MTMS.01.099 Mathematical Statistics

Lecture 7

Tõnu Kollo

Fall 2017

Reminder: The Method of Maximum Likelihood

$$L(\theta) = \begin{cases} f(x_1;\theta) \times f(x_2;\theta) \times \ldots \times f(x_n;\theta), & \text{continuous case} \\ p(x_1;\theta) \times p(x_2;\theta) \times \ldots \times p(x_n;\theta), & \text{discrete case.} \end{cases}$$

Definition

The value $\hat{\theta}_{ML}$ from sample space A for which $L(\theta)$ obtains its largest value within A is called the **ML estimate** of θ:

$$L(\hat{\theta}_{ML}) = max_{\theta \in A} L(\theta).$$

The likelihood function $L(\theta) = L(\mathbf{x}, \theta)$ gives the likelihood of θ, given the data $\mathbf{x}$. A maximum likelihood estimate $\hat{\theta}_{ML}$ is a value of θ that maximizes the likelihood.

Reminder: The Method of Least Squares

Let $x_1, x_2, \ldots, x_n$ be a sample from a distribution with mean $E(X) = \mu(\theta)$, where $\mu(\theta)$ is a known function and θ an unknown parameter with parameter space A. Let

$$Q(\theta) = \sum_{i=1}^{n} [x_i - \mu(\theta)]^2$$

be the sum of the squares of the deviations of the observations from $\mu(\theta)$.

Definition

The value $\hat{\theta}_{LS}$, for which $Q(\theta)$ assumes its least possible value within A, is called the **LS estimate** of θ:

$$Q(\hat{\theta}_{LS}) = min_{\theta \in A} Q(\theta)$$

Reminder: Method of Moments

Let $X_1, X_2, \ldots, X_n$ denote a sample from a distribution with r unknown parameters, $(\theta_1, \theta_2, \ldots, \theta_r)$. Let $f(x; \theta_1, \ldots, \theta_r)$ denote the probability density function of this distribution.

The method of moments procedure equates k-th theoretical moments with the corresponding k-th sample moments, $k = 1, \ldots, r$, to obtain a system of r equations in r unknowns,

$$\mu_k(\theta_1, \ldots, \theta_r) = m_k, \quad k = 1, \ldots, r, \quad (*)$$

where $m_k = \frac{1}{n}\sum_{i=1}^{n} x^k$ is a ***k*-th sample moment** and $\mu_k(\theta_1, \ldots, \theta_r) = EX^k$ is a ***k*-th theoretical moment.**

Definition

The solution of the system of equations $(*)$, $(\hat{\theta}_{1_{MM}}, \ldots, \hat{\theta}_{r_{MM}})$, are then the **method of moments estimates** of $\theta_1, \ldots, \theta_r$.

Application to the Normal Distribution

1. A Single Sample

Let $x_1, \ldots, x_n$ be a sample from $N(\mu, \sigma^2)$, where both parameters are unknown. Let us find ML estimates for these parameters.

Density function: $f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Likelihood function:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

Logarithmic likelihood function:

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

Application to the Normal Distribution (2)

Logarithmic likelihood function:

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

Differentiate:

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$$

Equate to zero and it follows

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \quad \longleftarrow \text{unbiased}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \quad \longleftarrow \text{biased}$$

Application to the Normal Distribution (2)

Remark

If μ is known, we obtain from equation $\frac{\partial l}{\partial \sigma^2} = 0$ the following ML estimate for σ^2:

