Dowel type connections in laminated bamboo with multiple slotted-in steel plates

Annex F – Analysis of result distribution

Auteur, studentnr:	JJB Debije, 4309995
Datum:	06-02-2017
Bedrijf:	Technische Universiteit Delft
Adres:	Stevinweg 1
	2628 CN Delft
Afdeling:	Faculteit Civiele Techniek en Geowetenschappen
Opleiding:	Civiele Techniek
Hoofdprofiel:	Structural Engineering
Uitstroomprofiel:	Staal en houtconstructie
Begeleiders:	prof.dr.ir. J.W.G. Van de Kuilen
	ir. P.A. de Vries
	drs. W.F. Gard
	drs. W.N.J. Ursem
	ing. A. van der Vegte

Table of Contents

1 Introduction

During this thesis research laminated bamboo connections were tested to measure the capacity of doweled connections and evaluate the probability to predict this capacity using design formulas already available for timber. Although laminated bamboo is an engineered product, the base material (bamboo) remains a natural product and a variance in test results is to be expected. In order to get an accurate measurement of the actual capacity of a connection in laminated bamboo and also determine the standard deviation of these test pieces, a large amount of test pieces need to be fabricated and tested.

To test and fabricate a large number of test pieces of course takes a lot of time and material. For this, a solution was sought and found in testing more than one connection per test piece. Usually a test piece is, on one end, equipped with a connection that will be loaded until failure and on the other with a connection that will be made stronger. This way, one knows on beforehand which connection will fail and one can place all measuring equipment so that only the weaker connection is examined. This is of course the easiest way to go about testing the capacity and deformation of a connection. However, this is also costly since every test piece can be used only once. To increase the efficiency of all tests and to make better use of the available material, all test pieces used in this research were designed and equipped with two identical connections. This decision lead to a doubling in the amount of connections that could be tested using the same amount of testing material. That same decision however also had two downsides. The first one being that, on beforehand, it was not known which of the two connections would break and so, per test piece, two connections had to be measured. The second, and maybe more problematic one, being that when one of the two connections failed, the test ended and only the capacity of the first failed (and thus weaker) connection was known. When taking the average capacity of all these weaker connections without taking into account that half of all connections are, at least, stronger, one could be misled into thinking the tested material is weaker than in reality. A method should thus be conceived by which the stronger half of the connections is also taken into account when determining the capacity of the connections from the obtained test results.

1.1 Goal

For the determination of the average capacity of all connections, both the stronger half and the weaker half of all connections need to be considered. Since the actual capacity of the stronger half cannot be measured and the only knowledge about this capacity is that it will, at least, be higher than the capacity of the weaker half (i.e. the failed connections), it is necessary to perform a study into statistics and probabilistic design. Through this study a method is sought by which the average capacity can be determined based solely upon the weaker half of all tested connections.

1.2 Plan of action

To be able to evaluate the average capacity of all connections by using only the test results that were obtained from the weaker half of all connections, knowledge about statistical analysis and probabilistic design is necessary. This knowledge is to be obtained by performing a literature study. For this literature study use will be made of the available courses and corresponding study material at TU Delft. By researching statistics, a way will be sought by which the average capacity of all connections can be determined.

1.3 Reading guide

First 2 - **Train of thought** will explain the exact problem faced when analysing the test results more clearly and will give an idea on how to overcome this problem. Using this idea, chapter 3 - **Determination of the actual average capacity** will start off by making a few assumptions after which a formula is devised by which the problem could be solved. Since the formula is based on assumptions, the calculated values will not be entirely correct. Chapter 4 - **Estimation error** will explain why. Knowing the incorrectness of the formula, chapter 5 - **Simulation using Excel** will perform a simulation of the tests done in this research and give a final remark on the way the test results should be interpreted and analysed.

2 Train of thought

By testing all test pieces to failure, exactly half of all connections will break and data will be obtained from which the capacity of the weaker half of the connections can be determined.
For the determination of that capacity all test results are assumed normally distributed. This normal distribution will then have an expectancy μ_z ('z' as is 'zwak') that will be taken as the average of the measured maximum forces during testing and a standard deviation that will also be calculated from the same test results. In doing this a distribution is obtained that is valid for the weaker half of all connections. It has to be noted that this approach is not entirely realistic. The reason for this will be discussed in 4 - **Estimation error**.
Knowing that the found average capacity is based upon exactly half of all connections and that all 'missing' data belongs to stronger connections it is possible, by making a few assumptions, to make an estimate of the average capacity of all connections (i.e. it is possible to shift the normal distribution of the weaker connections in such a way that a distribution for all connections can be obtained).
A visual display of the described normal distributions is given in Figure 1 - Normal distributions of all connections (Red: weak, Blue: total, Green: strong).

