Chapter 7: Parameter Estimation

ST2334 Probability and Statistics[1]
(Academic Year 2014/15, Semester 1)

Department of Statistics & Applied Probability (DSAP)
National University of Singapore (NUS)[1]

Outline

Introduction I

Learning Outcomes

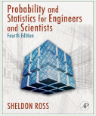
Questions to Address: ♦ Construction of likelihood function based on a sample ♦ How to derive the maximum likelihood (ML) estimator ♦ Difference between point estimation & interval estimation ♦ Formal & informal conclusion of a confidence interval (c.i.) ♦ Derivation of a c.i. by inverting a probability statement of an ML estimator ♦ Construction of c.i.'s for mean of a norml population with known or unknown variance ♦ Construction of c.i.'s for variance of a normal population with unknown mean ♦ Construction of c.i.'s for difference in means of 2 independent normal populations with known variances or with unknown but equal variances ♦ Construction of approximate c.i. for the mean of a Bernoulli r.v.

Introduction II

Learning Outcomes (continued)

Concept & Terminology: ♦ target parameter ♦ parameter space ♦ statistic ♦ point estimator ♦ likelihood ♦ maximum likelihood estimator/estimate ♦ point/interval estimation ♦ confidence coefficient/coverage probability ♦ confidence interval ♦ pooled variance estimator

Mandatory Reading



✏ Section 7.1 – Section 7.5 (except Section 7.2.1, one-sided lower or upper confidence intervals, & sample size determination problem)

Introduction III

- **Recall** (in Ch.6):
  - use a statistic constructed by a sample from a population of interest
  - understanding the sampling distribution of the statistic helps to estimate the unkown parameter of the population
  - mention that $\overline{X}$ & S^2 are appropriate statistics in estimating a population mean & a population variance, respectively
- In estimating any other parameters:

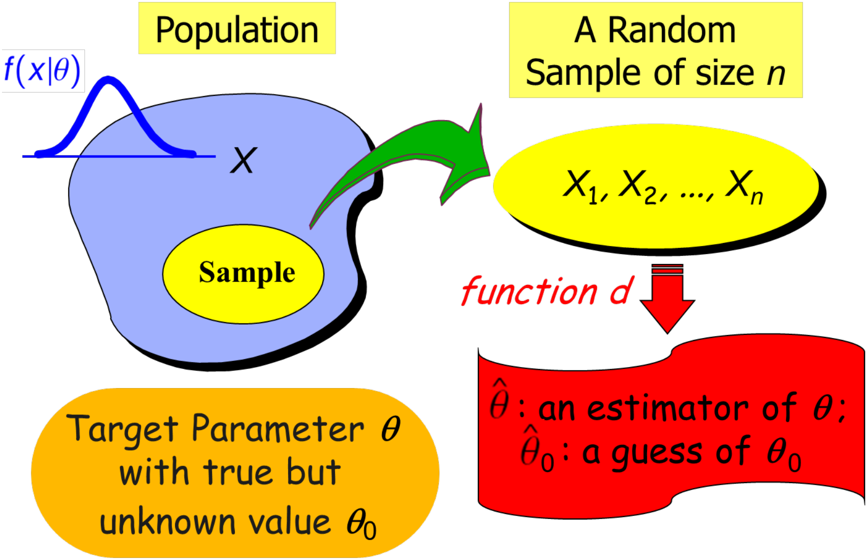
In general, how to find an appropriate statistic?

Introduction IV

Here are some *notation & terminologies* in parameter estimation:

$f(x) \equiv f(x\|\theta)$	The density (pdf/pmf) of the underlying distribution F, which is parametrized/characterized by θ with an unknown value
θ	The *target parameter*
Θ	The *set* of all possible values of θ, called the *parameter space*
θ_0	The true but unknown *value* of the parameter
$\hat{\theta} = d(X_1, \ldots, X_n)$	A **statistic** called the *point estimator of θ*, which is a **function d** of $X_1, \ldots, X_n$ observed from f/F
$\hat{\theta}_0 = d(x_1, \ldots, x_n)$	A particular *value* obtained based on n known values, $x_1, \ldots, x_n$, from a sample as a guess of the unknown value θ_0

Introduction V



Maximum Likelihood Estimators I

- Assume that we observe data from f which is characterized by a parameter θ with unknown value θ_0

Definition

The *likelihood* function of θ is equivalent to the joint density of $X_1, \ldots, X_n$ evaluated at $x_1, \ldots, x_n$:

$$\text{lik}(\theta) = \begin{cases} f(x_1, \ldots, x_n|\theta), & X_i\text{'s are cont.} \\ p(x_1, \ldots, x_n|\theta), & X_i\text{'s are discrete} \end{cases}$$

