

# Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu  
chenle02@gmail.com

Emory University  
Atlanta, GA

Last updated on Spring 2021  
Last compiled on January 15, 2023

2021 Spring

Creative Commons License  
(CC By-NC-SA)

# Chapter 5. Estimation

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 5.6 Sufficient Estimators

§ 5.7 Consistency

§ 5.8 Bayesian Estimation

# Plan

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 5.6 Sufficient Estimators

§ 5.7 Consistency

§ 5.8 Bayesian Estimation

# Chapter 5. Estimation

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 5.6 Sufficient Estimators

§ 5.7 Consistency

§ 5.8 Bayesian Estimation

**Two methods** for estimating parameters

Corresponding estimator

1. Method of maximum likelihood.

MLE

2. Method of moments.

MME

**Two methods** for estimating parameters

Corresponding estimator

1. Method of maximum likelihood.

MLE

2. Method of moments.

MME

# Maximum Likelihood Estimation

**Definition 5.2.1.** For a random sample of size  $n$  from the discrete (resp. continuous) population/pdf  $p_X(k; \theta)$  (resp.  $f_Y(y; \theta)$ ), the **likelihood function**,  $L(\theta)$ , is the product of the pdf evaluated at  $X_i = k_i$  (resp.  $Y_i = y_i$ ), i.e.,

$$L(\theta) = \prod_{i=1}^n p_X(k_i; \theta) \quad \left( \text{resp. } L(\theta) = \prod_{i=1}^n f_Y(y_i; \theta) \right).$$

**Definition 5.2.2.** Let  $L(\theta)$  be as defined in Definition 5.2.1. If  $\theta_e$  is a value of the parameter such that  $L(\theta_e) \geq L(\theta)$  for all possible values of  $\theta$ , then we call  $\theta_e$  the **maximum likelihood estimate** for  $\theta$ .

# Maximum Likelihood Estimation

**Definition 5.2.1.** For a random sample of size  $n$  from the discrete (resp. continuous) population/pdf  $p_X(k; \theta)$  (resp.  $f_Y(y; \theta)$ ), the **likelihood function**,  $L(\theta)$ , is the product of the pdf evaluated at  $X_i = k_i$  (resp.  $Y_i = y_i$ ), i.e.,

$$L(\theta) = \prod_{i=1}^n p_X(k_i; \theta) \quad \left( \text{resp. } L(\theta) = \prod_{i=1}^n f_Y(y_i; \theta) \right).$$

**Definition 5.2.2.** Let  $L(\theta)$  be as defined in Definition 5.2.1. If  $\theta_e$  is a value of the parameter such that  $L(\theta_e) \geq L(\theta)$  for all possible values of  $\theta$ , then we call  $\theta_e$  the **maximum likelihood estimate** for  $\theta$ .



## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Longrightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

## Examples for MLE

Often but not always MLE can be obtained by setting the first derivative equal to zero:

E.g. 1. Poisson distribution:  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ,  $k = 0, 1, \dots$ .

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n k_i} \left( \prod_{i=1}^n k_i! \right)^{-1}.$$

$$\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^n k_i \right) \ln \lambda - \ln \left( \prod_{i=1}^n k_i! \right).$$

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n k_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{1}{n} \sum_{i=1}^n k_i =: \bar{k}}.$$

Comment: The critical point is indeed global maximum because

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n k_i < 0.$$

The following two cases are related to waiting time:

E.g. 2. Exponential distribution:  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ .

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Longrightarrow \quad \boxed{\lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}}}.$$

The following two cases are related to waiting time:

E.g. 2. Exponential distribution:  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ .

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Longrightarrow \quad \boxed{\lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}}}.$$



The following two cases are related to waiting time:

E.g. 2. Exponential distribution:  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ .

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \implies \quad \boxed{\lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}}}.$$

The following two cases are related to waiting time:

E.g. 2. Exponential distribution:  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ .

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Longrightarrow \quad \boxed{\lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}}}.$$

The following two cases are related to waiting time:

E.g. 2. Exponential distribution:  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ .

