The Republic of Iraq
Ministry of Higher Education and Practical Research
University of Babylon
College of Education for Pure Sciences
Mathematics department

Maximum Likelihood Estimation

Research Submitted to University of Babylon / College of Education for Pure Sciences / Mathematic Department as Part of the Requirements for The Degree of B.Sc. in Mathematical Science

Fatima Al-Zahra Muhammad Juma Al-Saidi

Supervised By
Dr. Ali Hussein Mahmood Al-Obaidi

2024 AD 1445 AH

This research paper contains a comprehensive overview of the topic of the research title (Maximum Likelihood Estimation)

بِسْمِ اللهِ الرَّحْمنِ الرَّحِيمِ

وَيَسْأَلُونَكَ عَنِ الرُّوحِ ۖ قُلِ الرُّوحُ مِنْ أَمْرِ رَبِّي وَمَا أُوتِيتُم مِّنَ الْعِلْمِ إِلَّا قَلِيلًا ٨٥

صدق الله العظيم

سورة الاسراء الاية ﴿٨٥﴾

<u>Dedication</u>: -

To those who taught me and still .... my **father**, may God have mercy on him

To whom her sincere prayers were the secret of my success ... my dear **mother**

my family

To everyone who helped me during my studies,

Fatima Al-Zahra Al-Saidi

Acknowledgements

In the Name of God Most Gracious Most Merciful

I thank God Almighty for His grace, as He enabled me to accomplish this work thanks to Him.

Then I thank all who helped me during this period, and especially my supervisor. Because he chose the topic of research and her help and great efforts to accomplish this work. My thanks and appreciation to the Deanship of the College of Education for Pure Sciences / University of Babylon and to the respected head and professors and staff member of the Mathematics department

Contents

Chapter 1

Introduction

1-1 Introduction to Maximum Likelihood Estimation

Maximum likelihood is a widely used technique for estimation with applications in many areas including time series modeling, panel data, discrete data, and even machine learning.

Maximum likelihood estimation is a statistical method for estimating the parameters of a model. In maximum likelihood estimation, the parameters are chosen to maximize the likelihood that the assumed model results in the observed data.

This implies that to implement maximum likelihood estimation we must:

1. Assume a model, also known as a data-generating process, for our data.
2. Be able to derive the likelihood function for our data, given our assumed model (we will discuss this more later).

Once the likelihood function is derived, maximum likelihood estimation is nothing more than a simple optimization problem.

At this point, you may be wondering why you should pick maximum likelihood estimation over other methods such as least squares regression or the generalized method of moments. The reality is that we shouldn't always choose maximum likelihood estimation. Like any estimation technique, maximum likelihood estimation has advantages and disadvantages.

1-2 What is the Likelihood Function

Maximum likelihood estimation hinges on the derivation of the likelihood function. For this reason, it is important to have a good understanding of what the likelihood function is and where it comes from.

Let's start with a very simple case where we have one series with 10 independent observations: 5, 0, 1, 1, 0, 3, 2, 3, 4, 1.

1-3 The Likelihood Function

Let $\boldsymbol{X}1,\ldots\ldots \boldsymbol{X}_n$ be an iid sample with pdf $f(xi;\theta)$, where θ is a $(k \times 1)$ vector

of parameters that characterize $f(xi;\theta)$, Let $X_i \sim N(\mu, \sigma^2)$ then

$$\begin{aligned} f(x_i;\theta) &= (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\ \theta &= (\mu, \sigma^2)' \end{aligned}$$

The joint density of the sample is, by independence, equal to the product of the marginal densities

$$f(x_1, \ldots, x_n; \theta) = f(x_1; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

1-4 Advantages of Maximum Likelihood Estimation

There are many advantages of maximum likelihood estimation:

- If the model is correctly assumed, the maximum likelihood estimator is the most efficient.
- It provides a consistent but flexible approach which makes it suitable for a wide variety of applications, including cases where assumptions of other models are violated.
- It results in unbiased estimates in larger samples.

1-5 Disadvantages of Maximum Likelihood Estimation

- It relies on the assumption of a model and the derivation of the likelihood function which is not always easy.
- Like other optimization problems, maximum likelihood estimation can be sensitive to the choice of starting values.
- Depending on the complexity of the likelihood function, the numerical estimation can be computationally expensive.
- Estimates can be biased in small samples.