- The likelihood is a function of θ: in practice, we are given the data as $X_i = x_i,\ i = 1, \ldots, n$, where x_i's are some known values
- *Meaning of lik(θ) in terms of prob*:
  - When $X_1, \ldots, X_n$ are discrete r.v.'s, lik(θ) is the joint prob of observing $\{X_1 = x_1, \ldots, X_n = x_n\}$ when the value of the parameter is θ
  - When $X_1, \ldots, X_n$ are cont. r.v.'s, lik(θ) is proportional to the joint prob of observing $\{X_1 = x_1, \ldots, X_n = x_n\}$ when the value of the parameter is θ

Maximum Likelihood Estimators II

Definition

The *maximum likelihood (ML) estimate*, denoted by $\hat{\theta}_0$, of θ is the value of θ that maximizes $\text{lik}(\theta)$, *i.e.,* makes the observed data "most probable" or "most likely"

- Usually *obtained by differentiation*
- Easier to maximize the natural logarithm of $\text{lik}(\theta)$ called the *log likelihood*, denoted by $\ell(\theta) \equiv \log[\text{lik}(\theta)]$
- After the maximization, $\hat{\theta}_0$ must take a form of $d(x_1, \ldots, x_n)$ for some function d

Definition

The *maximum likelihood estimator*, denoted by $\hat{\theta}$, of θ is a statistic defined by $\hat{\theta} = d(\mathbf{X}) = d(X_1, \ldots, X_n)$ for some function d

Example 1: ML Estimation for $Exp(\lambda)$

Let $X_1, \ldots, X_n$ be a random sample of size n from $Exp(\lambda)$ with pdf $f(x) = \lambda e^{-\lambda x}, x \geq 0$ for $\lambda \geq 0$. Find the ML estimator of λ

Solution:

1. The log likelihood of λ is

$$\ell(\lambda) = \sum_{i=1}^{n} [\log \lambda - \lambda x_i] = n \log \lambda - \lambda n \overline{x}$$

   where $\overline{x} = (1/n) \sum_{i=1}^{n} x_i$

2. Setting the 1st derivative of $\ell(\lambda)$ with respect to (w.r.t.) λ to zero yields

$$\frac{d\ell(\lambda)}{d\lambda} = \ell'(\lambda) = \frac{n}{\lambda} - n\overline{x} = 0$$

3. The solution yields the ML estimate of λ as $\hat{\lambda}_0 = \overline{x}^{-1}$

   $\Rrightarrow$ The ML estimator is $\hat{\lambda} = \overline{X}^{-1}$

Example 2: ML Estimation for $N(\mu, \sigma_0^2)$

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma_0^2)$ with a given variance $\sigma_0^2 > 0$. Find the ML estimator of μ

Solution:

1. The log likelihood of μ is

$$\ell(\mu) = -\frac{n}{2}\log\sigma_0^2 - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

2. Setting the 1st derivative of $\ell(\mu)$ w.r.t. μ to zero yields

$$\frac{d\ell(\mu)}{d\mu} = \ell'(\mu) = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

3. The solution yields the ML estimate of μ as $\hat{\mu}_0 = \overline{x}$
   $\Rrightarrow$ The ML estimator is $\hat{\mu} = \overline{X}$

Example 3: ML Estimation for $N(\mu, \sigma^2)$ I

1. If $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \tau \equiv \sigma^2)$ with both μ & τ unknown, their joint pdf is

$$f(x_1, \ldots, x_n | \mu, \tau) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{(x_i - \mu)^2}{2\tau}\right\}$$

2. The log likelihood of (μ, τ) is

$$\ell(\mu, \tau) = -\frac{n}{2}\log\tau - \frac{n}{2}\log(2\pi) - \frac{1}{2\tau}\sum_{i=1}^{n}(x_i - \mu)^2$$

3. The partials w.r.t. μ & τ are

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\tau}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{\partial \ell}{\partial \tau} = -\frac{n}{2\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Example 3: ML Estimation for $N(\mu, \sigma^2)$ II

4. Setting the partials to zero yield a set of 2 equations with 2 unknowns:

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

$$0 = -\frac{n}{2\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

5. The solutions give the ML estimates as

$$\hat{\mu}_0 = \overline{x} \qquad \& \qquad \hat{\tau}_0 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\overline{x})^2$$

The ML estimators are $\hat{\mu} = \overline{X}$ & $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\overline{X}\right)^2$

