Foundation of Data Science and Analytics

Sampling Distribution and Point Estimation of Parameters

Arun K. Timalsina

Material Adaptation : Applied Statistics and Probability for Engineers, by Montgomery and Runger. John Wiley & Sons,

Point Estimation

- A point estimate is a reasonable value of a population parameter.
- Data collected, X_1, X_2,…, X_n are random variables.
- Functions of these random variables, x-bar and s_2, are also random variables called statistics.
- Statistics have their unique distributions that are called sampling distributions.

Point Estimator

A point estimate of some population parameter θ is a single numerical value $\hat{\Theta}$.

The statistic $\hat{\Theta}$ is called the point estimator.

As an example, suppose the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ. That is, $\hat{\mu} = \bar{X}$. After the sample has been selected, the numerical value $\bar{x}$ is the point estimate of μ.

Thus if $x_1 = 25, x_2 = 30, x_3 = 29, x_4 = 31,$ the point estimate of μ is

$$\bar{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Some Parameters & Their Statistics

Parameter	Measure	Statistic
μ	Mean of a single population	x-bar
σ^2	Variance of a single population	s^2
σ	Standard deviation of a single population	s
p	Proportion of a single population	p-hat
$\mu_1 - \mu_2$	Difference in means of two populations	x bar$_1$ - x bar$_2$
$p_1 - p_2$	Difference in proportions of two populations	p hat$_1$ - p hat$_2$

- There could be choices for the point estimator of a parameter.
- To estimate the mean of a population, we could choose the:
  - Sample mean.
  - Sample median.
  - Average of the largest & smallest observations of the sample.
- We need to develop criteria to compare estimates using statistical properties.

Some Definitions

- The random variables $X_1, X_2, \ldots, X_n$ are a random sample of size n if:
  a) The X_i are independent random variables.
  b) Every X_i has the same probability distribution.
- A statistic is any function of the observations in a random sample.
- The probability distribution of a statistic is called a sampling distribution.

Sampling Distribution of the Sample Mean

- A random sample of size *n* is taken from a normal population with mean μ and variance σ^2.
- The observations, $X_1, X_2, \ldots, X_n$, are normally and independently distributed.
- A linear function (X-bar) of normal and independent random variables is itself normally distributed.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \text{ has a normal distribution}$$

$$\text{with mean } \mu_{\bar{X}} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

$$\text{and variance } \sigma^2_{\bar{X}} = \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2}$$

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n is taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad (7\text{-}1)$$

as $n \to \infty$, is the standard normal distribution.

Sampling Distributions of Sample Means

Figure 7-1 Distributions of average scores from throwing dice. Mean = 3.5

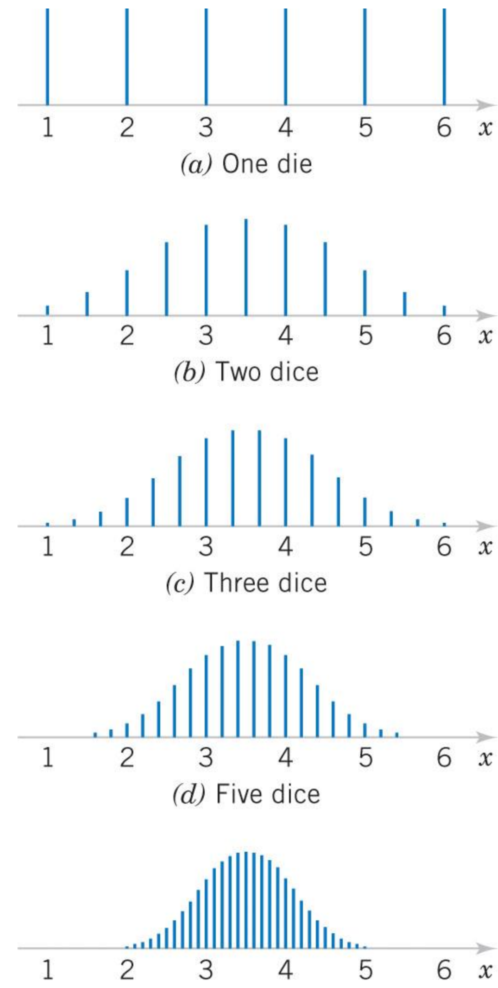
Formulas

$$\mu = \frac{b-a}{2}$$

$$\sigma_X^2 = \frac{(b-a+1)^2 - 1}{12}$$

$$\sigma_{\bar{X}}^2 = \sigma_X^2 / n$$

	n dies	var	std dev
a)	1	2.9	1.7
b)	2	1.5	1.2
c)	3	1.0	1.0
d)	5	0.6	0.8
e)	10	0.3	0.5
	a =	1	
	b =	6	



(a) One die
(b) Two dice
(c) Three dice
(d) Five dice
(e) Ten dice

Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. What is the probability that a random sample of *n* = 25 resistors will have an average resistance of less than 95 ohms?

