応用統計学 2017　第9回　仮説検定

2017年12月6日 (水)
清 智也　sei@mist.i.u-tokyo.ac.jp
http://ur0.pw/yTzt

「検定の論理は，確率的な背理法である。」（宮川 雅巳「統計技法」p.72）

- 仮説検定，帰無仮説，対立仮説，有意水準，棄却域，検出力，検定統計量，p-値，尤度比検定[1]。
- 正規線形モデルの仮説検定（t 検定，F 検定）は次回扱う。

演習問題

問題 9-1 (仮説検定の基礎)**.** $X_1,\ldots,X_n$ は独立で，平均 μ，分散 1 の正規分布に従うものとする。帰無仮説 H_0: $\mu=0$ のもとで，棄却域を $R=\{\boldsymbol{x}\in\mathbb{R}^n \mid |\bar{x}|\geq c\}$ とする検定方式を考える。有意水準が 0.05 となるような c を求めよ。棄却域を $R=\{\boldsymbol{x}\in\mathbb{R}^n \mid \bar{x}\geq c\}$ とした場合はどうか。なお，標準正規分布の上側 2.5%点は約 1.96，上側 5%点は約 1.64 である。

問題 9-2. 問題 9-1 において，対数尤度比検定統計量を求めよ。

問題 9-3. $X_1,\ldots,X_n$ は独立で，密度関数（または確率関数）$f(x;\theta)$ の分布に従うとする。以下の各場合について，帰無仮説を $\theta=\theta_0$，対立仮説を $\theta\in\Theta\setminus\{\theta_0\}$ として，対数尤度比検定統計量 (LLR) を $n,\theta_0,\hat{\theta}$ の式で表せ。ここで $\hat{\theta}\in\Theta$ は最尤推定量を表す。

(i) ベルヌーイ分布 $f(x;\theta)=\theta^x(1-\theta)^{1-x}$，$x\in\{0,1\}$，$\Theta=(0,1)$。

(ii) ポアソン分布 $f(x;\theta)=(\theta^x/x!)e^{-\theta}$，$x\in\{0,1,\cdots\}$，$\theta=(0,\infty)$。

(iii) 指数分布 $f(x;\theta)=\theta e^{-\theta x}$，$x\geq 0$，$\Theta=(0,\infty)$。

(iv) 分散未知の正規分布 $f(x;\mu,\sigma^2)=(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$，$x\in\mathbb{R}$，$\Theta=\mathbb{R}\times(0,\infty)$。

注意：指数型分布族以外の場合，LLR が $n,\theta_0,\hat{\theta}$ だけの関数になるとは限らない。

[1] hypothesis testing, null hypothesis, alternative hypothesis, significance level, rejection region, power, test statistic, p-value, likelihood ratio test.

問題 9-4 (ROC 曲線). $X \sim \mathrm{N}(\mu, \sigma^2)$ とする。帰無仮説を $\mu = 0$，対立仮説を $\mu = \delta > 0$ とし，棄却域を $R = \{x \in \mathbb{R} \mid |x| \geq c\}$ とする検定方式を考える。$c \geq 0$ を媒介変数とし，有意水準と検出力の関係を表すグラフを描け。棄却域を $R = \{x \in \mathbb{R} \mid x \geq c\}$ とした場合はどうなるか。

問題 9-5 (ネイマン・ピアソンの基本定理). 統計モデル $f_n(\boldsymbol{x};\theta)$ を考え，帰無仮説と対立仮説は単純仮説（つまり 1 点集合）で，それぞれ $\theta = \theta_0$，$\theta = \theta_1$ とする。また，次の条件を仮定する：

$$\mathrm{P}_{\theta_0}(f_n(\boldsymbol{X};\theta_1)/f_n(\boldsymbol{X};\theta_0) = c) = 0, \quad \forall c \geq 0. \tag{$*$}$$

このとき，検定方式

$$R = \{\boldsymbol{x} \mid f_n(\boldsymbol{x};\theta_1)/f_n(\boldsymbol{x};\theta_0) \geq c\}, \quad \mathrm{P}_{\theta_0}(\boldsymbol{X} \in R) = \alpha,$$

は有意水準 α の検定方式の中で最も検出力が大きいことを示せ。

補足：条件 $(*)$ は離散分布の場合は満たされない。離散分布の場合，検出力が最大となる検定方式を求める問題はナップサック問題[2]と等価であり，一般に解くのは難しい。

問題 9-6 (難しい). 指数型分布族の場合について，対数尤度比検定統計量が χ^2 分布に分布収束することを示せ[3]。χ^2 分布の定義については第 2 回レポート課題を参照せよ。

宿題 9

問題 9-7. (i) カテゴリー数 3，サンプルサイズ n の多項分布モデル

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^{3} \theta_i^{x_i}, \quad x_1 + x_2 + x_3 = n, \quad \theta_1 + \theta_2 + \theta_3 = 1$$

を考える。帰無仮説を $\theta_1 = \theta_3$ とするとき，対数尤度比検定統計量を求めよ。

(ii) 新しく開発したビールが従来のビールに比べ美味しいか否かを調べるため，被験者 40 人に対し，2 つのビールを（被験者には銘柄を隠し，かつランダムな順番で）飲んでもらい，どちらが美味しいかを答えてもらったところ，以下の表のようであったとする。新しいビールの方が美味しいと言えるかどうか，有意水準 $\alpha = 0.05$ で検定せよ。

	新しいビール	変わらない	従来のビール
人数	17	10	13

[2] 例えば「数理工学事典」朝倉書店を参照。
[3] 一般の場合の証明については，例えば竹村彰通「現代数理統計学」を参照せよ。