Review of Probability and Statistics

Prob. for DES

Radu T. Trîmbițaș

UBB

1st Semester 2010-2011

Purpose and Overview

- The world the model-builder sees is probabilistic rather than deterministic.
  - Some statistical model might well describe the variations.
- An appropriate model can be developed by sampling the phenomenon of interest:
  - Select a known distribution through educated guesses
  - Make estimate of the parameter(s)
  - Test for goodness of fit
- Intention:
  - Review several important probability distributions
  - Present some typical application of these models

Sample Space and Events

- *trial* or *experiment*: term to describe any process or oucome whose outcome is not known in advance (i.e. it has a random behaviour)
- A *sample space*, S, is the set of all possible outcomes of an experiment
- $x \in S =$ *elementary event*.
- $A \subseteq S$ *event*
- if $A \cap B = \emptyset$, events are called *mutually exclusive*

Sigma-field

- A collection $\mathcal{K}$ of events from S is a *σ-field* (*σ-algebra*)
  1. $\mathcal{K} \neq \emptyset$
  2. $A \in \mathcal{K} \Longrightarrow \overline{A} \in \mathcal{K}$
  3. $A_n \in \mathcal{K}, \forall n \in \mathbb{N} \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{K}$
- $(S, \mathcal{K})$, $\mathcal{K}$ σ-field in the sample space S is called *measurable space*
- $(A_i)_{i \in I}$, $A_i \in \mathcal{K}$ *partition* of S if $A_i \cap A_j = \emptyset$ and $\bigcup_{i \in I} A_i = S$

Probability

Definition

$\mathcal{K}$ σ-field in S, $P : \mathcal{K} \to \mathbb{R}$ *probability* if

1. $P(S) = 1$
2. $P(A) \geq 0$, for every $A \in \mathcal{K}$
3. for any sequence $(A_n)_{n \in \mathbb{N}}$ of mutually exclusive events from $\mathcal{K}$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \qquad (\sigma\text{-additive})$$

$(S, \mathcal{K}, P)$ where $(S, \mathcal{K})$ measurable space, P probability - *probability space*

Conditional Probability

- $(S, \mathcal{K}, P)$ probability space, $A, B \in \mathcal{K}$; the *conditional probability* of A given B is
$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$
provided $P(B) > 0$.
- $(S, \mathcal{K}, P)$ probability space s. t. $P(B) > 0$. Then $(S, \mathcal{K}, P(.|B))$ is a probability space
- (*Bayes' formula*) $(S, \mathcal{K}, P)$ probability space, $(A_i)_{i \in I}$ a partition of S, $P(A_i) > 0$, $i \in I$, $A \in \mathcal{K}$, s.t. $P(A) > 0$
$$P(A_j|A) = \frac{P(A_j)P(A|A_j)}{\sum\limits_{i \in I} P(A_i)P(A|A_i)}, \ \forall j \in I$$

Independence

- A is *independent* of B if

$$P(A \cap B) = P(A)P(B)$$

- $(A_n)_{n \in \mathbb{N}}$, $A_n \in \mathcal{K}$ is a *sequence of independent events* if

$$P(A_{i_1} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$$

  for each finite subset $\{i_1, \ldots, i_n\} \subset \mathbb{N}$

- $(A_n)_{n \in \mathbb{N}}$, $A_n \in \mathcal{K}$ is a *sequence of pairwise independent events* if

$$P(A_i \cap A_j) = P(A_i) \cap P(A_j), \quad i \neq j$$

- $(S, \mathcal{K}, P)$ probability space, $A, B \in \mathcal{K}$ A, B independent $\Longleftrightarrow$ $\overline{A}, B$ independent$\Longleftrightarrow$ $A, \overline{B}$ independent$\Longleftrightarrow$ $\overline{A}, \overline{B}$ independent

Random Variables

- $(\mathbb{R}, \mathcal{B})$ $((\mathbb{R}^n, \mathcal{B}^n))$ the measurable space $\mathbb{R}$ $(\mathbb{R}^n)$ endowed with the σ-field generated by open sets