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Longrightarrow \quad \boxed{\lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}}}.$$

A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

The maximum likelihood estimator of  $\lambda$  is  $\hat{\lambda} = \frac{r n}{\sum_{i=1}^n y_i}$ . This estimator is unbiased, i.e.  $E(\hat{\lambda}) = \lambda$ . The variance of  $\hat{\lambda}$  is  $\text{Var}(\hat{\lambda}) = \frac{\lambda}{n}$ . The Fisher information is  $I(\lambda) = \frac{n}{\lambda}$ . The Cramér-Rao lower bound is  $\frac{1}{I(\lambda)} = \lambda$ . The variance of  $\hat{\lambda}$  is  $\frac{\lambda}{n}$ .

A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

The maximum likelihood estimate of  $\lambda$  is  $\lambda_e = \frac{r}{\bar{y}}$ . This is the value of  $\lambda$  that maximizes the likelihood function  $L(\lambda)$ . The maximum likelihood estimate of  $\lambda$  is the value of  $\lambda$  that maximizes the likelihood function  $L(\lambda)$ . The maximum likelihood estimate of  $\lambda$  is the value of  $\lambda$  that maximizes the likelihood function  $L(\lambda)$ .

A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

The maximum likelihood estimator of  $\lambda$  is  $\hat{\lambda}_n = \frac{r n}{\sum_{i=1}^n y_i}$ . The maximum likelihood estimator of  $\lambda$  is unbiased, i.e.  $E(\hat{\lambda}_n) = \lambda$ . The maximum likelihood estimator of  $\lambda$  is efficient, i.e.  $\text{Var}(\hat{\lambda}_n) = \frac{1}{n I(\lambda)}$ .

A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

For the Gamma distribution, the maximum likelihood estimator of  $\lambda$  is  $\hat{\lambda} = \frac{r}{\bar{y}}$ . This estimator is unbiased, i.e.  $E[\hat{\lambda}] = \lambda$ . The variance of  $\hat{\lambda}$  is  $\text{Var}(\hat{\lambda}) = \frac{\lambda}{r n}$ . The Fisher information is  $I(\lambda) = \frac{r n}{\lambda}$ .

A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

- When  $r = 1$ , this reduces to the exponential distribution case.
- If  $r$  is also unknown, it will be much more complicated.  
No closed-form solution. One needs numerical solver<sup>2</sup>.  
Try MME instead.

---

<sup>2</sup>[DW, Example 7.2.25]



A random sample of size  $n$  from the following population:

E.g. 3. Gamma distribution:  $f_Y(y; \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$  with  $r > 1$  known.

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} e^{-\lambda y_i} = \lambda^{r n} \Gamma(r)^{-n} \left( \prod_{i=1}^n y_i^{r-1} \right) \exp \left( -\lambda \sum_{i=1}^n y_i \right)$$

$$\ln L(\lambda) = r n \ln \lambda - n \ln \Gamma(r) + \ln \left( \prod_{i=1}^n y_i^{r-1} \right) - \lambda \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{r n}{\lambda} - \sum_{i=1}^n y_i.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = 0 \quad \Rightarrow \quad \boxed{\lambda_e = \frac{r n}{\sum_{i=1}^n y_i} = \frac{r}{\bar{y}}}.$$

Comment:

- When  $r = 1$ , this reduces to the exponential distribution case.
- If  $r$  is also unknown, it will be much more complicated.  
No closed-form solution. One needs numerical solver<sup>2</sup>.  
Try MME instead.

---

<sup>2</sup>[DW, Example 7.2.25]

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1}p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i-1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Rightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1}p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i-1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Longrightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1}p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i-1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Longrightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1} p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i - 1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Longrightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1} p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i - 1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Longrightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

A detailed study with data:

E.g. 4. Geometric distribution:  $p_X(k; p) = (1 - p)^{k-1} p$ ,  $k = 1, 2, \dots$ .

$$L(p) = \prod_{i=1}^n (1 - p)^{k_i - 1} p = (1 - p)^{-n + \sum_{i=1}^n k_i} p^n.$$

$$\ln L(p) = \left( -n + \sum_{i=1}^n k_i \right) \ln(1 - p) + n \ln p.$$

$$\frac{d}{dp} \ln L(p) = -\frac{-n + \sum_{i=1}^n k_i}{1 - p} + \frac{n}{p}.$$

$$\frac{d}{dp} \ln L(p) = 0 \quad \Longrightarrow \quad \boxed{p_e = \frac{n}{\sum_{i=1}^n k_i} = \frac{1}{\bar{k}}}.$$

