Estimation, hypothesis testing, and linear regression (偵測與估計)(2-4 hours)

Mendenhall and Sincich, Statistics for engineering and the sciences, 5/e, Prentice Hall. (selected from ch7,8,10)

For cognitive radio, telehealth disease detection, object movement tracking, etc)

1

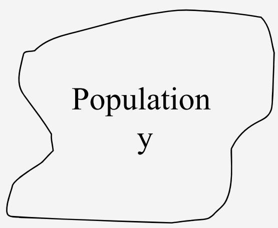
statistics are about estimation (infinite possibilities) /detection (finite possibilities) in communication systems.

- Point estimate: ML/MAP/MMSE estimation is common in communication systems such as Weiner filter/Kalman filter.
- Interval estimate: confidence level- (inherently assume two hypotheses: equal or not equal)

2

- Neyman-Pearson detector (one kind of hypothesis testing): in sync vs. not in sync, target in radar or not, spectrum sensing in cognitive radio, etc.
- ML or MAP detector (one kind of hypothesis testing): demodulation, channel decoding etc.
- (optional) Linear regression: find best–fit (in minimum mean square error, MMSE, sense) line for path loss exponent etc.
- Weiner/Kalman filer

Chapter 7
General Problem Statement



Where y is rv with unknown parameters (e.g., μ, σ^2, α, β - the parameters in question depend on y's distribution)

Naming convention:

$\Theta \equiv$ a parameter of a pdf

$\hat{\Theta} \equiv$ estimate for Θ

Characteristics of Good Estimators

$$\text{Let } \Theta = \mu_y \quad \Rightarrow \quad \hat{\Theta} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$\bar{y}$實驗值 少一degree of freedom

What is $E(\bar{y})$? Is the result a desirable characteristic?

In general :

if $E(\hat{\Theta}) = \Theta$ then $\hat{\Theta}$ is unbiased estimator

if $E(\hat{\Theta}) \neq \Theta$ then $\hat{\Theta}$ is biased estimator

Sample variance (n-1) degree of freedom

Note: 1) $\hat{\Theta}$ is a rv

2) sample (實驗) variance

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$ is an unbiased estimator of σ^2 (population理論 variance)

proof : see example 7.1

Derivation :

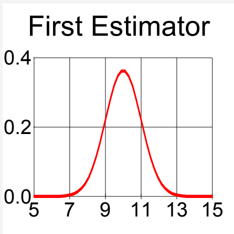
$$\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - 2\bar{y}n\bar{y} + n\bar{y}^2}{n-1}$$

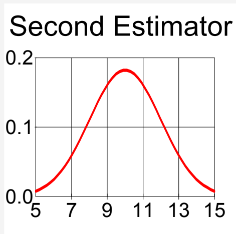
$$= \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - n\left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2}{n-1} = \boxed{\frac{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n-1} = s^2 \text{ (sample variance)}}$$

Easier to substitute

Efficiency of Estimator

Which is the better estimator for θ?





An efficient estimator has the minimum variance.
Here the first estimator is more efficient than the second

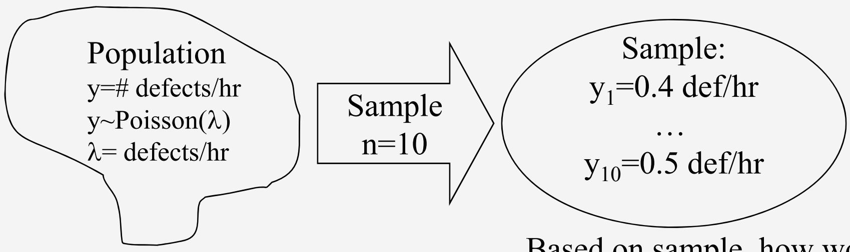
7

Finding Estimator Values
Where We Are Going

- Make estimate based on a sample of n observations
- Point estimate (ML, MAP, MMSE, etc.)
  - A single value estimate for Θ
    - For example, sample 10 Bernoulli trials of which 3 are S then 0.3 most likely value of p
- Interval estimate
  - A range estimate for Θ, with associated probability
    - For example, say that estimate for p is [0.25, 0.35] with 95% confidence

8

Formal Methods of Point Estimators Though Results Intuitive



What does the parameter λ represent?

Based on sample, how would you estimate λ?

- Section 7.2 presents formal confirmation of intuition
  - Moment estimators
  - Method of maximum likelihood

9

Method of Moments

- Have an rv y with assumed distribution
  - Take sample of n observations $(y_1, \ldots, y_n)$
  - Parameters of distribution (i.e., $\Theta_1, \ldots, \Theta_j$) found by solving

  k^{th} population moment= k^{th} sample moment

10

Example Problem 7.8
Illustrate Method

- Assume that y~Bernoulli(p)
  - Took sample of n (y_1, …, y_n)

  1^{st} population moment= 1^{st} sample moment

  (one equation solve one unknown)

$$E(y) = p \text{ set} = \hat{p} \qquad \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \text{ (pronouced as y\_bar)}$$

$$\therefore \hat{p} = \bar{y}$$

A constant found from the sample

P.S. 2 unknown=> use 1^{st} and 2^{nd} moments

Maximum Likelihood Approach

- Assume y has some distribution characterized by $\theta_1, \ldots, \theta_j$
- Took sample of n observations (y_1, …, y_n)
- What was probability I got sample (y_1, …, y_n)?
  - L = $p(y_1,) \, p(y_2) \ldots p(y_n)$
  - Cannot solve since do not know $\theta_1, \ldots, \theta_j$
- Find $\theta_1, \ldots, \theta_j$ that max L (the probability that I got what I got (my sample of n observations)

