Examples Of Maximum Likelihood Estimators For Data That Comes From A Discrete Distribution, The Likelihood

Examples Of Maximum Likelihood Estimators For Data That Comes From A Discrete Distribution, The Likelihood

Understanding the foundational concepts of statistical inference is crucial for data scientists, statisticians, and researchers working with real-world data. One of the most powerful methods for parameter estimation in statistical models is the Maximum Likelihood Estimation (MLE). When dealing with data originating from discrete distributions—such as Bernoulli, Binomial, Poisson, or Geometric distributions—the likelihood function plays a pivotal role in deriving estimators that best explain the observed data.

This article provides a comprehensive overview of maximum likelihood estimators (MLEs) for data from discrete distributions. We will explore the concept of likelihood, delve into specific examples with detailed derivations, and highlight the significance of these estimators in statistical modeling. Whether you are a student, researcher, or practitioner, understanding these concepts will enhance your ability to analyze categorical and count data efficiently and accurately.

Understanding Likelihood and Maximum Likelihood Estimation

What Is Likelihood?

Likelihood is a fundamental concept in statistical inference that measures the plausibility of a parameter value given observed data. Unlike probability, which considers the likelihood of data given parameters, likelihood treats parameters as fixed and data as random. Mathematically, for a set of observed data points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and a parameter θ, the likelihood function is defined as:

$$
L(\theta; \mathbf{x}) = P(\mathbf{x} \mid \theta)
$$

For discrete distributions, this involves probability mass functions (pmfs).

What Is Maximum Likelihood Estimation?

Maximum Likelihood Estimation seeks the value of the parameter $\hat{\theta}$ that maximizes the likelihood function:

$$
\hat{\theta} = \arg \max_{\theta} L(\theta; \mathbf{x})
$$

In practice, it is often easier to work with the log-likelihood:

$$
\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x})
$$

because the logarithm converts products into sums, simplifying differentiation and optimization.

Examples of Discrete Distributions and Their MLEs

In this section, we explore common discrete distributions, derive their likelihood functions, and obtain the corresponding maximum likelihood estimators.

1. Bernoulli Distribution

Distribution Overview

The Bernoulli distribution models binary outcomes—success (1) or failure (0). It depends on a single parameter p, the probability of success:

$$
P(X = x) = p^x (1 - p)^{1 - x}, \quad x \in \{0, 1\}
$$

Likelihood Function

Given a sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of independent Bernoulli trials, the joint likelihood is:

$$
L(p; \mathbf{x}) = \prod_{i=1}^n p^{x_i} (1 - p)^{1 - x_i}
$$

The log-likelihood becomes:

$$
\ell(p; \mathbf{x}) = \sum_{i=1}^n \left[x_i \log p + (1 - x_i) \log (1 - p) \right]
$$

Deriving the MLE

Differentiate the log-likelihood with respect to p:

$$
\frac{\partial \ell(p)}{\partial p} = \sum_{i=1}^n \left[\frac{x_i}{p} - \frac{1 - x_i}{1 - p} \right]
$$

Set derivative to zero for maximization:

$$
\frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1 - x_i)}{1 - p} = 0
$$

Simplify:

$$
\frac{\sum x_i}{p} = \frac{n - \sum x_i}{1 - p}
$$

Cross-multiplied:

$$
(1 - p) \sum x_i = p (n - \sum x_i)
$$

Solve for p:

$$
\sum x_i - p \sum x_i = p n - p \sum x_i
$$
$$
\sum x_i = p n
$$

Therefore, the MLE for p is:

$$
\boxed{
\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i
}
$$

which is simply the sample proportion of successes.

---

2. Binomial Distribution

Distribution Overview

The Binomial distribution models the number of successes in n independent Bernoulli trials with success probability p:

$$
P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}
$$

where $k = 0, 1, \ldots, n$.

Likelihood Function

Given observed data k, the likelihood in terms of p:

$$
L(p; k) = \binom{n}{k} p^k (1 - p)^{n - k}
$$

The log-likelihood:

$$
\ell(p; k) = \log \binom{n}{k} + k \log p + (n - k) \log (1 - p)
$$

Since $\binom{n}{k}$ does not depend on p, it can be treated as a constant during maximization.

Deriving the MLE

Differentiate:

$$
\frac{\partial \ell(p)}{\partial p} = \frac{k}{p} - \frac{n - k}{1 - p}
$$

Set equal to zero:

$$
\frac{k}{p} = \frac{n - k}{1 - p}
$$

Solve for p:

$$
k (1 - p) = p (n - k)
$$

$$
k - k p = p n - p k
$$

Rearranged:

$$
k = p n
$$

Thus, the MLE for p:

$$
\boxed{
\hat{p} = \frac{k}{n}
}
$$

which matches the sample proportion of successes.

---

3. Poisson Distribution

Distribution Overview

The Poisson distribution models count data representing the number of events occurring in a fixed interval, with parameter λ indicating the average rate:

$$
P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \ldots
$$

Likelihood Function

Given data points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, the likelihood:

$$
L(\lambda; \mathbf{x}) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}
$$

The log-likelihood:

\[

$$
\ell(\lambda; \mathbf{x}) = -n \lambda + \left(\sum_{i=1}^n x_i\right) \log \lambda - \sum_{i=1}^n \log x_i!
$$

The sum over factorials is constant with respect to λ.