Comment: Its cousin distribution, the negative binomial distribution can be worked out similarly (See Ex 5.2.14).

k	Observed frequency	Predicted frequency
1	72	74.14
2	35	31.2
3	11	13.13
4	6	5.52
5	2	2.32
6	2	0.98



```

1 # The example from the book.
2 library (pracma) # Load the library "Practical Numerical Math Functions"
3 k<-c(72, 35, 11, 6, 2, 2) # observed freq.
4 a=1:6
5 pe=sum(k)/dot(k,a) # MLE for p.
6 f=a
7 for (i in 1:6) {
8   f[i] = round((1-pe)^(i-1) * pe * sum(k),2)
9 }
10 # Initialize the table
11 d <-matrix(1:18, nrow = 6, ncol = 3)
12 # Now adding the column names
13 colnames(d) <- c("k",
14                 "Observed freq.",
15                 "Predicted freq.")
16 d[1:6,1]<-a
17 d[1:6,2]<-k
18 d[1:6,3]<-f
19 grid.table(d) # Show the table
20 PlotResults("unknown", pe, d, "Geometric.pdf") # Output the results using a user defined function

```

k	Observed frequency	Predicted frequency
1	42	40.96
2	31	27.85
3	15	18.94
4	11	12.88
5	9	8.76
6	5	5.96
7	7	4.05
8	2	2.75
9	1	1.87
10	2	1.27
11	1	0.87
13	1	0.59
14	1	0.4

```

1 # Now let's generate random samples from a Geometric distribution with  $p=1/3$  with the same size
  of the sample.
2 p = 1/3
3 n = 128
4 gdata<-rgeom(n, p)+1 # Generate random samples
5 g<- table(gdata) # Count frequency of your data.
6 g<- t(rbind(as.numeric(rownames(g)), g)) # Transpose and combine two columns.
7 pe=n/dot(g[,1],g[,2]) # MLE for p.
8 f <- g[,1] # Initialize f
9 for (i in 1:nrow(g)) {
10   f[i] = round((1-pe)^(i-1) * pe * n,2)
11 } # Compute the expected frequency
12 g<-cbind(g,f) # Add one columns to your matrix.
13 colnames(g) <- c("k",
14                 "Observed freq.",
15                 "Predicted freq.") # Specify the column names.
16 d_df <- as.data.frame(d) # One can use data frame to store data
17 d_df # Show data on your terminal
18 PlotResults(p, pe, g, "Geometric2.pdf") # Output the results using a user defined function

```

k	Observed frequency	Predicted frequency
1	99	105.88
2	69	68.51
3	47	44.33
4	28	28.69
5	27	18.56
6	9	12.01
7	8	7.77
8	5	5.03
9	5	3.25
10	3	2.11

In case we have several parameters:

E.g. 5. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0 \end{cases}$$

$\implies$

$$\begin{cases} \mu_e = \bar{y} \\ \sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{cases}$$

In case we have several parameters:

E.g. 5. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0 \end{cases}$$

$\implies$

$$\begin{cases} \mu_e = \bar{y} \\ \sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{cases}$$

In case we have several parameters:

E.g. 5. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0 \end{cases}$$

$\Rightarrow$

$$\begin{cases} \mu_{\text{e}} = \bar{y} \\ \sigma_{\text{e}}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{cases}$$

In case we have several parameters:

E.g. 5. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0 \end{cases}$$

$\Rightarrow$

$$\begin{cases} \mu_{\text{e}} = \bar{y} \\ \sigma_{\text{e}}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{cases}$$



In case we have several parameters:

E.g. 5. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{cases}$$

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0 \end{cases}$$

$\Rightarrow$

$$\begin{cases} \mu_e = \bar{y} \\ \sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{cases}$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7.  $f_Y(y) = \frac{1}{b-a}$  for  $y \in [a, b]$

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$$a_e = y_{\min}$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7.  $f_Y(y) = \frac{1}{b-a}$  for  $y \in [a, b]$

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise} \end{cases}$$

$$a_e = y_{\min}$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7. Exponential distribution

