Dependent variables and change of variables

MA13-Probability and statistics: Courses 05-06

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

Question in the last lesson

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: $S = (X \cap Y) \cup Z$.

- Compute the law of S if $S = 0$ (failure)

Solution:

2.

$$Pr(S = 1) \quad = \quad Pr\{[(X \cap Y) \cup Z] = 1\} =?$$

$$
\begin{aligned}
Pr(S=1) &= Pr\{[(X\cap Y)\cup Z]=1\}\\
&= Pr(X=1,Y=1)+Pr(Z=1)\\
&\quad -Pr(X=1,Y=1,Z=1)\\
&= pq+r-pqr
\end{aligned}
$$

Remark

Disjoint sets:

$$Pr(A\cup B)=Pr(A)+Pr(B)$$

Otherwise,

$$Pr(A\cup B)=Pr(A)+Pr(B)-Pr(A\cap B)$$

Homework–20140916

1. Project: Generate some samples of zero mean Gaussian random variable X with variance $\sigma^2 = 1$. Utilizing these samples to plot the probabilistic density function, the real and ideal characteristic functions. Furthermore, generate the samples obeying the lognormal distribution with mean $\mu = 3$ and variance $\sigma^2 = 100$ by employing these samples.

2. Law computation

Suppose that X and Y have the following joint p.d.f:

$$f_{XY}(x,y) = \begin{cases} 2(x+y) & 0 < x < y < 1 \\ 0 & otherwise \end{cases}$$

1. $Pr(X < \frac{1}{2})$
2. the marginal p.d.f. of X
3. the conditional p.d.f. of Y given $X = x$.

3. About uniform law

Let U a random variable with uniform law on $[0,1]$.

1. Compute the law of $\log U$.
2. Compute the law of U^2.
3. Compute the law of $\tan(\pi U - \frac{\pi}{2})$

4. About normal law

Let X be a random variable with normal law $N(0,\sigma)$.

1. Compute the law of e^X, and its expectation and variance.
2. Compute the law of X^2, and its expectation and variance.

5. Student law

Let X a random variable with a normal law $N(0,1)$, S be a χ^2 random variable with n freedom degrees. X and S are independent. What is the density of $T = \frac{X}{\sqrt{S/n}}$?

Dependent variables

If the probability density function of a random variable X is given as $f_X(x)$, it is possible to calculate the probability density function of some variable $Y = g(X)$.

This is also called a “**change of variable**” and is in practice used to generate a random variable of arbitrary shape $f_{g(X)}(\cdot) = f_Y(\cdot)$ using a known (for instance uniform) random number generator.

关于系统的基本概念

Definition

如果信号$x_1(t)$和$x_2(t)$经过一个系统$L[\cdot]$的输出分别为$y_1(t)$
$y_2(t)$，且满足

- （**叠加性**）$y_1(t)+y_2(t)=L[x_1(t)+x_2(t)]$
- （**比例性**）$ay_1(t)=L[ax_1(t)]$

则该系统被称为线性系统。

不满足上述两个条件的系统被称为非线性系统。

时变和时不变系统或因果系统与非因果系统

时不变系统：$y(t+t_0)=L[x(t+t_0)]$
因果系统：$y(t)=L[x(t)]$，若$t<0$时$x(t)=0$，则$y(t)=0$。

激励是产生响应的原因，响应是激励的结果（因果性）

分析系统的方法

时域分析

设系统的冲激响应为$h(t)$（稳定性条件：$h(t) \xrightarrow[t\to\infty]{} 0$），则

$$y(t) = x(t) * h(t)$$

频域分析

设系统函数为$H(\omega)$，则

$$Y(\omega) = X(\omega) * H(\omega)$$

稳定性和可实现性

可实现性

因果系统：$y(t)=x(t)*h(t)$，$\tau, h(\tau)=0$。

稳定性

$h(t)\xrightarrow[t\to\infty]{}0$：　任意有界输入的响应有界。

$$|y(t)|=\left|\int_{-\infty}^{\infty}h(\tau)x(t-\tau)\,d\tau\right|\leq\int_{-\infty}^{\infty}|h(\tau)|\cdot|x(t-\tau)|\,d\tau$$

若$|x(t)|\leq a\leq\infty$（有界输入），则$|x(t)|\leq a\int_{-\infty}^{\infty}|h(\tau)|\,d\tau$。
若$\int_{-\infty}^{\infty}|h(\tau)|\,d\tau<\infty$，冲激响应绝对可积，系统稳定。

复频域

$$s = \sigma + j\omega$$
$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}\,dt$$

$H(s)$在复平面的右半平面是解析的（极点在右半平面），系统稳定且可实现。

Probabilistic distribution of the random variable function

已知随机变量X和Y满足$Y = g(X)$，且已知$f_X(x)$，求$f_Y(y)$。

Example

已知：$X \sim N(0,1)$和$Y = X^3$，求$f_Y(y)$。

核心思想：**等概率原则**

$$\begin{aligned} f_Y(y)|dy| &= f_X(x)|dx| \\ f_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| = f_X\{g^{-1}(y)\}\left|\frac{d\{g^{-1}(y)\}}{dy}\right| \end{aligned}$$

非单调情况

Example

已知：$X \sim N(0,1)$和$Y = X^2$，求$f_Y(y)$。

核心思想：**等概率原则 + 单调区间**

$$f_Y(y)|dy| \quad = \quad f_X(x_1)dx_1 + f_X(x_2)dx_2 + \cdots$$

多维随机变量情况

Example

已知：$f_{X_1X_2}(x_1,x_2)$和$\begin{cases} Y_1 = a_{11}X_1 + a_{12}X_2 \\ Y_2 = a_{21}X_1 + a_{22}X_2 \end{cases}$，

