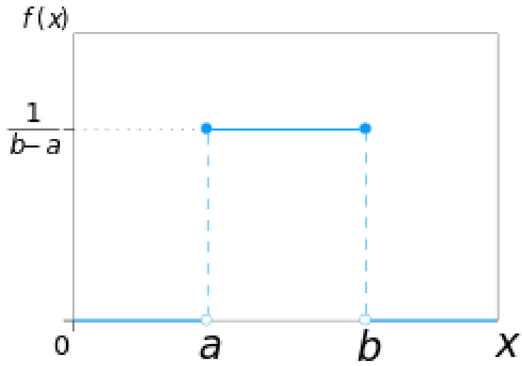
Uniform distribution (continuous)

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **continuous uniform distribution** or **rectangular distribution** is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's support are equally probable. The support is defined by the two parameters, *a* and *b*, which are its minimum and maximum values. The distribution is often abbreviated *U*(*a*,*b*). It is the maximum entropy probability distribution for a random variate *X* under no constraint other than that it is contained in the distribution's support.[1]

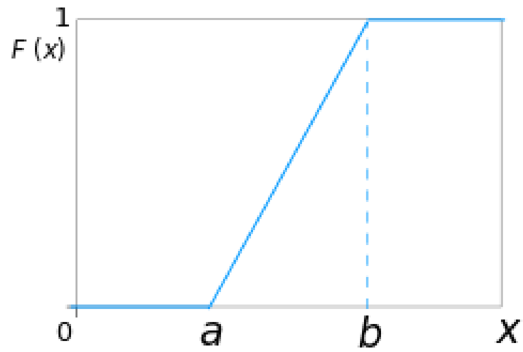
Uniform

Probability density function



Using maximum convention

Cumulative distribution function



Notation	$\mathcal{U}(a,b)$ or $\mathrm{unif}(a,b)$
Parameters	$-\infty < a < b < \infty$
Support	$x \in [a,b]$
PDF	$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$
CDF	$\begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a,b) \\ 1 & \text{for } x \ge b \end{cases}$
Mean	$\frac{1}{2}(a+b)$
Median	$\frac{1}{2}(a+b)$
Mode	any value in (a,b)
Variance	$\frac{1}{12}(b-a)^2$
Skewness	0
Ex. kurtosis	$-\frac{6}{5}$
Entropy	$\log(b-a)$

Contents

MGF	$\begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0 \end{cases}$
CF	$\dfrac{e^{itb} - e^{ita}}{it(b-a)}$

Characterization

Probability density function

The probability density function of the continuous uniform distribution is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

The values of $f(x)$ at the two boundaries a and b are usually unimportant because they do not alter the values of the integrals of $f(x)\,dx$ over any interval, nor of $x\,f(x)\,dx$ or any higher moment. Sometimes they are chosen to be zero, and sometimes chosen to be $1/(b-a)$. The latter is appropriate in the context of estimation by the method of maximum likelihood. In the context of Fourier analysis, one may take the value of $f(a)$ or $f(b)$ to be $1/(2(b-a))$, since then the inverse transform of many integral transforms of this uniform function will yield back the function itself, rather than a function which is equal "almost everywhere", i.e. except on a set of points with zero measure. Also, it is consistent with the sign function which has no such ambiguity.

In terms of mean μ and variance σ^2, the probability density may be written as:

$$f(x) = \begin{cases} \frac{1}{2\sigma\sqrt{3}} & \text{for } -\sigma\sqrt{3} \le x - \mu \le \sigma\sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

Cumulative distribution function

The cumulative distribution function is:

$$F(x) = \begin{cases} 0 & \text{for } x < a \\[8pt] \frac{x-a}{b-a} & \text{for } a \le x \le b \\[8pt] 1 & \text{for } x > b \end{cases}$$

Its inverse is:

$$F^{-1}(p) = a + p(b-a) \quad \text{for } 0 < p < 1$$

In mean and variance notation, the cumulative distribution function is:

$$F(x) = \begin{cases} 0 & \text{for } x - \mu < -\sigma\sqrt{3} \\ \frac{1}{2}\left(\frac{x-\mu}{\sigma\sqrt{3}} + 1\right) & \text{for } -\sigma\sqrt{3} \le x - \mu < \sigma\sqrt{3} \\ 1 & \text{for } x - \mu \ge \sigma\sqrt{3} \end{cases}$$

and the inverse is:

$$F^{-1}(p) = \sigma\sqrt{3}(2p-1) + \mu \quad \text{for } 0 \le p \le 1$$

Generating functions

Moment-generating function

The moment-generating function is:[2]

$$M_x = E(e^{tx}) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

from which we may calculate the raw moments m_k

$$m_1 = \frac{a+b}{2},$$

$$m_2 = \frac{a^2 + ab + b^2}{3},$$

$$m_k = \frac{1}{k+1} \sum_{i=0}^{k} a^i b^{k-i}.$$

For a random variable following this distribution, the expected value is then $m_1 = (a + b)/2$ and the variance is $m_2 - m_1^2 = (b - a)^2/12$.