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} e^{-\frac{y_i - a}{b-a}} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7.  $f_Y(y; \theta) = \frac{2y}{\theta^2}$  for  $y \in [0, \theta]$ .

$$L(\theta) = \begin{cases} \prod_{i=1}^n \frac{2y_i}{\theta^2} = 2^n \theta^{-2n} \prod_{i=1}^n y_i & \text{if } 0 \leq y_1, \dots, y_n \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

$\Downarrow$

$$\theta_e = y_{\max}.$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7.  $f_Y(y; \theta) = \frac{2y}{\theta^2}$  for  $y \in [0, \theta]$ .

$$L(\theta) = \begin{cases} \prod_{i=1}^n \frac{2y_i}{\theta^2} = 2^n \theta^{-2n} \prod_{i=1}^n y_i & \text{if } 0 \leq y_1, \dots, y_n \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

$\Downarrow$

$$\theta_e = y_{\max}.$$

In case when the parameters determine the support of the density:  
(Non regular case)

E.g. 6. Uniform distribution on  $[a, b]$  with  $a < b$ :  $f_Y(y; a, b) = \frac{1}{b-a}$  if  $y \in [a, b]$ .

$$L(a, b) = \begin{cases} \prod_{i=1}^n \frac{1}{b-a} = \frac{1}{(b-a)^n} & \text{if } a \leq y_1, \dots, y_n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$L(a, b)$  is monotone increasing in  $a$  and decreasing in  $b$ . Hence, in order to maximize  $L(a, b)$ , one needs to choose

$$a_e = y_{\min} \quad \text{and} \quad b_e = y_{\max}.$$

E.g. 7.  $f_Y(y; \theta) = \frac{2y}{\theta^2}$  for  $y \in [0, \theta]$ .

$$L(\theta) = \begin{cases} \prod_{i=1}^n \frac{2y_i}{\theta^2} = 2^n \theta^{-2n} \prod_{i=1}^n y_i & \text{if } 0 \leq y_1, \dots, y_n \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

$\Downarrow$

$$\theta_e = y_{\max}.$$

In case of discrete parameter:

E.g. 8. **Wildlife sampling.** Capture-tag-recapture.... In the history,  $a$  tags have been put. In order to estimate the population size  $N$ , one randomly captures  $n$  animals, and there are  $k$  tagged. Find the MLE for  $N$ .

**Sol.** The population follows hypergeometric distr.:  $p_X(k; N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}.$

$$L(N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

How to maximize  $L(N)$ ?

```
1 > a=10
2 > k=5
3 > n=20
4 > N=seq(a,a+100)
5 > p=choose(a,k)*choose(N-a,n-k)/
   choose(N,n)
6 > plot(N,p,type = "p")
7 > print(paste("The MLE is", n*a/k))
8 [1] "The MLE is 40"
```



In case of discrete parameter:

E.g. 8. **Wildlife sampling.** Capture-tag-recapture.... In the history,  $a$  tags have been put. In order to estimate the population size  $N$ , one randomly captures  $n$  animals, and there are  $k$  tagged. Find the MLE for  $N$ .

**Sol.** The population follows hypergeometric distr.:  $p_X(k; N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$ .

$$L(N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

How to maximize  $L(N)$ ?

```
1 > a=10
2 > k=5
3 > n=20
4 > N=seq(a,a+100)
5 > p=choose(a,k)*choose(N-a,n-k)/
   choose(N,n)
6 > plot(N,p,type = "p")
7 > print(paste("The MLE is", n*a/k))
8 [1] "The MLE is 40"
```

In case of discrete parameter:

E.g. 8. **Wildlife sampling.** Capture-tag-recapture.... In the history,  $a$  tags have been put. In order to estimate the population size  $N$ , one randomly captures  $n$  animals, and there are  $k$  tagged. Find the MLE for  $N$ .

**Sol.** The population follows hypergeometric distr.:  $p_X(k; N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$ .

$$L(N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

How to maximize  $L(N)$ ?

```
1 > a=10
2 > k=5
3 > n=20
4 > N=seq(a,a+100)
5 > p=choose(a,k)*choose(N-a,n-k)/
   choose(N,n)
6 > plot(N,p,type = "p")
7 > print(paste("The MLE is", n*a/k))
8 [1] "The MLE is 40"
```

In case of discrete parameter:

**E.g. 8. Wildlife sampling.** Capture-tag-recapture.... In the history,  $a$  tags have been put. In order to estimate the population size  $N$ , one randomly captures  $n$  animals, and there are  $k$  tagged. Find the MLE for  $N$ .

