

Discrete uniform distribution

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **discrete uniform distribution** is a symmetric probability distribution whereby a finite number of values are equally likely to be observed; every one of *n* values has equal probability *1*/*n*. Another way of saying "discrete uniform distribution" would be "a known, finite number of outcomes equally likely to happen".

A simple example of the discrete uniform distribution is throwing a fair die. The possible values are 1, 2, 3, 4, 5, 6, and each time the die is thrown the probability of a given score is 1/6. If two dice are thrown and their values added, the resulting distribution is no longer uniform since not all sums have equal probability.

The discrete uniform distribution itself is inherently non-parametric. It is convenient, however, to represent its values generally by an integer interval [*a*,*b*], so that *a*,*b* become the main parameters of the distribution (often one simply considers the interval [1,*n*] with the single parameter *n*). With these conventions, the cumulative distribution function (CDF) of the discrete uniform distribution can be expressed, for any *k* ∈ [*a*,*b*], as

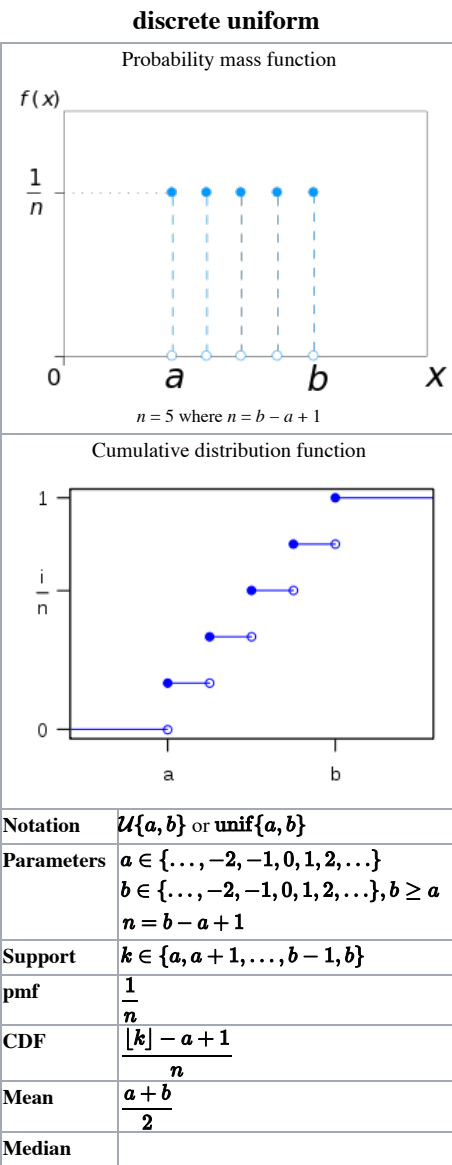
$$F(k;a,b) = \frac{\lfloor k \rfloor - a + 1}{b - a + 1}$$

Contents

- 1 Estimation of maximum
- 2 Random permutation
- 3 See also
- 4 Notes
- 5 References

Estimation of maximum

This example is described by saying that a sample of *k* observations is obtained from a uniform distribution on the integers **1, 2, ..., N**, with the problem being to



estimate the unknown maximum *N*. This problem is commonly known as the German tank problem, following the application of maximum estimation to estimates of German tank production during World War II.

The UMVU estimator for the maximum is given by

$$\hat{N} = \frac{k+1}{k}m - 1 = m + \frac{m}{k} - 1$$

where *m* is the sample maximum and *k* is the sample size, sampling without replacement.^{[1][2]} This can be seen as a very simple case of maximum spacing estimation.

The formula may be understood intuitively as:

"The sample maximum plus the average gap between observations in the sample",

the gap being added to compensate for the negative bias of the sample maximum as an estimator for the population maximum.^[notes 1]

This has a variance of^[1]

$$\frac{1}{k} \frac{(N - k)(N + 1)}{(k + 2)} \approx \frac{N^2}{k^2} \text{ for small samples } k \ll N$$

so a standard deviation of approximately $\frac{N}{k}$, the (population) average size of a gap between samples; compare $\frac{m}{k}$ above.

The sample maximum is the maximum likelihood estimator for the population maximum, but, as discussed above, it is biased.

If samples are not numbered but are recognizable or markable, one can instead estimate population size via the capture-recapture method.

Random permutation

See rencontres numbers for an account of the probability distribution of the number of fixed points of a uniformly distributed random permutation.

See also

- Delta distribution
- Uniform distribution (continuous)