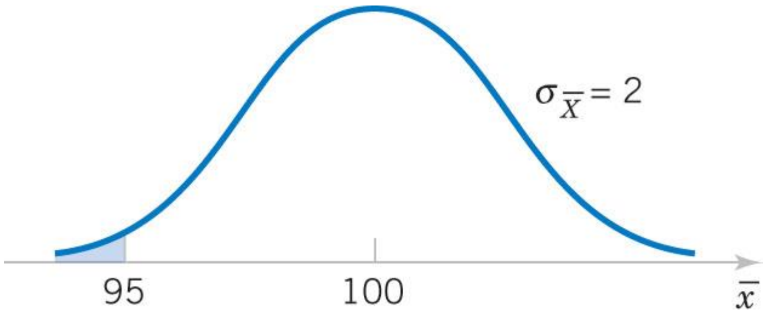


Figure 7-2 Desired probability is shaded

Answer:

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2.0$$

$$\Phi\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}\right) = \Phi\left(\frac{95-200}{2}\right)$$

$$= \Phi(-2.5) = 0.0062$$

0.0062	= NORMSDIST(-2.5)
A rare event at less than 1%.	

Example 7-2: Central Limit Theorem

Suppose that a random variable X has a continuous uniform distribution:

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size n = 40.

Distribution is normal by the CLT.

$$\mu = \frac{b+a}{2} = \frac{6+4}{2} = 5.0$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(6-4)^2}{12} 1/3$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{1/3}{40} = \frac{1}{120}$$

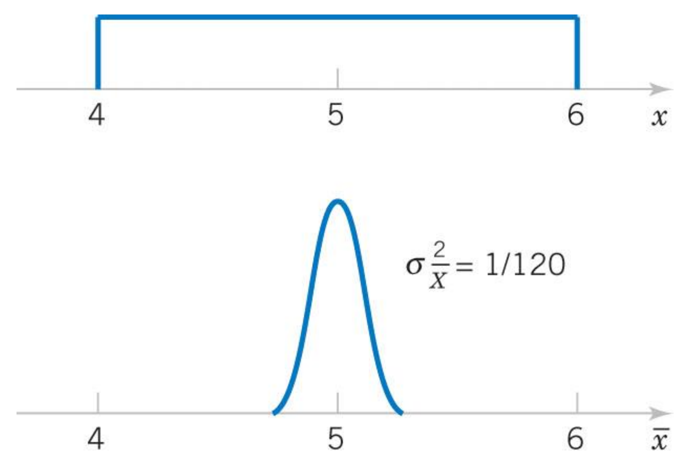


Figure 7-3 Distributions of *X* and *X*-bar

Two Populations

We have two independent normal populations. What is the distribution of the difference of the sample means?

The sampling distribution of $\bar{X}_1 - \bar{X}_2$ is:

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} - \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

The distribution of $\bar{X}_1 - \bar{X}_2$ is normal if:

(1) n_1 and n_2 are both greater than 30,

regardless of the distributions of X_1 and X_2.

(2) n_1 and n_2 are less than 30,

while the distributions are somewhat normal.

Sampling Distribution of a Difference in Sample Means

- If we have two independent populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2,
- And if X-bar$_1$ and X-bar$_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations:
- Then the sampling distribution of:

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \qquad (7\text{-}4)$$

  is approximately standard normal, if the conditions of the central limit theorem apply.
- If the two populations are normal, then the sampling distribution is exactly standard normal.

Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a normal-distributed random variable with parameters shown (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters as shown (new).

Random samples are selected from the "old" process and "new" process as shown.

What is the probability the difference in the two sample means is at least 25 hours?

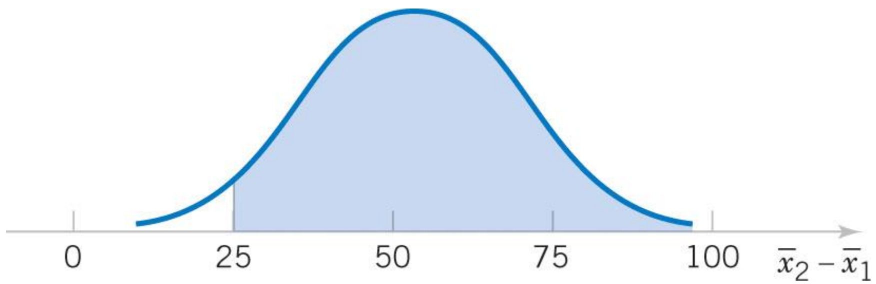


Figure 7-4 Sampling distribution of the sample mean difference.

	Process		
	Old (1)	New (2)	Diff (2-1)
x-bar =	5,000	5,050	50
s =	40	30	50
n =	16	25	
Calculations			
$s / \sqrt{n}$ =	10	6	11.7
		z =	-2.14
P(xbar$_2$-xbar$_1$ > 25) = P(Z > z) =			0.9840
		= 1 - NORMSDIST(z)	

General Concepts of Point Estimation

- We want point estimators that are:
  - Are unbiased.
  - Have a minimal variance.
- We use the standard error of the estimator to calculate its mean square error.

