Gauss vectors and statistical tests

EM13-Probability and statistics: Courses 09-10

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–20140930

Hypothesis testing of normal laws: Suppose that nine observations are selected at random from a normal distribution for which both the mean μ and the variance σ^2 are unknown. For these nine observations, the empirical mean and the empirical variance are respectively $\bar{X} = 22$ and $\sum_{i=1}^{n}(x_i - \bar{X})^2 = 72$

1. Carry out a test of the following hypotheses at the level of significance 0.05:

$$H_0: \quad \mu \leqslant 20, \qquad\qquad H_1: \quad \mu > 20$$

2. Carry out a test the following hypotheses at the level of the significance 0.05 by using the two-sided test:

$$H_0: \quad \mu = 20, \qquad\qquad H_1: \quad \mu \neq 20$$

3. From the data, construct a confidence interval for μ with confidence coefficient 0.95.

Definition and properties of Guassian vectors

高斯向量定义及性质

Recall on Gaussian real laws

- The normal law $N(m, \sigma^2)$is a probability law on $\mathbf{R}$ with density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(x-m)^2}{2\sigma^2}]$$

  - Its expectation is m
  - its variance is σ^2
  - its characteristic function is $\phi(t) = \exp[jtm - \frac{\sigma^2 t^2}{2}]$.

- The Gaussian family is stable par affine transformation:
  if X is Gaussian, $aX + b$ is Gaussian

Definition of Gaussian vectors

Definition

A random vector $\mathbf{X}$ in $\mathbf{R}^d$ is said Gaussian
if $\forall \mathbf{u} \in \mathbf{R}^d$,$\mathbf{u}^H\mathbf{X}$is a Gaussian real random variable.

Then, we note its law $\mathbf{X} \sim N(\mathbf{E}(\mathbf{X}), Cov(\mathbf{X}))$.

Proposition

Let $\mathbf{X}$ be a random vector with expectation μ and covariance matrix $\boldsymbol{\Gamma}$. $\mathbf{X}$ is Gaussian if its characteristic function is

$$\phi(\mathbf{u}) = \exp[j(\mathbf{u}^H\mu) - \frac{1}{2}(\mathbf{u}^H\Gamma\mathbf{u})]$$

Proof One has $\mathbf{E}(\mathbf{u}^H\mathbf{X}) = (\mathbf{u}^H\mu)$ and $Var(\mathbf{u}^H\mathbf{X}) = (\mathbf{u}^H\boldsymbol{\Gamma}\mathbf{u})$. The proof follows from the expression of the characteristic function of a Gaussian real random variable.

Main properties of Gaussian vectors

高斯向量主要特性

Linear stability/线性不变性

Theorem

The Gaussian character is stable par translation and linear transform.

Proof Let $\mathbf{X}$ be a Gaussian vector of law $N(\mu, \mathbf{\Gamma})$, its characteristic function is

$$\phi_{\mathbf{X}}(\mathbf{u}) = \exp[j(\mathbf{u}^H\mu) - \frac{1}{2}(\mathbf{u}^H\mathbf{\Gamma}\mathbf{u})]$$

Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}$, then

$$\begin{aligned} \phi_{\mathbf{Y}}(\mathbf{u}) &= \mathbf{E}\{\exp[j\mathbf{u}^H(\mathbf{A}\mathbf{X} + \mathbf{B})]\} = \exp[j(\mathbf{u}^H\mathbf{B})]\phi_{\mathbf{X}}(\mathbf{A}^H\mathbf{u}) \\ \phi_{\mathbf{Y}}(\mathbf{u}) &= \exp[j\mathbf{u}^H(\mathbf{A}\mu + \mathbf{B}) - \frac{1}{2}(\mathbf{u}^H\mathbf{A}\mathbf{\Gamma}\mathbf{A}^H\mathbf{u})] \end{aligned}$$

Canonical form and reduction of Gaussian vector

- The law of a Gaussian vector is completely defined by its expectation and its covariance matrix.
- Let $\mathbf{X} = (X_1, \dots, X_d)^T$ an i.i.d. sample of $N(0,1)$. Then X is a Gauss random vector with law $N(\mathbf{0}, \mathbf{I}_d)$.
- Two uncorrelated components of a Gaussian vector are independent (uncorrelated $\Rightarrow$ independent).
- Let $\mathbf{X} \sim N(\mu, \mathbf{\Gamma})$ be a Gaussian vector and $\mathbf{A}$ and $\mathbf{B}$ two matrixes such that $\mathbf{A\Gamma B}^T = 0$, then $\mathbf{AX}$ and $\mathbf{BX}$ are independent Gaussian vectors.
- Let $\mathbf{X} \sim N(\mu, \mathbf{\Gamma})$ be a Gaussian vector and U a matrix such that $\mathbf{U\Gamma U}^T$ is diagonal, then $\mathbf{UX}$ is a Gaussian vector with independent components.

