What is the difference? The cost of ignoring the knowledge of μ is that the second confidence interval will be typically larger, although for large n the difference is slight. On the other hand, if our knowledge of μ was inaccurate, then the first confidence interval is invalid (we have no idea what its level of confidence is!) which is more serious. In realistic situations, it is unlikely that we will know one of the parameters, but not the other. Hence, most often one just uses the confidence interval based on s_n^2.

2.4. **Estimating μ when σ^2 is unknown.** The earlier confidence interval $[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}]$ cannot be used (as we do not know the value of σ).

A natural idea would be to use the estimate $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ in place of σ^2. However, recall that the earlier confidence interval (in particular, the cut-off values $z_{\alpha/2}$ in the CI) was an outcome of the fact that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1).$$

Is it true if σ is replaced by s_n? Actually no, but we have a different distribution called *Student's t-distribution*. From Theorem 3, we know that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$, $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$, and the two are independent. Taking these random variables into account, we can conclude that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n}$ has t_{n-1} distribution from (5) of Result 1.

Fix $0 < \alpha < 1$. In particular, by $t_n(\alpha)$ we mean the $1-\alpha$ quantile of the t-distribution with n degrees of freedom. Then of course, the α quantile is $-t_n(\alpha)$ as the t-distribution is symmetric about zero (the density at t and at $-t$ are the same). What we need to know is that there are tables from which we can read off specific quantiles of the t distribution (use R!). Returning to the problem of the confidence interval, from the fact stated above (use T_n to indicate a random variable having t-distribution with n degrees of freedom), we see that

$$\begin{aligned}
&\mathbf{P}\left(\bar{X}_n - \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n + \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right)\right) \\
&= \mathbf{P}\left(-t_{n-1}\left(\frac{\alpha}{2}\right) \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \leq t_{n-1}\left(\frac{\alpha}{2}\right)\right) \\
&= \mathbf{P}\left(-t_{n-1}\left(\frac{\alpha}{2}\right) \leq T_{n-1} \leq t_{n-1}\left(\frac{\alpha}{2}\right)\right) \\
&= 1 - \alpha.
\end{aligned}$$

Hence, our $(1-\alpha)$-confidence interval for μ is $\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right), \bar{X}_n + \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right)\right]$.

3. APPROXIMATE CONFIDENCE INTERVAL FOR THE MEAN

Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables from some distribution with mean μ and variance σ^2, both unknown. How can we construct a confidence interval for μ?

Fix $0 < \alpha < 1$. In case of normal distribution, recall that the $(1-\alpha)$-CI that we gave was

$$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right), \bar{X}_n + \frac{s_n}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right)\right].$$

Is this a valid confidence interval in general? The answer is "No". If X_i for $1 \leq i \leq n$ are from some general distribution, then the distributions of $\sqrt{n}(\bar{X}_n - \mu)/s_n$ and $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ are very complicated to find. Even if X_is come from binomial or exponential, these distributions will depend on the parameters in a complex way (in particular, the distributions are *not free from the parameters*, which is important in constructing confidence intervals).

But, suppose that n is large. Then, the sample variance is close to population variance and hence $s_n^2 \xrightarrow{P} \sigma^2$ (Proof?). Using continuous mapping theorem, we can show that $1/s_n \xrightarrow{P} 1/\sigma$. Further, by CLT we know that $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has approximately $N(0,1)$ distribution. Using Slutsky's lemma, we finally see that

$$\sqrt{n}(\bar{X}_n - \mu)/s_n \xrightarrow{D} Z \sim N(0,1) \text{ as } n \to \infty.$$

Therefore, when n is large, we may as well use

$$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{s_n}{\sqrt{n}} z_{\alpha/2}\right].$$

In other words, we have

$$\mathbf{P}\left\{-z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \leq z_{\alpha/2}\right\} \approx \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha.$$

Consequently, we may say that

$$\mathbf{P}\left\{\bar{X}_n - \frac{s_n}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{X}_n + \frac{s_n}{\sqrt{n}} z_{\alpha/2}\right\} \approx 1 - \alpha.$$

Thus, $\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{s_n}{\sqrt{n}} z_{\alpha/2}\right]$ is an approximate $(1-\alpha)$-confidence interval for μ. Further, when n is large, the difference between $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ and $V_n := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is small (indeed, $s_n^2 = (n/(n-1))V_n$ and we can prove that $s_n^2 - V_n \xrightarrow{P} 0$ as $n \to \infty$). Hence, it is also okay to use $\left[\bar{X}_n - \frac{\sqrt{V_n}}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{\sqrt{V_n}}{\sqrt{n}} z_{\alpha/2}\right]$ as an approximate $(1-\alpha)$-confidence interval for μ when n is large.