Unbiased Estimators Defined

The point estimator $\hat{\Theta}$ is an unbiased estimator for the parameter θ if:

$$E\left(\hat{\Theta}\right) = \theta \qquad (7\text{-}5)$$

If the estimator is not unbiased, then the difference:

$$E\left(\hat{\Theta}\right) - \theta \qquad (7\text{-}6)$$

is called the bias of the estimator $\hat{\Theta}$.

The mean of the sampling distribution of $\hat{\Theta}$ is equal to θ.

Example 7-4: Sample Man & Variance Are Unbiased-1

- X is a random variable with mean μ and variance σ^2. Let X_1, $X_2, \ldots, X_n$ be a random sample of size n.
- Show that the sample mean (X-bar) is an unbiased estimator of μ.

$$E\left(\bar{X}\right) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right] = \frac{n\mu}{n} = \mu$$

Example 7-4: Sample Man & Variance Are Unbiased-2

Show that the sample variance (S^2) is a unbiased estimator of σ².

$$E\left(S^2\right)=E\left(\frac{\sum_{i=1}^{n}\left(X-\bar{X}\right)^2}{n-1}\right)=\frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i^2+\bar{X}^2-2\bar{X}X_i\right)\right]$$

$$=\frac{1}{n-1}\left[E\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\right]=\frac{1}{n-1}\left[\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\bar{X}^2\right)\right]$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\sigma^2/n\right)\right]$$

$$=\frac{1}{n-1}\left[n\mu^2+n\sigma^2-n\mu^2-\sigma^2\right]=\frac{1}{n-1}\left[\left(n-1\right)\sigma^2\right]=\sigma^2$$

Other Unbiased Estimators of the Population Mean

$$\text{Mean} = \bar{X} = \frac{110.4}{10} = 11.04$$

$$\text{Median} = \tilde{X} = \frac{10.3 + 11.6}{2} = 10.95$$

$$\text{Trimmed mean} = \frac{110.04 - 8.5 - 14.1}{8} = 10.81$$

- All three statistics are unbiased.
  - Do you see why?
- Which is best?
  - We want the most reliable one.

i	x_i	x_i'
1	12.8	8.5
2	9.4	8.7
3	8.7	9.4
4	11.6	9.8
5	13.1	10.3
6	9.8	11.6
7	14.1	12.1
8	8.5	12.8
9	12.1	13.1
10	10.3	14.1
Σ	110.4	

Choosing Among Unbiased Estimators

Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of θ.

The variance of $\hat{\Theta}_1$ is less than the variance of $\hat{\Theta}_2$.

$\therefore$ $\hat{\Theta}_1$ is preferable.

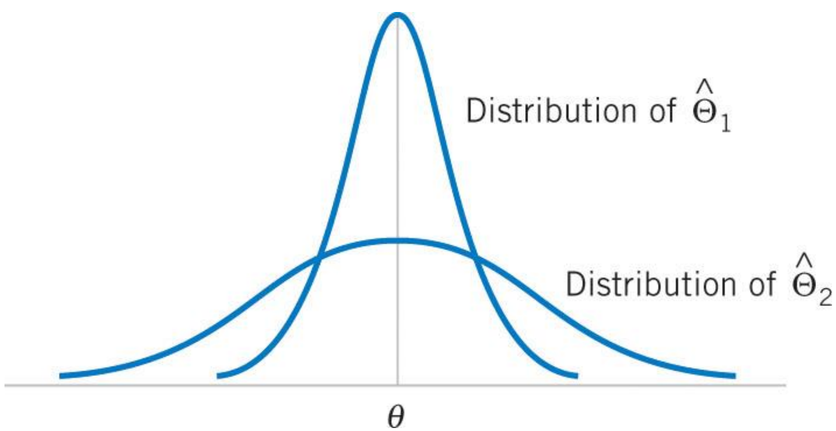


Figure 7-5 The sampling distributions of two unbiased estimators.

Minimum Variance Unbiased Estimators

- If we consider all unbiased estimators of θ, the one with the smallest variance is called the minimum variance unbiased estimator (MVUE).
- If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, then the sample X-bar is the MVUE for μ.
- The sample mean and a single observation are unbiased estimators of μ. The variance of the:
  - Sample mean is σ^2/n
  - Single observation is σ^2
  - Since $\sigma^2/n \le \sigma^2$, the sample mean is preferred.

Standard Error of an Estimator

The standard error of an estimator Θ is its standard deviation, given by

$$\sigma_{\Theta} = \sqrt{V(\Theta)}.$$

If the standard error involves unknown parameters that can be estimated,

substitution of these values into $\sigma_{\hat{\Theta}}$

produces an estimated standard error, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Equivalent notation: $\hat{\sigma}_{\hat{\Theta}} = s_{\hat{\Theta}} = se(\hat{\Theta})$

If the X_i are $\sim N(\mu, \sigma)$, then $\bar{X}$ is normally distributed,

and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$. If σ is not known, then $\hat{\sigma}_{\bar{X}} = \dfrac{s}{\sqrt{n}}$.