Example 3: *ML Estimation for* $N(\mu, \sigma^2)$ *III*

Suppose that *an available random sample of size 10*, denoted by some values, $x_1, \ldots, x_{10}$, gives values of the sample moments as

$$\frac{1}{10}\sum_{i=1}^{10} x_i = .52 \qquad \& \qquad \frac{1}{10}\sum_{i=1}^{10} x_i^2 = .313$$

We conclude, *under the framework of point estimation*, that

> When we assume a normal distribution $N(\mu, \sigma^2)$ to be the underlying distribution from which the sample of size 10 are realized, we estimate μ to be .52 & σ^2 to be $.313 - .52^2 = .2064^2$ based on the 10 observations

Are the true values of μ & σ^2 really given by $.52$ & $.2064^2$

Interval Estimation I

The answer to the question at the previous page is

- We "expect" $\hat{\theta}_0$ to be close to the true values θ_0, but it is almost *impossible* that $\hat{\theta}_0 = \theta_0$
- Estimates in 2 extremes:
  1. a single value $\hat{\theta}_0$ (which misses θ_0 for sure)
  2. a set/an interval identical to Θ (which includes θ_0 for sure)
- A compromise: An interval $(L, U) \neq \Theta$ such that (s.t.) we are

  $100(1-\alpha)\%$ sure that $\theta_0 \in (L, U)$

  for some *pre–specified* $0 < \alpha < 1$
- $1-\alpha$: *confidence coefficient* or *coverage probability* which measures our *degree of certainty* about $L < \theta_0 < U$

Definition

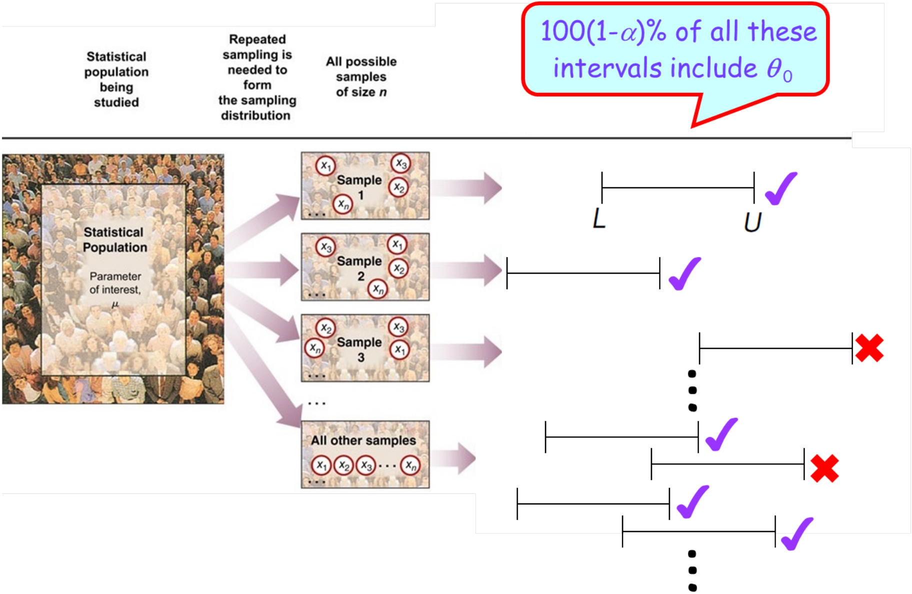
A *100$(1-\alpha)$% confidence interval (c.i.)* for a parameter θ with true but unknown value θ_0 is a random interval, (L, U), where

$$L = d_1(X_1, \ldots, X_n) \qquad \& \qquad U = d_2(X_1, \ldots, X_n)$$

are *2 statistics* (*i.e.,* functions of r.v.'s, $X_1, \ldots, X_n$, with joint density depending on θ) s.t. if we were to take all possible random samples of size n & form an interval from each set of samples, $100(1-\alpha)$% of all the resulting intervals would contain the true value θ_0

- *Remark*: Here, we *ONLY* look at c.i. of the form (L, U), which is called a *two–sided c.i.*. There exist *one–sided c.i.'s* of the form (L, ∞) or $(-\infty, U)$, but they are out of our syllabus

Interval Estimation III



Interval Estimation IV

- **An informal conclusion**: In practice, based on known values, $x_1, \ldots, x_n$, from 1 sample of size n, we can obtain only 1 particular c.i. of 2 fixed endpoints, *i.e.,*

$$(d_1(x_1, \ldots, x_n), d_2(x_1, \ldots, x_n)) = (a, b)$$

$\Rrightarrow$ There is 100(1 – α)% chance that such a c.i. is 1 of all those intervals that contain θ_0

$\Rrightarrow$ Informally, we *conclude that (a, b) contains the true value of θ*

- *Remark*: The above conclusion is *informal* or, indeed, *illegitimate*, as any interval either contain or does not contain θ_0

Finding The c.i.