Example
Method of Maximum Likelihood

- Recall L = f($\theta_1, \ldots, \theta_j$)
  - Want to find $\theta_1, \ldots, \theta_j$ that maximizes L
- Consider case where y~Binomial
  - Take sample n = 30, 18 people bought product

What is L as function of $\hat{p}$?

$$L = \binom{30}{18} \hat{p}^{18} (1 - \hat{p})^{12}$$

Now solve for $\hat{p}$ that max L => $\boxed{\partial \mathbf{L} / \partial \hat{\mathbf{p}} = \mathbf{0}}$

13

Moving Beyond Point Estimates
Confidence Intervals

- Interval estimates
  - estimate of population parameter within a range
- Confidence coefficient
  - probability that parameter will lie within interval
- Confidence interval (combi of 2 above)
- Pivotal statistic
  - way of finding confidence interval
  - function only of sample observations & parameters of interest

14

General Form
Confidence Interval

- Have y, a random variable with unknown distribution
  - Unknown distribution characterized by θ

Let $\hat{\Theta}$ be estimator for Θ.

If $\hat{\Theta} \sim N(\mu_{\hat{\Theta}}, \sigma_{\hat{\Theta}}^2)$ then

$[\mu_{\hat{\Theta}} - \sigma_{\hat{\Theta}}, \mu_{\hat{\Theta}} + \sigma_{\hat{\Theta}})$ is 68.3% CI

$[\mu_{\hat{\Theta}} - 2\sigma_{\hat{\Theta}}, \mu_{\hat{\Theta}} + 2\sigma_{\hat{\Theta}})$ is 95.4% CI

民調95%信心水準

$[\mu_{\hat{\Theta}} - 3\sigma_{\hat{\Theta}}, \mu_{\hat{\Theta}} + 3\sigma_{\hat{\Theta}})$ is 99.7% CI

15

Variance of sample mean

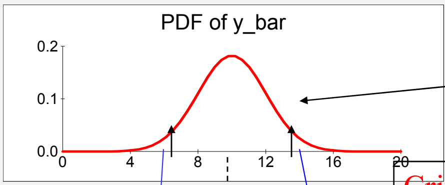
$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \text{ assume } y_i \quad \text{i.i.d.}$$

$$\sigma_{\bar{y}}^2 = \frac{n\sigma_y^2}{n^2} = \frac{\sigma_y^2}{n}$$

$$\boxed{\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}}$$

16

Estimating the Mean of a Population



$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}$$

Critical values (boundary of acceptance region) associated with $z_{\alpha/2}$

$\mu_{\hat{y}} = \mu_y$

$(1-\alpha)100\%$ CI for μ_y Area Prob=$\alpha/2$ Area Prob=$\alpha/2$

$$[\bar{y}_0 - z_{\alpha/2}\sigma_{\bar{y}}, \bar{y}_0 + z_{\alpha/2}\sigma_{\bar{y}}]$$

$$\boxed{[\bar{y}_0 - z_{\alpha/2}\frac{\sigma_y}{\sqrt{n}}, \bar{y}_0 + z_{\alpha/2}\frac{\sigma_y}{\sqrt{n}}]}$$

17

Form of CI & Correct Interpretation (large sample=>CLT)

$$(1-\alpha)100\% \text{ CI for } \mu_y = (\text{LCL}, \text{UCL})$$

$$= [\bar{y}_0 - z_{\alpha/2}\frac{s_0}{\sqrt{n}}, \bar{y}_0 + z_{\alpha/2}\frac{s_0}{\sqrt{n}}]$$

When the population standard deviation σ_y is unknown, use sample standard deviation s_0 (p.6) instead.

Practical interpretation of a Confidence Interval: (p.278 text)

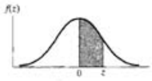
If a $(1-\alpha)100\%$ for a parameter is (LCL, UCL), then we are $(1-\alpha)100\%$ confident that the parameter fall between LCL and UCL

18

z-table

908 Appendix B Useful Statistical Tables

TABLE 5 Normal Curve Areas



z=0.11

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table 1 of Hald, A. *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald and the publisher John Wiley & Sons, Inc.

(pp.20-29 optional)
Small sample: CLT doe not apply

- In previous statement for CI, large sample means Gaussian distribution by central limit theorem (CLT)
- Now in small sample case, student's T distribution with degree of freedom n-1 replaces N(0,1) distribution, and sample variance replaces variance.

Small Sample CI

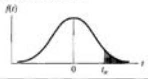
$(1 - \alpha)100\%$ CI for μ_y

$$[\bar{y}_0 - t_{\alpha/2}\frac{s_0}{\sqrt{n}}, \bar{y}_0 + t_{\alpha/2}\frac{s_0}{\sqrt{n}}]$$

910 Appendix B Useful Statistical Tables

t-table

TABLE 7 Critical Values for Student's *T*



v	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from Pearson, E. S., and Hartley, H. O. (eds.) *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., *Biometrika*, 1966.

