Finding the best parameter estimators using the negative log-likelihood method is a fundamental concept in statistics and machine learning. This method is commonly used for maximum likelihood estimation (MLE), which involves finding the parameter values that maximize the likelihood function. Here are the steps to do this, explained clearly with an example and the relevant math:

\*\*Step 1: Understand the Problem\*\*

Before we dive into the details, let's set up the problem. Suppose you have a dataset with observations and you want to estimate the parameters of a statistical model (e.g., a probability distribution) that describes the data. The goal is to find the parameter values that make the observed data most likely.

\*\*Step 2: Define the Likelihood Function\*\*

The likelihood function, denoted as L(θ), represents the probability of observing your data given a set of parameters θ. Mathematically, it's the joint probability density or mass function of your data points:

L(θ) = f(x₁; θ) \* f(x₂; θ) \* ... \* f(xₙ; θ)

Where:

- L(θ): Likelihood function

- θ: Parameter vector (the values you want to estimate)

- x₁, x₂, ..., xₙ: Data points

- f(x; θ): Probability density or mass function of your model

\*\*Step 3: Take the Logarithm\*\*

To simplify the calculations and avoid numerical issues, it's common to work with the natural logarithm of the likelihood function, which gives us the log-likelihood function, denoted as ℓ(θ):

ℓ(θ) = ln[L(θ)]

\*\*Step 4: Find the Negative Log-Likelihood\*\*

To turn the problem into a minimization problem (which is often easier to solve), we negate the log-likelihood function:

-ℓ(θ) = -ln[L(θ)]

\*\*Step 5: Maximize the Negative Log-Likelihood\*\*

Now, your task is to find the values of θ that minimize -ℓ(θ). You can do this using optimization techniques, such as gradient descent or numerical solvers. Differentiating -ℓ(θ) with respect to θ and setting the derivative equal to zero will give you the maximum likelihood estimators (MLEs) for the parameters θ:

∇(-ℓ(θ)) = 0

Solving this equation will yield the parameter estimates that maximize the likelihood of observing your data.

\*\*Step 6: Example\*\*

Let's illustrate this process with a simple example: estimating the mean and standard deviation of a Gaussian (normal) distribution from a set of observations {x₁, x₂, ..., xₙ}.

- Likelihood function:

L(μ, σ) = (1 / (σ√(2π))) \* exp(-(x₁ - μ)² / (2σ²)) \* exp(-(x₂ - μ)² / (2σ²)) \* ... \* exp(-(xₙ - μ)² / (2σ²))

- Log-likelihood function:

ℓ(μ, σ) = -n \* ln(σ√(2π)) - Σ((xᵢ - μ)²) / (2σ²)

- Negative log-likelihood:

-ℓ(μ, σ) = n \* ln(σ√(2π)) + Σ((xᵢ - μ)²) / (2σ²)

Now, you would differentiate -ℓ(μ, σ) with respect to μ and σ, set the derivatives equal to zero, and solve for μ and σ to find the MLEs. The actual calculations can be complex depending on the model, but the process remains the same.

That's it! You've now learned how to find the best parameter estimators using the negative log-likelihood method. Remember that this method is applicable to a wide range of statistical models, not just the Gaussian distribution, and the details of the calculations may vary accordingly.