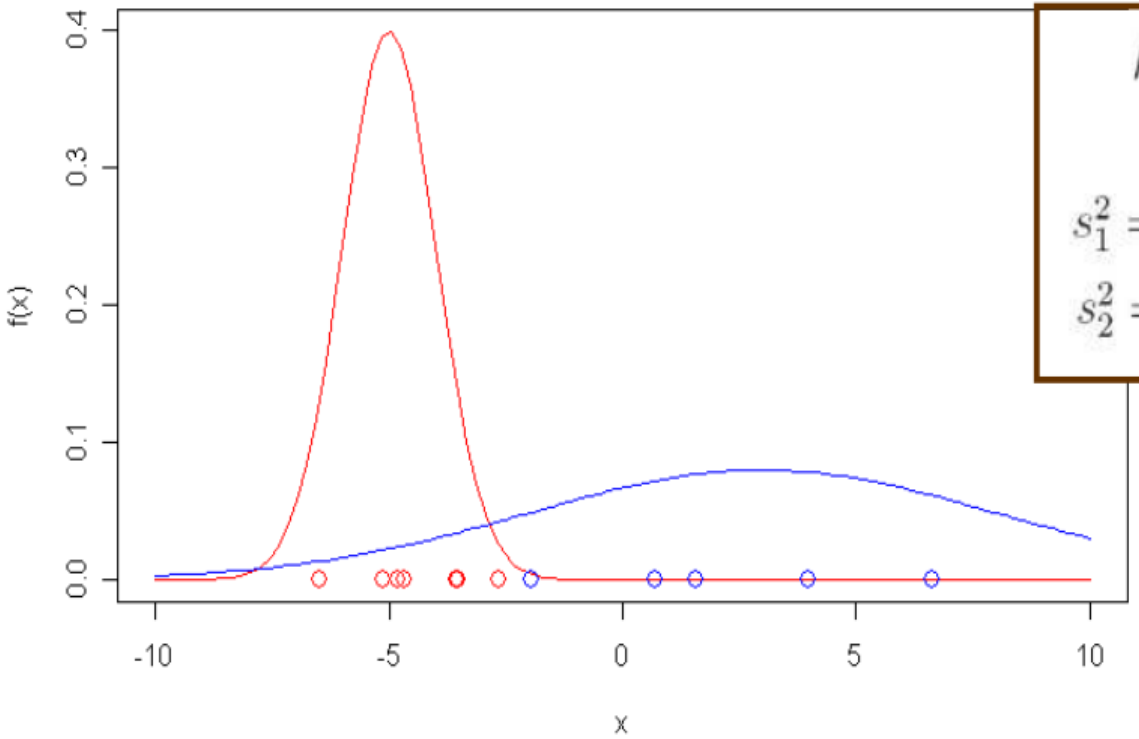
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2.$$

This estimate is unbiased.

Application to the Normal Distribution (3)

2. Two Samples

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be two independent samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. If all four parameters are unknown, each sample can be treated according to the the method in Single Sample case.