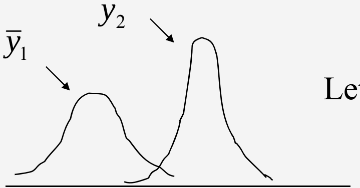
Z table

Extend Concepts:
Estimate (μ_{y1}- μ_{y2}) The new is better than the old?

- Desire CI for (μ_{y1}- μ_{y2})
  - What happens if we take mean of sample of n_1 & n_2 observations? (Assume both samples > 30)

$\bar{y}_2$

$\bar{y}_1$



Let $w = \bar{y}_1 - \bar{y}_2$

PDF of w

What are μ & σ^2 of above?

What is μ & σ^2 of above?

Confidence Interval of (μ_{y1}- μ_{y2})

$$(1-\alpha)100\% \text{ CI for } (\mu_{y_1} - \mu_{y_2})$$

$$[(\bar{y}_1) - (\bar{y}_2) \pm z_{\alpha/2} \, \sigma_{\bar{y}_1 - \bar{y}_2}]$$

What are these?

What is this variance?

How can we get this value?

$$\boxed{(1-\alpha)100\% \text{ CI for } (\mu_{y_1} - \mu_{y_2}) \quad [(\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2} (\boxed{\frac{s_{y_1}^2}{n_1} + \frac{s_{y_2}^2}{n_2}})^{1/2}]}$$

Assumes a large sample

What If n_1 & n_2 Not Large?

The rv w~N only if n_1 and n_2 are >30, what if not true?

Substitute $t_{\alpha/2}$ for $z_{\alpha/2}$, based on previous discussions:
- What assumption is needed about y_1 and y_2?
- Note that degrees of freedom (i.e., υ) now play a role

Two cases possible: → Why worry? Since have define υ & have two samples: n_1 and n_2

1) assume $\sigma_1 = \sigma_2$ → Use a pooled estimator for $\sigma_y = \sigma_1 = \sigma_2$

2) cannot assume $\sigma_1 = \sigma_2$ → Have to come up with a value of υ based on n_1 and n_2 (skipped)

Confidence Interval of (μ_{y1}- μ_{y2})
Assume σ_1 = σ_2 (實際標準差) s:實驗標準差 (樣本不足)

$$\text{Small Sample } (1-\alpha)100\% \text{ CI for } (\mu_{y_1} - \mu_{y_2})$$

$$[(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}(\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2})^{1/2}]$$

$$\nu = n_1 + n_2 - 2$$

$$s_p^2 = \frac{(n_1 - 1)s_{y_1}^2 + (n_2 - 1)s_{y_2}^2}{n_1 + n_2 - 2}$$

Degrees of freedom

n1-1+n2-1 (combine)

Sample<population, so -1

25

鋁合金

7.37 High-strength aluminum alloys. Mechanical engineers have developed a new high-strength aluminum alloy for use in antisubmarine aircraft, tankers, and long-range bombers. (*JOM*, Jan. 2003.) The new alloy is obtained by applying a retrogression and reaging (RAA) heat treatment to the current strongest aluminum alloy. A series of strength tests were conducted to compare the new RAA alloy to the current strongest alloy. Three specimens of each type of aluminum alloy were produced and the yield strength (measured in mega-pascals, MPa) of each specimen determined. The results are summarized in the table.

26

	Alloy Type	
	RAA	Current
Number of specimens	3	3
Mean yield strength (MPa)	641.0	592.7
Standard deviation	19.3 s_{y1} s_{y2} given	12.4

a. Estimate the difference between the mean yield strengths of the two alloys using a 95% confidence interval.

b. The researchers concluded that the RAA-processed aluminum alloy is superior to the current strongest aluminum alloy with respect to yield strength. Do you agree?

27

ANS: (small sample variance given, p25

a. Let μ_1 = mean yield strength of the RAA alloy and μ_2 = mean yield strength of the current alloy. The large sample confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(3-1)(19.3)^2 + (3-1)(12.4)^2}{3+3-2} = 263.125$

$s_p = \sqrt{s_p^2} = \sqrt{263.125} = 16.221$

28

ANS:

For confidence coefficient .95, $\alpha = 1 - .95 = .05$ and $\alpha/2 = .05/2 = .025$. From Table 7 in Appendix B, with df $= n_1 + n_2 - 2 = 3 + 3 - 2 = 4$, $t_{.025} = 2.776$. The 95% confidence interval is:

$$(641.0 - 592.7) \pm 2.776(16.221)\sqrt{\frac{1}{3}+\frac{1}{3}} \Rightarrow 48.3 \pm 36.77 \Rightarrow (11.53,\ 85.07)$$

b. We agree with the researchers. All of the values in the confidence interval are above the value of 0. This indicates that the mean yield strength of the RAA alloy exceeds the mean yield strength of the current alloy.

