12.215 Modern Navigation

Thomas Herring

Review of last class

- Map Projections:
  - Why projections are needed
  - Types of map projections
    - Classification by type of projection
    - Classification by characteristics of projection
  - Mathematics of map projections

Today's Class

- Basic Statistics
  - Statistical description and parameters
    - Probability distributions
    - Descriptions: expectations, variances, moments
    - Covariances
    - Estimates of statistical parameters
- Propagation of variances
  - Methods for determining the statistical parameters of quantities derived other statistical variables

Basic Statistics

- Concept behind statistical description: Some processes involve so many small physical effects that deterministic model is not possible.

- For Example: Given a die, and knowing its original orientation before throwing; the lateral and rotational forces during throwing and the characteristics of the surface it falls on, it should be possible to calculate its orientation at the end of the throw. In practice any small deviations in forces and interactions makes the result unpredictable and so we find that each face comes up 1/6th of the time. A probabilistic description.

- In this case, any small imperfections in the die can make one face come up more often so that each probability is no longer 1/6th (but all outcomes must still add to a probability of 1)
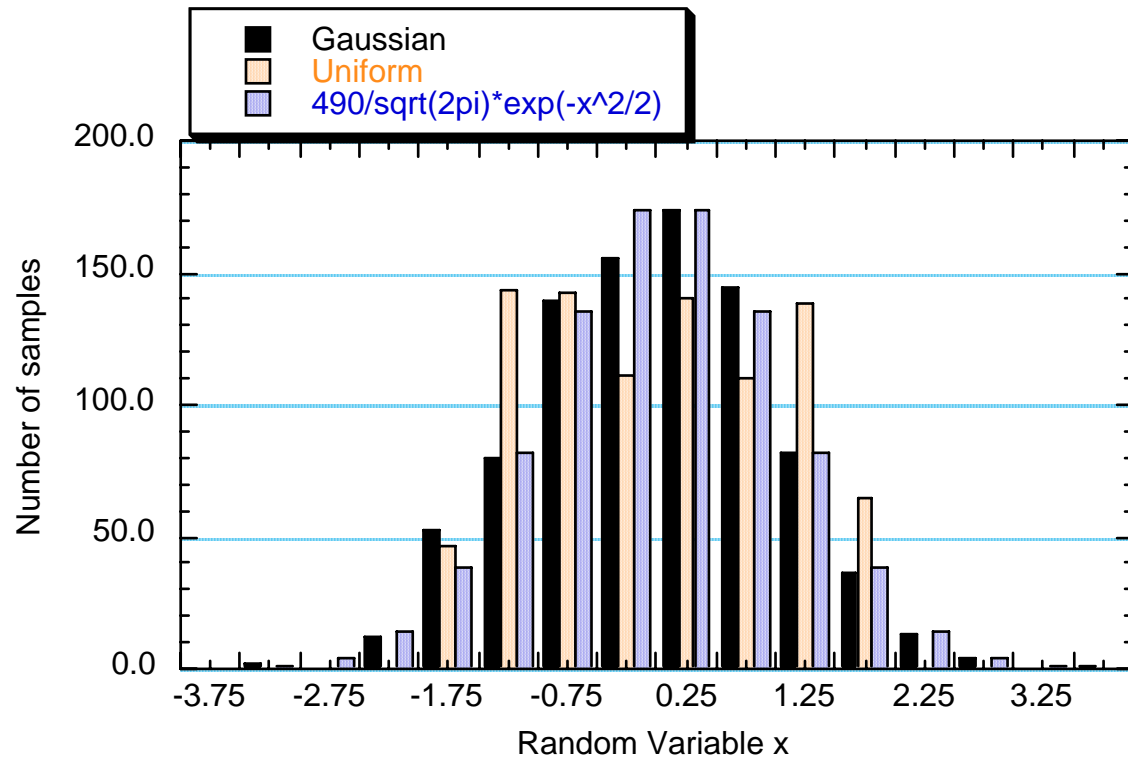
Probability descriptions

- For discrete processes we can assign probabilities of specific events occurring but for continuous random variables, this is not possible

- For continuous random variables, the most common description is a probability density function.

- A probability density function gives the probability that a random variable x will have a value between x and x+dx

- To find the probability of a random variable taking on a value between x_1 and x_2, the density function is integrated between these two values.

- Probability density functions can derived analytically for variables that are derived from other random variables with know probability density functions, or if the number of samples is large, then a histogram can be used to determine the density function (normally by fitting to a know class of density functions).

