

WIKIPEDIA

# Quantile normalization

In statistics, **quantile normalization** is a technique for making two distributions identical in statistical properties. To quantile-normalize a test distribution to a reference distribution of the same length, sort the test distribution and sort the reference distribution. The highest entry in the test distribution then takes the value of the highest entry in the reference distribution, the next highest entry in the reference distribution, and so on, until the test distribution is a perturbation of the reference distribution.

To quantile normalize two or more distributions to each other, without a reference distribution, sort as before, then set to the average (usually, arithmetic mean) of the distributions. So the highest value in all cases becomes the mean of the highest values, the second highest value becomes the mean of the second highest values, and so on.

Generally a reference distribution will be one of the standard statistical distributions such as the Gaussian distribution or the Poisson distribution. The reference distribution can be generated randomly or from taking regular samples from the cumulative distribution function of the distribution. However, any reference distribution can be used.

Quantile normalization is frequently used in microarray data analysis. It was introduced as **quantile standardization**<sup>[1]</sup> and then renamed as **quantile normalization**.<sup>[2]</sup>

## Example

A quick illustration of such normalizing on a very small dataset:

Arrays 1 to 3, genes A to D

A	5	4	3
B	2	1	4
C	3	4	6
D	4	2	8

For each column determine a rank from lowest to highest and assign number i-iv

A	iv	iii	i
B	i	i	ii
C	ii	iii	iii
D	iii	ii	iv

These rank values are set aside to use later. Go back to the first set of data. Rearrange that first set of column values so each column is in order going lowest to highest value. (First column consists of 5,2,3,4. This is rearranged to 2,3,4,5. Second Column 4,1,4,2 is rearranged to 1,2,4,4, and column 3 consisting of 3,4,6,8 stays the same because it is already in order from lowest to highest value.) The result is:

A	5	4	3	becomes A	2	1	3
B	2	1	4	becomes B	3	2	4
C	3	4	6	becomes C	4	4	6
D	4	2	8	becomes D	5	4	8

Now find the mean for each row to determine the ranks

$$\begin{aligned} A \ (2 \ 1 \ 3)/3 &= 2.00 = \text{rank i} \\ B \ (3 \ 2 \ 4)/3 &= 3.00 = \text{rank ii} \\ C \ (4 \ 4 \ 6)/3 &= 4.67 = \text{rank iii} \\ D \ (5 \ 4 \ 8)/3 &= 5.67 = \text{rank iv} \end{aligned}$$

Now take the ranking order and substitute in new values

A	iv	iii	i
B	i	i	ii
C	ii	iii	iii
D	iii	ii	iv

becomes:

A	5.67	4.67	2.00
B	2.00	2.00	3.00
C	3.00	4.67	4.67
D	4.67	3.00	5.67

These are the new normalized values.

However, note that when, as in column two, values are tied in rank, they should instead be assigned the mean of the values. So in column two, we assign the two tied rank iii entries the mean of 4.67 and 5.67, arriving at this set of normalized values:

A	5.67	5.17	2.00
B	2.00	2.00	3.00
C	3.00	5.17	4.67
D	4.67	3.00	5.67

The new values have the same distribution and can now be easily compared. Here are the summary statistics for each of the three columns:

Min.	:2.000	Min.	:2.000	Min.	:2.000
1st Qu.	:2.750	1st Qu.	:2.750	1st Qu.	:2.750
Median	:3.833	Median	:4.083	Median	:3.833
Mean	:3.833	Mean	:3.833	Mean	:3.833
3rd Qu.	:4.917	3rd Qu.	:5.167	3rd Qu.	:4.917
Max.	:5.667	Max.	:5.167	Max.	:5.667

## References

1. Amaratunga, D.; Cabrera, J. (2001). "Analysis of Data from Viral DNA Microchips". *Journal of the American Statistical Association*. **96** (456): 1161. doi:10.1198/016214501753381814 (<https://doi.org/10.1198%2F016214501753381814>).
2. Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias". *Bioinformatics*. **19** (2): 185–193. doi:10.1093/bioinformatics/19.2.185 (<https://doi.org/10.1093%2Fbioinformatics%2F19.2.185>). PMID 12538238 (<https://www.ncbi.nlm.nih.gov/pubmed/12538238>).

## External links

- Normalization of Affymetrix Chips (<http://www.bea.ki.se/staff/reimers/Web.Pages/Affymetrix.Normalization.htm>)

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Quantile\\_normalization&oldid=928243616](https://en.wikipedia.org/w/index.php?title=Quantile_normalization&oldid=928243616)"

This page was last edited on 27 November 2019, at 19:47 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.