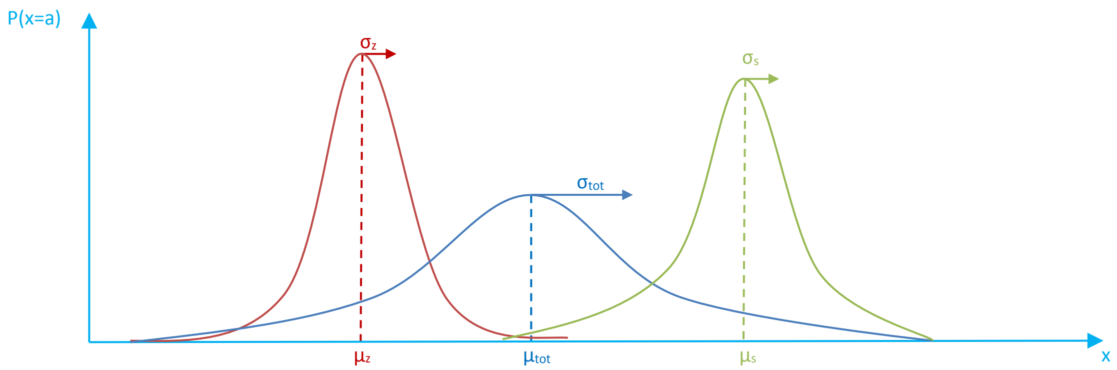


Figure 1 - Normal distributions of all connections (Red: weak, Blue: total, Green: strong)

In the figure above a sketch is given of the normal distributions that could result from testing. On the horizontal axis the capacity of a connection is shown and on the vertical axis the chance that a connection will have that capacity. The red graph represents the data received from the tests done in this research. The green graph shows the missing data that would have been received from the strong connections if they were loaded until failure. When combined, the two graphs would result in the blue graph that represents the normal distribution of all connections.
Usually to obtain this blue graph one would test all connections and just calculate the mean and standard deviation from the test results. Since in this research only the weak connections can be loaded until failure and thus only the red graph can be obtained, another way to obtain the blue graph has to be devised. Also, as can be seen in the figure, the standard deviation σ does not necessarily have to be the same for all three graphs.

3 Determination of the actual average capacity

In the previous chapter the difficulties in determining the average capacity for all connections have been explained. In this chapter a solution to this problem is given and a way to determine the average capacity of all connections from test data, based upon only the weaker half of all connections, is devised. It should again be noted that the assumptions made to come to this solution are not entirely correct (this is discussed in 4 - **Estimation error**). The solution given here is only exactly correct when the measurements from testing would be exactly normally distributed and the coefficient of variation would be a constant (which is not the case, as can be seen in the simulation done in 5 - **Simulation using Excel**).

When determining the normal distribution of the weaker half of all connections (the red line in Figure 1 - Normal distributions of all connections (Red: weak, Blue: total, Green: strong)), the expectancy μ_z is taken as the average capacity of all weaker connections. This means that exactly 50% of all weaker connections have a capacity lower than the value μ_z and exactly 50% will have a value that is higher than μ_z. Having this information and knowing that the weaker half of the connections makes up for exactly 50% of all fabricated connections one can conclude that, when looking at the average capacity of all connections (μ_{tot}), the value of μ_z will be such that exactly 25% of all connections will have a lower capacity than μ_z and 75% will have a higher capacity. The same can be said for the distribution and the average capacity μ_z of the strong connections (only this value will have a 25% chance of exceedance instead). A visual explanation of this is made in Figure 2 – Cumulative distribution function of all connections (Red: weak, Blue: total, Green: strong) where the distribution function of all connections is shown.

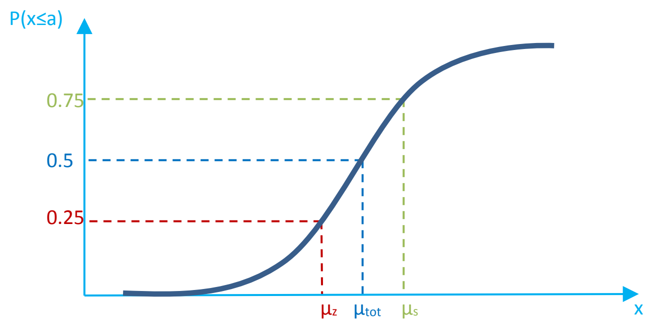


Figure 2 – Cumulative distribution function of all connections (Red: weak, Blue: total, Green: strong)

Given this information, now a way can be sought by which the known μ_z can be translated into a value for μ_{tot}. The following formula can be derived directly from the figure above and should indirectly give a relation between μ_z and μ_{tot}.