求$f_{Y_1Y_2}(y_1,y_2)$。

核心思想：**等概率原则**

$$
\begin{aligned}
f_{Y_1Y_2}(y_1,y_2)|\partial(y_1,y_2)| &= f_{X_1X_2}(x_1,x_2)|\partial(x_1,x_2)| \\
f_{Y_1Y_2}(y_1,y_2) &= f_{X_1X_2}(x_1,x_2)\left|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| \\
&= |\mathbf{J}| \cdot f_{X_1X_2}(x_1,x_2) \\
\mathbf{J} &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}
\end{aligned}
$$

Polar coordinates and isotropy

- The polar coordinates change is defined by

$$\begin{cases} x = \rho\cos\theta \\ y = \rho\sin\theta \end{cases} \Rightarrow \begin{cases} \rho = \sqrt{x^2+y^2} \\ \theta = \arctan(\frac{y}{x}) \end{cases}$$

- The density in polar coordinates is

$$f_{\rho,\theta}(\rho,\theta) = f_{XY}(x,y)\rho = f_{XY}(\rho\cos\theta,\rho\sin\theta)\rho$$

- If $f_{XY}(\rho\cos\theta,\rho\sin\theta)$ depends only on ρ, the density is isotropic, the argument is a random variable with uniform law on $[0,2\pi]$.

Sum of random variables

Proposition

If the probability density function of n real random variables $\mathbf{X} = (X_1, \ldots, X_n)$ is given as $f_{\mathbf{X}}(x_1, \ldots, x_n)$, the density of their sum is

$$f_S(s) = \int \ldots \int f_{\mathbf{X}}(y_1, \ldots, y_{n-1}, s - y_1 - \cdots - y_{n-1}) dy_1 \ldots dy_{n-1}$$

This is an immediate consequence of the diffeomorphism theorem since the jacobian of $\begin{cases} y_1 = x_1 \\ y_2 = x_2 \\ \ldots\ldots\ldots\ldots \\ y_{n-1} = x_{n-1} \\ s = x_1 + \cdots + x_{n-1} + x_n \end{cases}$ is 1.

Sum of independant random variables

- Another expression is

$$
\begin{aligned}
f_S(s) &= \int \ldots \int f_{\mathbf{X}}(y_1, y_2 - y_1, y_3 - y_2, \\
&\quad \ldots, y_{n-1} - y_{n-2}, s - y_{n-1}) dy_1 \ldots dy_{n-1}
\end{aligned}
$$

  which is associated to the linear change

$$
\begin{cases}
y_1 = x_1 \\
y_2 = x_1 + \cdots + x_2 \\
\ldots\ldots\ldots\ldots \\
y_{n-1} = x_1 + \cdots + x_{n-1} \\
s = x_1 + \cdots + x_{n-1} + x_n
\end{cases}
$$

- For two independent variables, of densities f_1 and f_2, the density of the sum amounts to $f_S(s) = \int f_1(x) f_2(s - x) dx$ which is the convolution product $f_1(s) * f_2(s)$.
- For n random independent variables, the density of the sum is the convolution product of the density of the terms.

Example: Bit transmission

We are interested in bit transmission.

- Let X the random binary variable representing the input bit: $X \in \{0, 1\}$,
- let B the random binary variable representing the emission process: $B = 1$ if the transmission is faithful, $B = 0$ if the transmission is wrong.
- Y is the value of the transmitted bit.

	$B = 1$	$B = 0$
$X = 1$	$Y = 1$	$Y = 0$
$X = 0$	$Y = 0$	$Y = 1$

- The state space Ω has four elements that can be labelled by the value of the couple (X, B).
- The law of Y is given by

$$\begin{aligned} P(Y=1) &= P(X=1,B=1)+P(X=0,B=0) \\ P(Y=0) &= P(X=1,B=0)+P(X=0,B=1) \end{aligned}$$

- These two probabilities give the different laws for X and B and same laws for Y:

	$X{=}1,B{=}1$	$X{=}1,B{=}0$	$X{=}0,B{=}1$	$X{=}0,B{=}0$
Case 1	0.5	0	0.5	0
Case 2	0.25	0.25	0.25	0.25

Principal component analysis

The spectral decomposition of covariance matrix allows to find an orthonormal basis of eigenvectors. Let us range it by decreasing positive eigenvalues, it gives $(\vec{u}_1, \ldots, \vec{u}_d)$ with $\lambda_1 \geq \cdots \geq \lambda_d \leq 0$ such that the random variables $(\vec{u}_i|\vec{X})$ are uncorrelated and that $Var(\vec{u}_i|\vec{X}) = \lambda_i$.

Definition

The spectral decomposition of the covariance matrix of a random vector allows to express it in terms of uncorrelated components with decreasing variance. Such a decomposition is called **principal analysis component**.

Principal analysis component is crucial for analysing the causes of a random phenomenon in social sciences but also in engineering. It often allows to reduce drastically the dimensionality of random variation (proper order decomposition (POD))

Definition of linear regression

Suppose we want to predict some data $\vec{Y}$ of a system using some known data X and the knowledge of the joint distribution $P_{(\vec{X},\vec{Y})}$. We limit here to the best affine prediction.

Theorem

The solution of $\min_{A,B} E(||\vec{Y} - A\vec{X} - B||^2)$ is

$$\begin{cases} \hat{A} = Cov(\vec{X},\vec{Y})Cov^{-1}(\vec{X}) \\ \hat{B} = \vec{\mathbb{Y}} - \hat{A}\vec{\mathbb{X}} \end{cases}$$

It is called the **linear regression** of $\vec{Y}$ on $\vec{X}$.
$\vec{Z} = \vec{Y} - \hat{A}\vec{X} - \hat{B}$ is called the **residue** of the regression.