Introduction: Descriptive statistics and probability for 1D model

EM13-Probability and statistics: Courses 01-02

September 2014

Manuel SAMUELIDES[1] Zhigang SU[2]

[1]Professor
Institut Supereur de l'Aeronautique et de l'Espace

[2]Professor
Sino-European Institute of Aviation Engineering
Civil Aviation University of China

The only certainty is uncertainty.

—— Piny the Elder (Historia Naturalis, 1535)

Objective of probabilistic methods

- Probabilistic methods(or stochastic methods) are used to predict the issue of uncertain events.
- It consists in endowing such event A with a number $P(A) \in [0,1]$. The greater $P(A)$ is, the more likely we consider the occurence of event A is.
- If the experience producing event A can be repeated many times, $(N \gg 1)$, we consider that the number of occurrences of A is about $N_A \approx N.P(A)$.
  It is the **objective** probability（客观概率）.
- Probability are currently used in more general settings (crash, profit and loss).
  It is called a **subjective** probability（主观概率）.

Probability and gambling

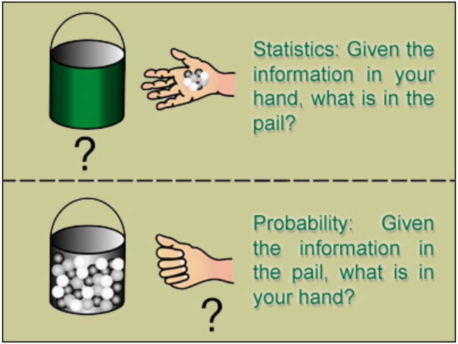
- A natural application of probabilistic methods by human is in gambling strategy.
- In that typical case, the event of interest: "I win" is realized if one among a set of elementary issues occurs. Probability computation amounts to list the favorable elementary issues.
- All elementary issues are considered as equivalent in a fair game.
- This use of probability was first in the XVII century developed mostly by Pascal(1654) and Huygens(1657).

机械加法计算机(1642)

Blaise Pascal (1623-1662)
法国数学、物理学家帕斯卡

Probability and statistics

- The science that fixes the probability from the past observations is called statistics.



Modern advances

- The development of probability leads to stochastic dynamics(随机动力学) where systems are no more stationary.
- Stochastic processes were developed by Markov (1913) and used in linguistics and biology.
- A rigorous framework based on measure theory was built up by Kolmogorov (1933).
- Probability is one of the basis of modern physics in quantum mechanics (Heisenberg relation, Bohr-Einstein controverse (1927)).
- Stochastic differential calculus was founded by Ito (1945) and is the basis of mathematical finance among others.

Using probability and statistics in aeronautical engineering

- Probability is currently used in signal theory and electrical engineering where transmission noise has to be taken into account.
- Elementary signal computation is done using linear assumption and gaussian laws.
- More recently, stochastic properties are taken into account in solid mechanics when using new materials.
- Probability is essential to compute reliability and safeness where discrete events are predicted.
- Probabilistic computation is also used in modelling transportation networks.

Objective of the course

- This course will deal mostly with probability computation in the framework of stationary phenomena.
- Discrete laws are considered as well as density probability.
- Gaussian laws will be at the heart of the course.
- Dynamical phenomena will be introduced.
- Elementary statistics will be presented to allow practical use.
- Today, we shall present simple quantitative probabilistic models on $\mathbb{R}$

Some basic conceptions

随机实验(Randomized experiment)

- 可在同一条件下重复进行
- 每次试验的结果不止一个，可以明确试验的所有结果
- 事先不能确定每次试验的结果

样本(sample)

随机试验的每一个可能结果为一个样本(sample)或一个样本点(sample point)，用小写字母e表示一个样本。

样本空间(sample space)

一个随机试验的所有可能结果的全体为样本空间(sample space)，用字母Ω表示样本空间。

事件(event)

