Random vectors

MA13-Probability and statistics: Courses 03-04

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

Homework–20140904

Compute expectation, variance and characteristic functions of the following discrete probability on $\mathbf{R}$

- Binomial law $B(n,p)$
  $P(X=k) = C_n^k p^k (1-p)^{n-k} \qquad 0 \leqslant k \leqslant n$
- Geometric law $G(p)$
  $P(X=k) = p(1-p)^{k-1} \qquad k \in N$
- Poisson law $P(\lambda)$
  $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$
- Gamma law $\Gamma(\alpha,\theta)$
  $f_\Gamma(x,\alpha,\theta) = \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\theta^\alpha \Gamma(\alpha)}$

Homework–20140909

1. Medical testing

Consider a medical testing of a disease. 1% of the population is ill. If a person is ill, the test is positive with probability 99%. If a person is safe, the test is positive with a probability of 1% (false positive). Suppose one person is tested positive. What is its probability to be ill ?

2. Simple reliability computation

We consider 3 elements which can break. Their life are defined by three random 0-1 independent variables X, Y, Z with respective parameters p, q, r. The operational character of the whole system is given by the simple Boolean formula: $S = (X \Cap Y) \Cup Z$.

1. Detail the eight-state space of the detailed system and give its probability law by computing the probability of each event
2. Compute the law of S
3. Detail the conditional probability if $S = 0$ (failure)

3. Never becoming old

A probability law of break-time T is said without memory if $Pr(T > t + s | T > s) = Pr(T > t)$, i.e. it is the original law of T shifted with a delay of s. Show that the exponential law

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \geqslant 0 \\ 0 & t < 0 \end{cases}$$

is without memory.

多维随机变量

联合概率分布

若X_1，X_2是二维随机变量，对于任意实数x_1，x_2，我们定义

$$F_{X_1X_2}(x_1,x_2) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2\}$$

为联合概率分布函数，或称随机变量X_1，X_2的联合分布。

联合概率密度

若X，Y是二维随机变量，它们的联合密度函数为

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

N维联合分布

$$F_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=P(X_1<x_1,X_2<x_2,\cdots,X_n<x_n)$$

$$f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=\frac{\partial^n F_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)}{\partial x_1\partial x_2\cdots\partial x_n}$$

离散型（Discrete）

$$F_{X_1X_2}(x_1,x_2) = \sum_{n=1}^{N}\sum_{m=1}^{M} P(X_1 = x_{1,n}, X_2 = x_{2,m}) \cdot U(x_1 - x_{1,n})U(x_2 - x_{2,m})$$

$$f_{XY}(x,y) = \sum_{n=1}^{N}\sum_{m=1}^{M} P(X = x_n, Y = y_m)\delta(x - x_n)\delta(y - y_m)$$

多维分布函数性质

1. $F_{X_1X_2}(-\infty,-\infty)=0, F_{X_1X_2}(x_1,-\infty)=0,$
   $F_{X_1X_2}(-\infty,x_2)=0$
2. $F_{X_1X_2}(\infty,\infty)=1$
3. $0\le F_{X_1X_2}(x_1,x_2)\le 1$
4. $F_{X_1X_2}(x_1,x_2)$是X_1，X_2的非减函数
5. $P\{x_1<X<x_2,y_1<Y<y_2\}\ge 0$，且
   $$P\{x_1<X<x_2,y_1<Y<y_2\}=$$
   $$F_{XY}(x_1,y_1)+F_{XY}(x_2,y_2)-F_{XY}(x_1,y_2)-F_{XY}(x_2,y_1)$$
6. $F_{XY}(x,\infty)=F_X(x)$，Y是必然事件
   $F_{XY}(\infty,y)=F_Y(y)$，X是必然事件

1,2,5是任一函数$G(x,y)$为二维分布函数的检验准则。

多维密度函数性质

1. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = 1$
2. $\int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)\,dxdy = F_{XY}(x,y)$
3. $\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = F_X(x)$
4. $\int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{XY}(x,y)\,dxdy = F_Y(y)$
5. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$
   $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$

Remark

性质1,2是检验一个函数$G(x,y)$是否为密度函数的准则。

边缘概率

边缘概率分布

若$F_{XY}(x,y)$中以一个随机变量趋向无穷大，所得到的一维随机变量的分布

$$\begin{aligned} F_X(x) &= F_{XY}(x,\infty) \\ F_Y(y) &= F_{XY}(\infty,y) \end{aligned}$$

称之为边缘分布函数。

Proof.

