**Researches about theory (R)**

**10\_R. Distributions of the order statistics: look on the web for the most simple (but still rigorous) and clear derivations of the distributions, explaining in your own words the methods used.**

In statistics, the k-th order statistic of a statistical sample is equal to its k-th smallest value. Together with rank statistics, order statistics are among the most fundamental tools in non-parametric statistics and inference.

Important special cases of the order statistics are the minimum and maximum value of a sample, and the sample median and other sample quantiles.

When using probability theory to analyze order statistics of random samples from a continuous distribution, the cumulative distribution function is used to reduce the analysis to the case of order statistics of the uniform distribution. (Wikipedia)

For example, suppose that four numbers are observed or recorded, resulting in a sample of size 5. If the sample values are

6, 9, 3, 8,12

the order statistics would be denoted

x(1) = 3,

x(2) = 6,

x(3) = 8,

x(4) = 9,

x(5) = 12

{\displaystyle x\_{(1)}=3,\ \ x\_{(2)}=6,\ \ x\_{(3)}=8,\ \ x\_{(4)}=9,\,}where the (i) enclosed in parentheses indicates the i-th order statistic of the sample.

The first order statistic (or smallest order statistic) is always the minimum of the sample, that is,

X(1) = Min{X1, …, Xn}

where, following a common convention, we use upper-case letters to refer to random variables, and lower-case letters (as above) to refer to their actual observed values.

Similarly, for a sample of size n, the n-th order statistic (or largest order statistic) is the maximum, that is,

X(n) = Max{X1, …, Xn}

For a random sample as above, with cumulative distribution Fx(x), the order statistics for that sample have cumulative distributions as follows (where *r* specifies which order statistic):

Text

Description automatically generated

Probability density functions of the order statistics for a sample of size n = 5 from an exponential distribution with unit scale parameter

A picture containing shape

Description automatically generated

**11\_R. Do a research about the general correlation coefficient for ranks and the most common indices that can be derived by it. Do one example of computation of these correlation coefficients for ranks.**   
What is Rank Order Statistics?

We can say that in Order Statistics we use the actual values to label the items. In the example above with some numbers (6, 9, 3, 8,12) we have discussed using these numbers referring to them and using their actual values, but we could also use some ‘ranks’, so something similar as categories. In the example above the order statistics would be (3, 6, 8, 9, 12) and the ranks could be (2, 4, 1, 3, 5).

Another example:

A picture containing text, whiteboard

Description automatically generated

The ranks order statistics is useful to compare the same experiments to see correlations using different points of view.

In statistics, a rank correlation is any of several statistics that measure an ordinal association - the relationship between rankings of different ordinal variables or different rankings of the same variable, where a "ranking" is the assignment of the ordering labels "first", "second", "third", etc. to different observations of a particular variable. A rank correlation coefficient measures the degree of similarity between two rankings and can be used to assess the significance of the relation between them. For example, two common nonparametric methods of significance that use rank correlation are the Mann–Whitney U test and the Wilcoxon signed-rank test.

If, for example, one variable is the identity of a college basketball program and another variable is the identity of a college football program, one could test for a relationship between the poll rankings of the two types of programs: do colleges with a higher-ranked basketball program tend to have a higher-ranked football program? A rank correlation coefficient can measure that relationship, and the measure of significance of the rank correlation coefficient can show whether the measured relationship is small enough to likely be a coincidence.

If there is only one variable, the identity of a college football program, but it is subject to two different poll rankings (say, one by coaches and one by sportswriters), then the similarity of the two different polls' rankings can be measured with a rank correlation coefficient.

As another example, in a contingency table with low income, medium income, and high income in the row variable and educational level - no high school, high school, university - in the column variable), a rank correlation measures the relationship between income and educational level. (Wikipedia)

Some of the more popular rank correlation statistics include:

1. Spearman’s p
2. Kendall’s t
3. Goodman and Kruskall’s y
4. Somers’ D

**Applications / Practice (A)     [work on this at least 30' a day, all days]**

**10\_A. Given a random variable, extract m samples of size n and plot the empirical distribution of its mean (histogram), the first and the last order statistics. Comment on what you see.  
  
11\_A. Discover a new important stochastic process by yourself! Consider the general scheme we have used so far to simulate some stochastic processes (such as the relative frequency of success in a sequence of trials, the sample mean and the random walk) and now add this new process to our process simulator.  
  
Same scheme as previous program (random walk), except changing the way to compute the values of the paths at each time. Starting from value 0 at time 0, for each of m paths, at each new time compute N(i) = N(i-1) + Random step(i), for i = 1, ..., n, where Random step(i) is now a Bernoulli random variable with success probability equal to λ \* (1/n)  (where λ is a user parameter, eg. 50, 100, ...).  
  
At time n (last time) and one (or more) other chosen inner time 1<j<n (j is a program parameter) create and represent with histogram the distribution of N(i).**

**Represent also the distributions of the following quantities (and any other quantity that you think of interest):  
- Distance (time elapsed) of individual jumps from the origin  
- Distance (time elapsed) between consecutive jumps (the so-called "holding times")**

Write here **Researches about applications (RA)**

**8\_RA. Find out on the web what you have just generated in the previous application. Can you find out about all the well known distributions that "naturally arise" in this process ?**

In probability theory and statistics, the Poisson distribution, named after French mathematician Siméon Denis Poisson, is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area, or volume.

**References**

<https://en.wikipedia.org/wiki/Order_statistic>

<https://en.wikipedia.org/wiki/Rank_correlation>