Cumulant-generating function

For $n \geq 2$, the nth cumulant of the uniform distribution on the interval [-1/2, 1/2] is b_n/n, where b_n is the nth Bernoulli number.[3]

Properties

Moments

The mean (first moment) of the distribution is:

$$E(X) = \frac{1}{2}(a + b).$$

The variance (second central moment) is:

$$V(X) = \frac{1}{12}(b - a)^2$$

Order statistics

Let $X_1, ..., X_n$ be an i.i.d. sample from $U(0,1)$. Let $X_{(k)}$ be the kth order statistic from this sample. Then the probability distribution of $X_{(k)}$ is a Beta distribution with parameters k and $n - k + 1$. The expected value is

$$\mathrm{E}(X_{(k)}) = \frac{k}{n+1}.$$

This fact is useful when making Q-Q plots.

The variances are

$$\mathrm{V}(X_{(k)}) = \frac{k(n-k+1)}{(n+1)^2(n+2)}.$$

Uniformity

The probability that a uniformly distributed random variable falls within any interval of fixed length is independent of the location of the interval itself (but it is dependent on the interval size), so long as the interval is contained in the distribution's support.

To see this, if $X \sim U(a,b)$ and $[x, x+d]$ is a subinterval of $[a,b]$ with fixed $d > 0$, then

$$P(X \in [x, x+d]) = \int_x^{x+d} \frac{\mathrm{d}y}{b-a} = \frac{d}{b-a}$$

which is independent of x. This fact motivates the distribution's name.

Generalization to Borel sets

This distribution can be generalized to more complicated sets than intervals. If S is a Borel set of positive, finite measure, the uniform probability distribution on S can be specified by defining the pdf to be zero outside S and constantly equal to $1/K$ on S, where K is the Lebesgue measure of S.

Standard uniform

Restricting $a = 0$ and $b = 1$, the resulting distribution $U(0,1)$ is called a **standard uniform distribution**.

One interesting property of the standard uniform distribution is that if u_1 has a standard uniform distribution, then so does $1-u_1$. This property can be used for generating antithetic variates, among other things.

Related distributions

- If X has a standard uniform distribution, then by the inverse transform sampling method, $Y = -\lambda^{-1}$ ln(X) has an exponential distribution with (rate) parameter λ.
- If X has a standard uniform distribution, then $Y = X^n$ has a beta distribution with parameters *1/n* and *1*. (Note this implies that the standard uniform distribution is a special case of the beta distribution, with parameters 1 and 1.)
- The Irwin–Hall distribution is the sum of n i.i.d. *U(0,1)* distributions.
- The sum of two independent, equally distributed, uniform distributions yields a symmetric triangular distribution.
- The distance between two i.i.d. uniform random variables also has a triangular distribution, although not symmetric.
- The uniform distribution can be thought of as a beta distribution with parameters (1,1).

Relationship to other functions

As long as the same conventions are followed at the transition points, the probability density function may also be expressed in terms of the Heaviside step function:

$$f(x) = \frac{\mathrm{H}(x-a) - \mathrm{H}(x-b)}{b-a},$$

or in terms of the rectangle function

$$f(x) = \frac{1}{b-a} \operatorname{rect}\left(\frac{x - \left(\frac{a+b}{2}\right)}{b-a}\right).$$

There is no ambiguity at the transition point of the sign function. Using the half-maximum convention at the transition points, the uniform distribution may be expressed in terms of the sign function as:

$$f(x) = \frac{\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)}{2(b-a)}.$$

Applications

In statistics, when a p-value is used as a test statistic for a simple null hypothesis, and the distribution of the test statistic is continuous, then the p-value is uniformly distributed between 0 and 1 if the null hypothesis is true.

