FARIBORZ MASEEH DEPARTMENT OF MATHEMATICS + STATISTICS
Portland State UNIVERSITY

Mth 410/510: Inverse Problems & Data Assimilation I

Instructor: Dacian N. Daescu

Review of Probability and Statistics Concepts

Probability Space

Definition

A probability space is a triplet $(\Omega, \mathcal{F}, P)$ consisting of

- Ω - the sample space, a non-empty set of possible outcomes of an experiment
- $\mathcal{F}$ - a σ-algebra of subsets of Ω whose elements are called events
  1. $\Omega \in \mathcal{F}$
  2. If $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$
  3. If $A_i \in \mathcal{F}, i = 1, 2, \ldots$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$
- A probability measure function $P : \mathcal{F} \rightarrow [0, 1]$ that assigns a value to each event,
  1. $P(\Omega) = 1$
  2. If $A_i \cap A_j = \emptyset,\ i \neq j,\ i, j = 1, 2, \ldots$

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Random Variable

Definition

A random variable (rv) is a function $X : \Omega \to \mathbb{R}$ such that

$$\forall x \in \mathbb{R}, \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

For each $\omega \in \Omega$, $X(\omega)$ is a *sample value or realization* of the rv X.

Thus a random variable X is a measurable function. For simplicity of notation, denote $(X \leq x) = \{\omega \in \Omega : X(\omega) \leq x\}$.

Definition

The distribution function of the rv X is the function defined as

$$F_X : \mathbb{R} \to [0, 1], \quad F_X(x) = P(X \leq x)$$

The distribution function has the following properties

- F_X is nondecreasing: $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
- $\lim_{x \to -\infty} F_X(x) = 0, \ \lim_{x \to +\infty} F_X(x) = 1$
- If $a < b$ then $P(a < X \leq b) = F_X(b) - F_X(a)$

Probability Density Function (pdf)

Definition

A probability density function is a function $p_X : \mathbb{R} \to [0, \infty)$ such that $p_X(x) \geq 0$ and $\int_{\mathbb{R}} p_X(x)\, dx = 1$

Definition

A rv X is called continuous if there is a pdf p_X such that

$$F_X(x) = \int_{-\infty}^{x} p_X(s)\, ds, \quad \forall x \in \mathbb{R}$$

The pdf satisfies the following properties

- If $x_1 < x_2$ then $F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} p_X(s)\, ds$
- $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p_X(s)\, ds$
- For Δx sufficiently small, $P(x \leq X \leq x + \Delta x) \approx p_X(x)\Delta x$

Examples: Uniform and Gaussian rv

Definition

A uniform rv on the interval $[a, b]$ has the distribution function and the pdf

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}, \quad p_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

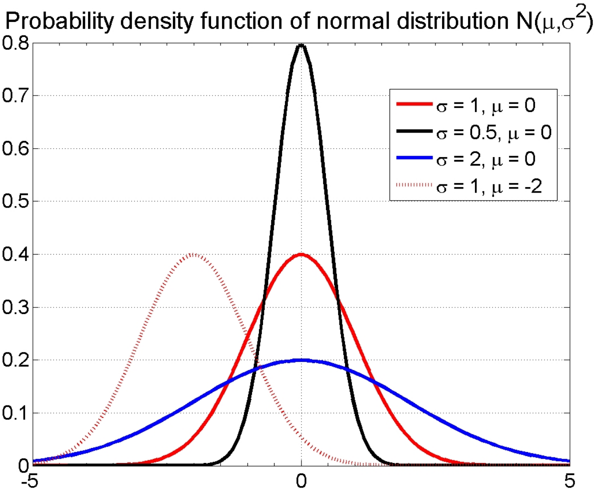
Definition

A rv X is called *normally distributed* or $Gaussian$ if it has the pdf

$$p_X(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We denote $X \sim N(\mu, \sigma^2)$. The *standard normal random variable* has $\mu = 0, \sigma = 1$.