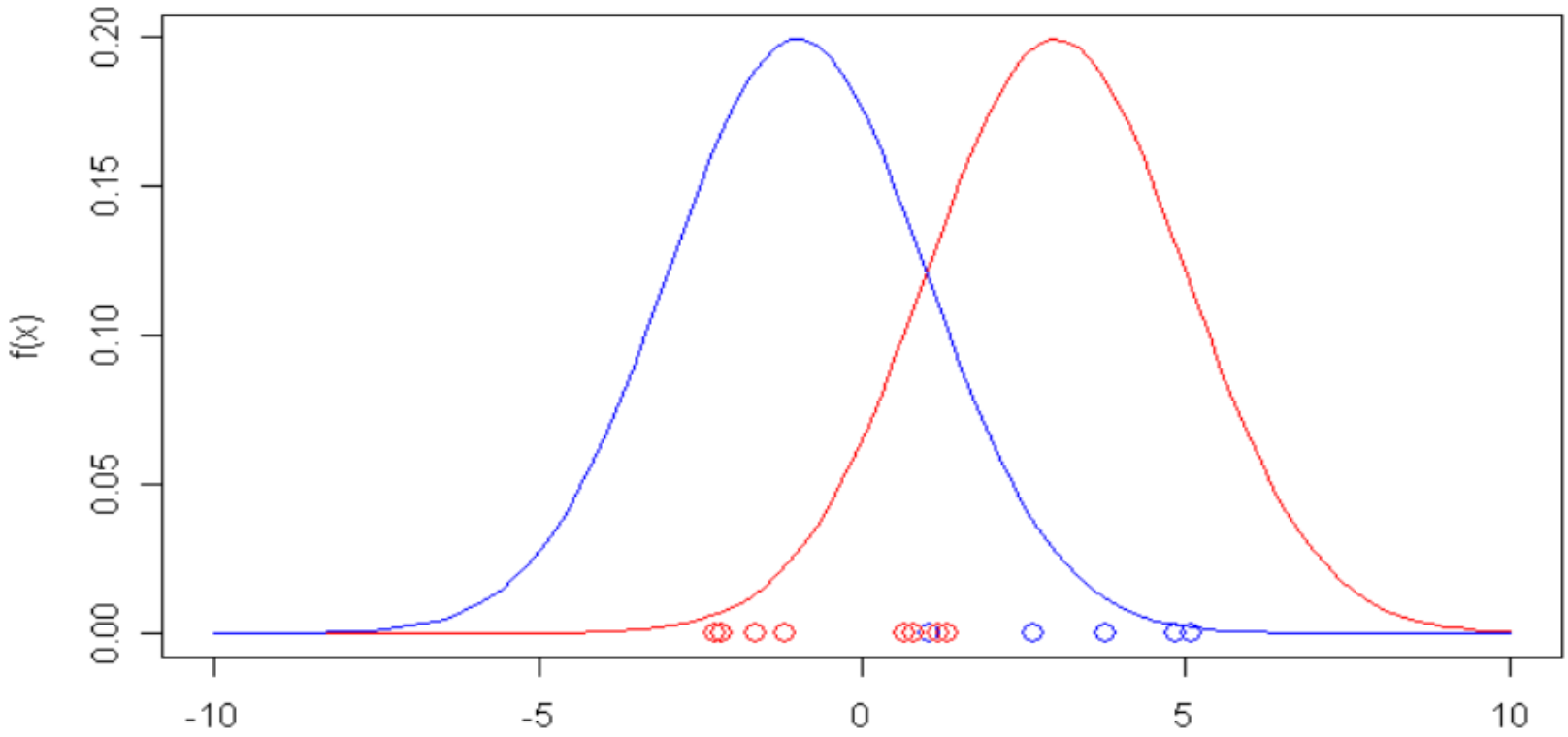


$$\hat{\mu}_1 = \bar{x}, \ \hat{\mu}_2 = \bar{y},$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum (x_i - \bar{x}^2),$$
$$s_2^2 = \frac{1}{n_2 - 1} \sum (y_i - \bar{y}^2).$$

Application to the Normal Distribution (4)

1. Two Samples



If we know that $\sigma_1^2 = \sigma_2^2 = \sigma$, then we can estimate σ using both samples.

Application to the Normal Distribution (4)

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be two independent samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. The joint likelihood function for two samples is

$$L(\mu_1, \mu_2, \sigma^2) = L(\mu_1, \sigma^2) \cdot L(\mu_2, \sigma^2),$$

where $L(\mu_1, \sigma^2) = (2\pi\sigma^2)^{-\frac{n_1}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(x_i-\mu_1)^2}$ (analogously for $L(\mu_2, \sigma^2)$), so that

$$L(\mu_1, \mu_2, \sigma^2) = (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{1}{2\sigma^2}B(\mu_1,\mu_2)},$$

where $B(\mu_1, \mu_2) = \sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \sum_{i=1}^{n_2}(y_i - \mu_2)^2$.

Application to the Normal Distribution (5)

Logarithmic likelihood function is:

$$l(\mu_1, \mu_2, \sigma^2) = -\frac{n_1 + n_2}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} B(\mu_1, \mu_2).$$

Differentiating and equating to zero, we get the following equation system:

$$\frac{\partial l}{\partial \mu_1} : -\frac{1}{2\sigma^2} \sum_{i=1}^{n_1} (x_i - \mu_1)(-1) = 0,$$

$$\frac{\partial l}{\partial \mu_2} : -\frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (y_i - \mu_2)(-1) = 0,$$

$$\frac{\partial l}{\partial \sigma^2} : -\frac{n_1 + n_2}{2} \frac{1}{2\pi\sigma^2} 2\pi + \frac{1}{2\sigma^4} B(\mu_1, \mu_2) = 0.$$

Application to the Normal Distribution (6)

Solutions of equation system are the ML estimates for the parameters μ_1, μ_2, σ^2:

$$\hat{\mu}_1 = \overline{x},$$

$$\hat{\mu}_2 = \overline{y},$$

$$\hat{\sigma}^2 = \frac{B(\mu_1, \mu_2)}{n_1 + n_2} = \frac{1}{n_1 + n_2}\left(\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \sum_{i=1}^{n_2}(y_i - \mu_2)^2\right) \Rightarrow$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}.$$

The first two estimates are unbiased, but the third one is biased. The corrected unbiased ML estimate is found to be

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Standard Error of an estimate

We have an unknown parameter θ. From sample we get a point estimate $\hat{\theta}$, which is a value of the estimator $\hat{\theta}(\boldsymbol{X})$ at $\boldsymbol{X} = \boldsymbol{x}$. The variance of the estimator $\hat{\theta}(\boldsymbol{X})$ is a constant, but the corresponding estimator is a random variable:

$$\theta \longleftrightarrow \hat{\theta}(\boldsymbol{X}), \quad Var(\hat{\theta}(\boldsymbol{X})) \longleftrightarrow \widehat{Var}(\hat{\theta}(\boldsymbol{X}))$$

Definition

An estimate of $\sigma_\theta = \sqrt{Var\hat{\theta}(\boldsymbol{X})}$ is called the **standard error** of $\hat{\theta}$

$$se(\hat{\theta}) = \sqrt{\widehat{Var}\hat{\theta}(\boldsymbol{x})}.$$

Standard Error

Example

The researcher wants to investigate the proportion of households who are following BBC News Channel. It is known that the standard error of the estimate cannot exceed 0.03. How many households need to be in the sample?

Standard Error

Example

The researcher wants to investigate the proportion of households who are following BBC News Channel. It is known that the standard error of the estimate cannot exceed 0.03. How many households need to be in the sample?