**Sol.** The population follows hypergeometric distr.:  $p_X(k; N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$ .

$$L(N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

How to maximize  $L(N)$ ?

```
1 > a=10
2 > k=5
3 > n=20
4 > N=seq(a,a+100)
5 > p=choose(a,k)*choose(N-a,n-k)/
   choose(N,n)
6 > plot(N,p,type = "p")
7 > print(paste("The MLE is", n*a/k))
8 [1] "The MLE is 40"
```

In case of discrete parameter:

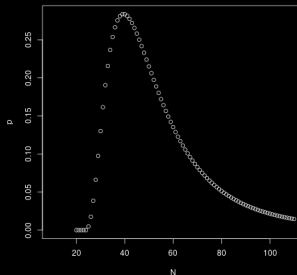
E.g. 8. **Wildlife sampling.** Capture-tag-recapture.... In the history,  $a$  tags have been put. In order to estimate the population size  $N$ , one randomly captures  $n$  animals, and there are  $k$  tagged. Find the MLE for  $N$ .

**Sol.** The population follows hypergeometric distr.:  $p_X(k; N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$ .

$$L(N) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$$

How to maximize  $L(N)$ ?

```
1 > a=10
2 > k=5
3 > n=20
4 > N=seq(a,a+100)
5 > p=choose(a,k)*choose(N-a,n-k)/
  choose(N,n)
6 > plot(N,p,type = "p")
7 > print(paste("The MLE is", n*a/k))
8 [1] "The MLE is 40"
```



The graph suggests to study the following quantity:

$$r(N) := \frac{L(N)}{L(N-1)} = \frac{N-n}{N} \times \frac{N-a}{N-a-n+k}$$

$$r(N) < 1 \iff na < Nk \quad \text{i.e., } N > \frac{na}{k}$$

$$N_e = \arg \max \left\{ L(N) : N = \left\lfloor \frac{na}{k} \right\rfloor, \left\lceil \frac{na}{k} \right\rceil \right\}.$$



The graph suggests to study the following quantity:

$$r(N) := \frac{L(N)}{L(N-1)} = \frac{N-n}{N} \times \frac{N-a}{N-a-n+k}$$

$$r(N) < 1 \iff na < Nk \quad \text{i.e., } N > \frac{na}{k}$$

$$N_e = \arg \max \left\{ L(N) : N = \left\lfloor \frac{na}{k} \right\rfloor, \left\lceil \frac{na}{k} \right\rceil \right\}.$$



The graph suggests to study the following quantity:

$$r(N) := \frac{L(N)}{L(N-1)} = \frac{N-n}{N} \times \frac{N-a}{N-a-n+k}$$

$$r(N) < 1 \iff na < Nk \text{ i.e., } N > \frac{na}{k}$$

$$N_e = \arg \max \left\{ L(N) : N = \left\lfloor \frac{na}{k} \right\rfloor, \left\lceil \frac{na}{k} \right\rceil \right\}.$$



The graph suggests to study the following quantity:

$$r(N) := \frac{L(N)}{L(N-1)} = \frac{N-n}{N} \times \frac{N-a}{N-a-n+k}$$

$$r(N) < 1 \iff na < Nk \text{ i.e., } N > \frac{na}{k}$$

$$N_e = \arg \max \left\{ L(N) : N = \left\lfloor \frac{na}{k} \right\rfloor, \left\lceil \frac{na}{k} \right\rceil \right\}.$$





# Method of Moments Estimation

**Rationale:** The population moments should be close to the sample moments, i.e.,

$$\mathbb{E}(Y^k) \approx \frac{1}{n} \sum_{i=1}^n y_i^k, \quad k = 1, 2, 3, \dots$$

**Definition 5.2.3.** For a random sample of size  $n$  from the discrete (resp. continuous) population/pdf  $p_X(k; \theta_1, \dots, \theta_s)$  (resp.  $f_Y(y; \theta_1, \dots, \theta_s)$ ), solutions to

$$\begin{cases} \mathbb{E}(Y) = \frac{1}{n} \sum_{i=1}^n y_i \\ \vdots \\ \mathbb{E}(Y^s) = \frac{1}{n} \sum_{i=1}^n y_i^s \end{cases}$$

which are denoted by  $\theta_{1e}, \dots, \theta_{se}$ , are called the **method of moments estimates** of  $\theta_1, \dots, \theta_s$ .