$$P_{\mu_{tot}}(x \leq \mu_z) = 0.25 \tag{1}$$

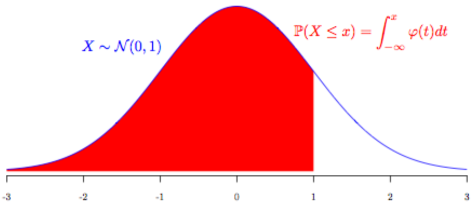
Now to further analyse this relation the normal distribution needs to be standardized. Usually this is done by subtracting the expectancy of the distribution (μ_{tot}) from the considered x-value (μ_z). After that one would divide by the standard deviation σ_{tot}. However, in this case σ_{tot} still remains unknown. In order to overcome this obstacle one has to consider that the desired average value μ_{tot} corresponds to the same material as the value μ_z. Although it is a different 'batch' with a different standard deviation, since the material is the same, it is assumed to have the same coefficient of variation. Because the coefficient of variation is a material parameter and not a 'batch' parameter, it should be the same for every connection (of the same type) made with the same material. The coefficient of variation gives a relation between the expected value μ and the standard deviation σ in the following manner:

$$COV = \frac{\sigma}{\mu} \tag{2}$$

By making use of the relation above the standard deviation σ_{tot} can be expressed as a function of the COV and the average capacity of all connections μ_{tot}. Entering this relation in the first formula, the normal distribution can be standardized.

$$P_{\mu_{tot}}\left(Z \leq \frac{\mu_z - \mu_{tot}}{COV * \mu_{tot}}\right) = 0.25 \tag{3}$$

Using tables for normal distributions a value for ‘Z’ can be sought that corresponds to the exceedance chance of 75% (i.e. the undershoot chance of 25%). This is done in Figure 3 - **75% exceedance value for 'Z'**.



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517

$$\left.\begin{aligned} P_{\mu_{tot}}(Z \leq 0.67) &= 0.7486 \\ P_{\mu_{tot}}(Z \leq 0.68) &= 0.7517 \end{aligned}\right\} \; 0.67 + \frac{0.68 - 0.67}{0.7517 - 0.7486}(0.75 - 0.7486) = 0.67452$$

Figure 3 - 75% exceedance value for 'Z'

The value for ‘Z’ that corresponds to the 75% exceedance chance is now determined (note that in the figure actually the 25% exceedance is determined so a minus sign needs to be incorporated into the calculations). Substituting the found value in the formula yields a relation in which μ_{tot} is the only unknown value remaining. The formula can now be solved for μ_{tot}.

$$-0.67452 \leq \frac{\mu_z - \mu_{tot}}{COV * \mu_{tot}}$$

$$-0.67452 * COV * \mu_{tot} \leq \mu_z - \mu_{tot}$$

$$1 - 0.67452 * COV * \mu_{tot} \leq \mu_z$$

$$\mu_{tot} \leq \frac{\mu_z}{1 - 0.67452 * COV} \tag{4}$$

Through the use of the obtained formula, the found capacity of the weaker connections can now be used to give an estimate for the capacity of all connections. A simulation using Excel showed that this formula gives a slight overestimation of μ_{tot}. The overestimation is however small for COV values below 0.1.

4 Estimation error

In the previous chapter a formula is derived by which the actual capacity for all connections could be estimated. This chapter will describe the correctness of this estimation.

In order to check the validity of the derived formula Mr. A. Hensbergen of the EWI faculty at TUDelft was consulted. Mr. Hensbergen teaches the course 'WI2031TH - Kansrekening en statistiek'. In this consult he showed that the estimation of the average capacity through the derived formula does not yield the exact value. The reason for this is the assumption that the measured capacities of the test pieces are normally distributed is not entirely correct. When looking at all connections, the capacities are normally distributed with expectation μ_{tot} and variation σ_{tot}. By equipping every test piece with two of these connections and only measuring the weaker of these two connections the test results are not normally distributed anymore. In other words: if connection 1 would be $N(\mu_{tot},\sigma_{tot})$ distributed and connection 2 would be $N(\mu_{tot},\sigma_{tot})$ distributed then $MIN(1,2) \neq N(\mu_z,\sigma_z)$.
A more realistic representation of the distribution of the measured test results is given in Figure 4 - Realistic distribution of test results.