设事件$A=\{X\le x\}$，$B=\{Y\le y\}$，它们的联合分布为：

$$F_{XY}(x,y)=P\{X\le x, Y\le y\}=P(A\cap B)$$

若$y\to\infty$，事件B为确定事件。

$$\therefore P(A\cap B)=P(A\cap\infty)=P(A)=P(X\le x)=F_X(x)$$

$\square$

边缘概率密度函数

$$
\begin{aligned}
f_X(x) &= \frac{dF_{XY}(x,\infty)}{dx} \\
f_Y(y) &= \frac{dF_{XY}(\infty,y)}{dy}
\end{aligned}
$$

关系

$$
\begin{aligned}
F_X(x) &= F_{XY}(x,\infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy \\
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \\
f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx
\end{aligned}
$$

N维随机变量的K维边缘分布

$$f_{X_1X_2\cdots X_K}(x_1,x_2,\cdots,x_K)=\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{N-K}f_{X_1X_2\cdots X_N}(x_1,x_2,\cdots,x_N)\,dx_{K+1}dx_{K+2}\cdots dx_N$$

Definition of conditional discrete probability

Definition

Let $(\Omega, \mathcal{F}, P)$ a probability space

- Let $B \in \mathcal{F}$ a random event such that $P(B) \neq 0$, then the conditional probability if B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability

若X,Y是二维随机变量，在Y发生的条件下，X出现的概率有两种可能：

- 区间的情况（$y_a \leq Y \leq y_b$）
- 点的情况（$Y = y$）

区间情况

若事件$B=\{y_a<Y\leq y_b\}$、且$P(B)=P\{y_a<Y\leq y_b\}\neq 0$

$$\begin{aligned} F_X(x|y_a<Y\leq y_b) &= \frac{P\{X<x\cap y_a<Y\leq y_b\}}{P(y_a<Y\leq y_b)} \\ &= \frac{F_{XY}(x\cap y_b)-F_{XY}(x\cap y_a)}{F_Y(y_b)-F_Y(y_a)} \\ &= \frac{\int_{-\infty}^{x}\int_{y_a}^{y_b} f_{XY}(x,y)\,dxdy}{\int_{-\infty}^{\infty}\int_{y_a}^{y_b} f_{XY}(x,y)\,dxdy} \end{aligned}$$

$$\begin{aligned} f_X(x|y_a<Y\leq y_b) &= \frac{F_X(x|y_a<Y\leq y_b)}{dx} \\ &= \frac{\int_{y_a}^{y_b} f_{XY}(x,y)\,dy}{\int_{-\infty}^{\infty}\int_{y_a}^{y_b} f_{XY}(x,y)\,dxdy} \end{aligned}$$

点情况

若$y_a \to y_b$，则间隔条件变为

$$
\begin{aligned}
F_X(x|Y=y_a) &= \lim_{y_b \to y_a} F_X(x|y_a < Y \le y_b) \\
&= \lim_{y_b \to y_a} \frac{F_{XY}(x \cap y_b) - F_{XY}(x \cap y_a)}{F_Y(y_b) - F_Y(y_a)} \\
&= \frac{\int_{-\infty}^{x} f_{XY}(x, y_a)\, dx}{f_Y(y_a)} \\
&= \frac{\int_{-\infty}^{x} f_{XY}(x, y_a)\, dx}{\int_{-\infty}^{\infty} f_{XY}(x, y_a)\, dx}
\end{aligned}
$$

Definition of conditional probability

Definition

Let $(\Omega, \mathcal{F}, P)$ a probability space and X be a random variable with values in E with probability law f_X on E. Then the conditional probability if X is the application (measurable and almost surely defined) $x \in E \rightarrow P(\cdot|X = x) \in P$ which checks the recomposition property

- $P(X = x|X = x) = 1$
- $P(B) = \int_{X \in E} P(B|X = x) f_X(x) dx$ or equivalently $\mathbf{E}(Y) = \int_{X \in E} \mathbf{E}(Y|X = x) f_X(x) dx$

