Learning Maximum Likelihood Estimation: A Practical Guide to MLE with Uniform Distributions

Authored by

Mohammed looti

November 9, 2025

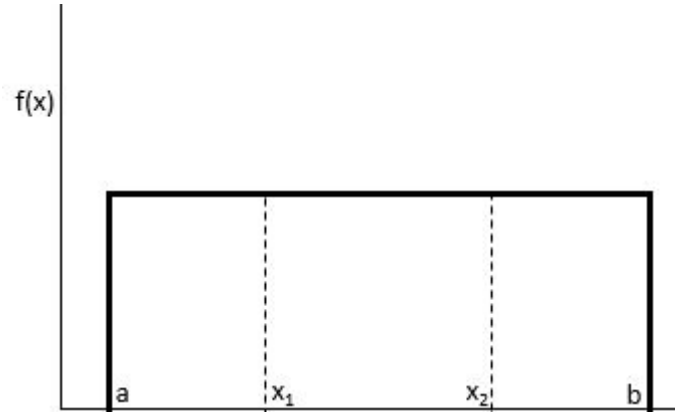
The Uniform Distribution stands as a foundational concept in **probability theory**, sometimes referred to descriptively as the rectangular distribution. It mathematically models scenarios where every outcome within a specified finite interval, defined by a lower bound, **a**, and an upper bound, **b**, possesses precisely the same probability of occurrence. This inherent simplicity makes the uniform distribution an ideal candidate for modeling processes characterized by constant, unbiased randomness across a given range. Before one attempts to apply statistical estimation techniques, a thorough understanding of this distribution's unique properties is essential, as the approach to estimating its parameters, **a** and **b**, diverges sharply from methods typically used for curves defined by continuous parameters, such as the Normal or Exponential distributions.

In formal terms, the probability of observing a random variable X within any sub-interval (x_1, x_2) that lies within the total support interval $$ is directly proportional to the relative length of that sub-interval compared to the total span $(b - a)$. This geometric proportionality is the defining characteristic of the uniform distribution, explaining why its plot forms a flat rectangle. If X is designated as a random variable following a uniform distribution on $$, its Probability Density Function (PDF) remains constant throughout this interval and is strictly zero everywhere else.

The computation of the probability that a random variable falls between two specific points, x_1 and x_2, given the known parameters a and b, relies purely on these geometric ratios. This confirms the intuitive definition where the probability mass corresponds exactly to the length. The calculation is summarized as follows:

P(obtain value between x1 and x2) = (x2 - x1) / (b - a)

This comprehensive guide is dedicated to exploring the methodology of Maximum Likelihood Estimation (MLE), a statistically potent framework used to derive estimators for unknown parameters. Our objective is to determine the optimal values for the bounds, **a** and **b**, of the uniform distribution based on an observed sample of data. Crucially, because the uniform distribution's PDF involves a severe discontinuity at its boundaries, the conventional calculus-based approach to MLE must be augmented by a rigorous consideration of the boundary constraints imposed by the parameters themselves.



The Nature of the Continuous Uniform Distribution

The continuous Uniform Distribution is mathematically unique because its Probability Density Function (PDF), denoted $f(x|a, b)$, is non-zero and fixed only across the support interval $$. Specifically, the function is defined as $f(x|a, b) = 1/(b-a)$ for $a le x le b$, and zero otherwise. The parameters **a** and **b** are the non-random values that define the absolute minimum and maximum possible values, respectively, that the random variable X can take. When we collect a sample of n observations, $X = {X_1, X_2, dots, X_n}$, our central goal in parameter estimation is to identify the specific values of a and b that were most likely to have generated the observed data set.

In contrast to distributions like the Gaussian, where parameters such as the mean and variance smoothly influence the shape and position of the curve, **a** and **b** in the uniform distribution function as abrupt, sharp cut-off points. This distinctive structural feature has profound consequences for estimation. The most significant consequence is the absolute constraint: if even a single observation X_i falls outside the hypothesized interval $$--meaning $X_i b$--the probability of observing that entire sample given the parameters (a, b) instantly becomes zero. This strict dependence on the exact boundaries is precisely what prevents the straightforward application of the differential calculus techniques commonly employed in other MLE problems.

Therefore, before proceeding with the formal steps of Maximum Likelihood Estimation, it is mandatory to internalize this critical geometric constraint: for any parameter set (a, b) to be considered plausible or to have a non-zero likelihood, it must satisfy the condition that $a le X_i le b$ for every single observed data point X_i in the sample. If this condition is violated even once, the assumed statistical model is incompatible with the data, rendering the likelihood associated with those parameters null. This boundary condition proves to be the dominant factor in determining the maximum likelihood estimators, overriding any conclusions drawn from derivative calculations alone.