Example of random variables
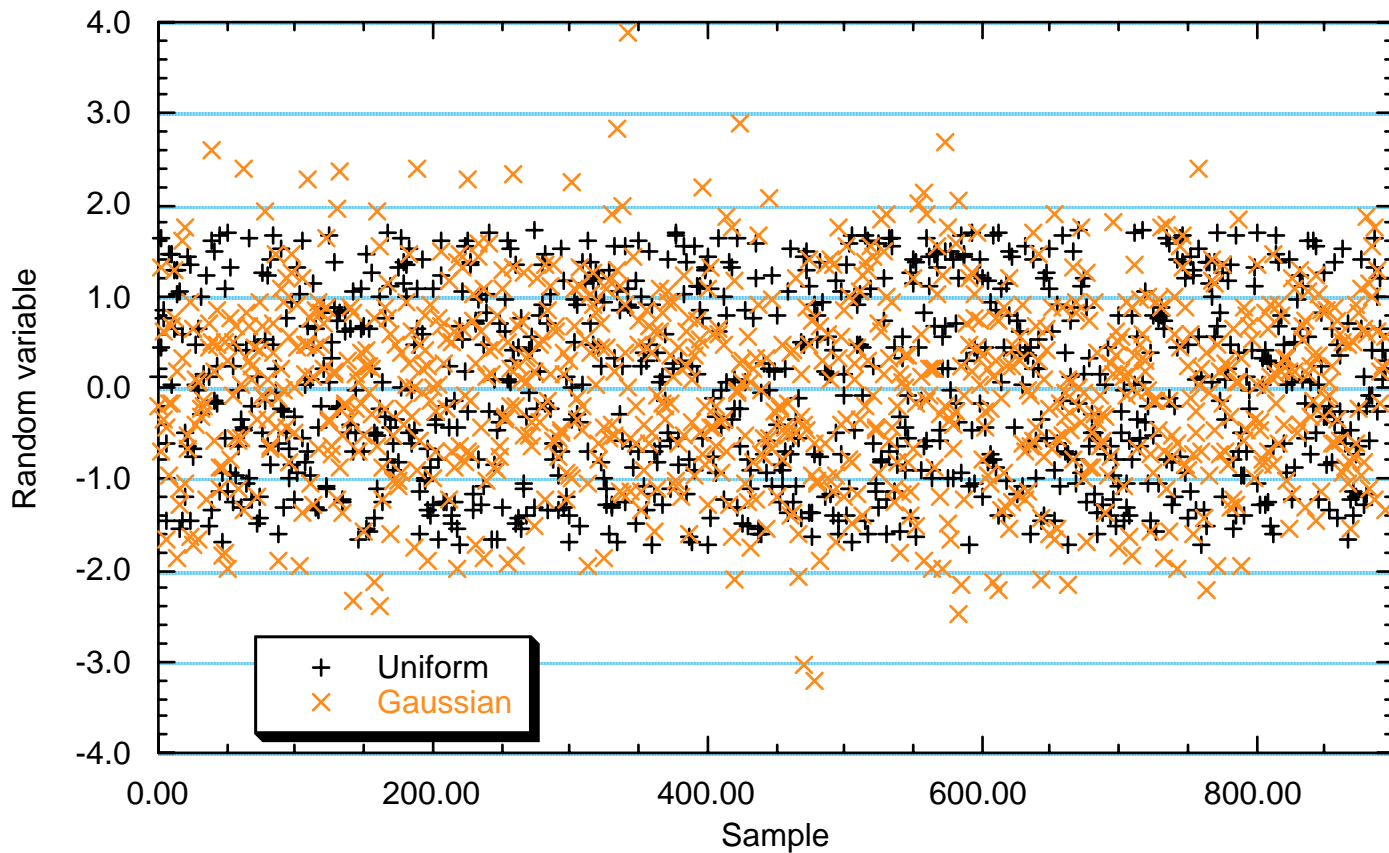
Histograms of random variables

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	$< h(x) >$	$\int h(x) f(x) dx$
Expectation	$< x >$	$\int x f(x) dx = \mu$
Variance	$< (x - \mu)^2 >$	$\int (x - \mu)^2 f(x) dx$
Moments	$< (x - \mu)^n >$	$\int (x - \mu)^n f(x) dx$

Square root of variance is called standard deviation

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:

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

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

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)

Probability distributions

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

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

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

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

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).

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.

Covariance matrices

- For large systems of random variables (such as GPS range measurements, position estimates etc.), the variances and covariances are arranged in a matrix called the variance-covariance matrix (or simply covariance matrix).
- This is the matrix used in multivariate Gaussian probability density function.

$$C = \begin{vmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{vmatrix}$$ Notice that matrix is symmetric

Properties of covariance matrices

- Covariance matrices are symmetric
- All of the diagonal elements are positive and usually non-zero since these are variances
- The off diagonal elements need to satisfy at least that $-1 <= \sigma_{ij}/(\sigma_i\sigma_j) <= 1$ where σ_i and σ_j are the standard deviations (square root of diagonal elements)
- The matrix has all positive (or zero) eigenvalues
- Matrix is positive definite (i.e., for all vectors x, $x^T V x$ is positive)

Propagation of Variance-Covariance matrices

- Given the covariance matrix of a set of random variables, the characteristics of expected values can be used to determine the covariance matrix of any linear combination of the measurements.

Given linear operation : $\mathbf{y} = \mathbf{Ax}$ with $\mathbf{V_{xx}}$ as covariance matrix of $\mathbf{x}$

$$\mathbf{V_{yy}} = <\mathbf{yy^T}> = <\mathbf{Axx^TA^T}> = \mathbf{A} <\mathbf{xx^T}> \mathbf{A^T}$$

$$\mathbf{V_{yy}} = \mathbf{AV_{xx}A^T}$$

Applications

- Propagation of variances is used extensively to "predict" the statistical character of quantities derived by linear operations on random variables

- Because of the Central Limit theorem we know that if the random variables used in the linear relationships are Gaussianly distributed, then the resultant random variables will also be Gaussianly distributed.

- Example:

$$C = \begin{vmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{vmatrix} \text{ for random variable } \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$y = x_1 \pm x_2 \text{ then } \sigma_y^2 = \sigma_1^2 + \sigma_2^2 \pm 2\sigma_{12}$$

Application

- Notice here that depending on σ_{12} (sign and magnitude), the sum or difference of two random variables can either be very well determined (small variance) or poorly determined (large variance)

- If σ_{12} is zero (uncorrelated in general and independent for Gaussian random variables (all moments are zero), then the sum and difference have the same variance.

- Even if the covariance is not known, then an upper limit can be placed on the variance of the sum or difference (since the value of σ_{12} is bounded).

- We return to covariance matrices after covering estimation.

Summary

- Basic Statistics
  - Statistical description and parameters
    - Probability distributions
    - Descriptions: expectations, variances, moments
    - Covariances
    - Estimates of statistical parameters
- Propagation of variances
  - Methods for determining the statistical parameters of quantities derived other statistical variables