29

Ch. 8 Hypothesis Testing: Format Problem as Two Hypothesis

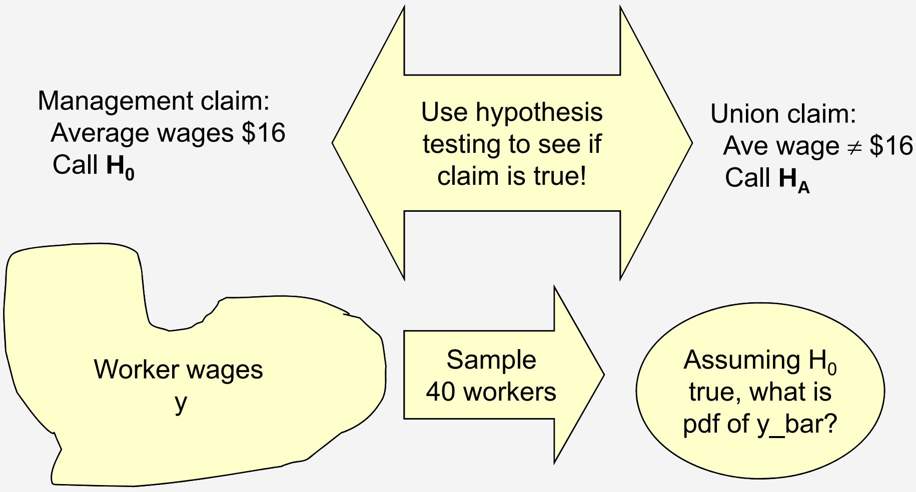
- Starting point of process
  - Claim you wish to test
    - Is a new design lowering warranty costs?
- Pose as two alternative hypothesis
  - Null hypothesis (H_0)
    - The status quo 現狀 (warranty costs stay the same)
  - Alternative hypothesis (H_A)
    - The research question (warranty costs are reduced
- Burden of proof is to support HA. *Is this similar to our system of law (you are assumed innocent 政府要證明頂新油品有害人體)?*
- Ch7 and ch8 are related:

95% $(1-\alpha)$ confidence interval (H_0 vs. H_A =not H_0) =5% false alarm (α)

Politics, industrial engineering radar, communications

30

Simple Illustration of Concept

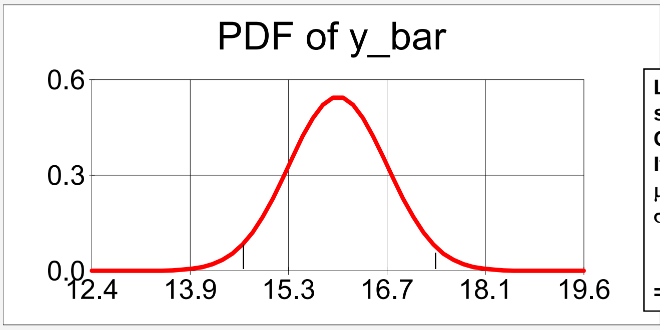


Continuing With Example: Assume Some Results

Sample results:
n = 40
mean = 14.50
standard deviation=s= 4.50

Real (population) mean and standard deviation are unknown,
So use sample mean and sample standard deviation instead.

Testing Validity of Assumption: H_0 is True



Large sample=> Gaussian
If H_0 true:
$\mu = 16$
$\sigma^2 = \sigma^2/n$
$=4.5^2/n$
$= 0.506$
$=(0.712)^2$

Rejection region
Area Prob 0.025

Like confidence interval in ch7

Rejection region
Area Prob 0.025

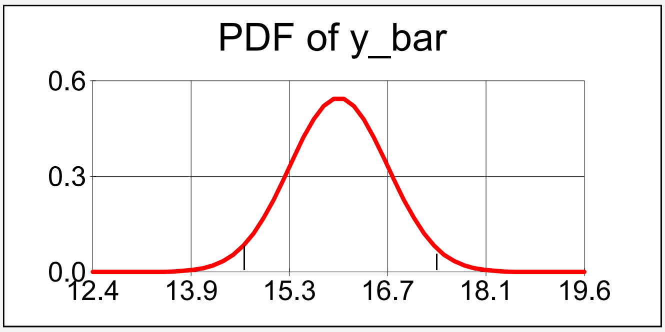
Assume $\alpha \equiv$ Type I error (probability of false alarm)= 0.05

Create region for rejecting management claim H_0

33

PDF of y_bar

Are you 100% sure of conclusions?



Rejection region

Rejection region

14.60

17.40

Let's restate these

Given sample mean = 14.50, what do you conclude? H0 is false

34

Rejection Region

$$\boxed{z_{\alpha/2}} = 1.96 = \frac{y_0 - \mu_{\bar{y}}}{\sigma_{\bar{y}}}$$

$$\boxed{\approx \frac{y_0 - \mu_{\bar{y}}}{s/\sqrt{n}}}$$

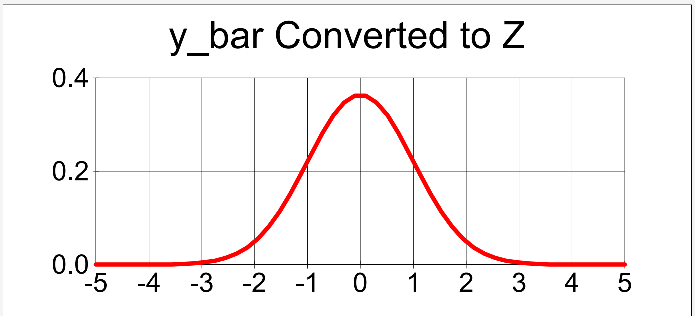
$$= \frac{y_0 - 16}{0.712}$$

$$y_0 = (0.712)(1.96) + 16 = 17.40$$

Textbook backcover z table
P(0~1.96)=0.475
=> P(-∞~1.96)=0.975=1-α/2

35

Rejection Region In Terms of Z



Rejection region → ← Rejection region

$-z_{\alpha/2}$= -1.96 $z_{\alpha/2}$=1.96

Next, convert sample statistic to equivalent z value
& see if falls in rejection region.