The Maximum Likelihood Estimation Framework

The core principle underpinning Maximum Likelihood Estimation (MLE) is to identify the parameter values that maximize the probability, or likelihood, of having observed the specific sample data that we have collected. Assuming that our observations $X_1, X_2, dots, X_n$ are independent and identically distributed (i.i.d.), the joint probability density function is calculated as the product of the individual PDFs. This resulting function, when treated as a function of the parameters **a** and **b** given the fixed data X, is formally designated as the **Likelihood function**.

For the majority of continuous distributions, the MLE procedure involves two primary stages: first, rigorously defining the likelihood function $L(a, b | X)$; and second, maximizing this function. Maximization is typically achieved by calculating the partial derivatives with respect to each parameter, setting these derivatives to zero, and solving the resulting system of equations to find the critical points. While this procedure is standard, the discontinuous nature introduced by the indicator function within the uniform distribution's PDF introduces a non-differentiable constraint. This necessitates a maximization strategy that deviates from standard calculus and relies heavily on analyzing the geometric boundaries.

The ultimate goal, however, remains fixed: we seek the pair of estimated parameters $(hat{a}_{MLE}, hat{b}_{MLE})$ that assigns the highest possible likelihood value to the observed sample X. Because the core algebraic component of the likelihood function for the uniform distribution is constant (i.e., not dependent on the specific value of x), maximizing the likelihood requires minimizing the denominator $(b-a)$ while simultaneously ensuring that the support boundaries are never violated. This delicate balance--achieving maximum density by minimizing the interval width, while ensuring the interval is wide enough to contain every data point--is the defining conceptual challenge of MLE for the uniform distribution.

Formulating the Likelihood Function and Constraints

The foundational step in solving any MLE problem is the formal construction of the Likelihood function, $L(a, b | X)$. Given the assumption of independence among our n observations, the joint probability density is the product of the individual PDFs: $L(a, b | X) = prod_{i=1}^{n} f(X_i | a, b)$. For the uniform distribution, the PDF $f(X_i | a, b)$ equals $1/(b-a)$ only if $a le X_i le b$, and it equals zero otherwise.

This non-zero condition must be satisfied by all n data points concurrently. This means that a must be less than or equal to the smallest observation, and b must be greater than or equal to the largest observation. We denote the minimum observation in the sample as $X_{(1)}$ and the maximum observation as $X_{(n)}$. If the proposed parameters (a, b) fail to encompass the entire sample (i.e., if $a > X_{(1)}$ or $b < X_{(n)}$), the likelihood is instantaneously zero. This leads to the critical formulation of the likelihood function for the uniform distribution:

The likelihood function, $L(a, b | X)$, is expressed mathematically as $L(a, b | X) = left(frac{1}{b-a} right)^n$ multiplied by an indicator function, $I(a le X_{(1)} text{ and } X_{(n)} le b)$.

Formally, the likelihood function based on the observed sample X is written as:

$$\prod_{i=1}^{n} f(x_i; a, b) = \prod_{i=1}^{n} \frac{1}{(b-a)} = \frac{1}{(b-a)^n}$$

This formulation precisely captures the constraint: the likelihood only possesses a non-zero value if the hypothesized lower bound **a** is less than or equal to the minimum observed value $X_{(1)}$, AND the hypothesized upper bound **b** is greater than or equal to the maximum observed value $X_{(n)}$. Outside of these precise constraints, the likelihood function immediately drops to zero, confirming that any parameter set that does not fully enclose the data is deemed statistically impossible under this model.

The Log-Likelihood and the Failure of Calculus

In standard MLE procedures, the next step involves transforming the likelihood function $L(a, b | X)$ into the Log-likelihood function, $ell(a, b | X) = ln(L(a, b | X))$. This transformation is beneficial for simplifying mathematical analysis, as the logarithm converts the product of probabilities into a summation, and because the logarithm is a monotonically increasing function, maximizing the likelihood is mathematically equivalent to maximizing the log-likelihood. Applying the logarithm to the uniform likelihood function yields:

$$\log \prod_{i=1}^{n} f(x_i; a, b) = \log \prod_{i=1}^{n} \frac{1}{(b-a)} = \log \left((b-a)^{-n}\right) = -n \log (b-a)$$

The algebraic term, $-n cdot ln(b-a)$, is maximized precisely when the interval width $(b-a)$ is minimized, since the likelihood L is inversely proportional to the interval width raised to the power of n. However, this minimization must strictly occur only within the region defined by the indicator function, which imposes the essential constraints $a le X_{(1)}$ and $b ge X_{(n)}$. The non-zero region of the likelihood function is defined by these inequalities, which fundamentally complicate the traditional approach of using partial derivatives to locate a critical point.