1-6 The Maximum Likelihood Estimator

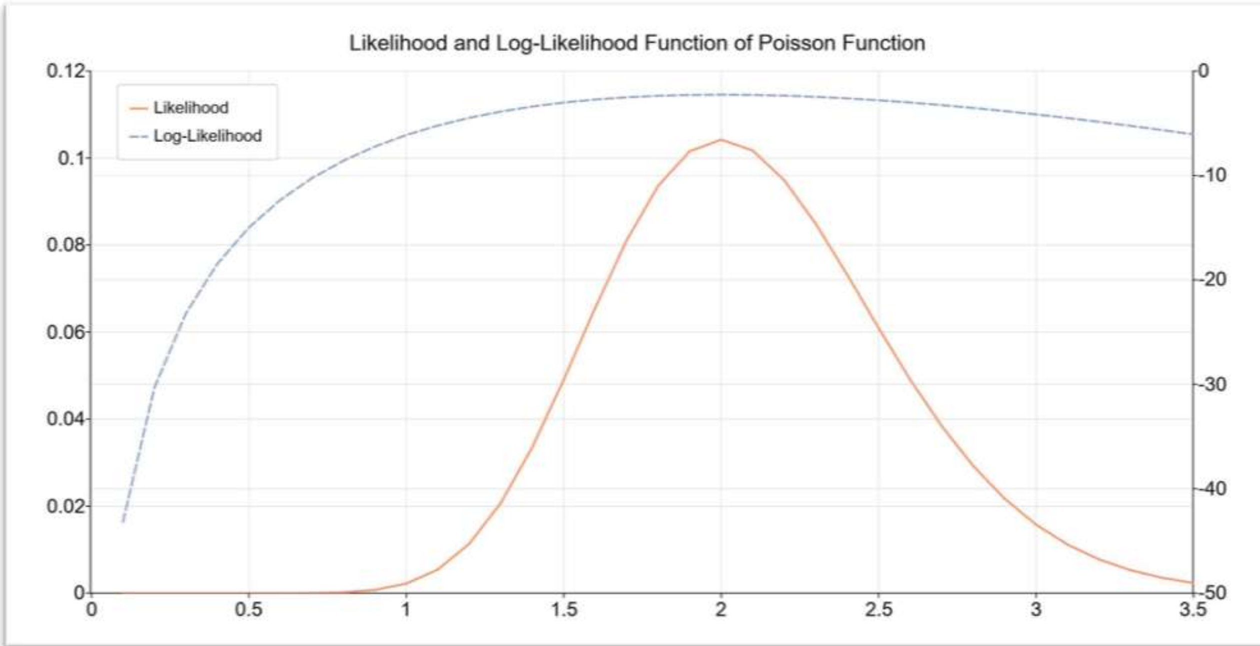


Figure (1-1) position of likelihood functions

Chapter 2

Properties, Applications.

2-1 Properties of Maximum Likelihood Estimators

Certainly! Let's delve into the **properties of Maximum Likelihood Estimators (MLE)**, a widely used statistical estimation method. MLE aims to estimate the parameters of a statistical model based on observed data. Here are the key properties:

1. **Consistency**:
    - MLE is a **consistent estimator**. As the sample size increases, the estimates obtained by MLE converge in probability to their true values, provided certain conditions are met(1).
    - In other words, as we collect more data, the MLE becomes more reliable and approaches the actual parameter values.
2. **Asymptotic Normality**:
    - MLE follows the **asymptotic normality** property. As the sample size grows, the distribution of the MLE converges to a normal distribution.
    - This property is crucial for statistical inference, as it allows us to use normal-based approximations for hypothesis testing and confidence intervals(2).
3. **Efficiency**:
    - An **efficient estimator** achieves equality with the **Cramér–Rao Lower Bound (CRLB)**. The CRLB provides a lower bound on the variance of any unbiased estimator(3).
    - In simple terms, an efficient estimator has the smallest possible variance among all unbiased estimators(4).
    - For MLE, the variance cannot be lower than the CRLB, making it an efficient choice(5).
4. **Invariance**:
    - MLE exhibits an **invariance property**. If we transform the parameter (e.g., take the logarithm), the MLE of the transformed parameter remains the MLE of the original parameter.
    - This property ensures that MLE is robust to changes in parameterization(6).
5. **Unbiasedness (in large samples)**:
    - Although MLE may be biased in small samples, it becomes **asymptotically unbiased** as the sample size increases.

- In large samples, MLE tends to have negligible bias, making it a desirable property(7).

6. **Unique (in some cases)**:
   - MLE is often unique for a given model and dataset. This uniqueness simplifies the estimation process.
   - However, uniqueness depends on the specific statistical model and the likelihood function.

In summary, Maximum Likelihood Estimators combine consistency, asymptotic normality, efficiency, invariance, and often uniqueness, making them powerful tools for parameter estimation in statistical modeling.

2-2 Applications of Maximum Likelihood Estimation

2-2-1 find the likelihood function

Example 2-1:

For the following random samples, find the likelihood function:

- Xi∼Binomial(3,θ)$Xi{\sim}Binomial(3,\theta)$, and we have observed (x1,x2,x3,x4)=(1,3,2,2)$(x1,x2,x3,x4)$=(1,3,2,2).
- Xi∼Binomial(3,θ)$Xi{\sim}Binomial(3,\theta)$, and we have observed (x1,x2,x3,x4)=(1,3,2,2)$(x1,x2,x3,x4)$=(1,3,2,2).

Solution:

Remember that when we have a random sample, X_i's are i.i.d., so we can obtain the joint PMF and PDF by multiplying the marginal (individual) PMFs and PDFs.