某类结果的全体为一个事件(event)。因此一个事件就是样本空间的一个子集。用大写拉丁字母A，B，C等表示事件。

Example

抛掷一颗骰子，统计一次抛掷得到的点数。

- 样本空间$\Omega = \{1,2,3,4,5,6\}$由6个样本组成
- 事件"一次抛掷得到的点数是偶数"可表示为$A = \{2,4,6\}$
- 如果试验得到的样本$e \in A$，则称在这次试验中A发生了
- 如果试验得到的样本$e \notin A$，则称A未发生。

Remark

- 可以用**集合论**的相关概念作为建立随机模型的基本元素。
- 通常情况下，我们只能通过事件来了解试验的结果。因此在样本空间、样本、事件这三个概念中我们更关心**事件**。
- 事件类的选取不是唯一的!

Probability space / 概率空间

A probability space $(\Omega, \mathcal{F}, P)$ consists of three parts:

1. A sample space, Ω, which is the set of all possible outcomes.
2. A set of events $\mathcal{F}$, where each event is a set containing zero or more outcomes.
3. The assignment of probabilities to the events; that is, a function P from events to probabilities.

直观含义

事件A的概率$P(A)$的直观含义是一次随机试验中A发生的可能性的大小，如 何确定映射

$$P : \mathcal{F} \to R, \qquad A \mapsto P(A)$$

Remark

$(\Omega, \mathcal{F})$ is the measurable space.

Definition

A probability space $(\Omega, \mathcal{F}, P)$ is a measure space such that the measure of the whole space is equal to one (i.e. $P(\Omega) = 1$).

- the sample space Ω — an arbitrary non-empty set,
- the σ -algebra $\mathcal{F}$ (also called σ -field) — a set of subsets of Ω, called events, such that:
  - $\Omega \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$
  - $\mathcal{F}$ is closed under complements: if $A \in \mathcal{F}$, then also $\bar{A} \in \mathcal{F}$
  - $\mathcal{F}$ is closed under countable unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, ...$, then also $\cup_i A_i \in \mathcal{F}$
  - $\mathcal{F}$ is the collection of all the random events of interest. The set operations $\cup, \cap, ^-$ are associate to the logical operations "OR", "AND", "NOT".
- the probability measure $P : \mathcal{F} \to [0, 1]$ — a function on $\mathcal{F}$ such that:
  - P is countably additive: if $\{A_i\} \in \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $P(\cup A_i) = \sum P(A_i)$
  - the measure of entire sample space is equal to one: $P(\Omega) = 1$.

Remark

$\mathcal{F}$是Ω上的一个σ代数或σ域是指当且仅当它非空，并且对它的可数多个元素按交、并、余运算封闭。

Kolmogorov probabilistic model

设E为随机实验，Ω是它的样本空间，对于E的每一个事件A，赋于实数记为$P(A)$， 如果它满足下述三个条件：

1. 非负性：对每一个事件A，$0 \leq P(A) \leq 1$;
2. 规范性：对必然事件S，$P(S) = 1$;
3. 可数可加性：对任何互不相容的事件$A_1, A_2, \ldots, A_n, \ldots$,

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_1) + \cdots$$

则定义$P(A)$为事件A的概率。

概率为1的事件不一定发生

Example

从$[0,1]$中任取一个数（以取出的结果作为随机变量，就是$U[0,1]$）取出来是0的概率是0，因为单点的长度为0，所以其概率为0/1=0。 取出来不是0的概率是1（对立事件），但也有可能不发生。

为什么这和我们现实中不同呢？

在离散情况下为什么没有这样的情况呢？这样的问题的发生和概率定义中的公理有关系，实际上概率的公理最多只限制到可列个样本点的情况，对更多的约束不强。另一个方面，概率本身是认为赋予事件的一个数，某个角度来说他是很主观，所以有的时候和事实不符。

Properties of probability

性质1

当两个事件满足$A \subset B$时，

$$P(B) = P(A) + P(\bar{A}B)$$

$$P(\bar{A}B) = P(B) - P(A)$$

性质2

$0 \le P(A) \le P(S) = 1;$

性质3(有限可加性)