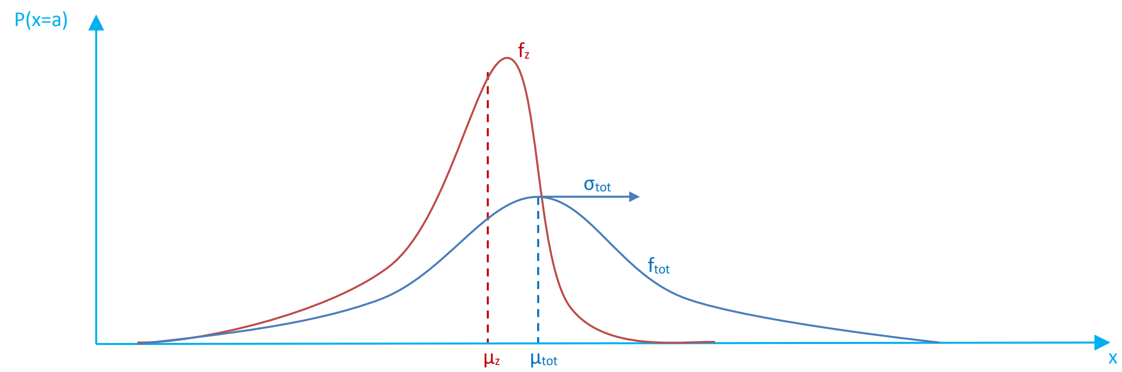


Figure 4 - Realistic distribution of test results

The above figure displays a distribution in which the weaker connections are no longer normally distributed. Since the test results are always based upon the minimum of two normally distributed connections, the distribution of these minima will show a long tail for values below μ_{tot} and a sharp drop for values above μ_{tot}.
What we can say about the test results is the amount of measurements that fall below (or above) the value for μ_{tot}. This is done by stating that the chance of a test piece (consisting of two connections) having a capacity below μ_{tot} is equal to 1 minus the chance of the capacity being higher than μ_{tot}. The capacity of a test piece is higher than μ_{tot} only if both connections are stronger than μ_{tot}. The chance of this happening is 50% times 50%.

$$P_z = 1 - P_s = 1 - 0.5*0.5 = 0.75$$

This means that 75% of the found test results are below μ_{tot} and 25% of the found results are above μ_{tot}.
This is displayed in Figure 5 - Distribution of test results.

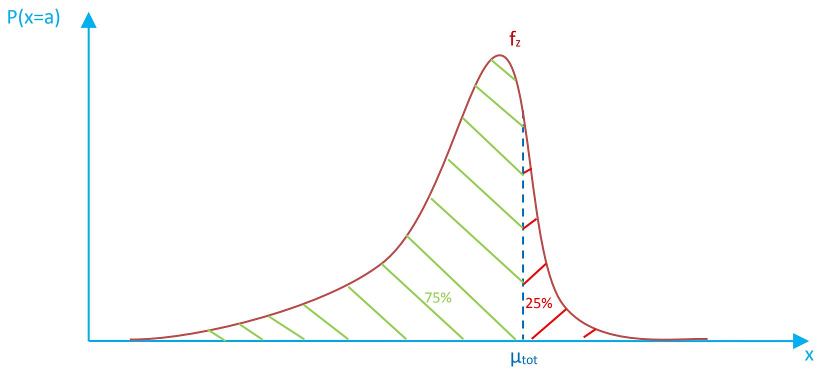


Figure 5 - Distribution of test results

Although it is known that 75% of all test results fall below the actual value of μ_{tot}, for the determination of μ_{tot} from the obtained test results the distribution formula f_z is still necessary. By assuming this formula to be a normal distribution in the previous chapter, a graph was assumed that is somewhat wider than the actual graph of f_z. This wider graph yields a larger coefficient of variation and will eventually give a slight overestimation of the value for $\mu_{tot.}$

5 Simulation using Excel

Since all measured capacities of the connections consist of the minimum value of two normally distributed capacities, the found values will no longer be normally distributed.
As already explained in the previous paragraph there is no exact way to determine the exact distribution of the found minimum values nor the missing maximum values. To get a better view of the available data and to find a way to solve the problem, a simulation was ran using Excel.
In this simulation the capacity of the upper and lower connection of a test piece was randomly generated from a normal distribution. This was done 1000 times. Next, the minimum value and the maximum value out of every connection were separated and sorted. In this way two compilations were obtained. One of which consisted of only the minimum values and one contained only the maximum values.
A picture of the simulation is given in Figure 6 - Simulation of two times 1000 random N(100,5) numbers.