条件概率的推广

概率乘法定理

若$P(A) > 0$，则

$$P(AB) = P(B|A) \cdot P(A)$$

全概公式

若$B_1, B_2, \cdots, B_n$是n个互不相容的事件，且$\sum_{i=1}^n P(B_i) = 1$，则

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i) = \sum_{i=1}^n P(AB_i)$$

Remarks

通常在应用全概公式时，样本空间的正划分是需要我们特意构造的，其原则是：

1. 每个B_k的情况清楚（$P(B_k)$容易被确定）；
2. 每个B_k对所研究的对象A的影响清楚（$P(A|B_k)$容易被确定，这往往是要在B_k发生的情况下选取经适当改造的模型)。

Bayes公式

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)}$$

Proof.

$$\begin{aligned} \because B_i \cap A &= A \cap B_i \\ P(B_i \cdot A) &= P(A \cdot B_i) \\ P(A|B_i)P(B_i) &= P(B_i|A)P(A) \\ \therefore P(B_i|A) &= \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} \end{aligned}$$

□

两点说明

1. 我们分别称$P(B_k)$和$P(B_k|A)$为B_k的先验概率和后验概率。
2. Bayes公式常被用来进行概率意义上的因果分析。

Example

B_k是某种疾病，$P(B_k)$是这种疾病的发病率，它的数值来自卫生部门的统计结果。 而A是一个病人的症状，$P(A|B_k)$是不 同疾病导致这种症状的可能性，医生的任务就是根据症状来确诊患者的病因(比如使$P(B_k|A)$ 明显大于其他条件概率值的B_k)。

统计独立性

定义

若两事件A，B有非零的概率，如果一个事件发生的概率不影响另一事件 发生的概率，则称这两个事件相互独立(或简称独立)，即：

$$P(AB) = P(A)P(B)$$

三个事件独立定义

设三个事件A，B，C，若满足

$$
\begin{aligned}
P(AB) &= P(A)P(B) \\
P(BC) &= P(B)P(C) \\
P(CA) &= P(C)P(A) \\
P(ABC) &= P(A)P(B)P(C)
\end{aligned}
$$

则事件A，B，C相互独立(或简称独立)。

Definition of independence

Definition

Random variables $(X_1, \dots, X_n)$ are independent if their joint law is the product of the marginal laws of the components

$$f_{(X_1,\dots,X_n)} = f_{X_1} \cdots f_{X_n}$$

Proposition

Random variables $(X_1, \dots, X_n)$ are independent iff for all $(g_1, \dots, g_n)$ measurable and bounded one has

$$E[g_1(X_1) \cdots g_n(X_n)] = E[g_1(X_1)] \cdots E[g_n(X_n)]$$

Example: In case 2 of example 1, the input bit and the transmission noise are independent

统计独立

随机变量X,Y统计独立的充要条件

若将X,Y两随机变量为统计独立时

$$
\begin{aligned}
F_X(x|Y \leqslant y) &= F_X(x) \\
F_Y(y|X \leqslant x) &= F_Y(y) \\
\because F_{XY}(x,y) &= F_X(X|Y \leqslant y)F_Y(y) \\
&= F_Y(Y|X \leqslant x)F_X(x) \\
\therefore F_{XY}(x,y) &= F_X(x)F_Y(y) \\
f_{XY}(x,y) &= f_X(x)f_Y(y)
\end{aligned}
$$

推广

N维随机变量统计独立的充要条件为：

$$
\begin{aligned}
F_{X_1X_2\cdots X_N}(x_1,x_2,\cdots,x_N) &= F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_N}(x_N)\\
f_{X_1X_2\cdots X_N}(x_1,x_2,\cdots,x_N) &= f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_N}(x_N)
\end{aligned}
$$

Properties of independence

- From Fourier characterization of measures, it comes that Random variables are independent iff the characteristic function of their joint law is the tensor product of the characteristic functions of the component

$$\begin{aligned}\phi_{X_1\cdots X_n}(t_1,\ldots,t_n) &= E(\exp(jt_1X_1+\cdots+jt_nX_n)) \\ &= \phi_{X_1}(t_1)\cdots\phi_{X_n}(t_n)\end{aligned}$$

- In the continuous case, the density of the joint law is the product of the marginal densities.