1. If $X_i \sim Binomial(3,\theta)$, then

$$P_{X_i}(x;\theta) = \binom{3}{x}\theta^x(1-\theta)^{3-x}$$

Thus,

$$L(x_1,x_2,x_3,x_4;\theta) = P_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4;\theta)$$

$$= \binom{3}{x_1}\binom{3}{x_2}\binom{3}{x_3}\binom{3}{x_4}\theta^{x_1+x_2+x_3+x_4}(1-\theta)^{12-(x_1+x_2+x_3+x_4)}.$$

Since we have observed $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$, we have

$$L(1,3,2,2;\theta) = \binom{3}{1}\binom{3}{3}\binom{3}{2}\binom{3}{2}\theta^8(1-\theta)^4$$
$$= 27 \qquad \theta^8(1-\theta)^4.$$

2. If $X_i \sim Exponential(\theta)$, then

$$f_{X_i}(x;\theta) = \theta e^{-\theta x}u(x),$$

Now that we have defined the likelihood function, we are ready to define maximum likelihood estimation. Let X_1X1, X_2X2, X_3X3, ……, X_nXn be a random sample from a distribution with a parameter $\theta\theta$. Suppose that we have observed $X_1=x_1X1=x1$, $X_2=x_2X2=x2$, ⋯⋯, $X_n=x_nXn=xn$. The maximum likelihood estimate of $\theta\theta$, shown by $\hat{\theta}_{ML}\theta\text{^}ML$ is the value that maximizes the likelihood function

$$L(x_1,x_2,\cdots,x_n;\theta).L(x1,x2,\cdots,xn;\theta).$$

Figure 2.1 illustrates finding the maximum likelihood estimate as the maximizing value of $\theta\theta$ for the likelihood function. There are two cases shown in the figure: In the first graph, $\theta\theta$ is a discrete-valued parameter, such as the one in Example 2.1 . In the second one, $\theta\theta$ is a continuous-valued parameter, such as the ones in Example 8.8. In both cases, the maximum likelihood estimate of $\theta\theta$ is the value that maximizes the likelihood function.

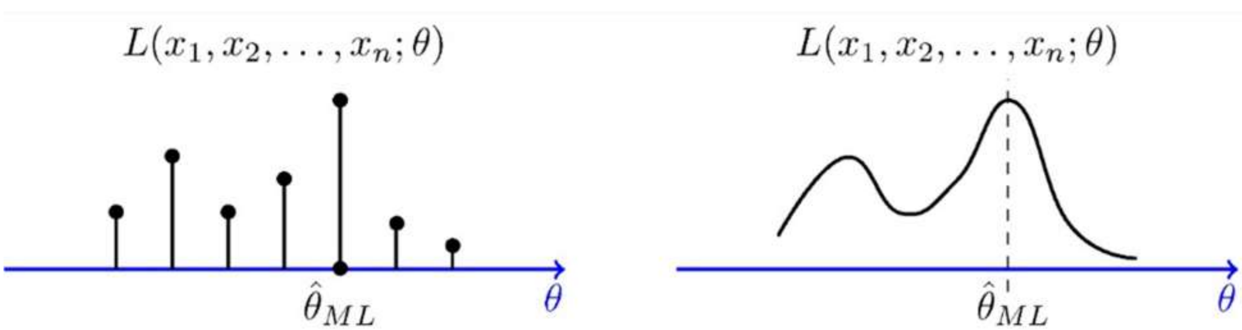


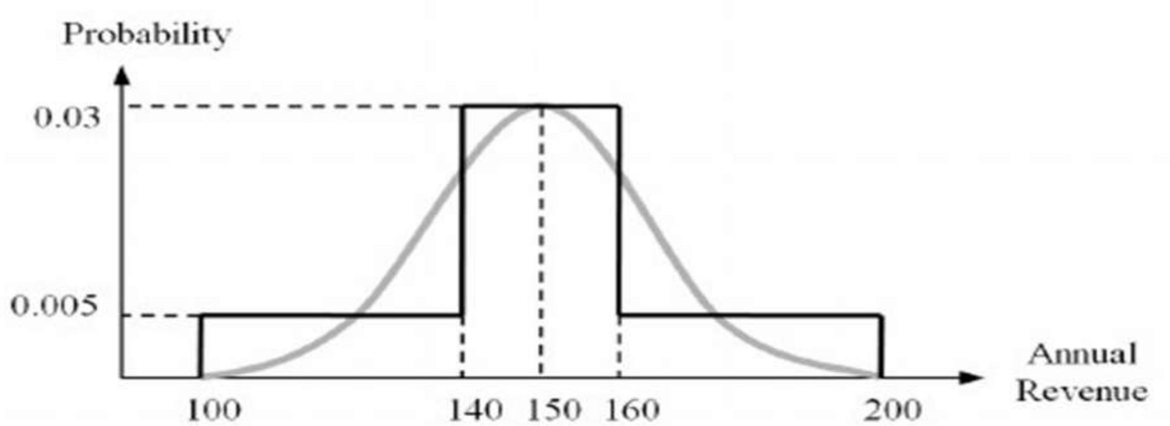
Figure 2.1 - The maximum likelihood estimate for $\theta\theta$.

