Sampling Distributions

(Corresponds to sections 8.1 and 8.2 in the text.)

We model the sampling process as follows: $X_1, X_2, X_3, \ldots, X_n$ are n independent copies (or instances) of the random variable X. This represents taking a sample of size n from the population represented by the distribution of X.

Example 1: Suppose X represents the weight of the coffee in a can of coffee (marked "8 oz") which is produced in my factory. If the cans are mechanically filled, the actual weights of the coffee in the cans are likely to have a normal distribution (or very close to it) with some mean μ and standard deviation σ. Here the population consists of all of the"8-oz" cans of coffee produced in my factory - we generally assume this population to be infinite. (See Note on the finite population case, below.) I would be interested in knowing the mean weight of the coffee in the cans I produce, and also the standard deviation (to know how consistently the cans are being filled). Taking a sample of 25 cans and weighing them means that I have found 25 values of the random variable $X : X_1, X_{,2}, \ldots, X_{25}$.

Example 2: Suppose that X represents the number of heads which show when I flip a coin. (In other words, $X = 0$ if tails shows and $X = 1$ if heads shows.) This is a random variable with a binomial distribution $b(x; 1, \theta)$ where θ represents the probability that heads shows when I flip the coin. This distribution has mean θ and standard deviation $\sigma = \sqrt{\theta(1-\theta)}$. Taking a sample in this case would mean that I repeat the experiment (of flipping the coin) a number of times, say 100 times. This will give me 100 values for $X : X_1, X_2, \ldots, X_{100}$, each of them being either a 0 or a 1. The population in this case represents all possible flips of the coin. The population is infinite.

Example 3: Suppose that X represents the number of heads which show when I flip a coin 10 times. This is a random variable with a binomial distribution $b(x; n, \theta)$ with mean 10θ and standard deviation $\sigma = \sqrt{10\theta(1-\theta)}$. Taking a sample in this case would mean that I repeat the experiment (of flipping the coin 10 times) a number of times, say 36 times. This will give me 36 values for $X : X_1, X_2, \ldots, X_{36}$ which are all numbers between 0 and 10. The population in this case represents all possible repetitions of the experiment (flip the coin ten times). The population is infinite.

Sampling from an infinite population means that we assume that we can sample as many times as we like from the distribution X: there is no upper bound on the size of the sample. That applies in the situations described in the examples above. What about sampling from a finite population?

Example: In real life, when taking samples, we frequently encounter the following types of situations:

- Take a sample from a shipment of circuit boards, and see how many boards in the sample meet specifications. We want to use that information to estimate how many boards in the whole shipment will meet specifications. The population here consists of all of the circuit boards in the shipment.
- Take a sample of registered voters in a certain area, and count how many of them say they intend to vote for candidate A. We want to use that information to estimate the percentage of voters who intend to vote for that candidate. The population here consists of all the registered voters in that area.
- Take a sample of adult males in the USA, and measure their heights. We want to use that information to estimate the average height of adult males in the USA. The population here consists of all the adult males in the USA.

What these three examples have in common is that the population from which we are sampling is finite in each case. The random variable X will give values which result in a hypergeometric distribution in the first two cases,; in the third case, the distribution depends on how the heights of all the adult males in the USA are distributed.

The following theorems apply to the distribution of sample mean for random samples taken from a random variable X, sampling from an infinite population with mean μ and standard deviation σ.

Theorem 1. *If $X_1, X_2, \ldots, X_n$ is a random sample of size n, then the distribution of the sample mean $\bar{X}$ has mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.*

Note: The standard deviation of the sample means, $\sigma_{\bar{X}}$, is also called the **standard error of the mean**.

Theorem 2. *Central Limit Theorem*

If $X_1, X_2, \ldots, X_n$ is a random sample of size n, then as $n \to \infty$, the distribution of the sample mean $\bar{X}$ approaches a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Another way to say the same thing: The statistic

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

has a distribution which approaches the standard normal distribution as $n \to \infty$.

Theorem 3. *Samples taken from a normally distributed population If X has a normal distribution and $X_1, X_2, \ldots, X_n$ is a random sample of size n, then the sample mean $\bar{X}$ has a normal distribution with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.*

Another way to say the same thing: If X has a normal distribution, then the statistic

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

has a standard normal distribution.

Return to Example 1 to illustrate what these say:
Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and let us assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 1 tells us that for samples of size 25 taken from this population, the sample means will have mean $\mu_{\bar{X}} = 8.0$ and standard deviation $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{25}} = 0.06$. We can think of this as saying that the average sample mean is the same as the average weight of the coffee in the cans, and its standard deviation is small (only one-fifth of the standard deviation of the weights). Also, if we would increase the sample size to 100, the standard deviation of the sample means (standard error of the mean) would decrease to $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{100}} = 0.03$ So by taking larger samples we can make the standard error of the mean as small as desired.