Example 7-5: Thermal Conductivity

- These observations are 10 measurements of thermal conductivity of Armco iron.
- Since σ is not known, we use s to calculate the standard error.
- Since the standard error is 0.2% of the mean, the mean estimate is fairly precise. We can be very confident that the true population mean is 41.924 ± 2(0.0898).

x_i	
41.60	
41.48	
42.34	
41.95	
41.86	
42.18	
41.72	
42.26	
41.81	
42.04	
41.924	= Mean
0.284	= Std dev (s)
0.0898	= Std error

Mean Squared Error

The mean squared error of an estimator Θ of the parameter θ is defined as:

$$\mathrm{MSE}\left(\Theta\right) = E\left(\Theta - \theta\right)^2 \qquad (7\text{-}7)$$

Can be rewritten as $= E\left[\Theta - E\left(\Theta\right)\right]^2 + \left[\theta - E\left(\Theta\right)\right]^2$

$$= V\left(\Theta\right) + \left(\text{bias}\right)^2$$

Conclusion: The mean squared error (MSE) of the estimator is equal to the variance of the estimator plus the bias squared. It measures both characteristics.

Relative Efficiency

- The MSE is an important criterion for comparing two estimators.

$$\text{Relative efficiency} = \frac{\text{MSE}\left(\hat{\Theta}_1\right)}{\text{MSE}\left(\hat{\Theta}_2\right)}$$

- If the relative efficiency is less than 1, we conclude that the 1st estimator is superior to the 2nd estimator.

Optimal Estimator

- A biased estimator can be preferred to an unbiased estimator if it has a smaller MSE.
- Biased estimators are occasionally used in linear regression.
- An estimator whose MSE is smaller than that of any other estimator is called an optimal estimator.

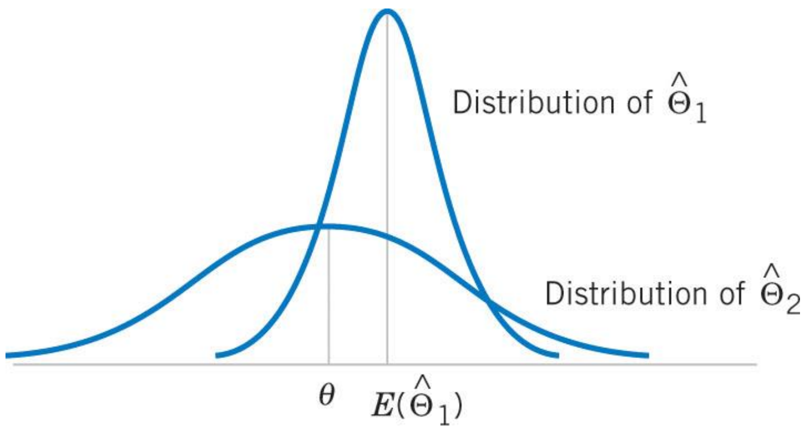


Figure 7-6 A biased estimator has a smaller variance than the unbiased estimator.

Methods of Point Estimation

- There are three methodologies to create point estimates of a population parameter.
  - Method of moments
  - Method of maximum likelihood
  - Bayesian estimation of parameters
- Each approach can be used to create estimators with varying degrees of biasedness and relative MSE efficiencies.

Method of Moments

- A "moment" is a kind of an expected value of a random variable.
- A population moment relates to the entire population or its representative function.
- A sample moment is calculated like its associated population moments.

Moments Defined

- Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability $f(x)$, where $f(x)$ can be either a:
  - Discrete probability mass function, or
  - Continuous probability density function
- The k^{th} population moment (or distribution moment) is $E(X^k)$, k = 1, 2, ….
- The k^{th} sample moment is $(1/n)\Sigma X^k$, k = 1, 2, ….
- If k = 1 (called the first moment), then:
  - Population moment is μ.
  - Sample moment is x-bar.
- The sample mean is the moment estimator of the population mean.

Moment Estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or a probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$.

The moment estimators $\Theta_1, \Theta_2, \ldots, \Theta_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting simultaneous equations for the unknown parameters.

Example 7-6: Exponential Moment Estimator-1

- Suppose that X_1, X_2, …, X_n is a random sample from an exponential distribution with parameter λ.
- There is only one parameter to estimate, so equating population and sample first moments, we have $E(X)$ = X-bar.
- $E(X) = 1/\lambda$ = x-bar
- $\lambda = 1/x$-bar is the moment estimator.

Example 7-6: Exponential Moment Estimator-2

- As an example, the time to failure of an electronic module is exponentially distributed.
- Eight units are randomly selected and tested. Their times to failure are shown.
- The moment estimate of the λ parameter is 0.04620.

x_i	
11.96	
5.03	
67.40	
16.07	
31.50	
7.73	
11.10	
22.38	
21.646	= Mean
0.04620	= λ est

Example 7-7: Normal Moment Estimators

Suppose that X_1, X_2, …, X_n is a random sample from a normal distribution with parameter μ and σ^2. So $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$.

$$\mu = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{n}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right] = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n} \text{ (biased)}$$