Sampling from a uniform distribution

There are many applications in which it is useful to run simulation experiments. Many programming languages have the ability to generate pseudo-random numbers which are effectively distributed according to the standard uniform distribution.

If u is a value sampled from the standard uniform distribution, then the value $a + (b - a)u$ follows the uniform distribution parametrised by a and b, as described above.

Sampling from an arbitrary distribution

The uniform distribution is useful for sampling from arbitrary distributions. A general method is the inverse transform sampling method, which uses the cumulative distribution function (CDF) of the target random variable. This method is very useful in theoretical work. Since simulations using this method require inverting the CDF of the target variable, alternative methods have been devised for the cases where the cdf is not known in closed form. One such method is rejection sampling.

The normal distribution is an important example where the inverse transform method is not efficient. However, there is an exact method, the Box–Muller transformation, which uses the inverse transform to convert two independent uniform random variables into two independent normally distributed random variables.

Quantization error

In analog-to-digital conversion a quantization error occurs. This error is either due to rounding or truncation. When the original signal is much larger than one least significant bit (LSB), the quantization error is not significantly correlated with the signal, and has an approximately uniform distribution. The RMS error therefore follows from the variance of this distribution.

Estimation

Estimation of maximum

Minimum-variance unbiased estimator

Given a uniform distribution on [0, *b*] with unknown *b,* the minimum-variance unbiased estimator (UMVU) estimator for the maximum is given by

$$\hat{b}_{UMVU} = \frac{k+1}{k} m = m + \frac{m}{k}$$

where *m* is the sample maximum and *k* is the sample size, sampling without replacement (though this distinction almost surely makes no difference for a continuous distribution). This follows for the same reasons as estimation for the discrete distribution, and can be seen as a very simple case of maximum spacing estimation. This problem is commonly known as the German tank problem, due to application of maximum estimation to estimates of German tank production during World War II.

Maximum Likelihood estimator

The maximum likelihood estimator is given by:

$$\hat{b}_{ML} = m$$

where *m* is the sample maximum, also denoted as $m = X_{(n)}$ the maximum order statistic of the sample.

Method of moment estimator

The method of moments estimator is given by:

$$\hat{b}_{MM} = 2\bar{X}$$

where $\bar{X}$ is the sample mean.

Estimation of midpoint

The midpoint of the distribution (*a* + *b*) / 2 is both the mean and the median of the uniform distribution. Although both the sample mean and the sample median are unbiased estimators of the midpoint, neither is as efficient as the sample mid-range, i.e. the arithmetic mean of the sample maximum and the sample minimum, which is the UMVU estimator of the midpoint (and also the maximum likelihood estimate).

Confidence interval for the maximum

Let $X_1, X_2, X_3, ..., X_n$ be a sample from $U(0, L)$ where *L* is the population maximum. Then $X_{(n)} = \max(X_1, X_2, X_3, ..., X_n)$ has the density[4]

$$f_n(X_{(n)}) = n\frac{1}{L}\left(\frac{X_{(n)}}{L}\right)^{n-1} = n\frac{X_{(n)}^{n-1}}{L^n}, 0 < X_n < L$$

The confidence interval for the estimated population maximum is then ($X_{(n)}$, $X_{(n)}$ / $\alpha^{1/n}$) where 100 (1 - α)% is the confidence level sought. In symbols

$$X_{(n)} \le L \le X_{(n)}/\alpha^{1/n}$$

See also

- Beta distribution
- Box–Muller transform
- Probability plot
- Q-Q plot
- Rectangular function
- Uniform distribution (discrete)

References

1. Park, Sung Y.; Bera, Anil K. (2009). "Maximum entropy autoregressive conditional heteroskedasticity model". *Journal of Econometrics* (Elsevier) **150** (2): 219–230. doi:10.1016/j.jeconom.2008.12.014.
2. Casella & Berger 2001, p. 626
3. https://galton.uchicago.edu/~wichura/Stat304/Handouts/L18.cumulants.pdf
4. Nechval KN, Nechval NA, Vasermanis EK, Makeev VY (2002) Constructing shortest-length confidence intervals. Transport and Telecommunication 3 (1) 95-103

Further reading

- Casella, George; Roger L. Berger (2001), *Statistical Inference* (2nd ed.), ISBN 0-534-24312-6, LCCN 2001025794

External links

- Online calculator of Uniform distribution (continuous) (http://www.elektro-energetika.cz/calculations/ro.php?language=english)

Categories: Continuous distributions | Probability distributions | Location-scale family probability distributions