## Examples for MME

MME is often the same as MLE:

E.g. 1. Normal distribution:  $f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,  $y \in \mathbb{R}$ .

$$\left\{ \begin{array}{l} \mu = \mathbb{E}(Y) = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y} \\ \sigma^2 + \mu^2 = \mathbb{E}(Y^2) = \frac{1}{n} \sum_{i=1}^n y_i^2 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \mu_e = \bar{y} \\ \sigma_e^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \mu_e^2 \\ \quad = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \end{array} \right.$$

More examples when MLE coincides with MME: Poisson, Exponential, Geometric.

MME is often much more tractable than MLE:

E.g. 2. Gamma distribution<sup>3</sup>:  $f_Y(y; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}$  for  $y \geq 0$ .

$$\begin{cases} \frac{r}{\lambda} = \mathbb{E}(Y) = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y} \\ \frac{r}{\lambda^2} + \frac{r^2}{\lambda^2} = \mathbb{E}(Y^2) = \frac{1}{n} \sum_{i=1}^n y_i^2 \end{cases} \Rightarrow \begin{cases} r_e = \frac{\bar{y}^2}{\hat{\sigma}^2} \\ \lambda_e = \frac{\bar{y}}{\hat{\sigma}^2} = \frac{r_e}{\bar{y}} \end{cases}$$

where  $\bar{y}$  is the sample mean and  $\hat{\sigma}^2$  is the sample variance:  
 $\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ .

Comments: MME for  $\lambda$  is consistent with MLE when  $r$  is known.

---

<sup>3</sup>Check Theorem 4.6.3 on p. 269 for mean and variance

Another tractable example for MME, while less tractable for MLE:

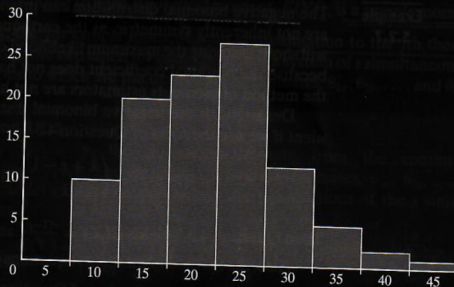
E.g. 3. Neg. binomial distribution:  $p_X(k; p, r) = \binom{k+r-1}{k} (1-p)^k p^r$ ,  $k = 0, 1, \dots$ .

$$\begin{cases} \frac{r(1-p)}{p} = \mathbb{E}(X) = \bar{k} \\ \frac{r(1-p)}{p^2} = \text{Var}(X) = \hat{\sigma}^2 \end{cases} \Rightarrow \begin{cases} p_e = \frac{\bar{k}}{\hat{\sigma}^2} \\ r_e = \frac{\bar{k}^2}{\hat{\sigma}^2 - \bar{k}} \end{cases}$$

Table 5.2.4

Number	Observed Frequency	Expected Frequency
0-5	0	0
6-10	10	7.7
11-15	20	21.4
16-20	23	28.4
21-25	27	22.4
26-30	12	12.3
31-35	5	5.3
36-40	2	1.8
> 40	1	0.7

Data from: <http://www.seattlecentral.edu/qelp/sets/039/039.html>



$$r_e = 12.74 \text{ and } p_e = 0.391.$$

E.g. 4.  $f_Y(y; \theta) = \frac{2y}{\theta^2}$  for  $y \in [0, \theta]$ .

$$\bar{y} = \mathbb{E}[Y] = \int_0^\theta \frac{2y^2}{\theta^2} dy = \frac{2}{3} \frac{y^3}{\theta^2} \Big|_{y=0}^{y=\theta} = \frac{2}{3} \theta.$$

$\Downarrow$

$$\boxed{\theta_e = \frac{3}{2} \bar{y}.$$