对有限多个互不相容的事件$A_1, A_2, \ldots, A_n$,

$$P(A_1 + A_2 + \cdots A_n) = P(A_1) + P(A_1) + \cdots P(A_n)$$

Recall that P is sigma-additive. That implies that if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.
In particular, $P(A) + P(\hat{A}) = 1$

性质4

对任意两个事件A, B（不必互不相容），

$$\begin{aligned} P(A+B) &= P(A)+P(B)-P(AB) \\ P(A+B) &\leq P(A)+P(B) \end{aligned}$$

性质5(概率的连续性)

一个事件序列的极限事件的概卑等于这列事件的概率的极限。即

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

Random variable / 随机变量

- 用数值来刻画各种结果出现的可能性（概率）
- 用数量的方式表示实验结果（随机变量）

Remark

随机变量是样本空间Ω映射到实数域R的函数$X : A \to R$

- 随机变量的数值$X(A)$按由事件A得出
- 事件A的随机性导致$X(A)$表现出不确定性

Definition

设随机实验的样本空间$\Omega = \{e\}$，如果对于每一个$e \in \Omega$有一个实数$X(e)$与之对应，这样就得到一个定义在Ω上的实值单值函数$X(e)$，称$X(e)$为随机变量。

进一步说明

设一变量$X(e)$，它能随机地取数值（但不能预言它将取什么数值），而对应于每 一数值（或某一范围的值）有相应的概率，称之为随机变量。

Cumulative distribution function / 概率分布函数

Definition

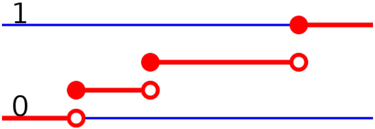
Cumulative distribution function:

$$\begin{aligned} F(x) &\triangleq P\{\omega : X(\omega) \le x\} \\ &= P\{\omega : X(\omega) \in (-\infty, x]\} \\ &= P\{X \le x\} \quad \forall x \in R \end{aligned}$$

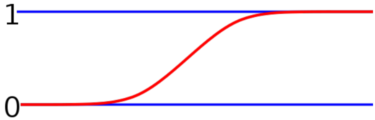
Because a probability distribution P on the real line is determined by the probability of a scalar random variable X being in a half-open interval $(-\infty, x]$, the probability distribution is completely characterized by its cumulative distribution function.

Cumulative distribution function / 概率分布函数

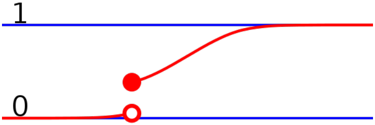
- Discrete random variable:



- Continuous random variable:



- Mixed random variable:



- Discrete random variable
  - If the value set of X is a countable set $\{x_i\}$, its probability law is defined by

$$P(X = x_k) = P_X(x_k) = p_k$$

  - Cumulative distribution function

$$F_X(x) = \sum_k p_k U(x - x_k)$$

- Continuous random variable
  - Cumulative distribution function

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

Probability density function(pdf) / 概率密度函数

- a function that describes the relative likelihood for this random variable to take on a given value.
- The probability for the random variable to fall within a particular region is given by the integral of this variable’s density over the region.
- The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.

Continuous distributions

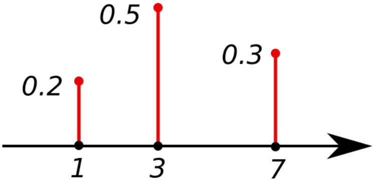
$$
\begin{aligned}
f(x) &= \frac{d}{dx}F(x) \\
dF(x) &= dP(x) \triangleq P(dx) = P\{\omega : x < X(\omega) \leq x + dx\}
\end{aligned}
$$

Discrete distributions

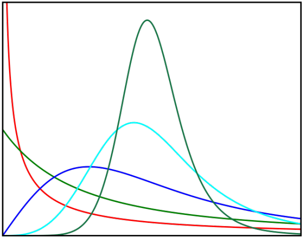
$$
f(x) = \sum_{i=1}^{n} p_i \delta(x - x_i)
$$

