

Uniform distribution

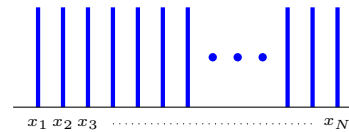
Statistics for Business

Josemari Sarasola - Gizapedia

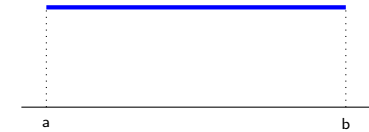


Uniform distribution

The **uniform distribution** is the probability distribution that gives the same probability to all possible values. There are two types of uniform distributions:



Discrete uniform distribution



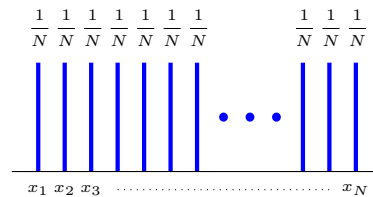
Continuous uniform distribution

We apply uniform distributions when there is **absolute uncertainty** about what will occur. We also use them **when we take a random sample from a population**, because in such cases all the elements have the same probability of being drawn. Finally, they are also used **to create random numbers**, as random numbers are those that have the same probability of being drawn.

Discrete uniform distribution

Probability function

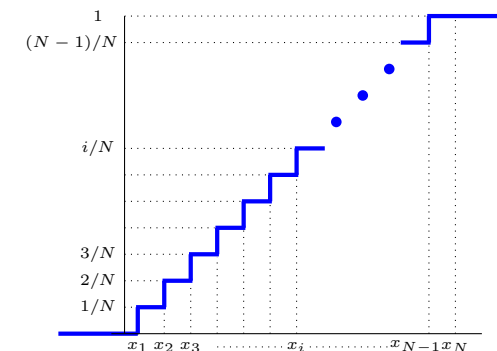
$$P[X = x] = \frac{1}{N}; x = x_1, x_2, \dots, x_N$$



Discrete uniform distribution

Distribution function

$$F(x) = P[X \leq x_i] = \frac{i}{N}; x_i = x_1, x_2, \dots, x_N$$



Discrete uniform distribution

Notation, mean and variance

$$X \sim U(x_1, x_2, \dots, x_N) \begin{cases} \mu = \frac{x_1 + x_N}{2} \\ \sigma^2 = \frac{(x_N - x_1 + 2)(x_N - x_1)}{12} \\ = \frac{(x_N - x_1 + 1)^2 - 1}{12} \end{cases}$$

Discrete uniform distribution

Distribution of the maximum

We draw a value from a uniform distribution n times. How is distributed the maximum among those n values?

- Among n values, **the maximum will be less than x_i** when all of them are less than x_i :

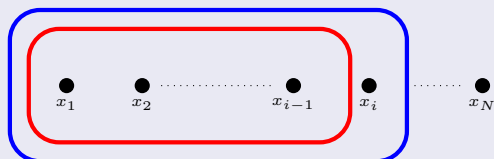
$$P[X_{max} \leq x_i] = \frac{i}{N} \times \frac{i}{N} \times \dots \times \frac{i}{N} = \left(\frac{i}{N}\right)^n$$

- Among n values, **the maximum will be less than x_{i-1}** when all of them are less than x_{i-1} :

$$P[X_{max} \leq x_{i-1}] = \left(\frac{i-1}{N}\right)^n$$

Discrete uniform distribution

Distribution of the maximum



Thus, among n values the probability of **the maximum being x_i** is:

$$P[X_{max} = x_i] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n$$

$$\text{[Blue oval]} - \text{[Red oval]}$$

Discrete uniform distribution

Distribution of the maximum: example

We throw a dice 4 times. We assume (logically) that the number of points follow a discrete uniform distribution. Calculate the probability of the maximum being 1, 2, 3, 4, 5 or 6:

- $P[X_{max} = 6] = \left(\frac{6}{6}\right)^4 - \left(\frac{5}{6}\right)^4 = 0.5177$
- $P[X_{max} = 5] = \left(\frac{5}{6}\right)^4 - \left(\frac{4}{6}\right)^4 = 0.2847$
- $P[X_{max} = 4] = \left(\frac{4}{6}\right)^4 - \left(\frac{3}{6}\right)^4 = 0.1350$
- $P[X_{max} = 3] = \left(\frac{3}{6}\right)^4 - \left(\frac{2}{6}\right)^4 = 0.0501$
- $P[X_{max} = 2] = \left(\frac{2}{6}\right)^4 - \left(\frac{1}{6}\right)^4 = 0.0115$
- $P[X_{max} = 1] = \left(\frac{1}{6}\right)^4 - \left(\frac{0}{6}\right)^4 = 0.0007$
- Probability decreases (logically)**

Discrete uniform distribution

Distribution of the minimum

Likewise,
among n values the probability of the smallest being x_i is:

$$\begin{aligned} P[X_{min} = x_i] &= P[X_{min} \geq x_i] - P[X_{min} \geq x_{i+1}] \\ &= \left(\frac{N - (i - 1)}{N}\right)^n - \left(\frac{N - i}{N}\right)^n \end{aligned}$$