One possible way to determine the 2 functions, d_1 & d_2, in the c.i. (L, U) for θ is by inverting a probability statement of an ML estimator of θ

c.i. For a Normal Mean (Known σ^2) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma_0^2)$ with unknown mean μ & a given variance $\sigma_0^2 > 0$

1. $\overline{X}$ is the ML estimator of μ (Example 2)
2. Sampling distribution of $\overline{X}$: $\overline{X} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right)$ (Equation (2) in Ch.6) as μ has a true but unknown value μ_0
3. Start with the following exact probability statement, for any $0 < \alpha < 1$,

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu_0}{\sqrt{\sigma_0^2/n}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\mu_0 - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} < \overline{X} < \mu_0 + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) = 1 - \alpha$$

c.i. For a Normal Mean (Known σ^2) II

4. Manipulate the inequalities to get

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} < \mu_0 < \overline{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) = 1 - \alpha$$

$\Rrightarrow$ The random interval $\left(\overline{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right)$ contains μ_0 with prob $1 - \alpha$

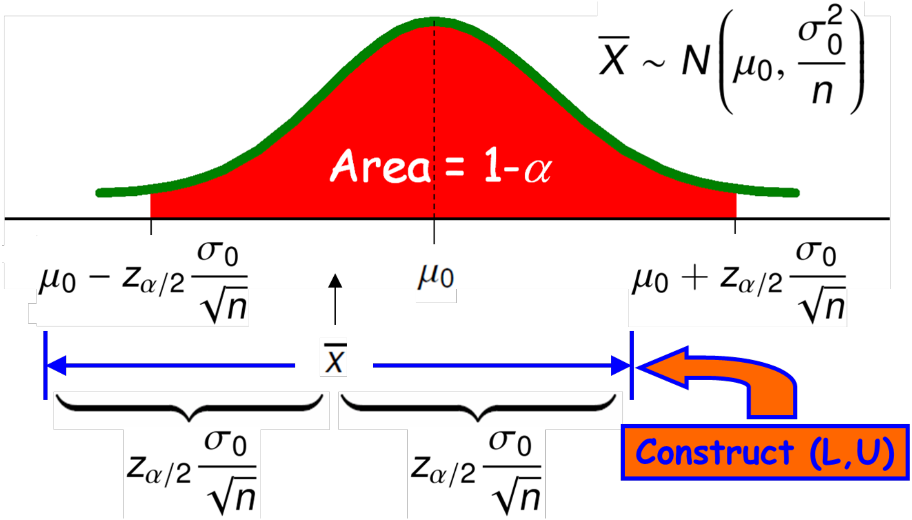
c.i. For a Normal Mean (Known Variance)

Based on a sample of size n from a $N(\mu, \sigma_0^2)$ population with unknown mean μ & known variance σ_0^2, a $100(1-\alpha)\%$ c.i. for μ is given by

$$(L, U) = \overline{x} \pm z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} = \left(\overline{x} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) \quad (1)$$

where $\overline{x}$ is the sample mean

c.i. For a Normal Mean (Known σ^2) III



Example 4: *Estimating a Normal Mean (Known σ^2) I*

The student body at many community colleges is considered a "commutor population". The student activities office wishes to obtain an answer to the question, "How far (one-way) does the average community college student commute to college each day?" (Typically the "average student's commute distance" is meant to be the "mean distance" commuted by all students who commute.) A random sample of 30 commuting students was identified, & the one-way distance each commuted was obtained. The resulting sample mean distance was 10.22 miles. Assume that the commute distance by all commuting students is normally distributed with 6 miles as standard deviation. Estimate the mean one-way distance μ commuted by all commuting students using a point estimate & a 95% c.i.

