

12.540 Principles of the Global Positioning System Lecture 10

Prof. Thomas Herring

03/12/03

12.540 Lec 10

1

Estimation: Introduction

- Homework review
- Overview
 - Basic concepts in estimation
 - Models: Mathematical and Statistical
 - Statistical concepts

03/12/03

12.540 Lec 10

2

Basic concepts

- Basic problem: We measure range and phase data that are related to the positions of the ground receiver, satellites and other quantities. How do we determine the “best” position for the receiver and other quantities.
- What do we mean by “best” estimate?
- Inferring parameters from measurements is estimation

03/12/03

12.540 Lec 10

3

Basic estimation

- Two styles of estimation (appropriate for geodetic type measurements)
 - Parametric estimation where the quantities to be estimated are the unknown variables in equations that express the observables
 - Condition estimation where conditions can be formulated among the observations. Rarely used, most common application is leveling where the sum of the height differences around closed circuits must be zero

03/12/03

12.540 Lec 10

4

Basics of parametric estimation

- All parametric estimation methods can be broken into a few main steps:
 - Observation equations: equations that relate the parameters to be estimated to the observed quantities (observables). Mathematical model.
 - Example: Relationship between pseudorange, receiver position, satellite position (implicit in ρ), clocks, atmospheric and ionosphere delays
 - Stochastic model: Statistical description that describes the random fluctuations in the measurements and maybe the parameters
 - Inversion that determines the parameters values from the mathematical model consistent with the statistical model.

03/12/03

12.540 Lec 10

5

Observation model

- Observation model are equations relating observables to parameters of model:
 - Observable = function (parameters)
 - Observables should not appear on right-hand-side of equation
- Often function is non-linear and most common method is linearization of function using Taylor series expansion.
- Sometimes log linearization for $f=a.b.c$ ie. Products fo parameters

03/12/03

12.540 Lec 10

6

Taylor series expansion

- In most common Taylor series approach:

$$y = f(x_1, x_2, x_3, x_4)$$
$$y_0 + \Delta y = f(\mathbf{x})|_{\mathbf{x}_0} + \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \Delta \mathbf{x} \quad \mathbf{x} = (x_1, x_2, x_3, x_4)$$

- The estimation is made using the difference between the observations and the expected values based on apriori values for the parameters.
- The estimation returns adjustments to apriori parameter values

03/12/03

12.540 Lec 10

7

Linearization

- Since the linearization is only an approximation, the estimation should be iterated until the adjustments to the parameter values are zero.
- For GPS estimation: Convergence rate is 100-1000:1 typically (ie., a 1 meter error in apriori coordinates could result in 1-10 mm of non-linearity error).

03/12/03

12.540 Lec 10

8

Estimation

- (Will return to statistical model shortly)
- Most common estimation method is “least-squares” in which the parameter estimates are the values that minimize the sum of the squares of the differences between the observations and modeled values based on parameter estimates.
- For linear estimation problems, direct matrix formulation for solution
- For non-linear problems: Linearization or search technique where parameter space is searched for minimum value
- Care with search methods that local minimum is not found (will not treat in this course)

03/12/03

12.540 Lec 10

9

Least squares estimation

- Originally formulated by Gauss.
- Basic equations: Δy is vector of observations; A is linear matrix relating parameters to observables; Δx is vector of parameters; v is residual

$$\Delta y = A\Delta x + v$$

minimize $(v^T v)$; superscript T means transpose

$$\Delta x = (A^T A)^{-1} A^T \Delta y$$

03/12/03

12.540 Lec 10

10

Weighted Least Squares

- In standard least squares, nothing is assumed about the residuals v except that they are zero mean.
- One often sees weight-least-squares in which a weight matrix is assigned to the residuals. Residuals with larger elements in W are given more weight.

$$\begin{aligned} &\text{minimize } (\mathbf{v}^T \mathbf{W} \mathbf{v}); \\ &\Delta \mathbf{x} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \Delta \mathbf{y} \end{aligned}$$

03/12/03

12.540 Lec 10

11

Statistical approach to least squares

- If the weight matrix used in weighted least squares is the inverse of the covariance matrix of the residuals, then weighted least squares is a maximum likelihood estimator for Gaussian distributed random errors.
- This latter form of least-squares is most statistically rigorous version.
- Sometimes weights are chosen empirically

03/12/03

12.540 Lec 10

12

Review of statistics

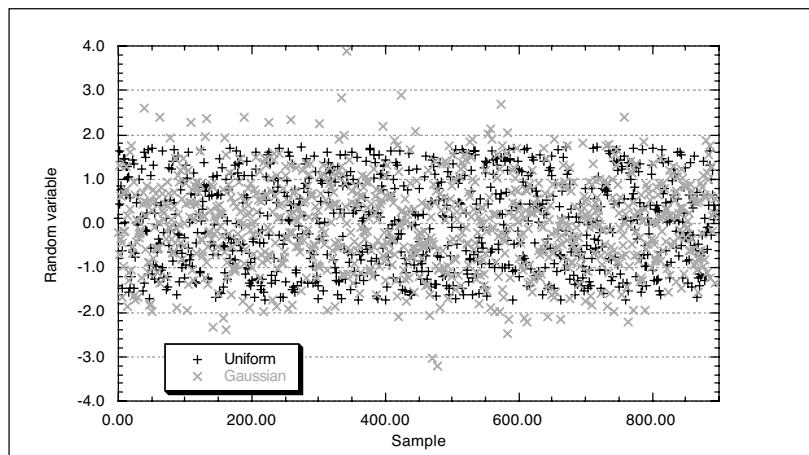
- Random errors in measurements are expressed with probability density functions that give the probability of values falling between x and $x+dx$.
- Integrating the probability density function gives the probability of value falling within a finite interval
- Given a large enough sample of the random variable, the density function can be deduced from a histogram of residuals.

