Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

4 Chap.4. Some Elementary Statistical Inferences

4.5 Introduction to Hypothesis Testing

点推定と信頼区間は便利な統計的推測手法である．別の形式の推測手法として頻繁に利用されるものとして，仮説検定があり，本節ではこれを紹介する．

先行研究や，事前の実験から，密度関数 $f(x;\theta)$, $\theta \in \Omega$ のパラメータ θ が，互いに排反な集合 ω_0, ω_1 のいずれかに属するとする．ここで $\omega_0 \cup \omega_1 = \Omega$ である．この時，それぞれの集合に属するとする仮説を，以下のように表す．

$$H_0:\ \theta \in \omega_0\ v.s.\ H_1:\ \theta \in \omega_1 \tag{4.5.1}$$

このように表した時，H_0 は帰無仮説 (null hypothesis)，H_1 は対立仮説 (alternative hypothesis) と呼ばれる．一般的に，帰無仮説は，「変化や差がない」ということを表し，対立仮説は，検定を行う人の仮説を表す．帰無仮説と対立仮説のいずれを採択するかは，X の分布からの標本 $X_1, X_2, \ldots, X_n$ に基づくため，決定を間違えることがある．実際は，帰無仮説が正しい時に，対立仮説を採択する間違いを**第 *1* 種の過誤**，逆に，実際は対立仮説が正しい時に，帰無仮説を採択する間違いを**第二種の過誤**と呼ぶ．第 1 種の過誤の例としては，第二種の過誤の例としては，

現実の状況

決定	H_0 が正しい	H_1 が正しい
H_0 を棄却	**第 *1* 種の誤り**	正しい決定
H_0 を採択	正しい決定	**第 *2* 種の誤り**

次に，どちらの仮説を採択するかのルールについて考える．標本の空間を $\mathcal{D} = \{(X_1, \ldots, X_n)\}$ で表す．H_1 に対する H_0 の検定は，$\mathcal{D}$ の部分集合 C に基づく．この集合 C は**棄却域** *(critical* *region)* と呼ばれ，いずれの仮説を採択するかのルールは，

$$\begin{aligned} &Reject\ H_0(Accept\ H_1)\ if\ (X_1,\ldots,X_n)\in C \\ &Retain\ H_0(Reject\ H_1)\ if\ (X_1,\ldots,X_n)\in C^c \end{aligned} \tag{4.5.3}$$

に対応する．棄却域は，第 1 種の過誤と第 2 種の過誤の起こる確率をできるだけ小さくするよう選ぶことができれば理想的である．しかし，これら二つの確率は，互いに影響しあっており，いずれかを小さくすると，他方が大きくなるという関係を持っている．しばしば，第 1 種の過誤の方が，第 2 種の過誤よりも良くないこととされているため，第 1 種の過誤が起こる確率を制限した棄却域の中から，第 2 種の過誤が起こる確率を最小にするような棄却域を選択する．

定義 4.5.1(大きさ α の棄却域)

棄却域 C の大きさが α であるとは，

$$\alpha = \max_{\theta\in\omega_0} P_\theta[(X_1,\ldots,X_n)\in C] \tag{4.5.4}$$

であることをいう．

全ての大きさ α の棄却域について，私たちは，第 2 種の過誤が起こりうる確率が最小となるような棄却域を考えたい．第二種の過誤の補集合 (ちょっと使い方おかしいです) は，対立仮説 H_1 が正しい時に，H_0 を棄却するという正しい判断である．第 2 種の過誤が起こりうる確率を最小化することは，この正しい判断が起こる確率を最大にすることと等価である．つまり，$\theta\in\omega_1$ について，

$$1 - P_\theta[\text{第 2 種の過誤}] = P_\theta[(X_1,\ldots,X_n)\in C]$$

を最大化すれば良い．この右辺の確率を θ における検定の**検出力** *(power)* と呼ぶ．この式から明らかなように，第 2 種の過誤の確率を最小化することは，検出力を最大化することと等しい．
次に，棄却域の**検出力関数** *(power function)* を

$$\gamma_C(\theta) = P_\theta[(X_1,\ldots,X_n)\in C];\theta\in\omega_1 \tag{4.5.5}$$

と定義する．よって，もし，同じ大きさ α の二つの棄却域 C_1, C_2 が存在する時，任意の $\theta\in\omega_1$ について，$\gamma_{C_1}(\theta)\geq\gamma_{C_2}(\theta)$ なら，C_1 の方が優れていると判断できる．
¡note¿