36

Two & one Tail Tests

- Previous hypothesis test called two tailed
  - Rejection region both tails of y_bar pdf
    - ✓ Assuming H_0 is true
- Two tailed test take form
  - H_0: $\mu = \mu_0$
  - H_A: $\mu \neq \mu_0$
- One tailed test take form
  - H_0: $\mu = \mu_0$
  - H_A: $\mu > \mu_0$ (or, $\mu < \mu_0$)

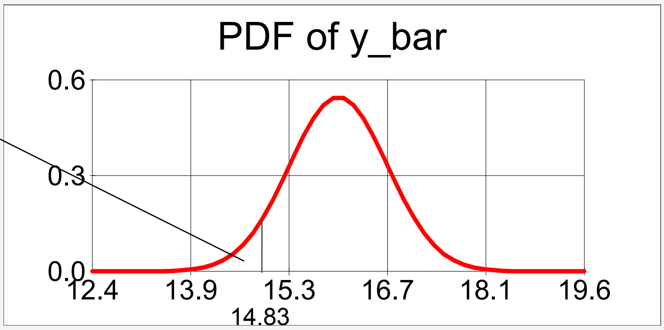
Modify previous example:
H_0: $\mu = 16$
H_A: $\mu < 16$

37

Modifying the Rejection Region: One Tail Test

Area
Prob 0.05

Now H_A: $\mu < 16$, why is rejection region only left tail?



Rejection region

The assumption that H_0 is true still holds, so pdf of y_bar is the same!

38

Computation

Given that α = 0.05, why is critical value for rejection = 14.83?

$$z_\alpha = 1.645 = \frac{y_0 - 16}{0.712}$$

$$y_0 = (0.712)(1.645) + 16 = 17.17$$

By symmetry, critical value (boundary) $= 2*16 - 17.17 = 14.83$

sample mean $= 14.5 < 14.83$, so reject H_0

Textbook backcover z table
P(0~1.645)=0.45
=> P(-∞~1.645)=0.95=1-α

39

Back to Decision: What Are Potential Errors?

Who defines α & β? *What are ideal values of α & β?*

What do you think happens to β as α decreases?

True States of Nature

Decision made		H_0 True	H_A True
	Not Reject H_0	OK Prob= 1- α	Type II error Prob= β
	Reject H_0	Type I error Prob= α	OK Prob= 1-β

This is defined as power of the test

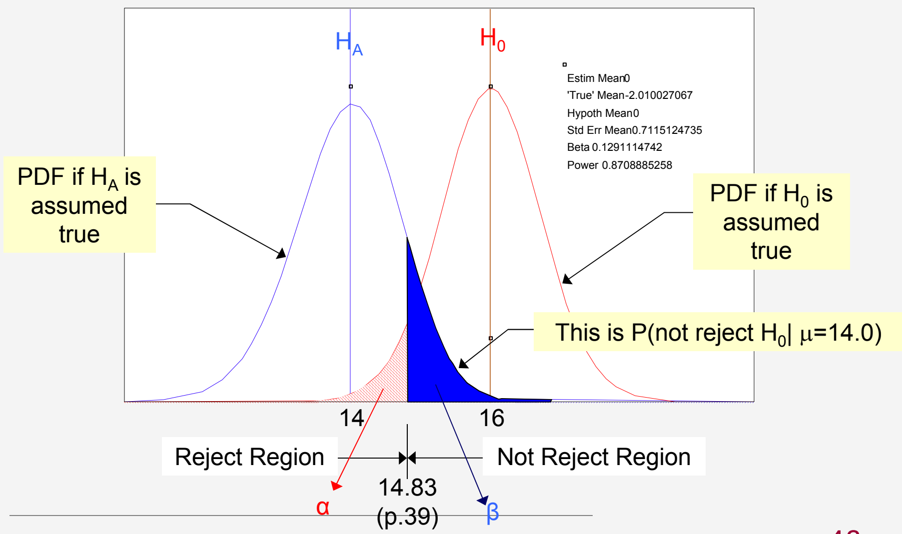
40

Hypothesis Testing: Designing the Experiment

- As experimenter can modify
  - n - the number of observations
  - α - the probability of Type I error (prob. of false alarm)
  - β - the probability of Type II error (prob. of miss)
- What are impacts of increasing n?
- What is impact of decreasing α?
- What is impact of decreasing β?