Deriving the MLE

Differentiate:

$$
\frac{\partial \ell(\lambda)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda}
$$

Set to zero:

$$
-n + \frac{\sum x_i}{\lambda} = 0
$$

Solve for λ:

$$
\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i
$$

Thus, the MLE for λ is the sample mean of the observed counts.

---

4. Geometric Distribution

Distribution Overview

The Geometric distribution models the number of Bernoulli trials needed to obtain the first success, with success probability p:

$$
P(X = x) = (1 - p)^{x - 1} p, \quad x = 1, 2, \ldots
$$

Likelihood Function

Given observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$:

$$
L(p; \mathbf{x}) = \prod_{i=1}^n (1 - p)^{x_i - 1} p = p^n (1 - p)^{\sum_{i=1}^n (x_i - 1)}
$$

\]

Log-likelihood:

\[
\ell(p; \mathbf{x}) = n \log p + \left(\

Frequently Asked Questions

What is a maximum likelihood estimator (MLE) in the context of discrete distributions?

A maximum likelihood estimator (MLE) is a parameter value that maximizes the likelihood function for a given set of discrete data, providing the most probable estimate of the underlying distribution parameters based on observed data.

Can you give an example of an MLE for a Bernoulli distribution?

Yes. For Bernoulli data, the MLE for the probability parameter p is the sample proportion of successes: $\hat{p}$= (number of successes) / (total number of trials).

What is the MLE for the parameter of a Poisson distribution based on observed counts?

The MLE for the Poisson distribution's rate parameter λ is the sample mean of the observed count data: $\hat{\lambda}$= (sum of observed counts) / (number of observations).

How does the likelihood function help in deriving MLEs for discrete distributions?

The likelihood function quantifies the probability of observed data given parameters. By finding the parameter values that maximize this function, we obtain the MLEs, which are the most probable parameters explaining the data.

Are MLEs for discrete distributions always unique and consistent?

Not always. While MLEs are often consistent and asymptotically normal, their uniqueness and properties depend on the distribution and data. Proper regularity conditions are required to ensure these properties hold.

Why is the likelihood function important in statistical inference for discrete data?

The likelihood function is central because it provides a basis for parameter estimation, hypothesis testing, and model comparison by quantifying how well different parameter values explain the observed discrete data.

Additional Resources

Maximum Likelihood Estimators (MLEs) for Discrete Distributions: An Expert Exploration

In the realm of statistical inference, Maximum Likelihood Estimation (MLE) stands as a cornerstone technique for deriving parameter estimates from observed data. When working with data that originate from discrete distributions—such as the binomial, Poisson, or geometric distributions—the application of MLE becomes both nuanced and powerful. This article provides an in-depth exploration of the examples of MLEs tailored for discrete data, emphasizing their formulation, interpretation, and practical significance.

Understanding the Foundations: Likelihood Function in Discrete Contexts

Before delving into specific estimators, it’s crucial to understand what the likelihood function represents in the context of discrete distributions.

Likelihood Function Defined:
For a set of observed data points $\mathbf{x} = (x_1, x_2, ..., x_n)$, and a discrete distribution characterized by a parameter θ, the likelihood function $L(\theta; \mathbf{x})$ measures the probability of observing this data given the parameter:

$$
L(\theta; \mathbf{x}) = \prod_{i=1}^n P(X = x_i |\theta)
$$

Since the data are discrete, the probability mass function (pmf) $P(X = x |\theta)$ assigns a probability to each possible outcome.

Maximizing the Likelihood:
The MLE of θ, denoted $\hat{\theta}$, is obtained by finding the value that maximizes $L(\theta; \mathbf{x})$. Often, because products can be unwieldy, it’s preferable to maximize the log-likelihood:

$$
\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}) = \sum_{i=1}^n \log P(X = x_i |\theta)
$$

Examples of MLEs for Common Discrete Distributions

The following sections examine key discrete distributions, their likelihood functions, and the derivation of their MLEs. Each example reflects real-world data analysis scenarios.

---

1. Binomial Distribution: Estimating the Probability of Success

Scenario:
Suppose you conduct a series of independent Bernoulli trials, such as flipping a coin multiple times, and observe the number of successes.

Distribution and Parameters:
The binomial distribution models the number of successes X in n trials:

$$
P(X = x | p) = \binom{n}{x} p^x (1 - p)^{n - x}
$$

where p is the probability of success in each trial.