例 4.1. 中心極限定理を用いた z 検定 $X \sim N(\mu, \sigma^2)$ とする．$X_1, X_2, \ldots, X_n$ をこの分布からの標本とする．以下の検定を考える．

$$H_0: \ \mu = \mu_0 \ v.s. \ H_1: \ \mu > \mu_0$$

ここで，μ_0 はある特定の値 *(2* とか *0.8* とか*)* である．$\mu = E(\bar{X})$ から，この検定に対する *decision rule* は

$$Reject \ H_0 \ in \ favor \ of \ H_1 \ if \ \bar{X} \ is \ much \ larger \ than \ \mu_0 \tag{4.5.10}$$

である．一般的に標本平均の分布は，閉じた形で得られない，X の分布が正規分布であるという強い過程のもとならば，検定を行える．今，中心極限定理から，標本数が多くなると，

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$

であり，これを用いると，*decision rule* は，

$$Reject \ H_0 \ in \ favor \ of \ H_1 \ if \ Z \geq z_\alpha$$

となる．帰無仮説を棄却するために，$\bar{X} \geq \mu_0 + z_\alpha \sigma/\sqrt{n}$ であるべきである．このことから，検出力関数は，

$$\begin{aligned} \gamma(\mu) &= P_\mu(\bar{X} \geq \mu_0 + z_\alpha \sigma/\sqrt{n}) \\ &= P_\mu\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{\mu_0 - \mu}{S/\sqrt{n}} + z_\alpha\right) \\ &\approx 1 - \Phi\left(z_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \\ &= \Phi\left(-z_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \end{aligned} \tag{4.5.12}$$

となる．もし，σ の値がわかっている場合は，この検出力関数を計算できる．この近似的な検出力関数は，μ に関して，狭義単調増加関数であるので，帰無仮説を，

$$H_0: \ \mu \leq \mu_0 \ v.s. \ H_1: \ \mu > \mu_0$$

のように変更できる．

この検定は，例えば，パソコンを用いた学習を行った生徒のテストの平均点と，そうでない生徒のテストの平均点を比べて，パソコンを用いた学習に効果があるのかを検証するなどの際に利用される．

例 **4.2.** t 検定

4.6 Additional Comments About Statistical Tests

仮説検定には，**片側検定** *(one-sided test)* と**両側検定** *(two-sided test)* がある．片側検定は，4.5 章で紹介した仮説検定問題のように，対立仮説が，$H_1:\ \mu > 3000$ となるようなものである．両側検定とは，対立仮説が，$H_1:\ \mu \neq 3000$ となるようなものである．両側検定では，有意水準が

$$\alpha = P_{H_0}[\bar{X} \leq h\ or\ \bar{X} \geq k] = P_{H_0}[\bar{X} \leq h] + P_{H_0}[\bar{X} \geq h]$$

となり，

$$P_{H_0}[\bar{X} \leq h] = P_{H_0}[\bar{X} \geq h] = \alpha/2 \tag{4.6.3}$$

と計算されることや，decision rule が

$$Reject\ H_0\ in\ favor\ of\ H_1\ if\ \left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| \geq z_{\alpha/2} \tag{4.6.4}$$

となることに注意．

最後に p 値について紹介する．p 値とは，帰無仮説のもとでの，検定統計量 T の分布の検定統計量の実現値 t より外側の確率，

$$p.value = P_{H_0}[T \geq t](onesided\ test)$$

のことである．

¡note¿

4.7 Chi-Square Tests

$X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ とし，X_i は互いに独立とする．ゆえに，これらの確率変数の同時密度は，

$$\frac{1}{\sigma_1 \sigma_2 \cdots \sigma_n} \exp\left[-\frac{1}{2}\sum_{i=1}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right],\ -\infty < x_i < \infty$$

となる．指数部の $\sum[(X_i - \mu_i)/\sigma_i]^2$ は，$\chi^2(n)$ に従う確率変数である．9.8 節では，同時密度の特定の指数部は $\chi^2(n)$ に従う確率変数となることを示すが，今回の節では，これを所与のものとする．このことは，カイ二乗検定の数理的な基盤となっている．

$X_1 \sim Bin(n, p_1)$ とした時，確率変数

$$Y = \frac{X_1 - np_1}{\sqrt{np_1(1 - p_1)}}$$

は，$n \to \infty$ で漸近的に $N(0, 1)$ に従う (定理 4.2.1)．さらに，Y^2 は漸近的に $\chi^2(1)$ に従う (Example5.3.6 を参照)．$X_2 = n - X_1, p_2 = 1 - p_1$ とし，$Q_1 = Y^2$ とする．この時，Q_1 は

$$\begin{aligned}
Q_1 &= \frac{(X_1 - np_1)^2}{np_1(1 - p_1)} \\
&= \frac{(1 - p_1 + p_1)(X_1 - np_1)^2}{np_1(1 - p_1)} \\
&= \frac{(X_1 - np_1)^2}{n(1 - p_1)} + \frac{(X_1 - np_1)^2}{np_1} \\
&= \frac{(n - X_2 - np_1)^2}{n(1 - p_1)} + \frac{(X_1 - np_1)^2}{np_1} \\
&= \frac{(-X_2 + n(1 - p_1))^2}{n(1 - p_1)} + \frac{(X_1 - np_1)^2}{np_1} \\
&= \frac{(X_2 - np_2)^2}{np_2} + \frac{(X_1 - np_1)^2}{np_1} \qquad (4.7.1)
\end{aligned}$$