- It is easy to see that the law of the sum of independent variables is the convolution product of the law of the terms

$$
\begin{aligned}
E(g(X_1+X_2)) &= \int\int g(x_1+x_2) f_{X_1X_2}(x_1,x_2)dx_1dx_2 \\
&= \int\int g(x_1+x_2) f_{X_1}(x_1) f_{X_2}(x_2)dx_1dx_2 \\
&= \int\int g(s) f_{X_1}(s-x_2) f_{X_2}(x_2)dsdx_2 \\
&= \int g(s)\left[\int f_{X_1}(s-x_2) f_{X_2}(x_2)dx_2\right]ds \\
&= \int g(s)[f_{X_1}(s) * f_{X_2}(s)]ds
\end{aligned}
$$

Example

(小概率原则)在一个随机试验中事件A发生的概率为$\varepsilon > 0$。无论$\varepsilon > 0$ 多小，只要我们不断的独立重复做该试验，A迟早会发生的概率为1，或者不严格地说，A迟早会出现。
设A_k表示在第k次试验时A发生。则$A_1, \ldots, A_n, \ldots$相互独立，$P(A_k) = \varepsilon$。于是

$$\begin{aligned}
P\left(\cup_{k=1}^{\infty} A_k\right) &= \lim_{n\to\infty} P\left(\cup_{k=1}^{n} A_k\right) \\
&= \lim_{n\to\infty} \left\{1 - \Pi_{k=1}^{n} P(\bar{A}_k)\right\} \\
&= \lim_{n\to\infty} \left\{1 - \Pi_{k=1}^{n} [1 - P(A_k)]\right\} \\
&= 1 - \lim_{n\to\infty} (1-\varepsilon)^n \\
&= 1
\end{aligned}$$

Remark

- 如果一只长生不老的猴子不知疲倦乱敲键盘，那么它几乎注定会写出莎翁的名著
- 不要说自己没有天份，只是你还努力得不够!!!

Definition of covariance

Definition

- The covariance of two random variables is the L^2 scalar product of their centered versions

$$
\begin{aligned}
Cov(X,Y) &= E\{[X-E(X)][Y-E(Y)]\} \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

- This definition is extended to random vectors where $Cov(\mathbf{X},\vec{Y})$ is the matrix with (i,j) term is $Cov(X_i,Y_j)$. In that sense covariance is sometimes called **cross-correlation**.

Properties of covariance

- Covariance is bilinear.
- From Schwarz inequality, one gets for real variables

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$$

- $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ is called the **correlation coefficient** of X and Y and $|\rho_{X,Y}| \leq 1$
- If $Cov(X,Y) = 0$, X and Y are said uncorrelated and one notes $X \perp Y$
- If X and Y are independent, they are uncorrelated but two variables may be uncorrelated without being independent.

Random vector

- Random vector $\mathbf{X} = \{X_1, \cdots, X_n\}$ is defined on a finite-dimensional vector space $\mathbb{R}^n$ or $\mathbb{C}^n$
- The expectation $E(\mathbf{X})$ is a vector which components are the expectations of the components (X_i)
- The above mentioned may be extended to an infinite-dimensional Hilbert space and this can be used to study stochastic process but it overcomes the level of the present course.

Covariance matrix of a random vector

Definition

The covariance matrix of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is the (d, d) symmetric, positive matrix with (i, j)-term $Cov(X_i, X_j)$ and is noted $Cov(\mathbf{X}) = \Gamma_{\mathbf{X}}$

Proposition

Let $\mathbf{A}$ be the matrix of a linear mapping defined on the value space of $\mathbf{X}$, then $\Gamma_{\mathbf{AX}} = \mathbf{A}\Gamma_{\mathbf{X}}\mathbf{A}^H$

As a special case, if

$$(\mathbf{u}_1, \mathbf{u}_2) \in \mathbb{R}^d \times \mathbb{R}^d, Cov((\mathbf{u}_1|\mathbf{X}), (\mathbf{u}_2|\mathbf{X})) = \mathbf{u}_1^H \Gamma_{\mathbf{X}} \mathbf{u}_2$$

Definition of conditional expectation

- Let X with values in E and let us consider the set $\{g(X)\}$ such that $\mathbf{E}[g(X)]$ exists. It is a Banach subspace of $L^1(\Omega, \mathcal{F}, \mathcal{P})$ which is denoted $L^1(\Omega, \mathcal{F}_X, \mathcal{P})$ and which is isomorphic to $L^1(E, \epsilon, \mathcal{P}_X)$.
- Let $Y \in L^1(\Omega, \mathcal{F}, \mathcal{P})$. From the recomposition property,

$$\mathbf{E}(Y) = \int \mathbf{E}(Y|X = x) f_X(x) dx$$

The function $X \in E \to \mathbf{E}(Y|X = x)$ is defined almost surely and is associated to a random variable function of X which is denoted $\mathbf{E}(Y|X)$ and which expectation is $\mathbf{E}(Y)$.