Example 7-8: Gamma Moment Estimators-1

Parameters = Statistics

$$\frac{r}{\lambda} = E(X) = \bar{X} \text{ is the mean}$$

$$\frac{r}{\lambda^2} = E(X^2) - E(X)^2 \text{ is the variance or}$$

$$\frac{r(r+1)}{\lambda^2} = E(X^2) \text{ and now solving for } r \text{ and } \lambda:$$

$$\hat{r} = \frac{\bar{X}^2}{(1/n)\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

$$\lambda = \frac{\bar{X}}{(1/n)\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Example 7-8: Gamma Moment Estimators-2

Using the exponential example data shown, we can estimate the parameters of the gamma distribution.

x_i	x_i^2
11.96	143.0416
5.03	25.3009
67.40	4542.7600
16.07	258.2449
31.50	992.2500
7.73	59.7529
11.10	123.2100
22.38	500.8644

x-bar =	21.646
ΣX^2 =	6645.4247

$$\hat{r} = \frac{\bar{X}^2}{(1/n)\sum_{i=1}^{n} X_i^2 - \bar{X}^2} = \frac{21.646^2}{(1/8)6645.4247 - 21.646^2} = 1.29$$

$$\lambda = \frac{\bar{X}}{(1/n)\sum_{i=1}^{n} X_i^2 - \bar{X}^2} = \frac{21.646}{(1/8)6645.4247 - 21.646^2} = 0.0598$$

Maximum Likelihood Estimators

- Suppose that X is a random variable with probability distribution $f(x{:}\theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the likelihood function of the sample is:

$$L(\theta) = f(x_1{:}\ \theta) \cdot f(x_2;\ \theta) \cdot \ldots \cdot f(x_n{:}\ \theta) \qquad (7\text{-}9)$$

- Note that the likelihood function is now a function of only the unknown parameter θ. The maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$.
- If X is a discrete random variable, then $L(\theta)$ is the probability of obtaining those sample values. The MLE is the θ that maximizes that probability.

Example 7-9: Bernoulli MLE

Let X be a Bernoulli random variable. The probability mass function is $f(x;p) = p^x(1-p)^{1-x}$, $x = 0, 1$ where P is the parameter to be estimated. The likelihood function of a random sample of size n is:

$$L(p) = p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdot \ldots \cdot p^{x_n}(1-p)^{1-x_n}$$

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

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

$$\frac{d\ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-p)} = 0$$

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

Example 7-10: Normal MLE for μ

Let X be a normal random variable with unknown mean μ and known variance σ^2. The likelihood function of a random sample of size n is:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

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

$$\frac{d\ln L(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X} \text{ (same as moment estimator)}$$

Example 7-11: Exponential MLE

Let X be a exponential random variable with parameter λ. The likelihood function of a random sample of size n is:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

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

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\lambda = n \Big/ \sum_{i=1}^{n} x_i = 1 \big/ \overline{X} \quad \text{(same as moment estimator)}$$

Why Does MLE Work?

- From Examples 7-6 & 11 using the 8 data observations, the plot of the ln $L(\lambda)$ function maximizes at $\lambda = 0.0462$. The curve is flat near max indicating estimator not precise.
- As the sample size increases, while maintaining the same x-bar, the curve maximums are the same, but sharper and more precise.
- Large samples are better☺