Randomly generated normally distributed N(100,5) numbers					MU	100		Sigma	5
	Number 1	**Number 2**	**Min(1,2)**	**Max(1,2)**	**Ave(1,2)**		**Ave Min(1,2)**		**Ave Max(1,2)**
0.001	100.2788	101.0738	100.2788	101.0738	99.94554		97.24624		102.6449
0.002	105.7763	109.3817	105.7763	109.3817					
0.003	97.59007	98.74393	97.59007	98.74393	**StDev(1,2)**		**StDev Min(1,2)**		**StDev Max(1,2)**
0.004	99.86479	102.5919	99.86479	102.5919	4.859265		4.020494		4.06166
0.005	96.39899	97.65812	96.39899	97.65812					
0.006	95.84925	108.1131	95.84925	108.1131	**COV(1,2)**		**COV Min(1,2)**		**COV Max(1,2)**
0.007	105.299	96.85128	96.85128	105.299	0.048619		0.041343		0.03957
0.008	93.33639	101.0277	93.33639	101.0277					
0.009	103.3101	103.4239	103.3101	103.4239	**5-perc(1,2)**		**5-perc Min(1,2)**		**5-perc Max(1,2)**
0.010	103.182	104.8783	103.182	104.8783	91.92776		90.61242		95.94311
0.011	97.97703	109.868	97.97703	109.868	107.9633		103.8801		109.3466
0.012	100.316	100.1269	100.1269	100.316					
0.013	101.006	98.13903	98.13903	101.006	**Kurt(1,2)**		**Kurt Min(1,2)**		**Kurt Max(1,2)**
0.014	100.2331	97.17794	97.17794	100.2331	0.037356		-0.00932		0.089236
0.015	100.1888	99.00291	99.00291	100.1888					
0.016	97.26622	101.3125	97.26622	101.3125	**Skew(1,2)**		**Skew Min(1,2)**		**Skew Max(1,2)**
0.017	99.31782	100.4408	99.31782	100.4408	0.0384		-0.11858		0.206428
0.997	97.46165	90.72377	90.72377	97.46165					
0.998	90.46008	97.29551	90.46008	97.29551					
0.999	97.83497	89.19316	89.19316	97.83497					
1.000	102.4579	94.9712	94.9712	102.4579					

Figure 6 - Simulation of two times 1000 random N(100,5) numbers

In the above figure the second and third column ('Number 1' and 'Number 2') are used to randomly generate a number from a normal distribution with expectancy 100 and a standard deviation of 5.
From these two columns the minimum and maximum value is taken and sorted in the next two columns ('Min(1,2)' and 'Max(1,2)'). Also the average of all columns was determined and is shown on the first row next to the columns (abbreviated by 'Ave').

To get and understanding of the distribution of the found minimum and maximum columns and to determine how much these distributions actually differ from the original normal distributions two terms are used. These terms are the skewness and the kurtosis of the found distributions. With these terms similarity between a random distribution and a normal distribution can be expressed. The following definition for these terms is used:

- Skewness quantifies how symmetrical a distribution is around its mean value.
    - A symmetrical distribution has a skewness of zero.
    - An asymmetrical distribution with a long tail to the right (higher values) has a positive skew.
    - An asymmetrical distribution with a long tail to the left (lower values) has a negative skew.
    - The skewness is unitless.
    - The threshold to determine whether a distribution has an acceptable skewness to be classified as normal-like distribution varies per research and depends on the application and the, by the researcher, desired accuracy. According to George and Mallery (2010), Skewness and Kurtosis values between -2 to +2 indicate acceptable measures (Ruegg, 2015).
    - To check whether a distribution is skewed, the mean of the sample can be subtracted from each sample value. The result will be positive for values greater than the mean, negative when smaller than the mean and 0 if they equal the mean.
        - To compute a unitless measure of the skew each of these differences is divided by the standard deviation. What is obtained is called a 'z' ratio. The standard deviation of these ratios is 1. For each value z^3 is computed. After that, the sum of these values is divided by n-1, where n is the number of values in the sample. If the distribution is symmetrical, the positive and negative values will balance each other, and the average will be close to zero. If the distribution is not symmetrical, the average will be positive if the distribution is skewed to the right and negative if skewed to the left. The formula Excel uses to compute the skew looks like this: $Skewness = \frac{\Sigma(X-\bar{x})^3}{(N-1)\sigma^3}$
- Kurtosis quantifies whether the shape of a distribution matches a normal distribution.
    - A normal distribution has a kurtosis of 0 (mesokurtic).
    - A flatter distribution has a negative kurtosis (platykurtic). Such a distribution has highly dispersed values.
    - A more peaked distribution has a positive kurtosis (Leptokurtic). Such a distribution shows a sharp peak with relatively fat tails.
    - The kurtosis is unitless.
    - The formula used by Excel to calculate kurtosis is somewhat more elaborate and looks like this:

$$Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\Sigma(X-\bar{x})^4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

- Skewness and Kurtosis are graphically shown in Figure 7 - Skewness and kurtosis.