41

Calculating α and β: Making Assumption for H_A



42

Calculating β:
Assuming H_A: $\mu = 14.0$

- β = p(not reject H_0 | H_A true)
  - Now assume that $\mu = 14.0$

  β_1 = p(not reject H_0 | $\mu = 14.0$)

  = p($\bar{y}_0 > 14.83$ | $\mu = 14.0$)

  = p($z > [(14.83 - 14.0) / 0.712]$ | $\mu = 14.0$)

  = p($z > 1.16$ | $\mu = 14.0$)

  = 0.122

43

(pp.44-62 optional)
Hypothesis Testing (Sec. 8.7~)

- Testing the difference between two population means (independent samples)
- Did we do this using confidence intervals?
  - What was the key value we were looking for in the CI?
  - Why?
- Let's walk through an example

44

Example 8.13 (large sample: n>30)

We want to see if new process an improvement over old

New Process	*Old Process*
$n_1 = 50$	$n_2 = 30$
$\bar{y}_1 = 1{,}255$ calories	$\bar{y}_2 = 1{,}330$ calories
$s_1 = 215$ calories	$s_2 = 238$ calories

Collect sample data

μ_1 = average calories per loaf using new process

μ_2 = average calories per loaf using old process

Interested in whether $\mu_1 - \mu_2$ is less than zero

45

Set Up Test of Hypothesis

Establish null and alternate hypotheses

$$H_0 : (\mu_1 - \mu_2) = 0 \ (\text{i.e., } D_0 = 0)$$

$$H_a : (\mu_1 - \mu_2) < 0 \ (\text{i.e., } \mu_1 < \mu_2)$$

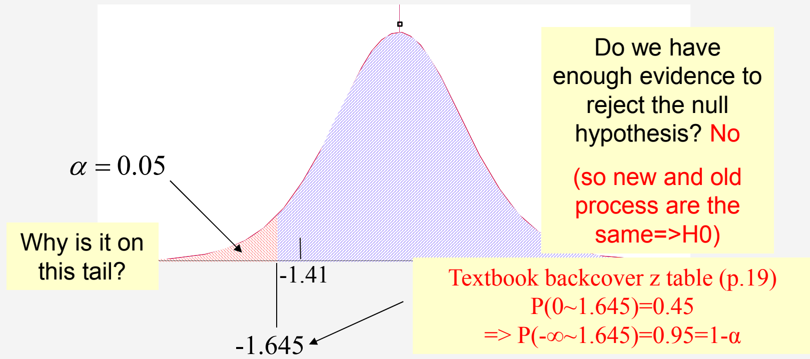
Calculate test statistic

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}} = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}}$$

Why "z"?

46

Establish Rejection Region



$$z = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}} = \frac{(1{,}255 - 1{,}330)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{-75}{\sqrt{\frac{(215)^2}{50} + \frac{(238)^2}{30}}} = -1.41$$

47

Difference in Population Means: Matched Pairs

- Large and small sample
- Note the assumption of normality when using small sample test
- Will work through an example problem manually and
- Discuss the observed value of the test statistic (p-value)

48

Example 8.15

- Cloud seeding (man-made rain) experiment
- Measuring monthly rainfall for two different areas
- Is rainfall higher in the seeded farm area?

Farm Area	1	2	3	4	5	6
Seeded	1.75	2.12	1.53	1.1	1.7	2.42
Unseeded	1.62	1.83	1.4	0.75	1.71	2.33
difference	0.13	0.29	0.13	0.35	-0.01	0.09

Seeded-unseeded mean and standard deviation

$$\mu_d = .1633$$
$$\sigma_d = .13307$$

Set Up Test of Hypothesis

$\mu_1 \equiv$ mean monthly precip for seeded areas

$\mu_2 \equiv$ mean monthly precip for unseeded areas

Establish null and alternate hypotheses

$$\mathrm{H}_0 : (\mu_1 - \mu_2) = 0$$
$$\mathrm{H}_a : (\mu_1 - \mu_2) > 0$$

(improved)

Calculate test statistic

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

Why "t"? Samp' size <30

Rejection Region

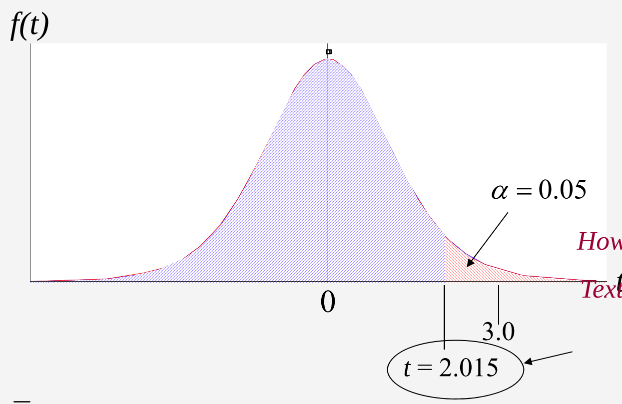
$$t > t_\alpha$$

What assumption s must we make?

t-distribution Assumptions

- Differences are approximately normally distributed
- How would we test for normality?
- What are the degrees of freedom for this calculation?
- Where can we go to get our t-critical value?
  - Remember: based on v=n-1=6-1=5 degrees of freedom

Establish Rejection Region



How do we get this?