Definition

The random variable $\mathbf{E}(Y|X)$is called the conditional expectation of Y if X is a random variable.

Projection property of conditional expectation

With the same notations, we have the following properties of conditional expectation on L^1

- $Y \in L^1 \to \mathbf{E}(Y|X) \in L^1$ is linear
- $|\mathbf{E}(Y|X)| \leq \mathbf{E}(|Y||X|)$ implies that $||\mathbf{E}(Y|X)||_1 \leq ||Y||_1$
- From the support of conditional probability property, we have $\mathbf{E}(g(X)Y|X=x) = g(x)\mathbf{E}(Y|X=x)$
  In other terms, $\mathbf{E}(g(X)Y|X) = g(X)\mathbf{E}(Y|X)$.
- Notably $\mathbf{E}([\mathbf{E}(Y|X)]|X) = \mathbf{E}(Y|X)$

This projection property suggests to consider the properties of conditional expectation with respect to 2nd order random variables.

Conditional expectation is an orthogonal projection

Theorem

Conditional expectation if X is the orthogonal projection of $L^2(\Omega, \mathcal{F}, \mathbf{P})$（表示概率空间$(\Omega, \mathcal{F}, \mathbf{P})$上平方可积的可测函数所组成的空间） onto $L^2(\Omega, \mathcal{F}_x, \mathbf{P})$. In other terms

$$\forall Z = g(X) \in L^2(\Omega, \mathcal{F}_x, \mathbf{P}), \mathbf{E}(Z\mathbf{E}(Y|X)) = \mathbf{E}(ZY)$$

Proof From the support property we have

$$\mathbf{E}(ZY|X) = Z\mathbf{E}(Y|X)$$

We integrate and get orthogonal projection condition.

Remark

Conditional expectation is the orthogonal projection onto the ∞-dimensional space of all the L^2 functions of X while linear regression is the more restrictive orthogonal projection onto the 2-dimensonal space of affine functions of X.

Conditional expectation and independence

Manuel SAMUELIDES, Zhigang SU

中国民航大学

Theorem

Random variables X and Y are independent iff

$$\mathbf{E}[g(Y)|X] = \mathbf{E}[g(Y)], \qquad \forall g \in L^2(E, \epsilon, P_Y)$$

Proof Independence of X and Y is equivalent to:

$$\forall f \in L^2, \forall g \in L^2, \mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$$

which amounts to

$$\forall f \in L^2, \forall g \in L^2, \mathbf{E}\{f(X)\mathbf{E}[g(Y)|X]\} = \mathbf{E}\{f(X)\mathbf{E}[g(Y)]\}$$

which is equivalent to the here above statement.

Conditional expectation and disjunction

Theorem

Let X_1 and X_2 be independent random elements and let Y be a random variable, then

$$\mathbf{E}(Y|X_1, X_2) = \mathbf{E}(Y|X_1) + \mathbf{E}(Y|X_2) - \mathbf{E}(Y)$$

Proof The orthogonal projection on the sum of two orthogonal spaces is the sum of the projections on each space.

Conditional variance

We want to quantify the influence of a factor X on a final result Y. It is possible if we have the joint law of (X, Y) through the conditional variance

Definition

The conditional variance of Y given X is

$$Var(Y|X) = \mathbf{E}\{[Y - \mathbf{E}(Y|X)]^2\} = \mathbf{E}(Y^2|X) - \mathbf{E}(Y|X)^2$$

Proof of the analysis of variance theorem:

Theorem

We have

$$Var(Y) = Var[\mathbf{E}(Y|X)] + \mathbf{E}[Var(Y|X)]$$

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

Remark

- The first term measures the variability of Y due to X.
- It is null when X is independent of Y because $\mathbf{E}(Y|X) = \mathbf{E}(Y)$
- The second one measures the residual variation.