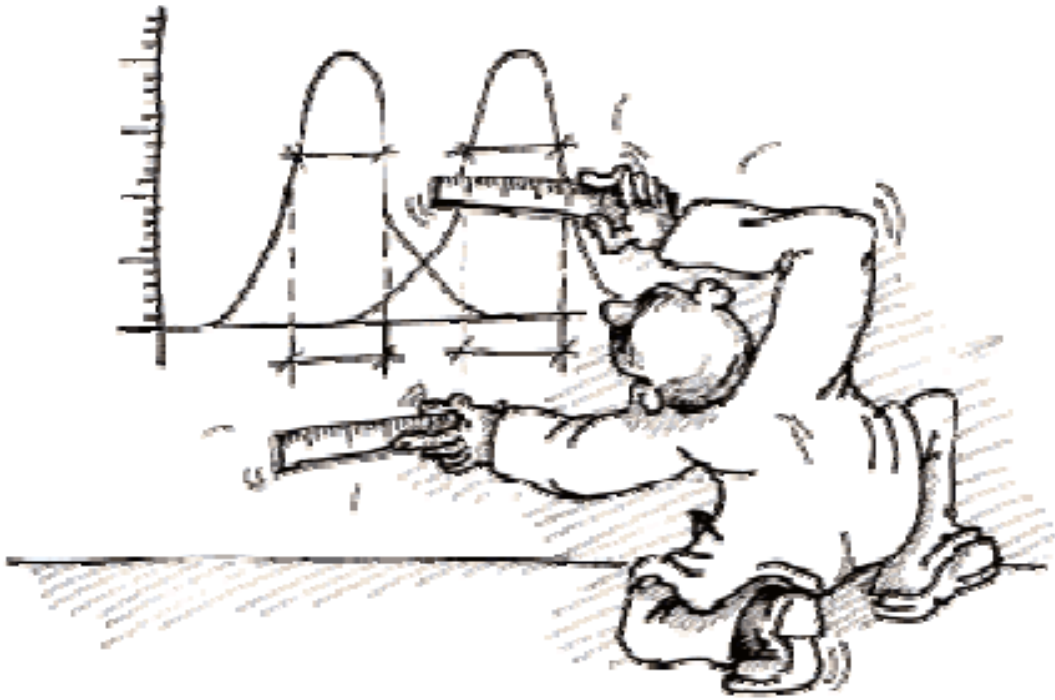
The number of BBC followers X is from binomial distribution: $X \sim Bin(n, p)$.

$$se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.03$$

If we compute this with respect to n, we get the asked sample size.

Estimation of Probability Function, Density Function and Distribution Function

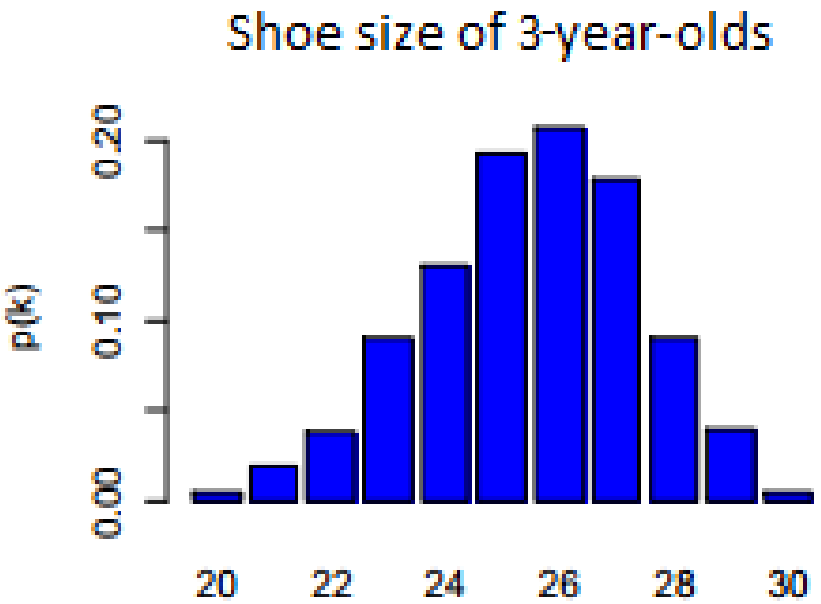
Have you thought about how (and based on what) SAS, R, SPSS etc are drawing the graphs (histograms) of empirical distributions?



Estimating the Probability Function

Let $p(k)$ be the unknown (discrete) probability function, which can be theoretically given by the following table $\{k, p(k)\}$:

k	k_1	k_2	...	k_m
$p(k)$	$p(k_1)$	$p(k_2)$	...	$p(k_m)$



Estimating the Probability Function (2)

Let us have a sample $x_1, x_2, \ldots, x_n$ from this distribution. Fix one concrete value k and find its frequency f_k.

$$f_k \sim B(n, p(k)) \Rightarrow \hat{p}_k = \frac{f_k}{n}.$$

Sample (n=20)
22,28,**26**,25,**26**,25,24,**26**,24,23,27,25,27,25,27,28,28,29,23,**26**
Fix one k (for example, $k = 26$) and find the corresponding f_k,