Solution:

1. The point estimate of μ is the sample mean $\overline{x} = 10.22$

Example 4: *Estimating a Normal Mean (Known σ^2) II*

2. Here, the population standard deviation $\sigma_0 = 6$ miles & the sample size $n = 30$. Then, with $\alpha = .05$, $z_{.025} = 1.96$ & a 95% c.i. for the mean one-way distance μ is given by

$$\begin{aligned} \bar{x} \pm z_{.025}\frac{\sigma_0}{\sqrt{n}} &= 10.22 \pm 1.96\frac{6}{\sqrt{30}} \\ &= 10.22 \pm 2.147 \\ &= (8.073, 12.367) \end{aligned}$$

3. An *informal conclusion*: we are 95% sure that the mean one-way commute distance is between 8.073 miles & 12.367 miles

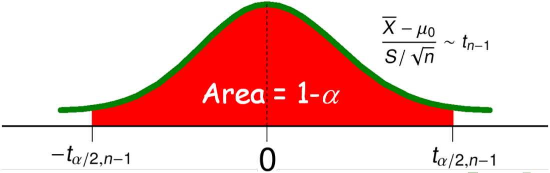
Formally, we conclude that if we collect a large number of samples of size 30 from this population, and construct 95% c.i.'s from each sample according to the above method/rule/formula, we would expect that 95% of all the constructed intervals contain the true value of the mean one-way commute distance by all commuting students

c.i. For a Normal Mean (Unknown σ^2) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$ with both mean μ & variance σ^2 unknown

1. $\overline{X}$ is the ML estimator of μ (Example 3)
2. $\dfrac{\overline{X} - \mu_0}{\sqrt{S^2/n}} \sim t_{n-1}$ (Equation (3) in Ch.6) as μ has a true but unknown value μ_0
3. Start with the following exact probability statement, for any $0 < \alpha < 1$,

$$P\left(-t_{\alpha/2,n-1} < \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}} < t_{\alpha/2,n-1}\right) = 1 - \alpha$$



c.i. For a Normal Mean (Unknown σ^2) II

4. Manipulate the inequalities to get

$$P\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu_0 < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$\Rightarrow$ The random interval $\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$ contains μ_0 with prob $1 - \alpha$

c.i. For a Normal Mean (Unknown Variance)

Based on a sample of size n from a $N(\mu, \sigma^2)$ population with unknown mean μ & unknown variance σ^2, a $100(1-\alpha)\%$ c.i. for μ is given by

$$(L, U) = \overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = \left(\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

where $\overline{x}$ & s^2 are the sample mean & the sample variance, respectively

Example 5: *Estimating a Normal Mean (Unknown σ^2)*

With the population variance σ^2 unknown, repeat Example 4 when the sample standard deviation for the 30 students' distances is given by 7.5 miles

Solution:

1. The point estimate of μ remains the same as $\overline{x} = 10.22$
2. Here, the population variance is unknown & the sample size $n = 30$. With $\alpha = .05$, $t_{.025,29} = 2.045$ & a 95% c.i. for the mean one-way distance μ is given by

$$\begin{aligned}\overline{x} \pm t_{.025,n-1}\frac{s}{\sqrt{n}} &\approx 10.22 \pm 2.045\frac{7.5}{\sqrt{30}} \\ &= 10.22 \pm 2.8 \\ &= (7.42, 13.02)\end{aligned}$$

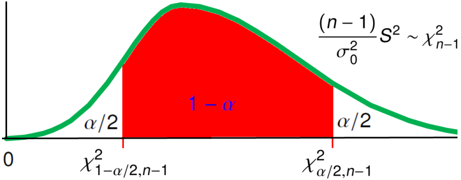
$\Rrightarrow$ We are 95% sure that the mean one-way commute distance is between 7.42 miles & 13.02 miles

c.i. For a Normal Variance (Unknown μ) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$ with both mean μ & variance σ^2 unknown

1. $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is the ML estimator of σ^2 (Example 3)
2. $\dfrac{n\hat{\sigma}^2}{\sigma_0^2} = \dfrac{(n-1)}{\sigma_0^2} S^2 \sim \chi^2_{n-1}$ (Property ❸ at Page 37 of Ch.6) as σ^2 has a true but unknown value σ_0^2
3. Start with the following exact probability statement, for any $0 < \alpha < 1$,

$$P\left(\chi^2_{1-\alpha/2, n-1} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{\alpha/2, n-1} \right)$$
$$= 1 - \alpha$$



c.i. For a Normal Variance (Unknown μ) II

4. Manipulate the inequalities to get

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma_0^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

$\Rrightarrow$ The random interval $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ contains σ_0^2 with prob $1-\alpha$

c.i. For a Normal Variance (Unknown Mean)

Based on a sample of size n from a $N(\mu, \sigma^2)$ population with unknown mean μ & unknown variance σ^2, a $100(1-\alpha)\%$ c.i. for σ^2 is given by

$$(L, U) = \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

where s^2 is the sample variance

Example 6: c.i. For a Normal Variance (Unknown μ)