2-2-2 piecewise distributions and find for me the maximum potential estimator.MLE.

In statistics problems, the maximum potential estimator. We will Choose five examples of piecewise distributions and find for me the maximum potential estimator.MLE.

Let's delve into piecewise distributions and explore some examples. Then, I'll explain how to find the maximum likelihood estimator (MLE) for these distributions.

A. Piecewise Uniform Distribution(8):

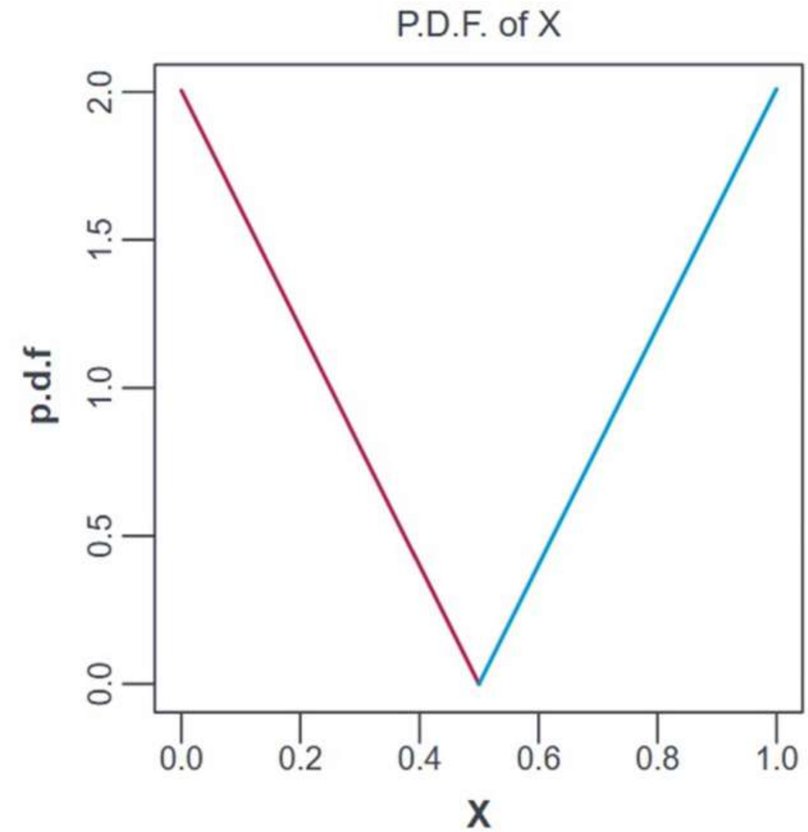


Piecewise-uniform distribution for an uncertain revenue, which is between 100 and 140 with 20% probability, between 140 and 160 with 60% probability, and between 160 and 200 with 20% probability. This distribution is an approximation of the normal distribution shown by the grey line.

- **Example 2-2** Some distributions are split into parts. They are not necessarily continuous, but they are continuous over particular intervals. These types of distributions are known as Piecewise distributions. Below is an example of this type of distribution

$$f(x) = \begin{cases} 2-4x, & x < 1/2 \\ 4x-2, & x \geq 1/2 \end{cases}$$

for $0<x<1$. The pdf of X is shown below.



The first step is to show this is a valid pdf. To show it is a valid pdf, we have to show the following:

The first step is to show this is a valid pdf. To show it is a valid pdf, we have to show the following:

1. $f(x) > 0$. We can see that $f(x)$ is greater than or equal to 0 for all values of X.
2. $\int_S f(x)dx = 1$.

$$\int_0^{1/2} 2 - 4xdx + \int_{1/2}^1 4x - 2dx$$
$$= 2\left(\frac{1}{2}\right) - 2\left(\frac{1}{4}\right) + 2 - 2 - \left[2\left(\frac{1}{4}\right) - 1\right]$$
$$= 1 - \left(\frac{1}{2}\right) + 2 - 2 - \left(\frac{1}{2}\right) + 1 = 2 - 1 = 1$$

3. If $(a, b) \subset S$, then $P(a < X < b) = \int_a^b f(x)dx$. Lets find the probability that X is between 0 and $2/3$.

$$P(X < 2/3) = \int_0^{1/2} 2 - 4xdx + \int_{1/2}^{2/3} 4x - 2dx = \frac{5}{9}$$

The next step is to know how to find expectations of piecewise distributions. If we know how to do this, we can find the mean, variance, etc of a random variable with this type of distribution. Suppose we want to find the expected value, $E(X)$.

$$E(X) = \int_0^{1/2} x(2-4x)dx + \int_{1/2}^1 x(4x-2)dx$$
$$= \left(x^2 - \frac{4}{3}x^3\right)\Big|_0^{1/2} + \left(\frac{4}{3}x^3 - x^2\right)\Big|_{1/2}^1 = \frac{1}{2}$$

The variance and other expectations can be found similarly.