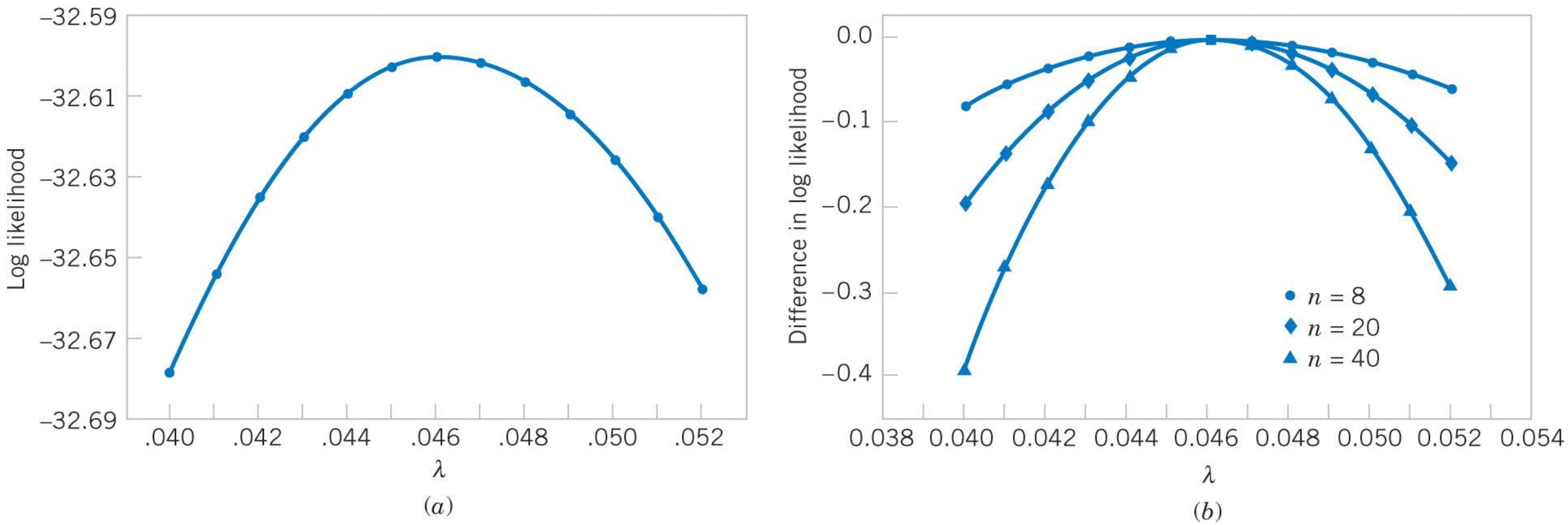


Figure 7-7 Log likelihood for exponential distribution. (a) n = 8, (b) n = 8, 20, 40.

Example 7-12: Normal MLEs for μ & σ^2

Let X be a normal random variable with both unknown mean μ and variance σ^2. The likelihood function of a random sample of size n is:

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

$$=\frac{1}{\left(2\pi\sigma^2\right)^{n/2}}e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

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

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)=0$$

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\left(\sigma^2\right)}=\frac{-n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2=0$$

$$\mu=\bar{X}\quad\text{and}\quad\sigma^2=\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n}$$

Properties of an MLE

Under very general and non-restrictive conditions,

when the sample size n is large and if Θ is the MLE of the parameter ,

(1) Θ is an approximately unbiased estimator for θ, i.e., $\left[E\left(\Theta\right)=\theta\right]$

(2) The variance of Θ is nearly as small as the variance that could be obtained with any other estimator, and

(3) Θ has an approximate normal distribution.

<u>Notes</u>:

- Mathematical statisticians will often prefer MLEs because of these properties. Properties (1) and (2) state that MLEs are MVUEs.
- To use MLEs, the distribution of the population must be known or assumed.

Importance of Large Sample Sizes

- Consider the MLE for σ^2 shown in Example 7-12:

$$E\left(\sigma^2\right) = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n} = \frac{n-1}{n}\sigma^2$$

Then the bias is:

$$E\left(\sigma^2\right) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

- Since the bias is negative, the MLE underestimates the true variance σ^2.
- The MLE is an asymptotically (large sample) unbiased estimator. The bias approaches zero as *n* increases.

Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators (MLEs) of the parameters $\theta_1, \theta_2, \ldots, \theta_k$.

Then the MLEs for any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$

This property is illustrated in Example 7-13.

Example 7-13: Invariance

For the normal distribution, the MLEs were:

$$\mu = \bar{X} \quad \text{and} \quad \sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}$$

To obtain the MLE of the function $h\left(\mu,\sigma^2\right) = \sqrt{\sigma^2} = \sigma$,

substitute the estimators μ and σ^2 into the function h:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}}$$

which is not the sample standard deviation s.

Complications of the MLE Method

The method of maximum likelihood is an excellent technique, however there are two complications:

1. It may not be easy to maximize the likelihood function because the derivative function set to zero may be difficult to solve algebraically.
2. The likelihood function may be impossible to solve, so numerical methods must be used.

The following two examples illustrate.

Example 7-14: Uniform Distribution MLE

Let X be uniformly distributed on the interval 0 to a.

$$f(x) = 1/a \text{ for } 0 \le x \le a$$

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n} = a^{-n} \text{ for } 0 \le x_i \le a$$

$$\frac{dL(a)}{da} = \frac{-n}{a^{n+1}} = -na^{-(n+1)}$$

$$a = \max(x_i)$$

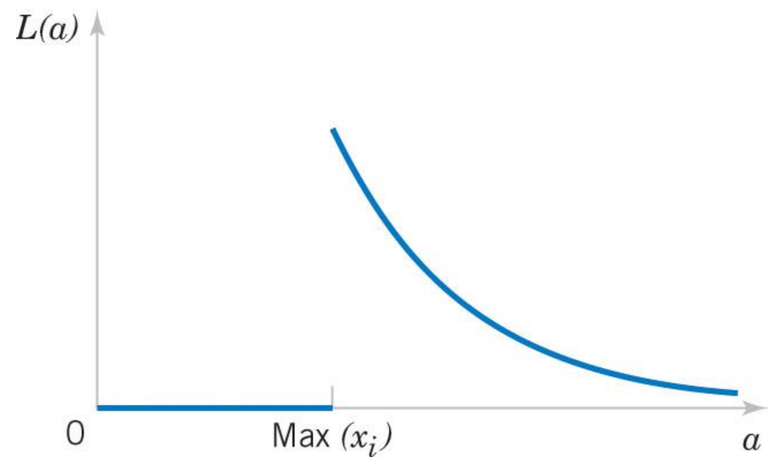


Figure 7-8 The likelihood function for this uniform distribution

Calculus methods don't work here because $L(a)$ is maximized at the discontinuity.

Clearly, a cannot be smaller than max(x_i), thus the MLE is max(x_i).