- Theorem 2 tells us that the sample means also have a distribution which is close to a normal distribution for large enough n. In practice, as a rule of thumb, we use this result when $n \geq 30$ (but this can be relaxed, especially if the distribution of X is not too far from normal).

- Theorem 3 tells us that since the distribution of the weights is normal, then the distribution of the sample means is also normal regardless of n - Theorem 2 is not needed in this case!

Some more illustrations of the significance of these theorems:

- What is probability that a randomly selected can of coffee will have weight 7.9 ounces or less if $\mu = 0.8$ and $\sigma = 0.3$? This is $P(z \le \frac{7.9-8.0}{0.3}) \approx P(z \le -0.33) \approx 0.3707$
- What is the probability that a random sample of 25 cans of coffee will have mean weight 7.9 ounces or less? Using CLT or Theorem. 3, this is $P(z \le \frac{7.9-8.0}{0.3/\sqrt{25}}) \approx P(z \le -1.67) \approx 0.0475$ Therefore a sample mean is much more likely to be close to the true value of the population mean than the weight of an individual can of coffee is.

Finite Population

Theorem 4. *If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a finite population of size N, then the distribution of the sample mean $\bar{X}$ has mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.*

Returning to our example of the cans of coffee, what would be different if we knew that we were taking the sample from a shipment (or limited production) of 500 cans?
Recall that X represents the actual weight of the coffee in an "8-ounce" can of coffee, and we assume that X has normal distribution. Suppose also that $\mu = \mu_X = 8.0$ and $\sigma = \sigma_X = 0.3$.

- Theorem 4 tells us that for samples of size 25 taken from this population, the sample means will have mean $\mu_{\bar{X}} = 8.0$ and standard deviation $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{25}}\sqrt{\frac{500-25}{500-1}} \approx 0.0585$. The mean is the same as it was in the infinite population case, but the standard error of the mean is slightly smaller. (Note that if we are using this in estimating the mean, the infinite population result is "conservative" in terms of estimating error.) Also, if we would increase the sample size to 100, the standard error of the mean would decrease to $\sigma_{\bar{X}} = \frac{0.3}{\sqrt{100}}\sqrt{\frac{500-100}{500-1}} \approx 0.027$. So be taking larger samples we can make the sample error of the mean as small as desired.
- Theorem 2 does not apply directly, since it assumes that the sampling is done from an infinite population. However, if the population size is very large compared to the sample size, we know that the hypergeometric distribution can be approximated by a binomial distribution (sampling with replacement), and so Theorem 2 holds approximately in this case. The usual rule of thumb is that the sample should not be more than 5% of the population. Under that condition, the sample means also have a distribution which is close to a normal distribution for large enough n. Yes, n has to be large but not too large! Recall that in practice, as a rule of thumb, we use $n \ge 30$ (but this can be relaxed, especially if the distribution of X is not too far from normal).
- Theorem 3 does not apply, since the population is finite (and a finite population cannot have a normal distribution - it has to have a discrete, in fact finite, distribution).

Note on the finite population case: As mentioned in the last bullet point, a finite population must have a discrete, in fact a finite, distribution. However, for populations which are finite but very large, it is often justifiable to treat them as if they were for all practical purposes infinite and their distributions were continuous. Even in our example where the population was 500, we saw that the finite population correction was not very much. If the population had been even larger, there would be no detectable difference between the finite population result and the infinite population result. And for a large enough population, the histogram representing its distribution can be very "smooth", approximating a smooth curve which we can test to see if it is likely to be close to normal. So in practice, it is more common to use the infinite population theorems even when sampling from a finite (but large) population.

Interval Estimate for a Population Mean μ

Corresponds to section 11.2 in the text.

Theorem 5. *If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normally distributed population, then*

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

is a $(1-\alpha)(100\%)$ confidence interval for μ.

The number $(1-\alpha)(100\%)$ is called the **level of confidence**. Common values of α in practice are 0.05, 0.02, and 0.01, corresponding to the 95%, 98%, and 99% levels of confidence respectively. We want the level of confidence to be as large as possible, which means that we want α to be very small, but there is a price to pay for making α small.

The numbers $\bar{x} - z_{\alpha 2}\frac{\sigma}{\sqrt{n}}$ and $\bar{x} + z_{\alpha 2}\frac{\sigma}{\sqrt{n}}$ are sometimes called the **confidence bounds** for μ.

The numbers $z_{\alpha/2}$ are called **critical values** of z. $z_{\alpha/2}$ is the z which cuts off a right-hand tail of area $\frac{\alpha}{2}$

The quantity $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ represents an estimate of the maximum amount of error that is expected when we use $\bar{X}$ as an estimate for μ.