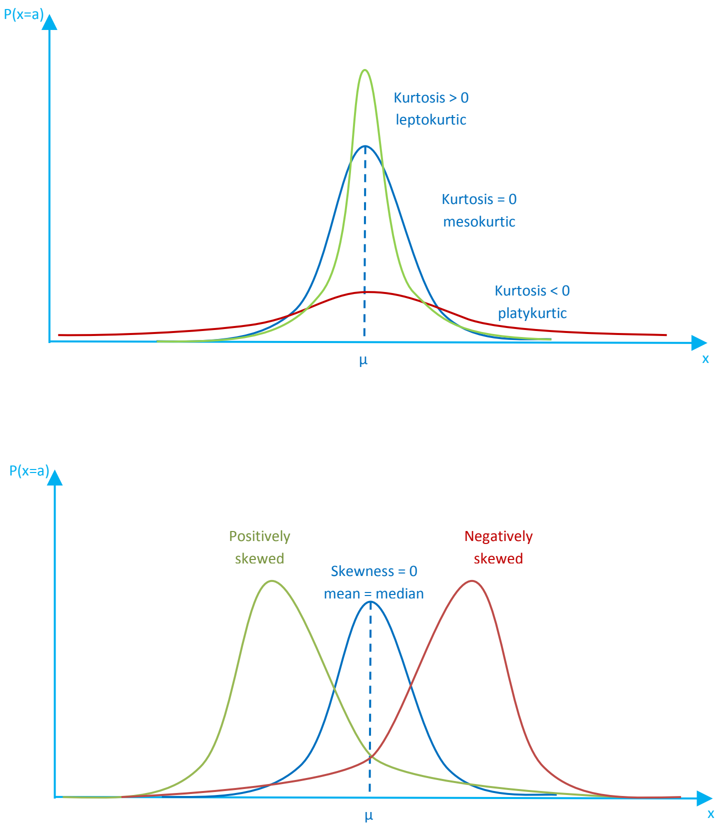


Figure 7 - Skewness and kurtosis

Given the above mentioned definition and limit values for the skewness and kurtosis the two columns with the minima and maxima can be further analysed. Using Excel, the skewness and kurtosis for both the maxima and minima columns and the original distribution of numbers 1 and 2 were computed. The obtained values are given in the bottom two rows of Figure 6 - Simulation of two times 1000 random N(100,5) numbers.
When looking at these numbers it can be seen that the minima and maxima distributions have a slight skew but it is well within the limits of -2 and +2 used by George and Mallery (2010). In terms of the simulated values for the kurtosis the difference between the Min. and Max. distributions and a normal distribution are negligible. In the simulation depicted here the value for kurtosis of the Min. distribution is even less than the kurtosis of the actual normally distributed numbers 1 and 2. The simulation thus shows that the obtained measurements from tests in this research (which are represented by the minima column) can be considered normally distributed.

Although it is now shown that the measurements are normally distributed, the average of the measured capacities (the minima column) is still lower than the average capacity for all connections (columns 1 and 2). So, in order to access the capacity of all connections by only looking at the minima column (the test results), another comparison between the three distributions was made. For this comparison the standard deviation for the Min., Max. and numbers 1 and 2 was calculated (since it has been shown that they are normally distributed this may now be done). After that, the values from all columns were taken and used to make the cumulative distribution graphs shown in Figure 8 - Cumulative distribution of the randomly generated N(100,5) numbers.

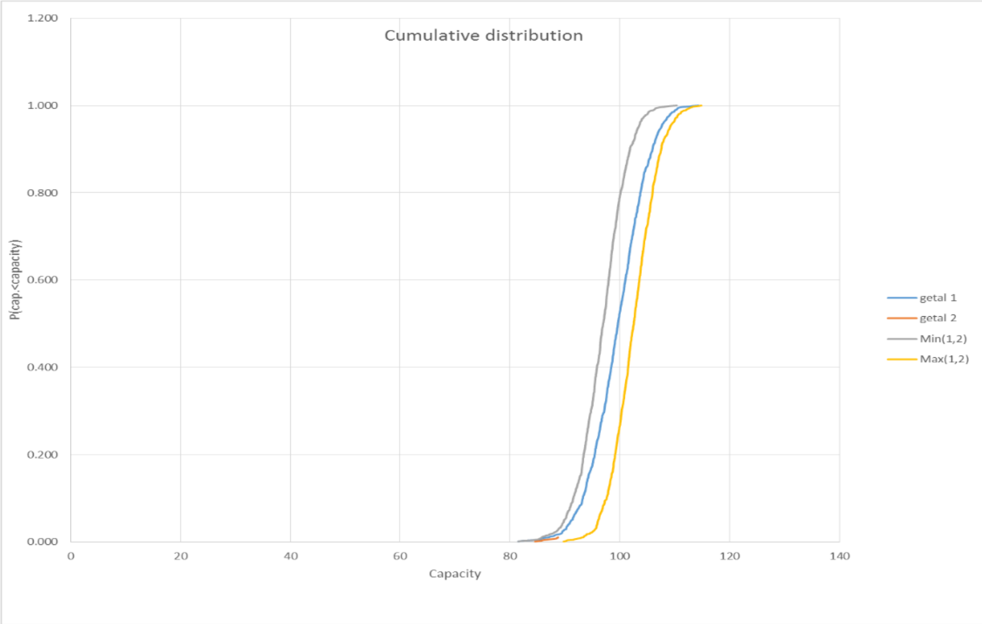


Figure 8 - Cumulative distribution of the randomly generated N(100,5) numbers

In this figure four cumulative distributions are shown. Since both columns 1 and 2 consist of numbers from the same normal distribution, the graphs from these columns (red and blue lines) are directly on top of each other. The graphs from the Min. and Max. columns (the grey and yellow lines) are translated horizontally and have a steeper inclination.