The final step is to find the cumulative distribution function. cdf. Recall the cdf of X is **$FX(t)=P(X\leq t)$.** Therefore, for **t<12,** we have

$$F_X(t) = \int_0^t 2-4xdx = 2x - x^2|_0^t = 2t - 2t^2$$

and for $t \geq \frac{1}{2}$ we have

$$F_X(t) = \int_0^{1/2} 2-4xdx + \int_{1/2}^t 4x-2dx = \frac{1}{2} + (2x^2-2x)|_{1/2}^t$$
$$= 2t^2 - 2t + 1$$

Thus, the cdf of X is

$$F_X(t) = \begin{cases} 2t-2t^2 & 0 < t < 1/2 \\ 2t^2-2t+1 & 1/2 \leq t < 1 \end{cases}$$

- The MLE for the parameter (in this case, there's no parameter) would be the maximum value of the likelihood function, which is 1.

B. Piecewise Linear Distribution(9):

- **Example 2-3** Fixed Number of Linear Segments

To find the piecewise-linear approximation with a fixed number T of linear segments that minimizes the squared error, the following problem is solved:

$$\min_{I} E = \sum_{t=1}^{T} E_t = \sum_{t=1}^{T} \sum_{k=i(t)}^{i(t+1)} \left(y_k - a_t x_k - b_t\right)^2 . \tag{1}$$

Models and Algorithms for Optimal Piecewise-Linear Function Approximation

$$F\left(j,t\right) = \min_{t \leq i \leq j} F\left(i, t-1\right) + E\left(i, j\right), \tag{2}$$

- Consider a random variable Y with the following:

$$f(y) = \begin{cases} 2y, & \text{if } 0 \leq y < 1 \\ 0, & \text{otherwise} \end{cases}$$

- The MLE for the parameter (again, no parameter here) would be the maximum value of the likelihood function, which occurs at $y = 1$.

C. Mixture Distribution(10)(11):

- **Example 2-4** Let's say we have a mixture of two distributions, Z_1 and Z_2, with pdfs $f_1(z)$ and $f_2(z)$, respectively. The overall pdf is given by:

$$f(z) = a f_1(z) + (1 - a) f_2(z)$$

where $0 \leq a \leq 1$.

- The MLE for the parameter a would be the value that maximizes the likelihood function based on the observed data.

D. Piecewise Exponential Distribution(12):

- Suppose we have a random variable T representing time until an event occurs. The pdf is:

$$
f(t) = \begin{cases}
\lambda_1 e^{-\lambda_1 t}, & \text{if } 0 \leq t < 1 \\
\lambda_2 e^{-\lambda_2 t}, & \text{if } t \geq 1
\end{cases}
$$

- Here, λ_1 and λ_2 are parameters. The MLE would involve maximizing the likelihood function with respect to these parameters.

E. Piecewise Normal Distribution(13):

- **Example 2-5** Consider a random variable X with the following:

$$
f(x) = \begin{cases}
\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x - \mu_1)^2}{2\sigma_1^2}}, & \text{if } x < 0 \\
\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x - \mu_2)^2}{2\sigma_2^2}}, & \text{if } x \geq 0
\end{cases}
$$

where μ_1, μ_2, σ_1, and σ_2 are parameters.

- The MLE involves finding the parameter values that maximize the likelihood function based on the observed data.

Remember that the MLE aims to find parameter values that make the observed data most probable under the assumed distribution.

2-2-3 continuous distribution in maximum likelihood estimation in statistic:

And her are the five example of continuous distribution in maximum likelihood estimation in statistic:

1. **Normal Distribution (Gaussian Distribution)**(14)**:**
   - The normal distribution is widely used in statistics due to its symmetry and applicability to various real-world scenarios.
   - MLE can be used to estimate the mean (μ) and variance (σ^2) of a normally distributed random variable based on observed data.
   - MLE can be used to estimate the mean (μ) and variance (σ^2) of a normally distributed random variable based on observed data.

Normal distribution is the most important distribution in probability and statistics. It has extremely rich structures and connections with other distributions. A random variable X is Gaussian with mean μ and variable σ^2, denoted by N (μ, σ^2

), if its pdf i

$$f_X(x) = 1 \surd\ 2\pi\sigma\ e\ -(x-\mu)\ 2/2\sigma\ 2\ ,\ x \in R.$$

In particular, if μ = 0 and σ = 1, we say X is standard Gaussian. One can verify

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2}\,\mathrm{d}x = 1$$

by using the trick from multivariate calculus. Let's verify E X = 0 and Var(X) = 1.

$$\mathbb{E}\,X = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-x^2/2}\,\mathrm{d}x = 0$$

since $xe^{-x\ 2/2}$ is an odd function. How about E^{X2} ?

$$\begin{aligned}\mathbb{E}\,X^2 &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-x^2/2}\,\mathrm{d}x \\ &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\,\mathrm{d}e^{-x^2/2} \\ &= -\frac{1}{\sqrt{2\pi}} x e^{-x^2/2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2}\,\mathrm{d}x = 1.\end{aligned}$$

Gaussian random variable is linearly invariant: suppose X ∼ N (μ, σ2), then aX + b is still Gaussian with mean aμ + b and variance a 2σ 2 , i.e., N (aμ + b, a2σ 2)

$$\mathbb{E}(aX+b) = a\mu + b, \quad \mathrm{Var}(aX+b) = \mathrm{Var}(aX) = a^2\,\mathrm{Var}(X) = a^2\sigma^2.$$

Moreover, suppose X ~ N (μX, σ2 X) and Y ~ N (μY , σ2 Y) are two independent random variables, then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

This can be extended to the sum of n independent Gaussian random variables. For example,

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(0, n)$$

if Xi ~ N (0, 1) are i.i.d. random variables.