Probability density function(pdf) / 概率密度函数

- Discrete random variable:



- Continuous random variable:



Characteristic function / 特征函数

Definition

The **characteristic function** ϕ_X of a random variable X is a version of the Fourier transform of its probability law and is defined by

$$\phi_X(t) = E[e^{jtX}] = \int e^{jtx} f_X(x) dx$$

$$f_X(x) = \frac{1}{2\pi} \int \phi_X(t) e^{-jtx} dt$$

Remark

Characteristic function also completely determines behavior and properties of the probability distribution of the random variable X

- $\phi_X(0) = 1$
- $\phi'_X(0) = jE(X)$
- $\phi"_X(0) = -E(X^2) = -Var(X) - E(X)^2$

- Discrete random variable

$$\phi_X(t) = \sum_k p_k e^{jtx_k}$$

- Continuous random variable

$$\phi_X(t) = \int f_X(x) e^{jtx} dx$$

Properties of characteristic function

- The characteristic function of a real-valued random variable always exists, since it is an integral of a bounded continuous function over a space whose measure is finite
- A characteristic function is uniformly continuous on the entire space
- It is non-vanishing in a region around zero: $\phi_X(0) = 1$
- It is bounded: $|\phi_X(t)| \leqslant 1$
- It is Hermitian: $\phi_X(-t) = \overline{\phi_X(t)}$. In particular, the characteristic function of a symmetric (around the origin) random variable is real-valued and even
- There is a bijection between distribution functions and characteristic functions. That is, for any two random variables X_1, X_2

$$F_{X_1} = F_{X_2} \Leftrightarrow \phi_{X_1}(t) = \phi_{X_2}(t)$$

- If a random variable X has moments up to k-th order, then the characteristic function ϕ_X is k times continuously differentiable on the entire real line. In this case

$$E[X^k] = (-i)^k \phi_X^{(k)}(0)$$

- f a characteristic function ϕ_X has a k-th derivative at zero, then the random variable X has all moments up to k if k is even, but only up to $k1$ if k is odd

$$\phi_X^{(k)}(0) = i^k E[X^k]$$

- If $X_1, \cdots, X_n$ are independent random variables, and $a_1, \cdots, a_n$ are some constants, then the characteristic function of the linear combination of the X_i 's is

$$\phi_{\{a_1 X_1 + \cdots + a_n X_n\}}(t) = \phi_{X_1}(a_1 t) \cdot \ldots \cdot \phi_{X_n}(a_n t)$$

One specific case is the sum of two independent random variables X_1 and X_2 in which case one has

$$\phi_{X_1 + X_2}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t)$$

Expectation / 期望、均值

Definition

Let $Y = g(X)$ be a real random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. If Y is P-integrable, i.e. $Y \in L^1(\Omega, \mathcal{F}, P)$, the integral $\int g(x)dP(x)$ is called **expectation** of Y and is denoted

$$E(Y) = \int g(x)dP(x) = \int g(x)f_X(x)dx$$

Discrete form

Let P be a discrete probability distribution on $\mathbb{R}$ with the previous notations which represents the variations of a "random variable" $\mathbf{X}$. We define its **expectation** by

$$E(X) = \sum_k p_k x_k$$

Variance and standard deviation / 方差和标准差

Definition

- The **variance** of a real random variable X is the positive quantity

$$\sigma_X^2 = Var(X) = E\{[X - E(X)]\}^2$$

- The **standard deviation** of X is the positive quantity

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

The standard deviation represents a good model of the dispersion of X. Whisker plot is more robust that expectation and standard deviation but these quantites are more tractable from a model point of view.

Theorem

$$Var(X) = E(X^2) - E(\cancel{Y})^2$$

Proof:

$$\begin{aligned} Var(X) = \{[X - E(X)]\}^2 &= E[X^2 - 2E(X)X + E(X)^2] \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Moments of a real random variable

Let X be a real random variable.