Discrete uniform distribution

Applications: sampling in finite populations

- When we draw a random sample from a finite population, all the elements of the population have the same probability. Thus, the model for the sampling should be the uniform discrete distribution.
- When random sampling is made with devolution, the probability of drawing a given sample of size n is:

$$P[x_1, x_2, \dots, x_n] = \frac{1}{N} \times \frac{1}{N} \times \dots \times \frac{1}{N} = \left(\frac{1}{N}\right)^n$$

- When random sampling is made without devolution, the probability of drawing a given sample of size n is given by the hypergeometric distribution:

$$P[x_1, x_2, \dots, x_n] = \frac{1}{\binom{N}{n}} = \frac{1}{N} \times \frac{1}{N-1} \times \dots \times \frac{1}{N-(n-1)} \times n!$$

Discrete uniform distribution

Application: German tank problem

- It's a classical probability problem, posed in World War II
- German tanks have a number from 1 to an unknown N number.
- Allies wanted to know that number. For that purpose, they collect the numbers of destroyed tanks.
- The model for numbers is the discrete uniform distribution: 1, 2, ..., N numbered tanks have all the same probability of being destroyed.
- Allies collect n random values from that distribution, they calculate how distributes the maximum of them and calculate the expected value, without devolution, and missing the N value:

$$E[X_{max}] = \frac{n(N+1)}{n+1}$$

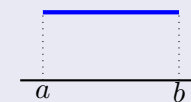
- They take the x_{max} maximum and equal to the expected value, in order to estimate the unknown N value

$$x_{max} = \frac{n(N+1)}{n+1} \rightarrow \hat{N} = x_{max} + \frac{x_{max} - n}{n}$$

- E.g., destroyed tank numbers: 82,123,345,614.
- The estimate for the number of tanks is: $\hat{N} = 614 + \frac{614 - 4}{4} = 766.5$.

Continuous uniform distribution

Density function



$$f(x) = \frac{1}{b-a}; a < x < b$$

$$F(x) = P[X < x] = \frac{x-a}{b-a}; a \leq x \leq b$$

Continuous uniform distribution

Notation, mean and variance

$$X \sim U(a, b)$$

$$X \sim U(a, b) \begin{cases} \mu = \frac{a+b}{2} \\ \sigma^2 = \frac{(b-a)^2}{12} \end{cases}$$

- The mean value is rather intuitive: as the possible values in the support have all the same probability, the mean value will be in the middle.
- The wider is the interval of the support, the bigger is the dispersion, and so the variance.

Continuous uniform distribution

Distribution of the maximum

Taken n values from a $U(a, b)$ distribution, which is the distribution of the maximum M ?

- Distribution function:

$$F(M = x) = P[M < x] = \left(\frac{x-a}{b-a}\right)^n ; a \leq x \leq b$$

- Mean value: $E[M] = a + \frac{n(b-a)}{n+1}$

- E.g., if we want to estimate the maximum of a $U(0, 10)$ distribution (that is, assuming we don't know that the maximum is 10) with 4 values, we should expect that with the maximum of those 4 values we would reach on average $0 + \frac{4 \times 10}{5} = 8$, that is, 80% of the true value. With 9 data we would reach 90%.

Continuous uniform distribution

Distribution of the minimum

Taken n values from a $U(a, b)$ distribution, which is the distribution of the minimum m ?

- Distribution function:

$$F(m = x) = P[m < x] = 1 - \left(\frac{b-x}{b-a}\right)^n ; a \leq x \leq b$$

- Mean value: $E[m] = a + \frac{b-a}{n+1}$
- E.g., if we want to estimate the minimum of a $U(0, 10)$ distribution (that is, assuming we don't know that the minimum is 0) with 4 values, we should expect that with the minimum of those 4 values we would reach on average $0 + \frac{10}{5} = 2$. With 10 data we would get 1 on average.

Continuous uniform distribution

Distribution of the range

Taken n values from a $U(a, b)$ distribution, which is the distribution of the range R (maximum - minimum)?

- Distribution function (for $a \leq x \leq b$):

$$F(R = x) = P[R < x] = n\left(\frac{x}{b-a}\right)^{n-1} \left(\frac{b-a-x}{b-a}\right) + \left(\frac{x}{b-a}\right)^n$$

- $E[R] = (b-a) \frac{n-1}{n+1}$

- E.g., if we want to estimate the range of a $U(0, 10)$ distribution (that is, assuming we don't know that the range is $10-0=10$) with 4 values, we should expect that with the sample range of those 4 values we would reach on average $(10-0) \times \frac{3}{5} = 6$, 60% from the true value ($10-0=10$). With 10 data we would get $9/11=81\%$ on average.

Continuous uniform distribution

Standard uniform distribution

$$X \sim U(0,1)$$

Random numbers from 0 to 1 come from this distribution. We can create (better, simulate) them, by typing SHIFT+RAN# in the calculator.

Stochastic simulation for uniform distributions

Stochastic simulation is artificially creating data, following a given distribution.

Simulating a continuous uniform distribution is very easy compared to other distributions: random numbers follow the $U(0,1)$ distribution, and to simulate $U(a,b)$ we just have to make this linear transform: $U(a,b) = a + (b-a)U(0,1)$

So, naming *sim* the simulated data: $sim_{U(a,b)} = a + (b-a)sim_{U(0,1)}$