Definition

$(\Omega, \mathcal{K})$, $(E, \mathcal{E})$ measurable spaces $F : \Omega \to E$ *$\mathcal{K}/\mathcal{E}$-measurable* if

$$F^{-1}(B) = \{\omega \in \Omega : F(\omega) \in B\} \in \mathcal{K} \ \forall B \in \mathcal{E}$$

Definitions

$X : \Omega \to \mathbb{R}$ *random variable* if it is $\mathcal{K}/\mathcal{B}$-measurable, i.e.

$$X^{-1}(B) = \{\omega \in \Omega : F(\omega) \in B\} \in \mathcal{K} \ \forall B \in \mathcal{B}$$

$X : \Omega \to \mathbb{R}$ *random vector* if it is $\mathcal{K}/\mathcal{B}^n$-measurable, i.e.

$$X^{-1}(B) = \{\omega \in \Omega : F(\omega) \in B\} \in \mathcal{K} \ \forall B \in \mathcal{B}^n$$

Definitions

the *indicator* of $A \in \mathcal{K}$, $I_A : \Omega \to \mathbb{R}$

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

X r.v. is a *discrete r.v.* if

$$X(\omega) = \sum_{i \in I} x_i I_{A_i}(\omega), \qquad \forall \omega \in \Omega$$

where $I \subseteq \mathbb{N}$, $(A_i)_{i \in I}$ partition of Ω, $A_i \in \mathcal{K}$, $x_i \in \mathbb{R}$. If I is a finite set X - *simple r.v.*

(Cumulative) Distribution Function

Definition

distribution function or *cumulative distribution function* of X, $F : \mathbb{R} \to \mathbb{R}$

$$F(x) = P(X \leq x)$$

1. F nondecreasing
2. F right continuous
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.
4. $P(a < X \leq b) = F(b) - F(a)$
5. $P(a \leq X) = 1 - F(a - 0)$

Density Function

Definitions

X r.v., $F:\mathbb{R}\to\mathbb{R}$ its distribution function. If there exists $f:\mathbb{R}\to\mathbb{R}$ s. t.

$$F(x)=\int_{-\infty}^{x} f(t)dt, \quad \forall x\in\mathbb{R}$$

f is called *(probability) density function* of X. X admits a density function X is called a *continuous r.v.*

X c.r.v, F cdf, f pdf

1. F absolute continuous $F'(x)=f(x)$ for a.e $x\in\mathbb{R}$
2. $f(x)\geq 0$ for a.e $x\in\mathbb{R}$
3. $\int_{\mathbb{R}} f(t)dt=1$
4. $P(X=b)=0$,

$$P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)$$

$$=P(a\leq X\leq b)=\int_a^b f(t)dt$$

Mass Probability Function

- X discrete r.v.

$$p(x) = P(X = x)$$

  p mass probability function

- X discrete r.v. with values $x_1, x_2, \ldots$

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

- X discrete r.v. with values $x_1, x_2, \ldots$. The table

$$\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

  where $p_i = p(x_i)$ the *distribution* of X

Joint Distribution and Joint Density

- $F : \mathbb{R}^n \to R$, $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$ *joint distribution function* of random vector (X_1, X_2)
- $P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) =$
  $F(b_1, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, a_2)$
- (X_1, X_2)
  $$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f\left(t_1, t_2\right) dt_1 dt_2$$
  f joint density function (X_1, X_2) continuous random vector
- $f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$

Independence

- *Marginal cdf*

$$F_X(x) = \lim_{y\to\infty} F(x,y)$$
$$F_Y(x) = \lim_{x\to\infty} F(x,y)$$

- *marginal pdf*

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy$$
$$f_Y(x) = \int_{\mathbb{R}} f(x,y)dx$$

- X, Y drv X, Y *independent* if

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

- X, Y crv X, Y *independent* if

$$f(x,y) = f_X(x)f_Y(y)$$

Expectation

- X r.v. F cdf — *expectation* (*mean value* or *expected value*)

$$E(X) = \int_{-\infty}^{\infty} x dF(x)$$

(if the integral is absolutely convergent!)

