Sampling Distributions

The Distribution of a Sample Mean: Part 2

We already know how the average of independent random measurements behaves when the individual measurements have normal distribution. In this document we investigate how averages behave, when the distribution that describes the behaviour of the individual measurements is not normal.

Here are two probability density functions:

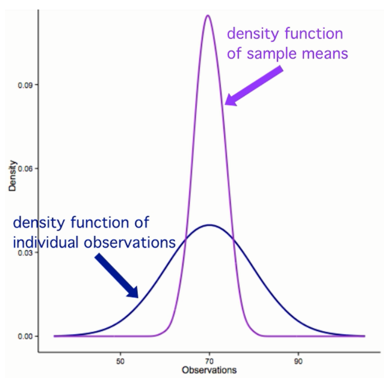

Figure 1: Distributions of individual and sample means of 9 measurements.

The blue density describes the behaviour of individual measurements. In purple we have the distribution of the averages we might get if we take a sample of 9 measurements from the blue density.

We saw that:

1. The probability distribution of the sample means is also a normal distribution.

2. The normal probability distributions of the individual observations and of the sample means have the same mean (the center of the distribution).

3. The standard deviation of the probability distribution of the sample means is smaller than the standard deviation of the distribution of the individual observations by a factor of 1 over the square root of the sample size.

Note that points **2** and **3** hold in all situations since they come from properties of expectation and variance that hold for any probability distribution (not just for Normal distributions). So we just have to consider a more general version of point **1**.

We start with a right-skewed distribution for our probability distribution that describes individual observations:

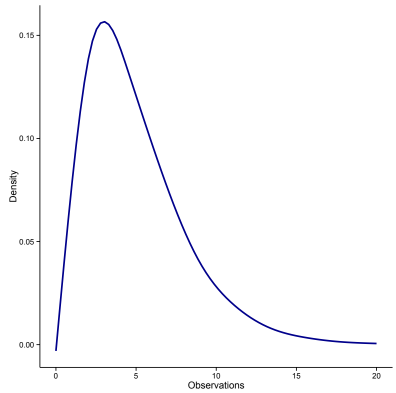

Figure 2: Density function of individual observations

The expected value for a measurement from this probability distribution is 5 ($E(X) = 5$). We have randomly generated 2000 individual observations from this probability distribution, and plotted them in following density histogram:

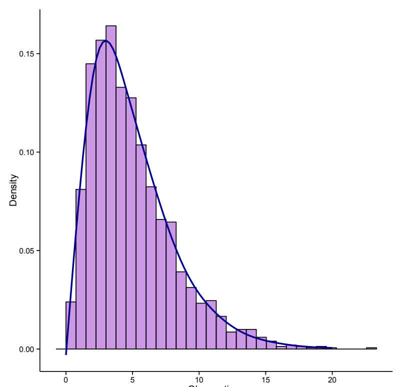

Figure 3: Density histogram of 2000 individual observations

Next we take several individual measurements, and find the mean. Here are two random samples of size 3:

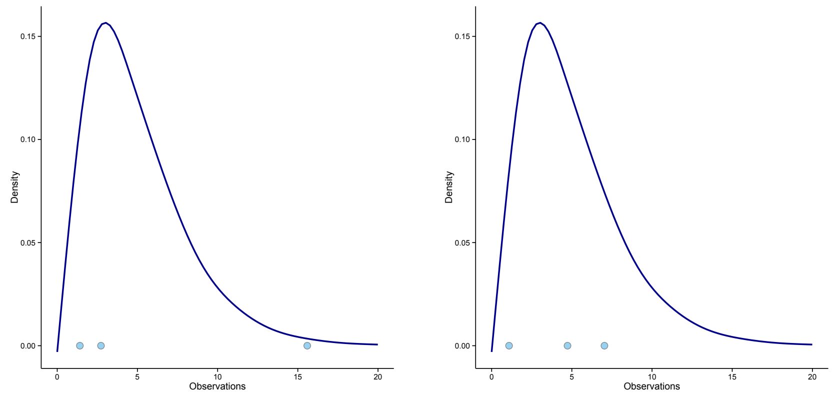

Figure 4: Two random samples of size 3

The first sample has an average or sample mean of 6.58, the second one has a sample mean of 4.28.
If we do this 2000 times and plot the 2000 resulting sample means in a histogram we get this:

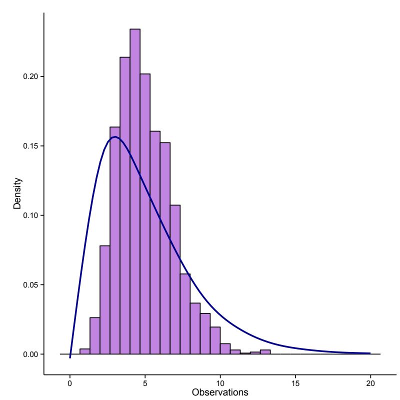

Figure 5: Histogram of 2000 sample means of size 3 with individual density function

Some things to note:

- The sample means are most likely to be close to the mean of the individual observations (5).
- The sample means have a smaller standard deviation than the individual observations.
- The histogram of sample means is less skewed than the density function of individual observations.

Lets repeat this process for another 2000 sample means, but this time, each of our samples will consist of 10 measurements. That is what we get:

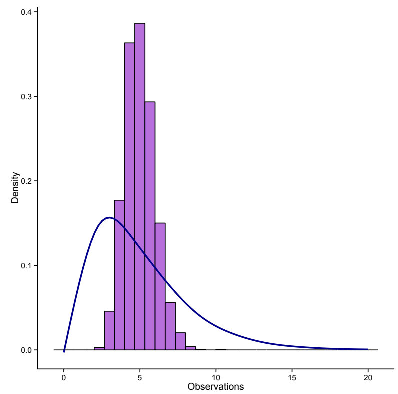

Figure 6: Histogram of 2000 sample means of size 10 with individual density function

The histogram of sample means is still centered at 5 (expectation), with even smaller spread, and except for a few larger values in the right tail, is almost symmetric.
Next we increase our sample size to 25. Here is the histogram of the means:

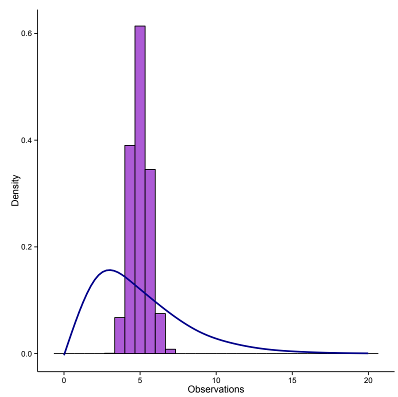

Figure 7: Histogram of 2000 sample means of size 25 with individual density function

It is still centred at 5, has even smaller spread with the larger sample size, and now has a symmetric bell-shaped distribution.
This is a central result of statistical theory. The result tells us that, **regardless** of the probability distribution of individual observations, averages of the observations have approximately a normal distribution, and the larger the sample size, the closer the means are to a normal distribution.

Central Limit Theorem: As more and more measurements are taken, the probability distribution for the possible averages of the measurements will converge to a Normal distribution.

We can now adjust our description of the distribution of sample means as follows:

1. For large enough sample size, the probability distribution of sample means is approximately a Normal distribution, regardless of the probability distribution of the individual measurements (by the Central Limit Theorem).

2. The normal probability distributions of the individual observations and of the sample means have the same expectation (the center of the distribution).

3. The standard deviation of the probability distribution of the sample means is smaller than the standard deviation of the probability distribution of the individual observations by a factor of 1 over the square root of the sample size.

We see that for point **1** we do not need individual observations to follow a normal distribution, as long as the sample size is large enough, the averages will have approximately normal distribution.