- The moments of X (if they exist) are the real $E(X^n)$.
- $E(aX+b) = aE(X) + b$
- $\sigma_{aX+b} = |a|\sigma_X$
- The expectation characterizes the trend
- the standard deviation characterizes the dispersion
- the shape is given by the standard reduction $\frac{X-E(X)}{\sigma_X}$

Bienayme-Chebycheff inequality

Theorem

Let X a real random variable with expectation and variance, then

$$P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}, \quad \forall \epsilon > 0$$

Proof:
We have from integral minoration

$$Var(X) = E\{[X - E(X)]^2\} \geq \epsilon^2 P(|X - E(X)| \geq \epsilon)$$

The quality of this majoration is bad but it shows the way to scale the deviation form the trend and it shows how the variance controls dispersion

Example of binomial law $\mathcal{B}(n,p)$ / 二项式分布

- X takes its value in the integer set $\{0,\ldots,n\}$ and

$$p_k = P(X=k) = C_n^k p^k (1-p)^{n-k}$$

- Its characteristic function is $\phi_X(t) = [pe^{jt} + (1-p)]^n$
- We get $E(X) = np \qquad Var(X) = np(1-p)$

Remark

For $n=1$ we get the Bernoulli law which is the generic probability law for any binary random variable.

Remark

Binomial laws are connected to discrete random walks (see exercise)

Example of binomial law $\mathcal{B}(n,p)$ / 二项式分布

- X takes its value in the integer set $\mathbb{N}^\star$ and

$$\forall k = 1, \ldots, \quad p_k = P(X = k) = (1-p)^{k-1}p)$$

- Its characteristic function is $\phi_X(t) = \frac{pe^{jt}}{1-(1-p)e^{jt}}$
- We get $E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$

Remark

Geometric law is commonly used to model waiting time (see exercise). It is a discretization of exponential law

Example of Poisson law $\mathcal{P}_\lambda$ / 泊松分布

- X takes its value in the integer set $\mathbb{N}$ and

$$p_k = P(X = k) = exp(-\lambda)\frac{\lambda^n}{n!}$$

- Its characteristic function is $\phi_X(t) = exp(\lambda(e^{jt} - 1))$
- We get $E(X) = \lambda \qquad Var(X) = \lambda$

Remark

Poisson laws are very important to model number of random events such as telephone calls, length of waiting lines ...(see exercise)

Example of uniform law $\mathcal{U}_{[a,b]}$/均匀分布

- The density is defined by $h(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(X)$
- Its characteristic function is $\phi_X(t) = \frac{e^{jtb}-e^{jta}}{jt(b-a)}$
- We get $E(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$

Remark

Uniform laws are the more common translation of "pure chance". Nevertheless, there is no uniform law on an unbounded set. Is it possible to choose randomly a point in the universe?

Example of Gauss law $\mathcal{N}(m, \sigma)$/高斯（正态）分布

- The density is defined by $h(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\{\frac{(x-m)^2}{2\sigma^2}\}$
- Its characteristic function is $\phi_X(t) = exp\{jtm - \frac{t^2\sigma^2}{2}\}$
- We get $E(X) = m \qquad Var(X) = \sigma^2$

Remark

The gaussian law is so common in physical modelling that its universality is questioning. Central limit theorem gives the key of the mystery.

Example of Gamma law $\Gamma(k,\theta)$/Γ分布

- The density is defined by $h(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\{\frac{-x}{\theta}\}$
- Its characteristic function is $\phi_X(t) = \exp\{jtm - \frac{t^2\sigma^2}{2}\}$
- We get $E(X) = k\theta \qquad Var(X) = k\theta^2$

Remark

Gamma laws are a large category that includes exponential laws ($k = 2$) and χ^2 laws ($k = \frac{n}{2}$) that are sum of square of independent centered Gaussian variables. Their use is basic for statistical techniques.

For Further Reading I

- Sylvie Méléard.
  Aléatoire.
  Ecole Polytechnique, 2009.
- Anastasios Papoulis.
  Probability, random variables and stochastic processes.
  Mac Graw Hill, 1991.
- Manuel Samuelides.
  Exercices corrigés de probabilités.
  SUPAERO, 1998.