Examples: PDF of a Gaussian random variable



Expected value and variance

Definition

The expected value (expectation, mean) of a continuous rv X is defined as

$$E[X] = \int_{-\infty}^{+\infty} x p_X(x)\, dx$$

If $g(\cdot)$ is a real-valued function then $g(X)$ is a random variable with mean

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) p_X(x)\, dx$$

Definition

The variance of a rv X is denoted by $Var(X)$ or σ_X^2 and is defined as

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 p_X(x)\, dx$$

The standard deviation (std) of X is denoted σ_X and is defined as

$$\sigma_X = \sqrt{Var(X)}$$

Interpretation, properties, fundamental examples

The variance and standard deviation provide measures of the spread (dispersion) of the rv about its mean. Notice that the std σ_X and the mean $E[X]$ have the same (physical) units.

The variance is the mean square minus the square mean

$$Var(X) = E[X^2] - E^2[X]$$

Example 1

If $X \sim N(\mu, \sigma^2)$ then $E[X] = \mu$ and $Var(X) = \sigma^2$

Example 2

If X has a uniform distribution in the interval $[a,\, b]$ then

$$E[X] = \frac{a+b}{2}, \quad Var(X) = \frac{1}{12}(b-a)^2$$

Jointly Distributed Random Variables

Definition

The random variables X and Y are said to be jointly distributed if they are defined on the same probability space. They may be characterized by the *joint distribution function*

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) \stackrel{def}{=} P\{(X \le x) \cap (Y \le y)\}$$

Definition

The joint density function is $p_{X,Y} : \mathbb{R} \times \mathbb{R} \to [0,\infty)$ such that

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{X,Y}(s,t)\, ds dt$$

Notice that

$$p_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

Vectors of random variables, expectation

Definition

If X_1, X_2, $\ldots X_n$ are jointly distributed rv's, then

$$X = [X_1 \, X_2 \ldots X_n]^{\mathrm{T}}$$

is a n-dimensional random vector with density function

$$p_X(x) = p_{X_1 \ldots X_n}(x_1, \ldots, x_n)$$

Definition

The expectation of the random vector X is the vector

$$E[X] = [E[X_1] \, E[X_2] \ldots E[X_n]]^{\mathrm{T}} \;\; where \; E[X_k] \stackrel{def}{=} \int_{\mathbb{R}^n} x_k p_X(x) \, dx$$

Expectation is a linear operator

If X and Y are jointly distributed rv's and a, b are fixed scalar constants,

$$E[aX + bY] = aE[X] + bE[Y]$$

Covariance matrix, correlation

Definition

The covariance matrix of the n-dimensional random vector X is defined as

$$Cov(X) = E\left[(X - E[X])(X - E[X])^{\mathrm{T}}\right] \in \mathbb{R}^{n \times n}$$

with entries

$$Cov(X)_{i,j} = E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$$

The covariance matrix $Cov(X)$ is symmetric and positive semi-definite

$$Cov(X)_{ii} = Var(X_i), \quad Cov(X)_{i,j} = E[X_i X_j] - E[X_i]E[X_j]$$

Definition

The correlation coefficient of the rv's components X_i and X_j is defined as

$$\rho(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma(X_i)\sigma(X_j)}$$

Independence and uncorrelated random variables

Definition

The jointly distributed rv $X_1, \dots, X_n$ are mutually independent if

$$p_{X_1,\dots X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$$

Definition

Two jointly distributed random variables X_i and X_j are uncorrelated if $Cov(X_i, X_j) = 0$ and therefore, their correlation coefficient is 0.

If the components of the rv vector X are pairwise uncorrelated then the covariance matrix $Cov(X)$ is diagonal

$$Cov(X) = diag(Var(X))$$

Theorem

If the jointly distributed rv's X_i and X_j are independent then they are uncorrelated. The converse is also true for Gaussian rv.

The multivariate normal distribution

Definition

The jointly distributed random variables $X_1, \dots X_n$ are said to be Gaussian or jointly normally distributed if their joint density function is

$$p_{X_1,\dots X_n}(x_1,\dots,x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(\mathbf{C})}} e^{-(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2}$$

where $\boldsymbol{\mu} = E[X] \in \mathbb{R}^n$ denotes the mean of the rv vector X and $\mathbf{C}$ is the covariance matrix of the random variables, $C_{i,j} = Cov(X_i, X_j)$.