Another way to write the confidence interval is

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Returning to Example 1, suppose that we take a sample of 25 cans of coffee at random and compute the mean weight of coffee in the 25 cans to be $\bar{x} = 7.8$. Somehow we know that $\sigma = 0.3$.

- Take $\alpha = 0.05$ so the level of confidence is 95%. This value is very common in the social sciences. Then we can look up $z_{\alpha/2} = z_{0.025} \approx 1.96$.

  The error estimate is then $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \approx 1.96\frac{0.3}{\sqrt{25}} = 0.1176$ which should be rounded to 0.12.

  The confidence bounds are $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 7.8 - 0.12 = 7.68$ and $\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 7.8 + 0.12 = 7.92$

  These results are usually stated as follows:

  With 95% confidence, the true mean weight of the coffee in the cans is not more than 0.12 ounces away from the sample mean 7.8 ounces.

  Or we can say:

  With 95% confidence, the true mean weight of the coffee in the cans is between 7.68 ounces and 7.92 ounces.

- If we try to increase the level of confidence to 99% (by reducing α to 0.01) what happens?

  Then the number $z_{\alpha/2} = z_{0.005} \approx 2.58$. This changes the error estimate to $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \approx 2.58\frac{0.3}{\sqrt{25}} = 0.1548$ which should be rounded to 0.15.

  The bounds of the confidence interval will be somewhat larger. In order to increase the level of confidence, we have to allow a larger interval for finding μ - in other words, we have to allow for a larger possible error.

The formula given in this theorem can be used in the following situations:

- When the random sample is taken from a population which is normally distributed, and σ is known.
- When n is "large" (usually we take $n \geq 30$) and σ is known, to give an approximate confidence interval via the Central Limit Theorem.

Finding sample size for a specified error bound

Another way we can make use of this formula is to answer the question: how large a sample should we take in order to make the margin of error small?

Returning to our example, trying to estimate the mean weight of coffee contained in the "8 ounce" cans, and where we assume that we know the population standard deviation is 0.3 ounces: suppose that we want the error term to be no more than 0.04 - in other words, we want to be able to estimate the mean weight of the coffee in the cans to within 0.04 ounce, with (let's say) 95% confidence.
The error term in the confidence interval is then $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \approx 1.96\frac{0.3}{\sqrt{n}}$ and we want this to be less than or equal to 0.04:
$1.96\frac{0.3}{\sqrt{n}} \leq 0.04$
$\frac{(1.96)(0.3)}{\sqrt{n}} \leq 0.04$
$(1.96)(0.3) \leq (0.04)\sqrt{n}$
$\frac{(1.96)(0.3)}{0.04} \leq \sqrt{n}$
Square both sides to get $n \geq (\frac{(1.96)(0.3)}{0.04})^2 = 216.09$. We should round this last number up (since n must be a whole number!) to 217. So by taking a sample of at least 217 cans at random, we can ensure that the error in estimating the mean weight of coffee in the "8-ounce" cans of coffee will be no more than 0.04 ounce, with 95% confidence.

From this example we can derive a formula for the sample size needed to have a specified lower bound on the error term: if we want the error term to be no more than ϵ with $(1-\alpha)(100\%)$ confidence then the sample size must be
$n \geq (\frac{z_{\alpha/2}\sigma}{\epsilon})^2$ and always round **up**.

What if we do not know the population standard deviation?

Obviously it will be unusual in the real world to know σ. So what do we do when σ is unknown?

If the sample is taken from a population which is normally distributed, we can use the fact that

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has the so-called Student's T-distribution with $n-1$ degrees of freedom.

What this means is that the formula given in Theorem 4 must be replaced with

$$\left(\bar{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

which gives a $(1-\alpha)(100\%)$ confidence interval for μ in case X has a normal distribution.

In practice it has been shown that this formula can be used even if the distribution of X is not exactly normal, as long as it is not too far away from normal.

For example, if we took a sample of 25 cans of coffee and found a sample mean weight $\bar{x} = 7.8$ ounces and sample standard deviation $S = 0.45$ ounces, taking again $\alpha = 0.05$, then $n-1 = 24$ so $t_{\alpha/2,24} = t_{0.025,24} \approx 2.064$ from Table IV. Therefore a 95% confidence interval for μ would be
$7.8 \pm 2.064(\frac{0.45}{\sqrt{25}}) = 7.8 \pm 1.8576 \approx 7.8 \pm 1.9$

Note that when n is 30 or more, the critical values $t_{\alpha/2.n-1}$ are virtually the same as the corresponding critical values $z_{\alpha/2}$ so it makes no practical difference which you use. It is common to use $z_{\alpha/2}$ when $n \geq 30$.