2. **Exponential Distribution**(15)**:**

- The exponential distribution models the time between events in a Poisson process (e.g., time between arrivals at a service center).
- MLE can be used to estimate the rate parameter (λ) of the exponential distribution based on observed inter-arrival times.
- Applications include reliability analysis and queueing theory.

Exponential distribution: X has an exponential distribution with parameter β, i.e., E(β) if

$$f(x) = \beta^{-1} e^{-x/\beta}, \quad x \geq 0$$

where β > 0.

$$\mathbb{P}(X \geq t + s | X \geq t) = \mathbb{P}(X \geq s), \quad \forall s \geq 0.$$

Example 2-6

Suppose we are interested in survival times, $T_1, T_2, \ldots, T_n$, but we don't observe T_i for all i. Instead, we observe

(U_i, δ_i), with $U_i = T_i$ and $\delta_i = 1$ if T_i is actually observed, and

(U_i, δ_i), with $U_i < T_i$ and $\delta_i = 0$ if all we know is that T_i is longer than U_i.

When $T_i > U_i, U_i$ is called the ***censoring time***.[7]

If the censoring times are all known constants, then the likelihood is

$$L = \prod_{i,\delta_i=1} f(u_i) \prod_{i,\delta_i=0} S(u_i)$$

where $f(u_i)$ = the probability density function evaluated at u_i,

and $S(u_i)$ = the probability that T_i is greater than u_i, called the *survival function*.

This can be simplified by defining the hazard function, the instantaneous force of mortality, as

$$\lambda(u) = f(u)/S(u)$$

so

$$f(u) = \lambda(u)S(u).$$

Then

$$L = \prod_i \lambda(u_i)^{\delta_i} S(u_i).$$

For the exponential distribution, this becomes even simpler, because the hazard rate, λ, is constant, and $S(u) = \exp(-\lambda u)$. Then:

$$L(\lambda) = \lambda^k \exp(-\lambda \sum u_i),$$

where $k = \sum \delta_i$.

From this we easily compute $\hat{\lambda}$, the maximum likelihood estimate (MLE) of λ, as follows:

$$l(\lambda) = \log(L(\lambda)) = k\log(\lambda) - \lambda \sum u_i.$$

Then

$$dl/d\lambda = k/\lambda - \sum u_i.$$

We set this to 0 and solve for λ to get:

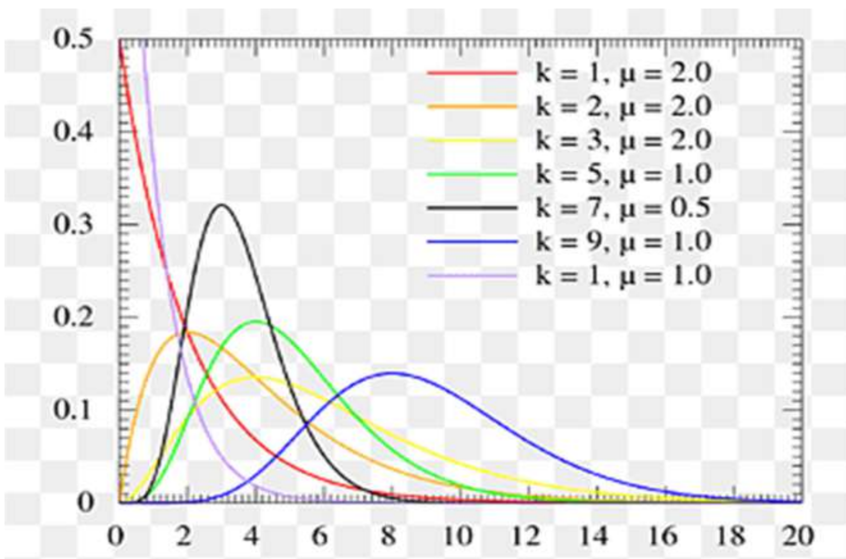
$$\hat{\lambda} = k/\sum u_i.$$

Equivalently, the mean time to failure is:

$$1/\hat{\lambda} = \sum u_i/k.$$

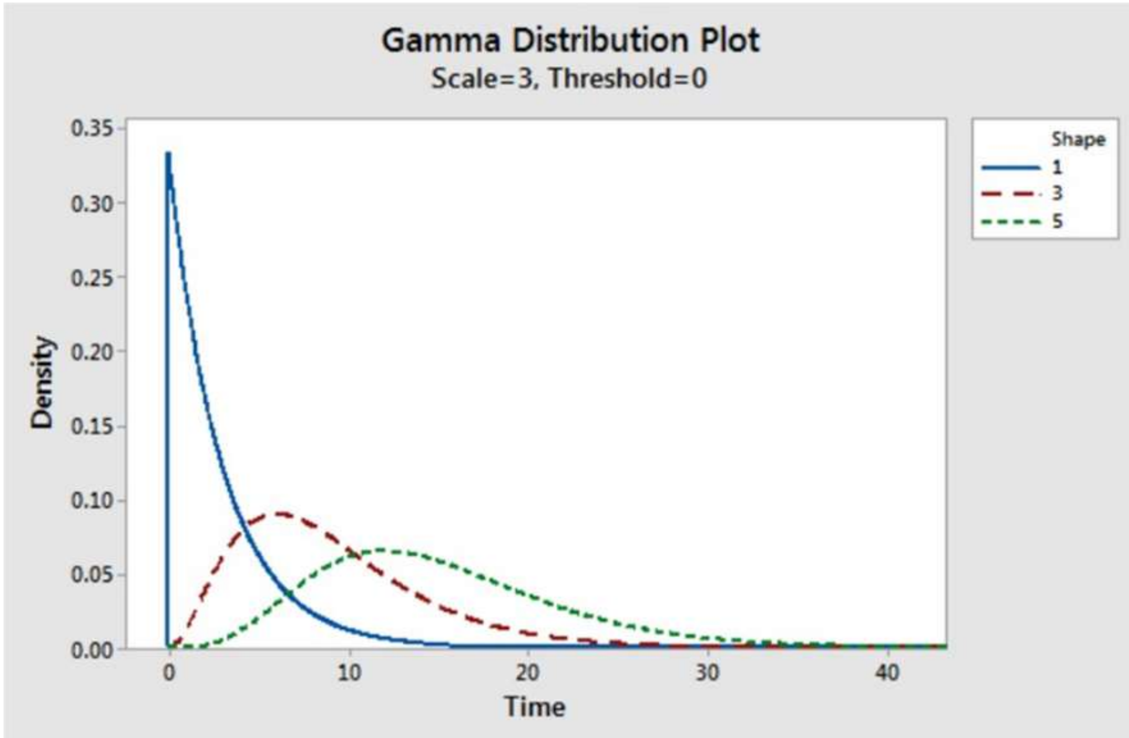
This differs from the standard MLE for the exponential distribution in that the any censored observations are considered only in the numerator.

3. **Weibull Distribution**(16)**:**



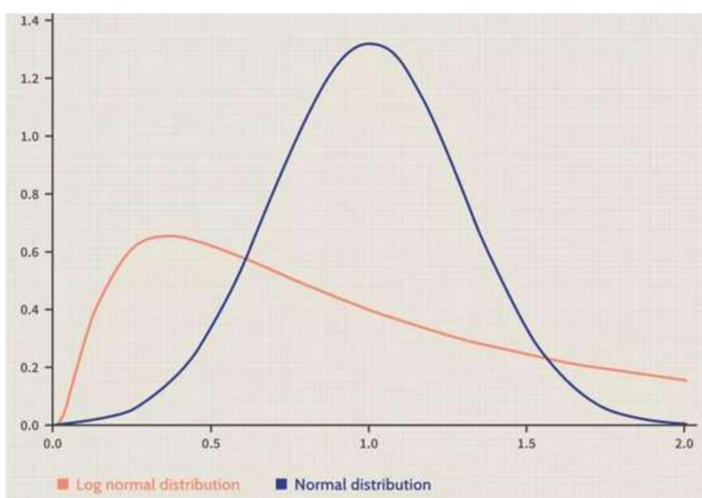
- The Weibull distribution is commonly used to model the lifetime of products or systems.
- MLE helps estimate the shape parameter (k) and scale parameter (λ) of the Weibull distribution.
- It is useful for reliability engineering and survival analysis.

4. **Gamma Distribution**(17)**:**



- The gamma distribution is versatile and can model various types of continuous data (e.g., waiting times, rainfall, insurance claims).
- MLE is used to estimate the shape parameter (α) and scale parameter (β) of the gamma distribution.
- It has applications in finance, hydrology, and quality control.

5. **Log-Normal Distribution**(18):



- The log-normal distribution is often used for modeling positive continuous data that follow a skewed pattern.
- MLE helps estimate the parameters of the underlying normal distribution (mean and standard deviation) after transforming the data to the log scale.
- Examples include modeling stock returns, income, and particle sizes.

We muest Remember that MLE aims to find parameter values that maximize the likelihood function, which represents the probability of observing the given data under a specific distribution.