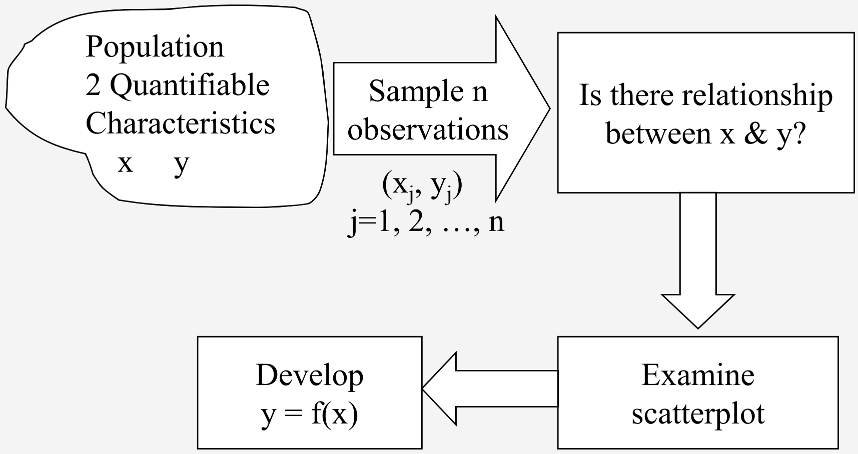
Textbook backcover

v=5, t0.050 =>2.015

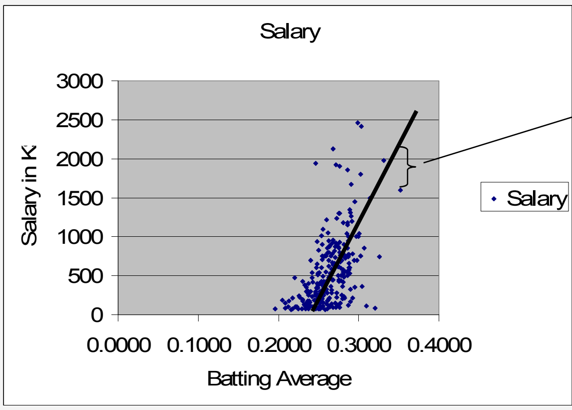
$$t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{.1633 - 0}{.1331/\sqrt{6}} = 3.005$$

What can we conclude from this test? Reject H0

Ch. 10 Linear Regression



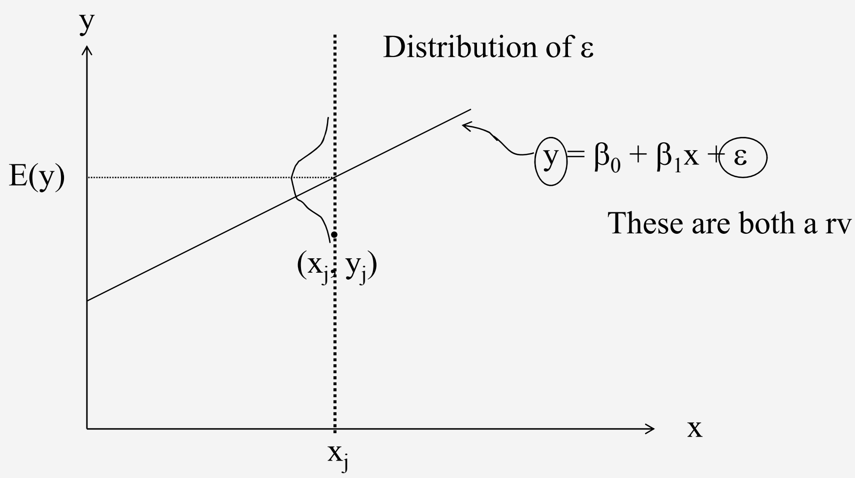
Creating Scatterplots Using Excel



ε_j
(epsilon-j)

$y = f(x) = \beta_0 + \beta_1 x + \varepsilon$ What are β_0 & β_1?

What Does Our Model Look Like?



55

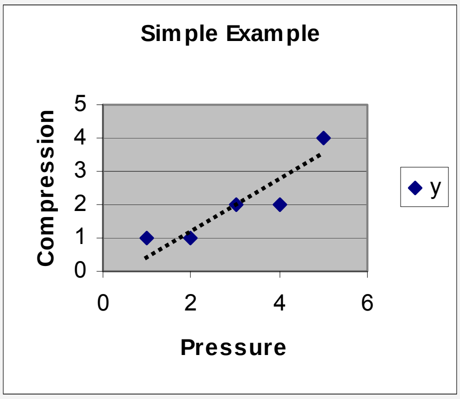
Finding the Model Parameters

Observation	Pressure	Compress.
	x	y
1	1	1
2	2	1
3	3	2
4	4	2
5	5	4

^

56

Method of least square (LS)



Can show only 1 line that minimizes $\Sigma\ \varepsilon^2$

Now need to find $\hat{\beta}_0$ & $\hat{\beta}_1$ such that you min $\Sigma\ \varepsilon^2$.

SSE: sum of squared errors

Let $SSE \equiv \Sigma\ \varepsilon^2 = \Sigma\ (y_i - \hat{y}_i)^2$

but $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_j$

$\therefore SSE = \Sigma\ (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_j)^2 = f(\hat{\beta}_0, \hat{\beta}_1)$

57

Finding Minima for SSE Using First Partial Derivatives

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = 0 = \frac{\partial \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\partial \hat{\beta}_0}$$

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = 0 = \frac{\partial \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\partial \hat{\beta}_1}$$

Get 2 equations in 2 unknowns $\Rightarrow$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x$$

58

Definitions

$$SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum_i x_i y_i - \frac{\sum_i x_i \sum_i y_i}{n}$$

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

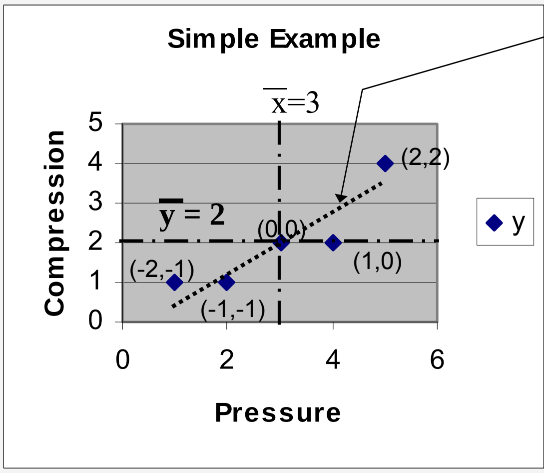
$$= \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}$$

$$\bar{y} = \frac{\sum_i y_i}{n}$$

$$\bar{x} = \frac{\sum_i x_i}{n}$$

59

What Do Terms Mean?