Theorem

Let X denote a multivariate Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ denote a matrix with constant coefficients and $\mathbf{c} \in \mathbb{R}^m$ a constant vector. The $Y = \mathbf{c} + \mathbf{A}X \in \mathbb{R}^m$ is also a multivariate normal vector with

$$E[Y] = \mathbf{c} + \mathbf{A}\boldsymbol{\mu}, \quad Cov(Y) = \mathbf{A}\mathbf{C}\mathbf{A}^{\mathrm{T}}$$

Bias and covariance of the error in the regularized solution

Consider the true (exact) data vector $\mathbf{b}_t \in \mathbb{R}^n$ and true solution $\mathbf{x}_t \in \mathbb{R}^n$,

$$\mathbf{A}\mathbf{x}_t = \mathbf{b}_t$$

expressed as

$$\mathbf{x}_t = \sum_{i=1}^{n} \frac{\mathbf{u}_i^{\mathrm{T}} \cdot \mathbf{b}_t}{\sigma_i} \mathbf{v}_i = \sum_{i=1}^{n} \left(\mathbf{x}_t^{\mathrm{T}} \cdot \mathbf{v}_i\right) \mathbf{v}_i$$

Assume that the errors $\boldsymbol{\xi} = \widehat{\mathbf{b}} - \mathbf{b}_t$ in the noisy data $\widehat{\mathbf{b}}$ have a multivariate normal distribution $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{C})$. Consider

$$\mathbf{A}\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$$

and a solution provided by a filtered SVD regularization (TSVD, Tikhonov)

$$\mathbf{x}_\lambda = \sum_{i=1}^{n} f_i(\lambda) \frac{\mathbf{u}_i^{\mathrm{T}} \cdot \widehat{\mathbf{b}}}{\sigma_i} \mathbf{v}_i$$

We are interested to estimate the mean and covariance of the error

$$\Delta \mathbf{x} = \mathbf{x}_\lambda - \mathbf{x}_t$$

Bias and covariance of the error in the regularized solution

We express the error in the regularized solution as

$$\Delta \mathbf{x} = \mathbf{x}_\lambda - \mathbf{x}_t = \underbrace{\sum_{i=1}^{n} \left(f_i(\lambda) - 1\right) \frac{\mathbf{u}_i^{\mathrm{T}} \cdot \mathbf{b}_t}{\sigma_i} \mathbf{v}_i}_{regularization\ error} + \underbrace{\sum_{i=1}^{n} f_i(\lambda) \frac{\mathbf{u}_i^{\mathrm{T}} \cdot \boldsymbol{\xi}}{\sigma_i} \mathbf{v}_i}_{error\ due\ to\ noisy\ data}$$

Taking the expectation, we obtain the mean of the error (bias)

$$E[\Delta \mathbf{x}] = \sum_{i=1}^{n} \left(f_i(\lambda) - 1\right) \frac{\mathbf{u}_i^{\mathrm{T}} \cdot \mathbf{b}_t}{\sigma_i} \mathbf{v}_i = \sum_{i=1}^{n} \left(f_i(\lambda) - 1\right) \left(\mathbf{x}_t^{\mathrm{T}} \cdot \mathbf{v}_i\right) \mathbf{v}_i \stackrel{def}{=} \Delta \mathbf{x}_{bias}$$

The covariance matrix of the estimation error is expressed as

$$Cov(\Delta x) = \left(\sum_{i=1}^{n} \frac{f_i(\lambda)}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^{\mathrm{T}} \right) \mathbf{C} \left(\sum_{i=1}^{n} \frac{f_i(\lambda)}{\sigma_i} \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}} \right)$$

In particular, if the noise in data is uncorrelated and $\mathbf{C} = \eta^2 \mathbf{I}$ then

$$Cov(\Delta x) = \eta^2 \sum_{i=1}^{n} \frac{f_i^2(\lambda)}{\sigma_i^2} \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$$