The horizontal translations are caused by a shift in the average value from every column (the average for the Min. is lower and the Max. is higher). The steeper inclinations are caused by a lower standard deviation the Min. and Max. columns. Both the horizontal translation and the steeper inclination could have been expected when looking at the values given in Figure 6 - Simulation of two times 1000 random N(100,5) numbers.

To show the shape of the distribution curves another simulation was ran using a total of 100 000 test values. Again for these values both the Min. and Max. values were gathered. Then all data was sorted into bins of width 1. The resulting amount of values within a bin was set out in a graph. This graph is shown in Figure 9 - Distribution of 100 000 randomly generated N(100,5) numbers.

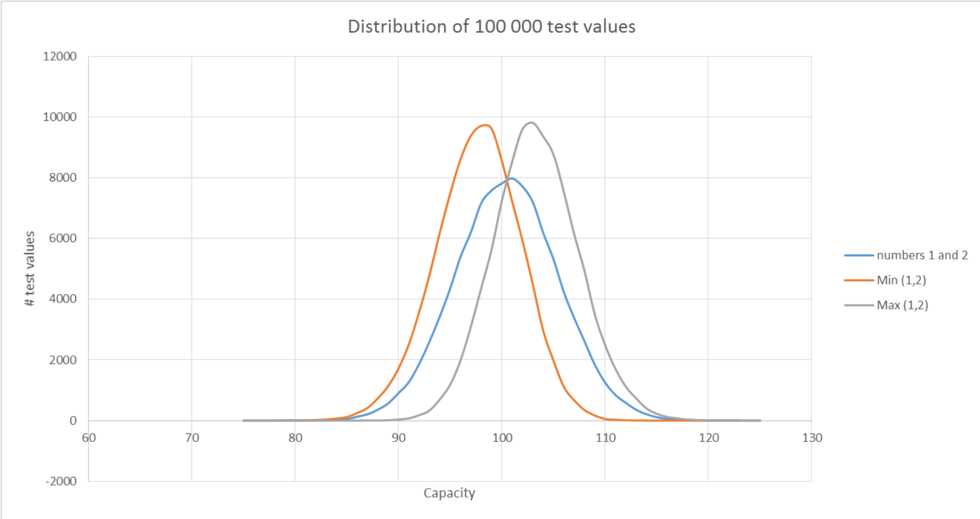


Figure 9 - Distribution of 100 000 randomly generated N(100,5) numbers

This distribution graph shows the same behaviour as the cumulative distribution graphs. The horizontal translation because of a difference in the found average values and the lower standard deviation of the Min. and Max. columns are clearly visible here.

Other things worth mentioning are the areas in which both the Min. and Max. graphs show close resemblance to the graphs of numbers 1 and 2. In the cumulative distributions the lower standard deviation causes the steeper inclination by rotating the Min. and Max. curves about their average value. Due to this rotating effect the Min. curve (grey line) will be close to the normal curve (blue line) in the lower end of the graph and further from the normal curve in the higher end. The Max. curve will be further away in the lower end and closer to the normal graph in the higher end of the curve. This could mean that, when only looking at the tails of the distributions (i.e. the 5-percentile values), the Min. column could be used to give an estimation (lower bound) for the lower 5-percentile of the normal distribution and the Max. column could be used to give an estimation (upper bound) of the upper 5-percentile of the normal distribution.
This effect can also be seen by looking at the values for the 5-percentiles in Figure 6 - Simulation of two times 1000 random N(100,5) numbers. Here, the first value is the lower 5-percentile and the second value is the higher 5-percentile. It can be seen that there is little difference between the lower 5-percentile value of the normal distribution and the 5-percentile of the Min. values. Also the higher 5-percentile of the normal distribution is close to that of the Max. values.
The 5-percentile value is calculated according to (NEN-EN 14358, 2007). In this norm a formula is given by which the 5-percentile value of a normal distribution can be calculated. This formula is the following:

$$x_k = \bar{x} \pm k_s * \sigma \quad (5)$$

In this formula $\bar{x}$ is the expectancy of the normal distribution and σ is the standard deviation. The factor k_s is determined by the amount of test data that is available. The norm employs a confidence level of 75%. Meaning that for any number of test series the 5-percentile value for 75% of all series will be equal to (or higher) than x_k. So when very little test data is available the factor k_s will have to be larger in order to be confident that 75% of all batches of test specimens have a 5-percentile value equal to or higher than x_k. The values for k_s prescribed by the norm are given in Figure 10 - k_s values.