Gaussian density

Theorem

If $\mathbf{\Gamma}$ is invertible, the normal law $N(\mu, \mathbf{\Gamma})$ is a continuous law with density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{\Gamma}|}} \exp\{-\frac{1}{2}[(\mathbf{x}-\mu)^H \mathbf{\Gamma}^{-1}(\mathbf{x}-\mu)]\} \qquad \mathbf{X} \in \mathbf{R}^d$$

Proof Let $\mathbf{A} = \sqrt{\mathbf{\Gamma}}$, let $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_d)$,the density of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^d}} \exp\{-\frac{(\mathbf{z}^H \mathbf{z})}{2}\}$$

Let $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu$ then $\mathbf{X} \sim N(\mu, \mathbf{\Gamma})$ and its density is

$$f_{\mathbf{X}}(\mathbf{x}) = |\mathbf{A}|^{-1} f_{\mathbf{Z}}\{\mathbf{A}^{-1}(\mathbf{x}-\mu)\}$$

Linear regression and independence

Theorem

- Let $\mathbf{X} \sim N(\mu_{\mathbf{X}}, \mathbf{\Gamma_X})$ and $\mathbf{Y} \sim N(\mu_{\mathbf{Y}}, \mathbf{\Gamma_Y})$ be two components or two affine transforms of a Gaussian vector
- and let $Cov(\mathbf{X}, \mathbf{Y})$ be their mutual covariance.

We recall that the linear regression of $\mathbf{Y}$ on $\mathbf{X}$ is

$$\hat{\mathbf{Y}} = \mathbf{\Lambda X} + \mathbf{\Theta}$$

where $\mathbf{\Lambda} = Cov(\mathbf{X}, \mathbf{Y})\mathbf{\Gamma_X}^{-1}$ and $\mathbf{\Theta} = \mu_{\mathbf{Y}} - \mathbf{\Lambda}\mu_{\mathbf{X}}$.

Then the residue $\mathbf{Z} = \mathbf{Y} - \mathbf{\Lambda X} - \mathbf{\Theta}$ is a gaussian law which is centered and independent of $\mathbf{X}$.

Proof Since that the residue $\mathbf{Z} = \mathbf{Y} - \mathbf{\Lambda X} - \mathbf{\Theta}$ is not correlated with $\mathbf{X}$ and the random vector $(\mathbf{X}, \mathbf{Z})$ is Gaussian, the result comes from the basic property of Gaussian law to be determined by second order moments.

Statistics is the study of the collection, analysis, interpretation, presentation and organization of data.
When analyzing data, it is possible to use one of two statistics methodologies:

- descriptive statistics(描述统计学)：是否可以摘要的说明数据的情形，不论是以数学或是图片表现，以用来代表母群体的性质？基础的数学描述包括了平均数和标准差等。图像的摘要则包含了许多种的表和图。主要是就说明数据的集中和离散情形。
- inferential statistics(推论统计学)：用来将数据中的数据模型化，计算它的概率并且做出对于母群体的推论。这个推论可能以对/错问题的答案所呈现（假设检定），对于数字特征量的估计（估计），对于未来观察的预测，关系性的预测（相关性），或是将关系模型化（回归）。其他的模型化技术包括变异数分析（ANOVA），时间串行（time series analysis），以及数据挖掘（data mining）。

A **statistical hypothesis test** is a method of statistical inference using data from a scientific study.

Statistical hypothesis testing is a key technique of both Frequentist inference and Bayesian inference although they have notable differences.

Several concepts — null and alternative hypotheses

Null hypothesis (H_0) / 零假设

A simple hypothesis associated with a contradiction to a theory one would like to prove.

在零假设中，所有因素对变量都不起任何作用。

Alternative hypothesis (H_1) / 备择假设、对立假设

A hypothesis (often composite) associated with a theory one would like to prove.

Example

法庭窘境：

- 零假设H_0认为被告是清白的，备择假设H_1认为被告有罪。
- 起诉是因为怀疑被告有罪。
- H_0（现状）与H_1对立并且被认可，除非能够证明其不成立。
- “无法排除H_0”并不能代表被告清白，只是说证据无法将其定罪。
- 陪审团没有必要在H_0“无法推翻”的情况下将其“接受”。
- 当零假设无法被“证明”时，可以通过强度检测判断假设是否近似成立，即进行第二型错误检测。

The hypothesis of innocence is only rejected when an error is very unlikely, because one doesn't want to convict an innocent defendant. Such an error is called error of the first kind (i.e., the conviction of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, the error of the second kind (acquitting a person who committed the crime), is often rather large.

	H_0 is true (Truly not guilty)	H_1 is true (Truly guilty)
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

A criminal trial can be regarded as either or both of two decision processes: guilty vs not guilty or evidence vs a threshold ("beyond a reasonable doubt"). In one view, the defendant is judged; in the other view the performance of the prosecution (which bears the burden of proof) is judged. A hypothesis test can be regarded as either a judgment of a hypothesis or as a judgment of evidence.

Several concepts — confidence level

Confidence level

H_0 is rejected by the test $\overline{X} \geq \xi$ at the confidence level $\alpha > 0$ iff

$$Pr_\theta(\overline{X} \geq \xi) \leq \alpha \qquad \forall\theta \in H_0$$

The rejection region of H_0 is called the critical region.