On the other hand, MLE is not as widely recognized among modelers in psychology, but it is a standard approach to parameter estimation and inference in statistics. MLE has many optimal properties in estimation: sufficiency (complete information about the parameter of interest contained in its MLE estimator); consistency (true parameter value that generated the data recovered asymptotically, i.e. for data of sufficiently large samples); efficiency (lowest-possible variance of parameter estimates achieved asymptotically); and parameterization invariance (same MLE solution obtained independent of the parametrization used). In contrast, no such things can be said about LSE. As such, most statisticians would not view LSE as a general method for parameter estimation, but rather as an approach that is primarily used with linear regression models. Further, many of the inference methods in statistics are developed based on MLE. For example, MLE is a prerequisite for the chi-square test, the G-square test, Bayesian methods, inference with missing data, modeling of random effects, and many model selection criteria such as the Akaike information criterion (Akaike, 1973) and the Bayesian information criteria (Schwarz, 1978).

In this tutorial paper, I introduce the maximum likelihood estimation method for mathematical modeling. The paper is written for researchers who are primarily involved in empirical work and publish in experimental journals (e.g. Journal of Experimental Psychology) but do modeling. The paper is intended to serve as a stepping stone for the modeler to move beyond the current practice of using LSE to more informed modeling analyses, thereby expanding his or her repertoire of statistical instruments, especially in non-linear modeling. The purpose of the paper is to provide a good conceptual understanding of the method with concrete examples. For in-depth, technically more rigorous treatment of the topic, the reader is directed to other sources(19) .

2-4 conclusion and recommendations

The purpose of this paper is to provide a good conceptual explanation of the method with illustrative examples so the reader can have a grasp of some of the basic principles.

The relationship between the gamma distribution and the piecewise distributions and find for me the maximum potential estimator.MLE.shows its importance in many fields, including statistics, probability, medical science, engineering, and others. Here are some benefits of this relationship.

The Weibull distribution is commonly used to model the lifetime of products or systems, MLE helps estimate the shape parameter (k) and scale parameter (λ) of the Weibull distribution.

References:

1. Pan J-X, Fang K-T, Pan J-X, Fang K-T. Maximum likelihood estimation. Growth curve Model Stat diagnostics. 2002;77–158.
2. Xie M, Singh K. Confidence distribution, the frequentist distribution estimator of a parameter: A review. Int Stat Rev. 2013;81(1):3–39.
3. Keramat A, Ghidaoui MS, Wang X, Louati M. Cramer-Rao lower bound for performance analysis of leak detection. J Hydraul Eng. 2019;145(6):4019018.
4. Huang J, Gu K, Wang Y, Zhang T, Liang J, Luo S. Connectivity-based localization in ultra-dense networks: CRLB, theoretical variance, and MLE. IEEE Access. 2020;8:35136–49.
5. Vankayalapati N, Kay S, Ding Q. TDOA based direct positioning maximum likelihood estimator and the Cramer-Rao bound. IEEE Trans Aerosp Electron Syst. 2014;50(3):1616–35.
6. He M, Chen J. Strong consistency of the MLE under two-parameter Gamma mixture models with a structural scale parameter. Adv Data Anal Classif. 2022;1–30.
7. Rainey C, McCaskey K. Estimating logit models with small samples. Polit Sci Res Methods. 2021;9(3):549–64.
8. Hasuike T, Katagiri H, Tsubaki H. A constructing algorithm for appropriate piecewise linear membership function based on statistics and information theory. Procedia Comput Sci. 2015;60:994–1003.
9. Chan S-O, Diakonikolas I, Servedio RA, Sun X. Efficient density estimation via piecewise polynomial approximation. In: Proceedings of the forty-sixth annual ACM symposium on Theory of computing. 2014. p. 604–13.
10. Arevalillo JM, Navarro J. Assessment of extreme records in

environmental data through the study of stochastic orders for scale mixtures of skew normal vectors. Environ Ecol Stat. 2024;1–29.

11. Arain QA, Deng Z, Memon I, Zubedi A, Mangi FA. RETRACTED ARTICLE: Location Privacy with Dynamic Pseudonym-Based Multiple Mix-Zones Generation over Road Networks. Wirel Pers Commun. 2017;97:3645–71.
12. Snyder DL, Miller MI. Random point processes in time and space. Springer Science & Business Media; 2012.
13. Milios D. Probability distributions as program variables. Edinburgh, UK Sch Informatics, Univ Edinburgh, MS Thesis. 2009;
14. Goodman NR. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). Ann Math Stat. 1963;34(1):152–77.
15. Arshadi L, Jahangir AH. An empirical study on TCP flow interarrival time distribution for normal and anomalous traffic. Int J Commun Syst. 2017;30(1):e2881.
16. Cheng Y, Sheu S. Robust estimation for weibull distribution in partially accelerated life tests with early failures. Qual Reliab Eng Int. 2016;32(7):2207–16.
17. Junmei Z, Liqin L. Estimating parameters of the gamma distribution easily and efficiently. Commun Stat Methods. 2023;1–9.
18. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001;20(4):461–94.
19. Baroni M, Evert S. Statistical methods for corpus exploitation. Corpus Linguist An Int Handb. 2009;2:777–803.