Likelihood Function:
Given observed successes x, the likelihood as a function of p:

$$
L(p; x) = \binom{n}{x} p^x (1 - p)^{n - x}
$$

The binomial coefficient is constant with respect to p, so the log-likelihood simplifies to:

$$
\ell(p; x) = x \log p + (n - x) \log (1 - p) + \text{constant}
$$

Deriving the MLE:
To find $\hat{p}$, differentiate with respect to p:

$$
\frac{d \ell}{d p} = \frac{x}{p} - \frac{n - x}{1 - p}
$$

Set to zero for maximization:

$$
\frac{x}{p} - \frac{n - x}{1 - p} = 0
$$

Solve for p:

$$
x (1 - p) = (n - x) p
$$
$$
x - x p = n p - x p
$$

Simplify:

$$
x = n p
$$
$$
\boxed{
\hat{p} = \frac{x}{n}
}
$$

Interpretation:
The MLE for the success probability p is simply the observed proportion of successes, making intuitive sense and providing an unbiased, consistent estimator.

---

2. Poisson Distribution: Estimating the Rate Parameter λ

Scenario:
Imagine counting the number of events, such as the number of emails received per hour, which follows a Poisson process.

Distribution and Parameters:
The Poisson pmf:

$$
P(X = x |\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}
$$

where $\lambda > 0$ is the average rate.

Likelihood Function:
Given data $\mathbf{x} = (x_1, x_2, ..., x_n)$ (independent observations):

$$
L(\lambda; \mathbf{x}) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}
$$

Log-likelihood:

$$
\ell(\lambda; \mathbf{x}) = \left(\sum_{i=1}^n x_i \right) \log \lambda - n \lambda - \sum_{i=1}^n \log x_i!
$$

The constant terms $\sum_{i=1}^n \log x_i!$ are independent of λ.

Derivation of the MLE:

Differentiate:

$$
\frac{d \ell}{d \lambda} = \frac{\sum x_i}{\lambda} - n
$$

Set equal to zero:

$$
\frac{\sum x_i}{\lambda} - n = 0
$$

Solve for λ:

$$
\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}
$$

Interpretation:
The MLE aligns with the sample mean, which is also the maximum likelihood estimator for the Poisson rate.

---

3. Geometric Distribution: Estimating the Success Probability p

Scenario:
Consider the number of trials until the first success in a sequence of Bernoulli trials, such as the number of coin flips until a head appears.

Distribution and Parameters:
The geometric pmf (assuming the number of trials until first success):

$$
P(X = x \mid p) = (1 - p)^{x - 1} p, \quad x = 1, 2, 3, ...
$$

where p is the success probability per trial.

Likelihood Function:

$$
L(p; \mathbf{x}) = \prod_{i=1}^n p (1 - p)^{x_i - 1} = p^n (1 - p)^{\sum_{i=1}^n (x_i - 1)}
$$

Log-likelihood:

$$
\ell(p; \mathbf{x}) = n \log p + \left(\sum_{i=1}^n x_i - n \right) \log (1 - p)
$$

Deriving the MLE:

Differentiate:

$$
\frac{d \ell}{d p} = \frac{n}{p} - \frac{\sum x_i - n}{1 - p}
$$

Set to zero:

$$
\frac{n}{p} = \frac{\sum x_i - n}{1 - p}
$$

Solve for p:

$$
n(1 - p) = p (\sum x_i - n)
$$

\[

$$
n - n p = p \sum x_i - n p
$$

Bring all to one side:

$$
n = p \sum x_i
$$

$$
\boxed{
\hat{p} = \frac{n}{\sum_{i=1}^n x_i}
}
$$

Interpretation:
The MLE for the success probability is the reciprocal of the average observed number of trials until success.

Practical Considerations and Common Pitfalls

While the derivations above are straightforward, practitioners should be aware of several key points:

- Boundary Solutions:
For some distributions, the MLE might lie on the boundary of the parameter space, leading to estimators like zero or one, which can be problematic.

- Sample Size and Consistency:
MLEs tend to be consistent and asymptotically normal, but small sample sizes can lead to biased or unstable estimates.

- Discrete Data Challenges:
Discrete distributions often involve integer counts, which may limit the variability of estimators and complicate confidence interval construction.

- Model Assumptions:
The validity of MLEs depends on correct model specification. Using a binomial model for data not truly binomial can lead to misleading estimates.

Extensions and Modern Applications

The concepts of MLE extend beyond these classical examples, especially with

```
the advent of computational tools:

- Composite Models:
Combining discrete distributions with covariates (e.g., Poisson regression)
allows modeling count data with predictors.

- Bayesian Alternatives:
While MLE provides point estimates, Bayesian methods incorporate prior
knowledge, which can be vital in sparse data scenarios.

- Simulation-Based Estimation:
For complex models, methods like Markov Chain Monte Carlo (MCMC) facilitate
```

Examples Of Maximum Likelihood Estimators For Data That Comes From A Discrete Distribution The Likelihood

Examples Of Maximum Likelihood Estimators For Data That Comes From A Discrete Distribution The Likelihood

Related Articles

- Please Answer 10-13Prog Lisard Sparrow Crocodile Dog Cat Salmon 10. Which Of The Following Is Most Closely
- How Did Lincoln's Views On His Ability To End Slavery Evolve Over Time, And To What Extent Is He Justified
- How Does A New Star Form? Place Each Stage In Order From The Beginning Of Star Formation To Star Formation

Back to Home