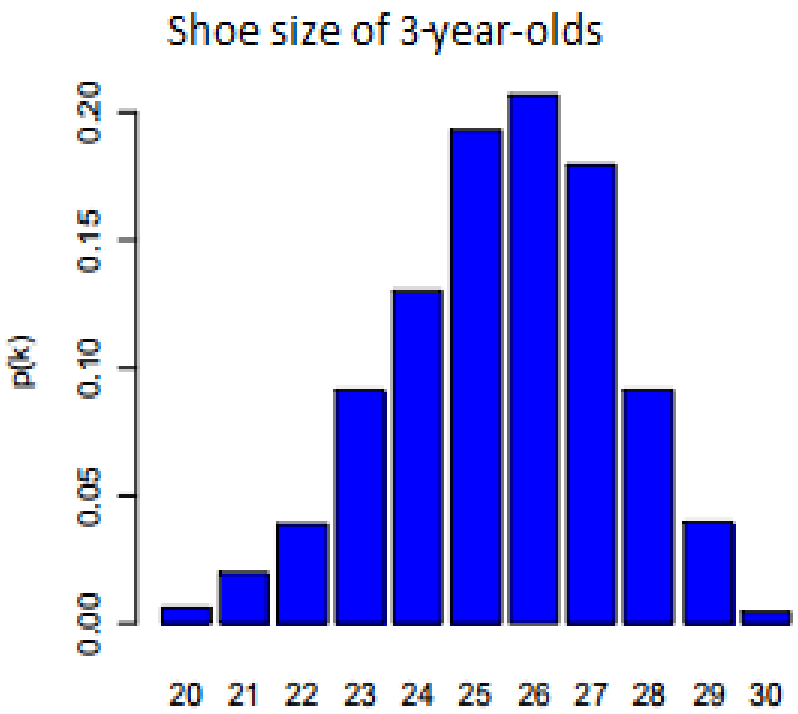
$$f_k = 4.$$

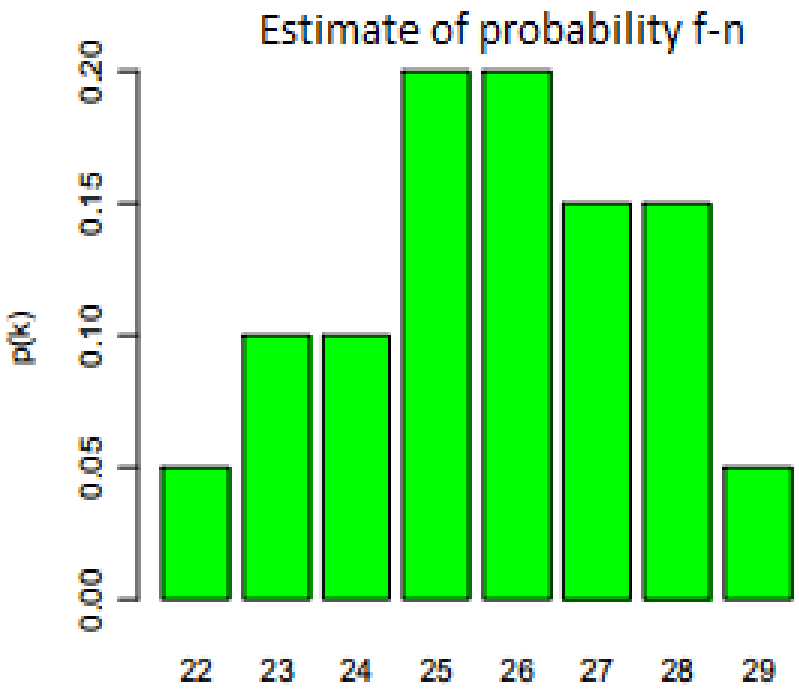
The estimate for the probability $p(k)$ is

$$\hat{p}(k) = 4/20 = 0.2$$

This is how we can estimate the probabilities of all values.