If one proceeds to take the partial derivatives of the log-likelihood function with respect to a and b, temporarily disregarding the indicator constraints, the resulting expressions depend only on the parameters and the sample size n. The partial derivative with respect to a is found to be:

$$\frac{n}{(b-a)}$$

Similarly, the partial derivative with respect to b is:

$$-\frac{n}{(b-a)}$$

The attempt to set these derivatives to zero fails to yield a meaningful solution for a and b, as the expressions simplify to $n/(b-a)$ and $-n/(b-a)$, which are non-zero provided $b ne a$. This mathematical result is crucial: it confirms definitively that the maximum of the log-likelihood function does not reside at an interior critical point discoverable by standard calculus techniques. Instead, the maximum must occur at the boundaries defined by the data itself. The likelihood function is, in essence, **monotonically decreasing** with respect to the interval width $(b-a)$, forcing us to seek the smallest possible interval that still manages to contain all the observed data points.

Deriving the Maximum Likelihood Estimators

Since the likelihood $L(a, b | X)$ is constant, $1/(b-a)^n$, within the permissible parameter region ($a le X_{(1)}$ and $b ge X_{(n)}$), the task of maximizing L translates directly into minimizing the denominator $(b-a)^n$. To achieve the absolute minimum possible value for the interval width $(b-a)$, we must strategically choose **a** to be as large as possible, and **b** to be as small as possible, while strictly adhering to the fundamental constraints defined by the sample data.

Considering the lower bound estimator $hat{a}$: the constraint requires that a must be less than or equal to the minimum observation, $X_{(1)}$. To maximize a while staying within this boundary condition, we must set a equal to the largest permissible value, which is precisely the minimum value observed in the sample. Therefore, the Maximum Likelihood Estimator for the parameter **a** is the minimum of the sample statistics:

$hat{a}_{MLE}$ = min(X1, X2, ... , Xn) = X_{(1)}

Conversely, considering the upper bound estimator $hat{b}$: the constraint dictates that b must be greater than or equal to the maximum observation, $X_{(n)}$. To minimize b while respecting this boundary, we must set b equal to the smallest permissible value, which is the maximum value observed in the sample. Consequently, the MLE for the parameter **b** is the maximum of the sample statistics:

$hat{b}_{MLE}$ = max(X1, X2, ... , Xn) = X_{(n)}

These estimators, derived from the sample minimum and sample maximum, define the tightest possible interval that can contain all the observed data points. Any choice of parameters resulting in a wider interval $(b-a)$ would necessarily yield a smaller likelihood value, whereas any choice resulting in a narrower interval would violate the strict data constraints and instantly yield a

likelihood of zero. Thus, the maximum likelihood is attained exactly when the estimated interval $$ aligns perfectly with the range of the observed data.

Practical Significance and Estimator Properties

The derivation of the maximum likelihood estimators for the uniform distribution provides a crucial lesson in statistical estimation theory: it demonstrates unequivocally that not all MLE problems can be solved through the routine process of setting derivatives to zero. For the uniform case, the maximization is intrinsically linked to observing and enforcing the stringent boundary conditions imposed by the sample data itself, resulting in estimators that are both mathematically sound and highly intuitive.

The major takeaway is the reliance on order statistics: the MLE for the lower bound **a** is the sample minimum, $\hat{a}_{MLE} = X_{(1)}$, and the MLE for the upper bound **b** is the sample maximum, $\hat{b}_{MLE} = X_{(n)}$. These estimators are vital in practical fields, such as quality control, manufacturing specifications, or physical modeling, where accurately defining the absolute limits of a process is paramount. While these estimators are intuitively appealing, it is statistically important to recognize that they are inherently biased. For example, the estimator $\hat{b}_{MLE}$ will, on average, underestimate the true maximum parameter b because the sample maximum $X_{(n)}$ must always be less than or equal to the true upper bound b. Similarly, $\hat{a}_{MLE}$ will tend to overestimate a.

Despite this bias in finite samples, these estimators are considered consistent. Consistency implies that as the sample size n tends toward infinity, the bias approaches zero, and the estimated values converge in probability to the true population parameters a and b. In conclusion, the methodology required for the uniform distribution underscores the paramount importance of the likelihood function's support constraints. Unlike distributions where the maximum occurs at an interior critical point, the maximum likelihood for the uniform distribution is found residing precisely on the boundary of the parameter space defined by the observed data, yielding the narrowest estimated range consistent with the observations.