- X d.r.v., p mass function

$$E(X) = \sum_{i \in I} x_i P(X = x_i) = \sum_{i \in I} x_i p(x_i)$$

(if the series is absolutely convergent)

- X c.r.v., f pdf

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Expectation – Properties

1. $h : \mathbb{R} \to \mathbb{R}$ $\mathcal{B}/\mathcal{B}$ measurable
$$E(h(x)) = \int_{-\infty}^{\infty} h(x)dF(x)$$
2. $$E(aX + b) = aE(X) + b$$
3. $$E(X + Y) = E(X) + E(Y)$$
4. X, Y independent r.v.
$$E(X \cdot Y) = E(X)E(Y)$$
5. $$X(\omega) \le Y(\omega),\ \omega \in \Omega \Longrightarrow E(X) \le E(Y)$$

Variance

- X r.v. with expectation $E(X)$ – *variance* (*dispersion*) of X
$$V(X) = E\left(X - E(X)\right)^2$$
(if $E\left(X - E(X)\right)^2$ exists) $\sqrt{V(X)}$ – *standard deviation*
- $$V(X) = E(X^2) - E(X)^2$$
- $$V(aX + b) = a^2 V(X)$$
- X, Y i.r.v.
$$V(X + Y) = V(X) + V(Y)$$
$$V(X \cdot Y) = V(X)V(Y) + E(X)^2 V(Y) + E(Y)^2 V(X)$$

Covariance and Correlation

X, Y r.v. *covariance* of X and Y

$$cov(X,Y) = E\left(X - E(X)\right) E(Y - E(Y))$$

correlation coefficient of X and Y

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

Properties

1. $cov(X, X) = V(X)$
2. $$cov(X, Y) = E(X \cdot Y) - E(X)E(Y)$$
3. X, Y independent $cov(X, Y) = \rho(X, Y) = 0$; the converse is false
4. $$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abcov(X, Y), \quad a, b \in \mathbb{R}$$
5. $$cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$$
6. $$-1 \leq \rho(X, Y) \leq 1$$
$$\rho(X, Y) = \pm 1 \iff \exists a, b \in \mathbb{R} : Y = aX + b$$

Moments

Definitions

$k \in \mathbb{N}$, X r.v.

1. $E(X^k)$ (if exists) the *moment of order* k of X
2. $E|X|^k$ (if exists) the *absolute moment of order* k of X
3. $E\left(X - E(X)\right)^k$ (if exists) the *central moment of order* k of X
4. a *quantile* of order α of (the distribution of) X ($\alpha \in (0,1)$) is the number q_α s.t.

$$P\left(X \leq q_\alpha\right) \leq \alpha \leq P\left(X < q_\alpha\right)$$

5. $\alpha = \frac{1}{2}$ *median*, $\alpha = \frac{1}{4}$ *quartiles*, $\alpha = \frac{1}{100}$ *percentiles*

q_α quantile of order α iff $F(q_\alpha - 0) \leq \alpha \leq F(q_\alpha)$
X continuous q_α quantile of order α iff $F(q_\alpha) = \alpha$

Inequalities

- **Markov's inequality**: If X r.v with expectation $E(x)$ and $a > 0$, then

$$P\left(|X| \geq a\right) \leq \frac{E(X)}{a}$$

- **Cebyshev's inequality**

$$P\left(|X - E(X)| \geq a\right) \leq \frac{1}{a^2} V(X)$$

- **Weak law of large numbers (WLLN)** - $(X_n)_{n\in\mathbb{N}}$ sequence of r.v. such that $E(X) < \infty$ for all n obeys the weak law of large numbers if

$$\frac{1}{n}\sum_{k=1}^{n}\left(X_k - E\left(X_k\right)\right) \overset{p}{\to} 0$$

- If $(X_n)_{n\in\mathbb{N}}$ sequence of pairwise i.r.v. s.t. $V(X_n) \leq L < \infty$, for all N, where L constant. Then $(X_n)_{n\in\mathbb{N}}$ obeys WLLN.