Example 7-15: Gamma Distribution MLE-1

Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution. The log of the likelihood function is:

$$\ln L(r,\lambda) = \ln\left(\prod_{i=1}^{n} \frac{\lambda^r x_i^{r-1} e^{-\lambda x_i}}{\Gamma(r)}\right)$$

$$= nr\ln(\lambda) + (r-1)\sum_{i=1}^{n}\ln(x_i) - n\ln[\Gamma(r)] - \lambda\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L(r,\lambda)}{\partial r} = n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i) - n\frac{\Gamma'(r)}{\Gamma(r)} = 0$$

$$\frac{\partial \ln L(r,\lambda)}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\lambda = \frac{\hat{r}}{\bar{x}} \quad \text{and} \quad n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i) = n\frac{\Gamma'(r)}{\Gamma(r)}$$

There is no closed solution for $\hat{r}$ and $\hat{\lambda}$.

Example 7-15: Gamma Distribution MLE-2

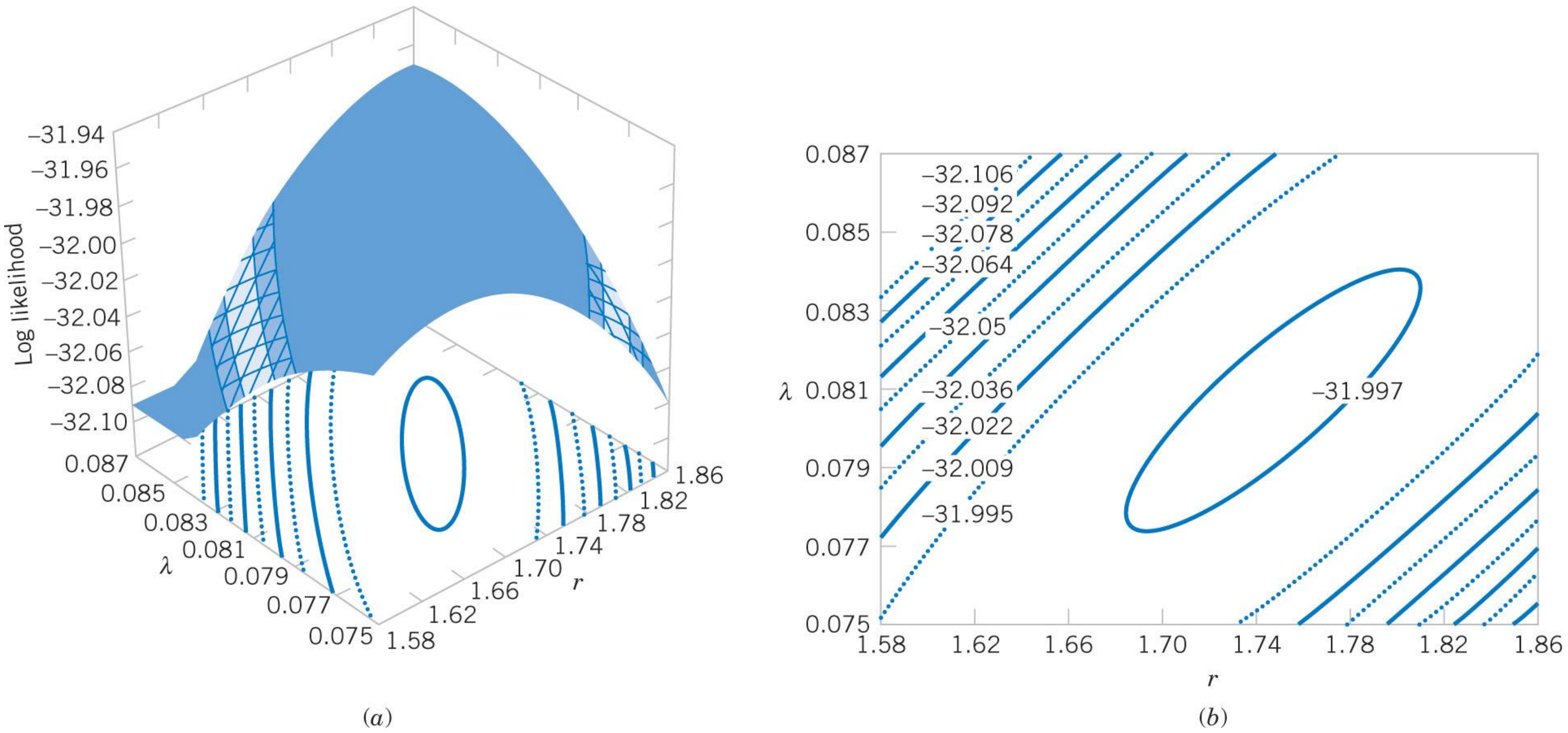


Figure 7-9 Log likelihood for the gamma distribution using the failure time data (n=8). (a) is the log likelihood surface. (b) is the contour plot. The log likelihood function is maximized at r = 1.75, λ = 0.08 using numerical methods. Note the imprecision of the MLEs inferred by the flat top of the function.