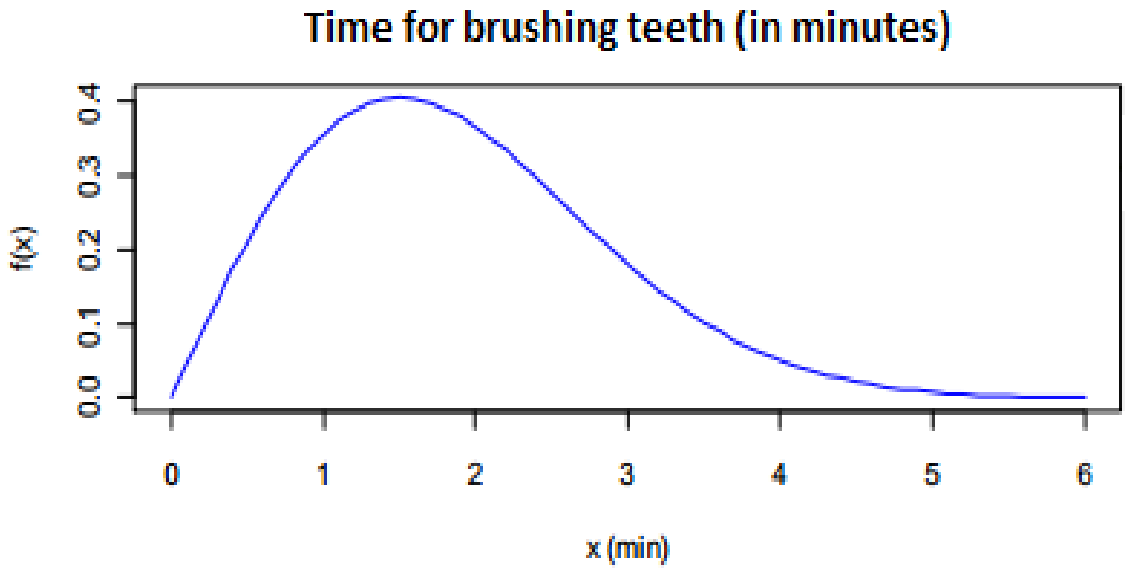
Estimating the Probability Function (3)





Estimating the Density Function

Let $f(x)$ be an unknown density function and consider sample $x_1, x_2, \ldots, x_n$ from this distribtuion.



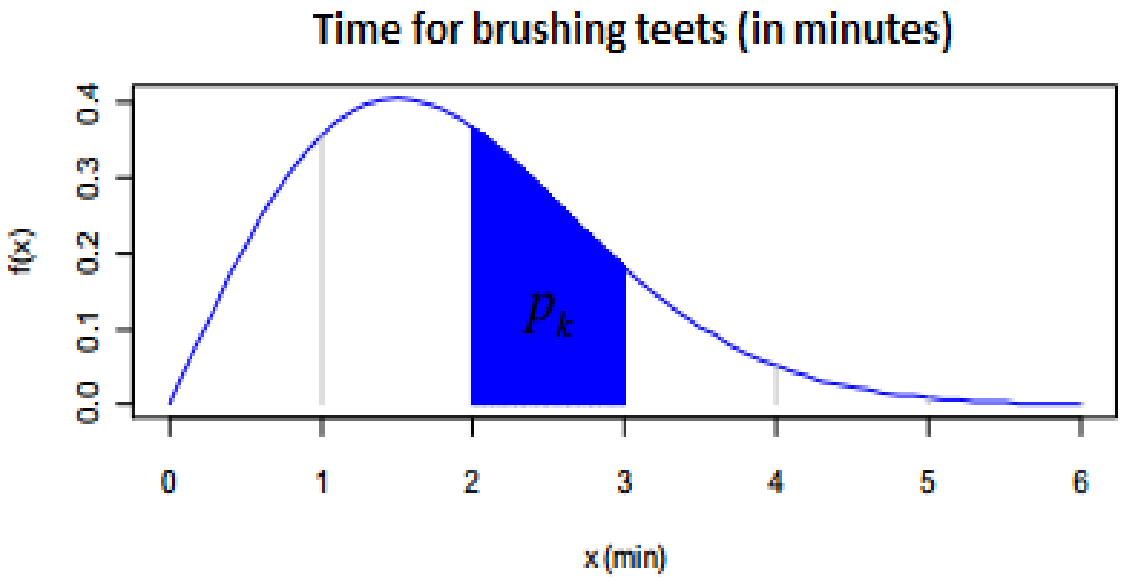
Sample

(n=20)

2.2,5.1,2.5,1.9,0.7,1.3,3.0,2.8,1.4,0.9,2.1,5.0,2.3,1.7,
0.5,1.4,3.3,2.7,1.8,0.9

Estimating the Density Function (2)

Group the observations into r classes. Let $p_1, \ldots, p_r$ be the unknown areas under the density function $f(x)$.



Need to estimate the probabilities $p_1, \ldots, p_r$.

Estimating the Density Function (3)

If f_i denotes the number of observations in i-th class, then

$$f_i \sim B(n, p_i) \quad \Rightarrow \quad \hat{p}_i = f_i/n.$$

Sample (n=20)