$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

60

Finding Coefficients
Simple Example

- $\boxed{\hat{\beta}_1 = SS_{xy} / SS_{xx}}$
  =(-2*-1+-1*-1+0*0+1*0+2*2)/(4+1+0+1+4)
  =7/ 10 = 0.7
- $\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}} = 2 - (0.7)(3) = -0.1$
- Resultant model
  $\hat{y} = -0.1 + 0.7\,x$

Next steps: trying to describe $\hat{\beta}_0$ & $\hat{\beta}_1$ and using Excel

61

Using Excel
Linear Regression Models

- Excel offers two linear regression approaches
  - TOOLS/DATA ANALYSIS/REGRESSION
  - Create scatterplot
    - ✓ double click on data set
    - ✓ Select CHART/Add Trendline
- Next step: describing estimators for β_0 & β_1

62

Introduction to Wiener filters and Kalman filters

(1 hour)

Shu-Ming 63

Orthogonality principle

Assume x is the column vector to be estimated,

and y is the observation column vector.

The linear estimator that minimize $E[(x\text{-}wy)^T(x-wy)]$

is the estimator w such that $E[(x\text{-}wy)y^T] = 0$

That is, the estimation error and observation (which the estimation is based on) is orthogonal

Also assume x, w, y are all real.

proof : Assume G is another estimator

$E[(Gy-x)^T(Gy-x)] = E[(Gy-wy+wy-x)^T(Gy-wy+wy-x)]$

$= E[(Gy-wy)^T(Gy-wy)] + E[(wy-x)^T(wy-x)]$

$$\left(\begin{array}{l} Note\, E[(wy-x)^T(Gy-wy)] = trace(E[(wy-x)(Gy-wy)^T]) \\ = trace(E[(wy-x)y^T](G-w)^T) = 0 \text{ by given condition} \end{array}\right)$$

$\geq E[(wy-x)^T(wy-x)]$ Q.E.D.

Shu-Ming 64

Wiener filter (linear MMSE)

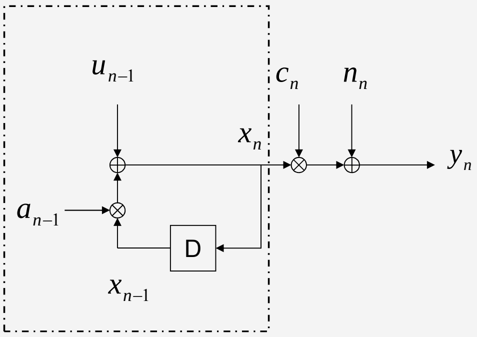
$$\hat{x} = wy$$

$$E[(x\text{-}wy)y^T] = 0$$

$$E[xy^T] = wE[yy^T]$$

$$w = R_{xy}R_{yy}^{-1}$$

System model for Kalman filter



1st-order auto-regressive
(AR1) model

Summary of Kalman filter (recursive linear MMSE, underlying AR1 model)

The model

$$\begin{cases} x_{n+1} = a_n x_n + u_n, & E[u_n]=0, E[u_n^2]=\sigma_u^2 \\ y_n = c_n x_n + n_n, & E[n_n]=0, E[n_n^2]=\sigma_n^2 \end{cases}$$

$$\alpha_n = y_n - \hat{y}_n,\ \ \hat{y}_n = \sum_{i=1}^{n-1} \frac{E[y_n\alpha_i]\alpha_i}{E[\alpha_i^2]}$$

$$\varepsilon_n = x_n - \hat{x}_n,\ \ \hat{x}_n = \sum_{i=1}^{n-1} \frac{E[x_n\alpha_i]\alpha_i}{E[\alpha_i^2]}$$

iteration

(A) $\alpha_n = y_n - c_n\hat{x}_n$

(B) $\begin{cases} g_n = \dfrac{a_n c_n E[\varepsilon_n^2]}{c_n^2 E[\varepsilon_n^2] + \sigma_n^2} \text{ Kalman gain} \\ \hat{x}_{n+1} = a_n\hat{x}_n + g_n\alpha_n \end{cases}$

(C) $E[\varepsilon_{n+1}^2] = a_n(a_n - g_n c_n)E[\varepsilon_n^2] + \sigma_u^2$

Comparisons

- Kalman filters: the values at time n+1 depend on the values at time n only (symbol by symbol processing) but assume additional model on signal x (AR underlying model)
- Wiener filter: block by block processing
- Thus Kalman filter is a popular adaptive filter.

Applications of Kalman filters

- Tracking the position of a moving target: The parameters are calculated from military exercise data.
- Tracking the channel coefficient of time-varying fading channels: The parameters are calculated from training sequences.