Consider Example 4 with both μ & σ^2 unknown. Find a 95% c.i. for σ^2 when the sample standard deviation for the 30 students' distances is given by 7.5 miles

Solution:

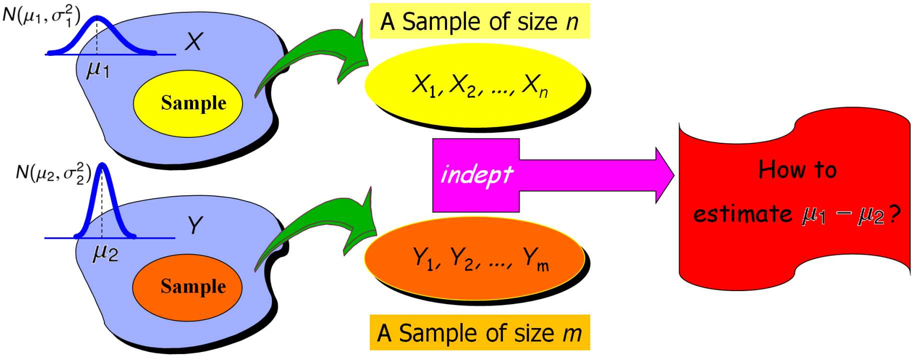
1. The point estimate of σ^2 is $\hat{\sigma}^2 = \frac{n-1}{n}s^2 = \frac{29}{30}7.5^2 = 54.375$
2. With $\alpha = .05$, obtain $\chi^2_{29,.975} = 16.047$ & $\chi^2_{29,.025} = 45.772$. Then, a 95% c.i. for the variance σ^2 of the one-way distance is given by

$$\left(\frac{(n-1)s^2}{\chi^2_{.025,n-1}}, \frac{(n-1)s^2}{\chi^2_{.975,n-1}}\right) = \left(\frac{(29)7.5^2}{45.772}, \frac{(29)7.5^2}{16.047}\right)$$
$$= (35.6386, 101.6545)$$

$\Rightarrow$ We are 95% sure that the variance of the one-way commute distance is between 35.6386 miles & 101.6545 miles

Estimation For The Difference in 2 Normal Means (Known Variances) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu_1, \sigma_1^2)$ & let $Y_1, \ldots, Y_m$ be a random sample of size m from $N(\mu_2, \sigma_2^2)$ with *both variances known*, & suppose that the 2 samples are indept of each other



Estimation For The Difference in 2 Normal Means (Known Variances) II

1. $\overline{X}$ & $\overline{Y} = m^{-1} \sum_{j=1}^{m} Y_j$ are the ML estimators of μ_1 & μ_2, respectively. Due to indep of the 2 samples, $\overline{X} - \overline{Y}$ is the *ML estimator of $\mu_1 - \mu_2$*
2. Sampling distributions of $\overline{X} - \overline{Y}$:
   1. $\overline{X} \sim N\left(\mu_{10}, \frac{\sigma_1^2}{n}\right)$ & $\overline{Y} \sim N\left(\mu_{20}, \frac{\sigma_2^2}{m}\right)$, where μ_k has a true but unknown value μ_{k0}, for $k = 1, 2$
   2. $\overline{X}$ & $\overline{Y}$ are indept
   3. $\overline{X} - \overline{Y} \sim N\left(\mu_{10} - \mu_{20}, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$ ($\because$ Normality of linear combination of indept normal r.v.'s at Page 45 (Ch.5))

Estimation For The Difference in 2 Normal Means (Known Variances) III

3. Start with the following exact probability statement, for any $0 < \alpha < 1$,

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_{10} - \mu_{20})}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} < z_{\alpha/2}\right) = 1 - \alpha$$

4. Manipulate the inequalities to get

$$P\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_{10} - \mu_{20} < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$
$$= 1 - \alpha$$

Estimation For The Difference in 2 Normal Means (Known Variances) IV

$\Rrightarrow$ The random interval $\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\right.$
$\left.\overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$ contains $\mu_{10} - \mu_{20}$ with prob $1 - \alpha$

c.i. For The Difference in 2 Normal Means (Known Variances)

Based on a sample of size n from a $N(\mu_1, \sigma_1^2)$ population, & a sample of size m from a $N(\mu_2, \sigma_2^2)$ population, where both population means are unknown & both population variances are known, a $100(1 - \alpha)\%$ c.i. for $\mu_1 - \mu_2$ is given by

$$(L, U) = \overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

where $\overline{x}$ & $\overline{y}$ are the sample means from the 2 populations