2.2,5.1,2.5,1.9,0.7,1.3,3.0,2.8,1.4,0.9,2.1,5.0,2.3,1.7,
0.5,1.4,3.3,2.7,1.8,0.9

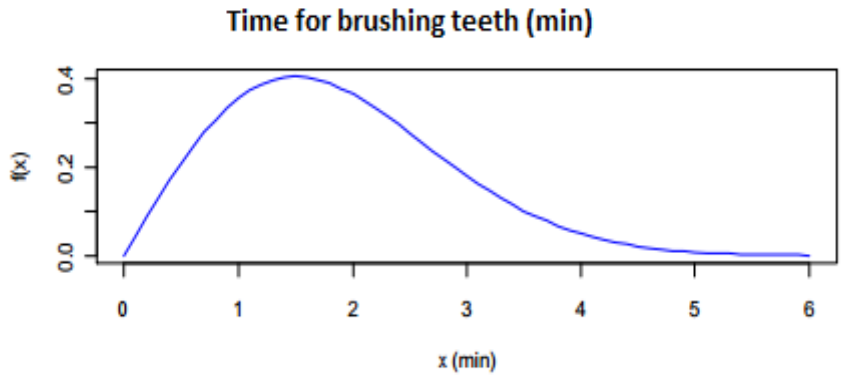
Group the classes:

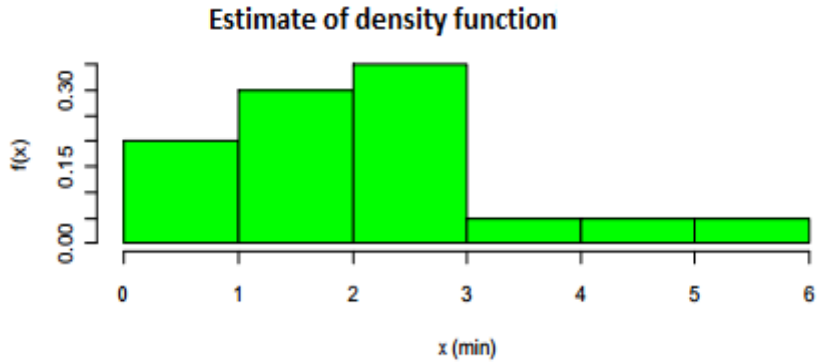
(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]
1	4	6	7	1	1

Corresponding estimates of the probabilities

(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]
1/20	4/20	6/20	7/20	1/20	1/20
0.05	0.20	0.30	0.35	0.05	0.05

Estimating the Density Function (4)



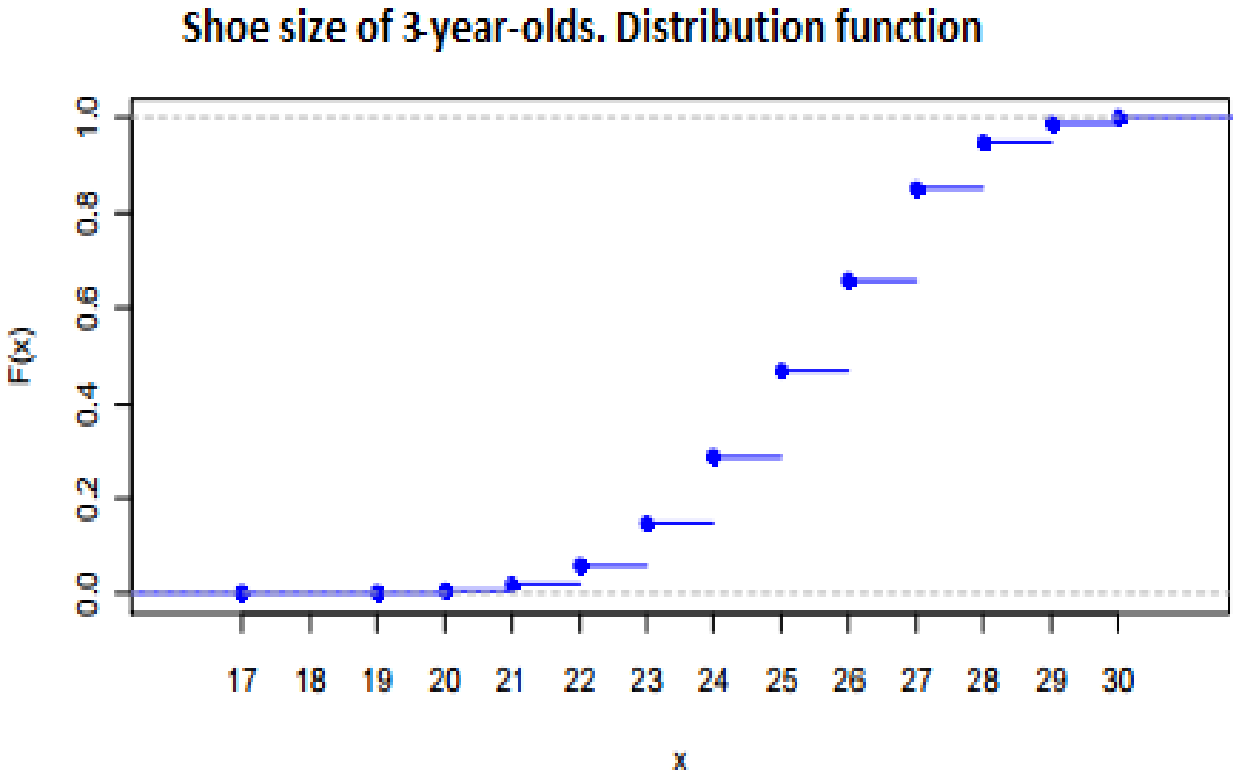


The conclusions are similar, but note that we do not obtain an estimate of the density function $f(x)$ itself, only estimates of certain areas under this function. By choosing r large (which requires a sufficiently large sample), we nevertheless obtain a good picture of the form of the density function.