A researcher will often "reject the null hypothesis" when the α-value (p-value) turns out to be less than a certain significance level.
An informal interpretation of a α-value (p-value), based on a significance level of about 10%, might be:

- $\alpha \leqslant 0.01$: very strong presumption against null hypothesis
- $0.01 < \alpha \leqslant 0.05$: strong presumption against null hypothesis
- $0.05 < \alpha \leqslant 0.1$: low presumption against null hypothesis
- $\alpha > 0.1$: no presumption against the null hypothesis

Such a result indicates that the observed result would be highly unlikely under the null hypothesis (that is, the observation is highly unlikely to be the result of random chance alone).

Several concepts — regions of acceptance and rejection

Region of acceptance / 接受区间

The set of values of the test statistic for which we fail to reject the null hypothesis.

Region of rejection / Critical region / 拒绝区间

The set of values of the test statistic for which the null hypothesis is rejected.

Computation of the critical region for a confidence level

Let $F_X(x)$ be the cumulative function of $N(0,1)$:

$$Pr_\theta(\overline{X} \geq \xi) = 1 - Pr_\theta(\overline{X} < \xi) = 1 - F(\frac{\sqrt{n}}{\sigma}\xi + \theta) \leq \alpha$$

So

$$\xi \geq sup_{\theta \in H_0} \frac{\sigma}{\sqrt{n}}[F^{-1}(1-\alpha) - \theta] = \frac{\sigma}{\sqrt{n}} F^{-1}(1-\alpha)$$

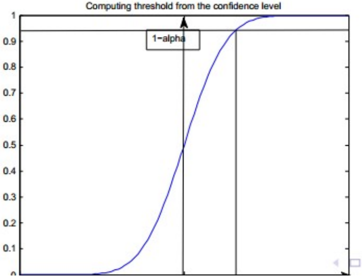

Figure: Computing threshold from the confidence level

Several concepts — Likelihood ratio

Likelihood function

Let X be a random variable with a probability distribution with density function f depending on a parameter θ. Then the function

$$\mathcal{L}(\theta|x) = f(x|\theta)$$

considered as a function of θ, is called the **likelihood function**.

Likelihood ratio

We want to test

- the null hypothesis $\theta \in H_0$
- against the alternative hypothesis $\theta \in H_1$.
- the test is based on the sample $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}$ with likelihood $f_{\mathbf{X}}(\mathbf{x}|\theta)$.

The likelihood ratio is defined by on $\mathcal{X} \times H_0 \times H_1$ by

$$R(\mathbf{x}, \theta_1, \theta_0) = \frac{f_{\mathbf{X}}(\mathbf{x}|\theta_1)}{f_{\mathbf{X}}(\mathbf{x}|\theta_0)}$$

Example

Mean of normal law $N(\theta, \sigma)$

$$
\begin{aligned}
R(x, \theta_1, \theta_0) &= \exp\left\{\frac{1}{2\sigma^2}\left[-\sum_{i=1}^{n}(x_i-\theta_1)^2+\sum_{i=1}^{n}(x_i-\theta_0)^2\right]\right\} \\
R(x, \theta_1, \theta_0) &= \exp\left\{\frac{1}{\sigma^2}\left[(\theta_1-\theta_0)(x_1+\cdots+x_n)-\frac{\theta_1^2-\theta_0^2}{2/n}\right]\right\}
\end{aligned}
$$

The testing process

The usual line of reasoning is as follows:

1. There is an initial research hypothesis of which the truth is unknown.
2. The first step is to state the relevant **null and alternative hypotheses**. This is important as mis-stating the hypotheses will muddy the rest of the process.
3. The second step is to consider the statistical assumptions being made about the sample in doing the test; for example, assumptions about the statistical independence or about the form of the distributions of the observations. This is equally important as invalid assumptions will mean that the results of the test are invalid.
4. Decide which test is appropriate, and state the relevant **test statistic** T.
5. Derive the distribution of the test statistic under the null hypothesis from the assumptions. In standard cases this will be a well-known result. For example the test statistic might follow a Student's t distribution or a normal distribution.

6. Select a significance level (α), a probability threshold below which the null hypothesis will be rejected. Common values are 5% and 1%.
7. The distribution of the test statistic under the null hypothesis partitions the possible values of T into those for which the null hypothesis is rejected, the so-called critical region, and those for which it is not. The probability of the critical region is α.
8. Compute from the observations the observed value tobs of the test statistic T.
9. Decide to either reject the null hypothesis in favor of the alternative or not reject it. The decision rule is to reject the null hypothesis H_0 if the observed value tobs is in the critical region, and to accept or "fail to reject" the hypothesis otherwise.

An alternative testing process

1. Compute from the observations the observed value t_{obs} of the test statistic T.
2. Calculate the α-value. This is the probability, under the null hypothesis, of sampling a test statistic at least as extreme as that which was observed.
3. Reject the null hypothesis, in favor of the alternative hypothesis, if and only if the α-value is less than the significance level (the selected probability) threshold.