Number of test specimens	Factor
n	k_s
3	3,15
5	2,46
10	2,10
15	1,99
20	1,93
30	1,87
50	1,81
100	1,76
500	1,71
∞	1,65

Figure 10 - k_s values **(NEN-EN 14358, 2007)**

In the table above can be seen that the value for k_s used to determine the 5-percentile value increases significantly for a small number of tests. For a number of 30 test specimens and up the value is relatively stable. To examine the effect of this larger value for k_s and to access whether or not it would be possible to say something about the 5-percentile value of all connections with measurements based solely upon the values from the Min. column, another simulation was run. This simulation represents exactly the tests that are done during this research. Each test specimen in this research consists of two connections ('Number 1' and 'Number 2') with a capacity that is normally distributed. Every specimen variant was tested a total of 5 times. Since the weaker connection breaks first, only the lower of the two capacities can be measured. These values are represented by the Min(1,2) column. The results of one of these simulations are shown in Figure 11 - Simulation of two times 5 random N(100,5) distributed numbers.

Randomly generated normally distributed N(100,5) numbers							MU	100		Sigma	5	
	Number 1	**Number 2**		**Min(1,2)**	**Max(1,2)**		**Ave(1,2)**		**Ave Min(1,2)**		**Ave Max(1,2)**	
0.001	100.74	102.16		100.74	102.16		96.51		94.98	-1.59%	98.04	1.59%
0.002	90.57	88.30		88.30	90.57							
0.003	97.29	95.02		95.02	97.29		**StDev(1,2)**		**StDev Min(1,2)**		**StDev Max(1,2)**	
0.004	102.72	97.41		97.41	102.72		4.77		4.64		4.89	
0.005	93.43	97.46		93.43	97.46							
							COV(1,2)		**COV Min(1,2)**		**COV Max(1,2)**	
							0.05		0.05		0.05	
							5-perc(1,2)		**5-perc Min(1,2)**		**5-perc Max(1,2)**	
							84.76		83.56	-1.42%	86.01	
							108.25		106.39		110.06	1.67%
							Kurt(1,2)		**Kurt Min(1,2)**		**Kurt Max(1,2)**	
							-0.68		0.45		0.50	
							Skew(1,2)		**Skew Min(1,2)**		**Skew Max(1,2)**	
							-0.38		-0.40		-0.87	

Figure 11 - Simulation of two times 5 random N(100,5) distributed numbers

In the figure the same values are displayed as for the earlier ran simulation using 1000 random numbers.
The average values of all columns are again listed on the first row. From a comparison can be seen that, even though only a small number of tests was performed, the found average capacities 'Ave Min(1,2)' are only slightly lower than the actual average capacity 'Ave(1,2)'. This also holds for other simulations ran using the same model.
When looking at differences in the 5-percentile values the gap between the numbers can still be considered small. The difference between the 5-percentile of the normally distributed numbers 1 and 2 and the 5-percentile of the Min. column (lower value) and the Max. column (upper value) are given in the form of percentages next to the 5-

percentile values. Since the simulation is only based on 5 test specimens, the shown percentage will, of course, vary somewhat when a new simulation is run. By doing a number of simulations could be seen that the difference between these 5-percentile values averages around 2% and the difference in the average values of the columns averages around 2.5%.

The difference between the average values and the standard deviations of symmetrically loaded specimens was also researched by Van Douwen et al. (1958) at the Stevin laboratory at TU Delft. They found that in an experiment where always the lower of the two values is found and the test results can be assumed normally distributed (as shown by the simulation from Excel), the following relations apply (Van de Kuilen, J.W.G., Blass, H.J, 2004):

$$x_1 = x_2 + c_1 * \sigma_2 \tag{6}$$
$$\sigma_1 = c_2 * \sigma_2 \tag{7}$$

In these relations x_1 and x_2 are the average of the whole and tested population, c_1 and c_2 the correction factors for the mean and standard deviation and σ_1 and σ_2 are the standard deviations of the whole and tested population.
For symmetrically loaded specimens the correction factors are $c_1 = 0.68$ and $c_2 = 1.21$ (Van de Kuilen, J.W.G., Blass, H.J, 2004). Using these values for the correction factors, the average and standard deviation of the Min. column of the Excel simulation can be adjusted to approximate the average and standard deviation of the original distributions. The given relations shall be used to determine the average capacity and standard deviation of all bamboo connections from the found test results.

6 References

F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester. (2005). *A Modern Introduction to Probability and Statistics.* London: Springer-Verlag London Limited 2005.

NEN-EN 14358, e. (2007). *houtconstructies, berekening van de karakteristieke 5-percentielwaarden en aanvaardingscriteria voor nde acceptatie van een steekproef.* NEN-EN.

Ruegg, R. (2015). *An Experiment in the Ability of Raters to Evaluate Lexis in Writing.* Akita International University, Japan: LIF - Language in Focus Journal, Volume 1, No: 1, 2015.

Van de Kuilen, J.W.G., Blass, H.J. (2004). *Mechanical properties of azobé (Lophira alata).* Holz aus Roh- und Werkstoff (2005) 63: 1-10.