Estimating the Distribution Function

Assume that we want to estimate $F(x)$ for a given value x:

$$F(x) = P(X \leq x).$$



Estimating the Distribution Function (2)

Assume that we want to estimate $F(x)$ for a given value x:

$$F(x) = P(X \leq x).$$

Let us have a sample with size n from this distribution. Denote

$$g_x = \#(x_i \leq x),$$

e.g the values in the sample that are $\leq x$.
Since g_x denotes the number of successes in a sequence of n,

$$g_x \sim B(n, F(x)),$$

where $F(x)$ is the probability, that the value in considered sequence is $\leq x$. It follows

$$\widehat{F}(x) = \frac{\#(x_i \leq x)}{n}.$$

Estimating the Distribution Function (3)

If the sample is small, or only moderately large, it is sometimes useful to perform the above calculation for all the n sample points $x_1, \ldots, x_n$. This implies that, proceeding from small to large x, we estimate $F(x)$ for $x = x_{(1)}, x = x_{(2)}$ and so on, where

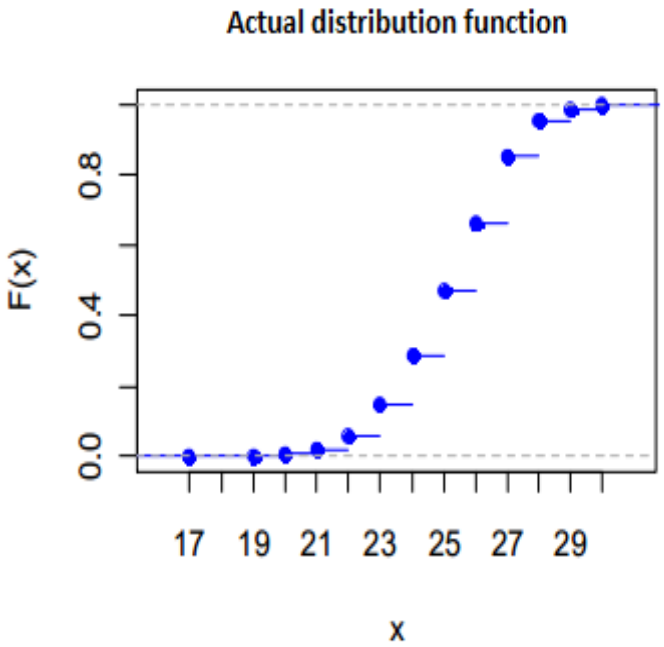
$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

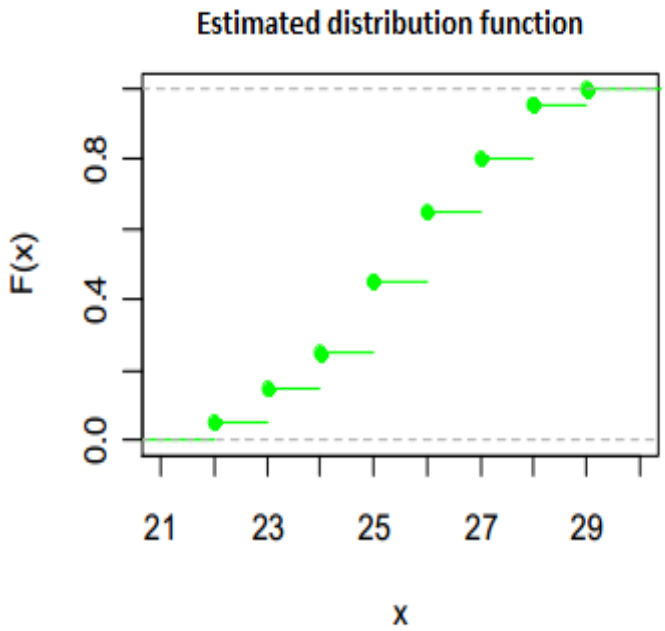
is the ordered sample.
When $x_{(i)} < x_{(i+1)}$, from $\widehat{F}(x) = \frac{\#(x_i \leq x)}{n}$. we get

$$\widehat{F}(x_{(i)}) = \frac{i}{n} \quad (i = 1, 2, \ldots, n)$$

In this way we obtain an estimate of the whole function $F(x)$ called the *empirical distribution function*.

Estimating the Distribution Function (4)





Estimating the Distribution Function (5)