Example 4. Let $X_1, \ldots, X_n$ be i.i.d. Ber(p). Consider the problem of finding a confidence interval for p. Since each X_i is 0 or 1, observe that

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 = \bar{X}_n - (\bar{X}_n)^2 = \bar{X}_n(1-\bar{X}_n).$$

Hence, an approximate $(1-\alpha)$-CI for p is given by

$$\left[\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right].$$

Remark 5. For the normal distribution, the t_{n-1} density in fact approaches the standard normal density as $n \to \infty$ (How? Are you convinced?). Hence, $t_{n-1}(\alpha)$ approaches z_α for any $0 < \alpha < 1$ (this can also be seen by looking at the t-table for large degrees of freedom) and we can use

$$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{s_n}{\sqrt{n}} z_{\alpha/2}\right]$$

as an approximate $(1-\alpha)$-CI for μ. Strictly speaking, the level of confidence is smaller than for the one with $t_{n-1}(\alpha/2)$ (discussed in the earlier module). However, for n large the level of confidence is quite close to $1-\alpha$.

4. ACTUAL CONFIDENCE BY SIMULATION

Fix $0 < \alpha < 1$. Suppose we have a candidate confidence interval whose confidence we do not know. For example, let us take the confidence interval

$$\left[\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]$$

for the parameter p of i.i.d. Ber(p) samples. We saw that for large n this has approximately $(1-\alpha)$ confidence. But, how large is large? One way to check this is by simulation. We explain how.

Take $p = 0.3$ and $n = 10$. Simulate $n = 10$ independent Ber(p) random variables and compute the confidence interval given above. Check whether it contains the true value of p (i.e., 0.3) or not. Repeat this exercise 10000 times and see what proportion of times it contains 0.3. That proportion is the true confidence, as opposed to $1-\alpha$ (which is valid only for large n). Repeat this experiment with $n = 20$, $n = 30$, etc. See how close the actual confidence is to $1-\alpha$. Repeat this experiment with different value of p. The n you need to get close to $1-\alpha$ will depend on p (in particular, on how close p is to $1/2$).

This was about checking the validity of a confidence interval that was specified. In a real situation, it may be that we can only get $n = 20$ samples. Then, what can we do? If we have an idea of the approximate value of p, we can first simulate Ber(p) random numbers. We compute the sample mean each time, and repeat 10000 times to get several values of the sample mean. Note that the histogram of these 10000 values tells us (approximately) the actual distribution of $\bar{X}_n$. Then, we can find t (numerically) such that $[\bar{X}_n - t, \bar{X}_n + t]$ contains the true value of p in $(1-\alpha)$-proportion of the 10000 trials. Then, $[\bar{X}_n - t, \bar{X}_n + t]$ is a $(1-\alpha)$-CI for p.

Alternatively, we may also try a CI of the form

$$\left[\bar{X}_n - t\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + t\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right],$$

where we choose t numerically to get $(1-\alpha)$ confidence.

R resources:

i. https://shiny.rit.albany.edu/stat/confidence/

ii. http://www.r-tutor.com/elementary-statistics/interval-estimation.

Summary: The gist of this discussion is as follows. In the neatly worked out examples of the previous modules, we got explicit confidence intervals. But, we assumed that we knew the data came from the $N(\mu, \sigma^2)$ distribution. What if that is not quite right? What if it is not any of the nicely studied distributions? The results also become invalid in such cases.

For large n, we can overcome this issue by using the weak law of large numbers (WLLN), central limit theorem (CLT) and Slutsky's lemma.

But, for small n? The point is that using simulations we can calculate probabilities, distributions, etc., numerically and approximately. That is often better, since it is more robust to assumptions.