03/12/03

12.540 Lec 10

13

Example of random variables

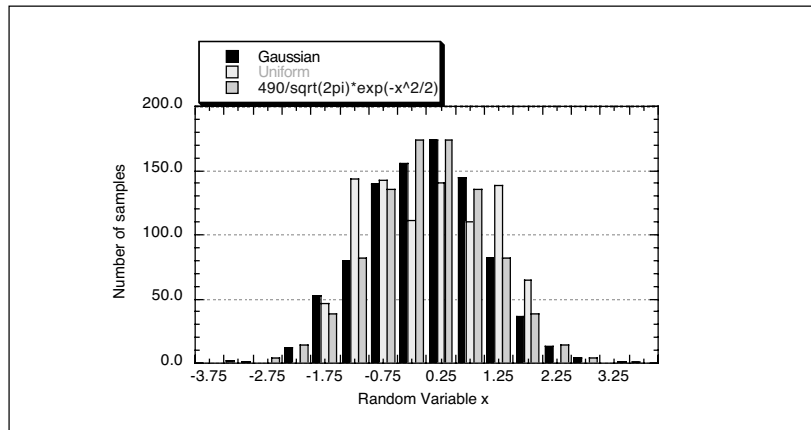


03/12/03

12.540 Lec 10

14

Histograms of random variables



03/12/03

12.540 Lec 10

15

Characterization Random Variables

- When the probability distribution is known, the following statistical descriptions are used for random variable x with density function $f(x)$:

Expected Value	$\langle h(x) \rangle$	$\int h(x)f(x)dx$
Expectation	$\langle x \rangle$	$\int xf(x)dx = \mu$
Variance	$\langle (x - \mu)^2 \rangle$	$\int (x - \mu)^2 f(x)dx$

Square root of variance is called standard deviation

03/12/03

12.540 Lec 10

16

Theorems for expectations

- For linear operations, the following theorems are used:
 - For a constant $\langle c \rangle = c$
 - Linear operator $\langle cH(x) \rangle = c\langle H(x) \rangle$
 - Summation $\langle g+h \rangle = \langle g \rangle + \langle h \rangle$
- Covariance: The relationship between random variables $f_{xy}(x,y)$ is joint probability distribution:

$$\sigma_{xy} = \langle (x - \mu_x)(y - \mu_y) \rangle = \int (x - \mu_x)(y - \mu_y) f_{xy}(x,y) dx dy$$

$$\text{Correlation: } \rho_{xy} = \sigma_{xy} / \sigma_x \sigma_y$$

03/12/03

12.540 Lec 10

17

Estimation on moments

- Expectation and variance are the first and second moments of a probability distribution

$$\hat{\mu}_x \approx \sum_{n=1}^N x_n / N \approx \frac{1}{T} \int x(t) dt$$

$$\hat{\sigma}_x^2 \approx \sum_{n=1}^N (x - \mu_x)^2 / N \approx \sum_{n=1}^N (x - \hat{\mu}_x)^2 / (N - 1)$$

- As N goes to infinity these expressions approach their expectations. (Note the N-1 in form which uses mean)

03/12/03

12.540 Lec 10

18

Probability distributions

- While there are many probability distributions there are only a couple that are common used:

-

$$\text{Gaussian } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{Multivariate } f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |V|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T V^{-1}(\mathbf{x}-\mu)}$$

$$\text{Chi-squared } \chi_r^2(x) = \frac{x^{r/2-1} e^{-x/2}}{\Gamma(r/2) 2^{r/2}}$$

03/12/03

12.540 Lec 10

19

Probability distributions

- The chi-squared distribution is the sum of the squares of r Gaussian random variables with expectation 0 and variance 1.
- With the probability density function known, the probability of events occurring can be determined. For Gaussian distribution in 1-D; $P(|x| < 1\sigma) = 0.68$; $P(|x| < 2\sigma) = 0.955$; $P(|x| < 3\sigma) = 0.9974$.
- Conceptually, people think of standard deviations in terms of probability of events occurring (ie. 68% of values should be within 1-sigma).

03/12/03

12.540 Lec 10

20

Central Limit Theorem

- Why is Gaussian distribution so common?
- “*The distribution of the sum of a large number of independent, identically distributed random variables is approximately Gaussian*”
- When the random errors in measurements are made up of many small contributing random errors, their sum will be Gaussian.
- Any linear operation on Gaussian distribution will generate another Gaussian. Not the case for other distributions which are derived by convolving the two density functions.

03/12/03

12.540 Lec 10

21

Summary

- Examined simple least squares and weighted least squares
- Examined probability distributions
- Next we pose estimation in a statistical framework

03/12/03

12.540 Lec 10

22