Example 7: c.i. For the Difference in 2 Normal Means (Known Variances) I

Suppose that we are interested in comparing the academic success of university students who have a religion with the academic sucess of those who do not have any. Random samples of size 40 are taken from each population. The average CAPs of the 2 samples are 3.03 & 3.21 for the religious group & the non-religious group, respectively. Assume that CAPs of the 2 groups of students are both normally distributed with standard deviation known to be .5. Find a 99% c.i. for $\mu_1 - \mu_2$, the difference in mean CAPs of students with religion & students without any religion

Solution:

1. The point estimate of $\mu_1 - \mu_2$ is the difference in sample means $\bar{x} - \bar{y} = 3.03 - 3.21 = -.18$

Example 7: c.i. For the Difference in 2 Normal Means (Known Variances) II

2. Here, the population variances are $\sigma_1^2 = \sigma_2^2 = .25$ & the sample sizes $n = m = 40$. Then, with $\alpha = .01$, $z_{.005} = 2.58$ & a 99% c.i. for $\mu_1 - \mu_2$ is given by

$$\begin{aligned} \bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} &= -.18 \pm 2.58\sqrt{\frac{.25}{40} + \frac{.25}{40}} \\ &= -.18 \pm .29 \\ &= (-.11, .47) \end{aligned}$$

$\Rrightarrow$ We are 99% sure that the difference in mean CAPs of students with religion & students without any religion is between -.11 & .47

Estimation For The Difference in 2 Normal Means (Unknown But Equal Variances) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu_1, \sigma^2)$ & let $Y_1, \ldots, Y_m$ be a random sample of size m from $N(\mu_2, \sigma^2)$ with *unknown but equal variances* σ^2, & suppose that the 2 samples are indept of each other

1. $\overline{X} - \overline{Y}$ is the *ML estimator of $\mu_1 - \mu_2$*
2. $\overline{X} - \overline{Y} \sim N\left(\mu_{10} - \mu_{20}, \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2\right)$ (Page 30 with $\sigma_1^2 = \sigma_2^2 = \sigma^2$)
3. $\frac{(n-1)}{\sigma^2} S_1^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$ &
   $\frac{(m-1)}{\sigma^2} S_2^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{\sigma^2} \sim \chi^2_{m-1}$; They are indept r.v.'s

Definition

Let $S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$ be the *pooled variance estimator* of σ^2

Estimation For The Difference in 2 Normal Means (Unknown But Equal Variances) II

4. $\dfrac{\overline{X} - \overline{Y} - (\mu_{10} - \mu_{20})}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2}} \sim N(0,1)$ & $\dfrac{(n+m-2)}{\sigma^2} S_p^2 \sim \chi^2_{n+m-2}$ are indept ($\because$ $\overline{X}, \overline{Y}, S_1^2, S_2^2$ are indept r.v.'s)

5. $$\frac{\dfrac{\overline{X} - \overline{Y} - (\mu_{10} - \mu_{20})}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2}}}{\sqrt{\dfrac{(n+m-2)}{\sigma^2} S_p^2/(n+m-2)}} = \boxed{\frac{\overline{X} - \overline{Y} - (\mu_{10} - \mu_{20})}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}}$$

Estimation For The Difference in 2 Normal Means (Unknown But Equal Variances) III

6. Start with the following exact probability statement, for any $0 < \alpha < 1$,

$$P\left(-t_{\alpha/2,n+m-2} < \frac{\overline{X} - \overline{Y} - (\mu_{10} - \mu_{20})}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} < t_{\alpha/2,n+m-2} \right) = 1 - \alpha$$

7. Manipulate the inequalities to get

$$P\left(\overline{X} - \overline{Y} - t_{\alpha/2,n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_{10} - \mu_{20} < \overline{X} - \overline{Y} + t_{\alpha/2,n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right) = 1 - \alpha$$

Estimation For The Difference in 2 Normal Means (Unknown But Equal Variances) IV

c.i. For The Difference in 2 Normal Means (Unknown But Equal Variances)

Based on a sample of size n from a $N(\mu_1, \sigma_1^2)$ population, & a sample of size m from a $N(\mu_2, \sigma_2^2)$ population, where both population means are unknown & both population variances are unknown but equal, a $100(1-\alpha)\%$ c.i. for $\mu_1 - \mu_2$ is given by

$$(L, U) = \bar{x} - \bar{y} \pm t_{\alpha/2, n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

where $\bar{x}$ & $\bar{y}$, & s_1^2 & s_2^2, are the sample means & the sample variances from the 2 populations, & $s_p^2 = \dfrac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$