Bayesian Estimation of Parameters-1

- The moment and likelihood methods interpret probabilities as relative frequencies and are called objective frequencies.
- The Bayesian method combines sample information with prior information.
- The random variable X has a probability distribution of parameter θ called $f(x|\theta)$. θ could be determined by classical methods.
- Additional information about θ can be expressed as $f(\theta)$, the prior distribution, with mean μ_0 and variance σ_0^2, with θ as the random variable. Probabilities associated with $f(\theta)$ are subjective probabilities.
- The joint distribution is $f(x_1, x_2, \ldots, x_n, \theta)$
- The posterior distribution is $f(\theta|x_1, x_2, \ldots, x_n)$ is our degree of belief regarding θ after gathering data

Bayesian Estimation of Parameters-2

- Now putting these together, the joint is:
  - $f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n \mid \theta) \cdot f(\theta)$
- The marginal is:

$$f\left(x_1, x_2, \ldots, x_n\right) = \begin{cases} \sum_{\theta} f\left(x_1, x_2, \ldots, x_n, \theta\right), & \text{for } \theta \text{ discrete} \\ \int_{-\infty}^{\infty} f\left(x_1, x_2, \ldots, x_n, \theta\right) d\theta, & \text{for } \theta \text{ continuous} \end{cases}$$

- The desired posterior distribution is:

$$f\left(\theta \mid x_1, x_2, \ldots, x_n\right) = \frac{f\left(x_1, x_2, \ldots, x_n, \theta\right)}{f\left(x_1, x_2, \ldots, x_n\right)}$$

- And the Bayesian estimator of θ is the expected value of the posterior distribution

Example 7-16: Bayes Estimator for a Normal Mean-1

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution unknown mean μ and known variance σ^2. Assume that the prior distribution for μ is:

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\mu-\mu_0)^2/2\sigma_0^2} = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\left(\mu^2-2\mu\mu_0+\mu_0^2\right)/2\sigma_0^2}$$

The joint distribution of the sample is:

$$f(x_1, x_2, \ldots, x_n \mid \mu) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\left(1/2\sigma^2\right)\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\left(1/2\sigma^2\right)\left(\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right)}$$

Example 7-16: Bayes Estimator for a Normal Mean-2

Now the joint distribution of the sample and μ is:

$$f\left(x_1, x_2, \ldots, x_n, \mu\right) = f\left(x_1, x_2, \ldots, x_n \mid \mu\right) \cdot f\left(\mu\right)$$

$$= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}\sqrt{2\pi\sigma_0^2}} e^u$$

$$\text{where } u = \left(\frac{-1}{2}\right)\left[\mu^2\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right) - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}\right) + \frac{\sum x_i^2}{\sigma^2} + \frac{\mu_0^2}{\sigma_0^2}\right]$$

$$= h_1\left(\cdot\right) e^{-(1/2)\left[\mu^2\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right) - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}\right)\right]} \text{ \& completing the square}$$

$$= h_2\left(\cdot\right) e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{\left(\sigma^2/n\right)\mu_0}{\sigma_0^2 + \sigma^2/n} + \frac{\bar{x}\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)\right]^2}$$

$$f\left(\mu \mid x_1, x_2, \ldots, x_n\right) = h_3\left(\cdot\right) e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{\left(\sigma^2/n\right)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2 + \sigma^2/n}\right)\right]} \text{ is the posterior distribution}$$

$$h_i\left(\cdot\right) = \text{ function to collect unneeded components (not } \mu\text{)}$$

Example 7-16: Bayes Estimator for a Normal Mean-3

- After all that algebra, the bottom line is:

$$E(\mu) = \mu = \frac{\left(\sigma^2/n\right)\mu_0 + \sigma_0^2 \overline{x}}{\sigma_0^2 + \sigma^2/n}$$

$$V(\mu) = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{-1} = \frac{\sigma_0^2\left(\sigma^2/n\right)}{\sigma_0^2 + \sigma^2/n}$$

- Observations:
  - Estimator is a weighted average of μ_0 and *x*-bar.
  - x-bar is the MLE for μ.
  - The importance of μ_0 decreases as *n* increases.

Example 7-16: Bayes Estimator for a Normal Mean-4

To illustrate:

- The prior parameters: $\mu_0 = 0$, $\sigma_0^2 = 1$
- Sample: $n = 10$, x-bar $= 0.75$, $\sigma^2 = 4$

$$\mu = \frac{\left(\sigma^2 / n\right)\mu_0 + \sigma_0^2 \overline{x}}{\sigma_0^2 + \sigma^2 / n}$$

$$= \frac{(4/10)0 + 1(0.75)}{1 + (4/10)} = 0.536$$

Important Terms & Concepts of Chapter 7

Bayes estimator
Bias in parameter estimation
Central limit theorem
Estimator vs. estimate
Likelihood function
Maximum likelihood estimator
Mean square error of an estimator
Minimum variance unbiased estimator
Moment estimator
Normal distribution as the sampling distribution of the:
- sample mean
- difference in two sample means

Parameter estimation
Point estimator
Population or distribution moments
Posterior distribution
Prior distribution
Sample moments
Sampling distribution
An estimator has a:
- Standard error
- Estimated standard error

Statistic
Statistical inference
Unbiased estimator