Example 8: c.i. For the Difference in 2 Normal Means (Unknown But Equal Variances) I

Most students have complained that the soft-drink vending machine in the student recreation room (A) dispenses a different amount of drink than machine in the faculty lounge (B). To test this beflief, a student randomly sampled several servings from each machine & carefully measured them, with the results shown below

Machine	Number	Mean	Standard Deviation
A	10	5.38	1.59
B	12	5.92	0.83

Example 8: c.i. For the Difference in 2 Normal Means (Unknown But Equal Variances) II

Assume that the amounts dispensed by both machines are normally distributed with equal variances, find a 95% c.i. for $\mu_1 - \mu_2$, difference in mean amount dispensed by machine A & that by machine B

Solution:

1. The point estimate of $\mu_1 - \mu_2$, difference in mean amount dispensed by machine A & that by machine B, is the difference in sample mean $\overline{x} - \overline{y} = 5.38 - 5.92 = -.54$
2. Here, the sample standard deviations are $s_1 = 1.59$ & $s_2 = .83$, & the sample sizes $n = 10$ & $m = 12$. This implies that

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} = \frac{9(1.59^2) + 11(.83^2)}{20} = 1.51654$$

Example 8: c.i. For the Difference in 2 Normal Means (Unknown But Equal Variances) III

3. With $\alpha = .05$, $t_{\alpha/2,n+m-2} = 2.086$, & a 95% c.i. for $\mu_1 - \mu_2$ is given by

$$\begin{aligned}
\overline{x} - \overline{y} \pm t_{\alpha/2,n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} &= -.54 \pm 2.086 \sqrt{1.51654\left(\frac{1}{10} + \frac{1}{12}\right)} \\
&= -.18 \pm 1.1 \\
&= (-1.28, .92)
\end{aligned}$$

$\Rightarrow$ We are 99% sure that the difference in mean amount dispensed by machine A & that by machine B is between -1.28 & .92

Approximate c.i. For the Mean of a Bernoulli r.v. I

Let $X_1, \ldots, X_n$ be a random sample of size n from $Ber(p)$ (*i.e.,* a binary population) with unknown p. We estimate the unknown proportion p as follows

1. $\hat{p} = n^{-1} \sum_{i=1}^{n} X_i$ (which is the sample mean) is the ML estimator of p (`textbook, Example 7.2a`)
2. Approximate sampling distribution of $\hat{p}$: $\hat{p} \dot{\sim} N\left(p_0, \frac{p_0 q_0}{n}\right)$ (`Page 30 (Ch.6)`) as p has a true but unknown value p_0 & $q_0 = 1 - p_0$
3. View $\hat{p} \dot{\sim} N\left(p_0, \frac{p_0 q_0}{n}\right)$ as $\overline{X} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right)$ in `Page 19` by assuming that $p_0 q_0 = \sigma_0^2$ is a fixed constant
4. Apply the result in (1) with the expression $\sigma_0 = \sqrt{p_0 q_0}$ replaced by $\sqrt{\hat{p}(1 - \hat{p})}$ ($\because$ the sample mean value $\hat{p} = n^{-1} \sum_{i=1}^{n} x_i$ is an ML estimate of the unknown value p_0)

Approximate c.i. For the Mean of a Bernoulli r.v. II

Approximate c.i. For the Mean of a Bernoulli r.v.

Based on a sample of size n from a binary population corresponding to a $Ber(p)$ r.v., an approximate $100(1-\alpha)\%$ c.i. for p is given by

$$(L, U) = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}$ is the sample mean

Example 9: c.i. For The Mean of a Bernoulli r.v.

In a discussion about the cars that fellow students drive, several statements were made about types, ages, makes, colors, & so on. Daniel decided he wanted to estimate p, the proportion of cars students drive that are convertibles, so he randomly identified 200 cars in the student parking lot & found 17 to be convertibles. Find a 90% c.i. for p

Solution:

1. The point estimate of p is the sample mean $\hat{p} = 17/200 = .085$
2. With $\alpha = .1$, $z_{.05} = 1.645$ & an approximate 90% c.i. for p is given by

$$
\begin{aligned}
\hat{p} \pm z_{.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= .085 \pm 1.645\sqrt{\frac{(.085)(.915)}{200}} \\
&= .085 \pm .032 \\
&= (.053, .117)
\end{aligned}
$$

$\Rrightarrow$ We are 90% sure that the proportion of cars students drive that are convertibles is between 5.3% & 11.7%