となる．この結果は以下のように一般化される．

$X_1, X_2, \ldots, X_{k-1}$ が多項分布 $M(n, p_1, p_2, \ldots, p_{k-1})$ に従うとし，$X_k = n - \sum_{i=1}^{k-1} X_i$，$p_k = 1 - \sum_{i=1}^{k-1} p_i$ とする．Q_{k-1} を

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

と定義する．これは，$n \to \infty$ の時，漸近的に $\chi^2(k-1)$ に従う．証明は，稲垣 (1990) に載っていたので，興味ある方は参照してください．n が十分に大きいことの判断要素としては，それぞれの np_i が 5 より大きくなることを目安とする．以下のような仮説を考える．

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, \ldots, p_{k-1} = p_{k-1,0}, (p_k = 1 - \sum_{i=1}^{k-1} p_i 0)$$

ここで，$p_{10}, p_{20}, \ldots, p_{k-1,0}$ は事前に特定した数字を表す．もし，帰無仮説 H_0 が正しい場合，漸近的に，

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_{i0})^2}{np_{i0}} \sim \chi^2(k-1)$$

である．帰無仮説が正しい時，np_{i0} は X_i の期待値であるため，直感的に，Q_{k-1} は大きい値を取らないということがわかる．有意水準 α のもとで，$P(Q_{k-1} \geq c) = \alpha$ となるような c をカイ二乗分布表などから求め，データから求めた統計量 Q_{k-1} が c を上回っている場合は，帰無仮説を棄却する．この検定様式は，**適合度の検定** *(goodness-of-fit test)* と呼ばれている．

例 4.3. サイコロが歪んでいるかどうかの適合度検定サイコロを *60* 回振った時，以下のようなデータが得られたとする．このサイコロがイカサマサイコロかどうかを，有意水準 5% で検定せ

	1	2	3	4	5	6	計
出た回数	13	19	11	8	5	4	60

よ．p 値も求めよ．

$Y \sim N(\mu, \sigma^2)$ とし，μ, σ^2 は未知パラメータとする．この時，互いに排反な事象 $A_1, \ldots, A_k$ の

それぞれの生起確率の式

$$p_i = \int_{A_i} \frac{1}{\sqrt{2\pi}\sigma} \exp[-(y-\mu)^2/2\sigma^2]dy,\ i = 1, \ldots, k$$

は μ, σ^2 の関数となる．この分布から，サイズ n の無作為標本 $Y_1, \ldots, Y_n$ を抽出するとし，各事象の頻度を $X_i, \sum_{i=1}^{k} X_i = n$ とする．確率変数

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

は，p_i, Q_{k-1} が未知パラメータ μ, σ^2 の関数であることから．$X_1, \ldots, X_k$ が一度観測されただけでは計算できない．そのため，Q_{k-1} を最小化するような μ, σ^2 を選択しなければならない．(なぜこの確率変数の値を最小化するような値を選ぶのかはわかってないので，わかる人教えてください．) これらの値は，$X_1 = x_1, \ldots, X_k = x_k$ の値に依存し，**最小カイ二乗推定量** *(minimum chi-square estimator)* と呼ばれる．これらの点推定値を用いて，数値的に p_i の推定量を計算することができる．

次に，**分割表** *(contingency table)* を取り扱う．二つの要因 A, B についてそれぞれ，水準が r, c 個ずつ存在するような観測度数の分割表を考える．

	A_1	A_2	$\cdots$	A_c	計
B_1	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1.}$
B_2	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2.}$
$\vdots$					$\vdots$
B_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r.}$
計	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.c}$	$n_{..}$

これに対応する．確率分布表は

	A_1	A_2	$\cdots$	A_c	計
B_1	p_{11}	p_{12}	$\cdots$	p_{1c}	$p_{1.}$
B_2	p_{21}	p_{22}	$\cdots$	p_{2c}	$p_{2.}$
$\vdots$					$\vdots$
B_r	p_{r1}	p_{r2}	$\cdots$	p_{rc}	$p_{r.}$
計	$p_{.1}$	$p_{.2}$	$\cdots$	$p_{.c}$	$p_{..}$

要因 A, B が独立であると仮定すると，

$$p_{ij} = P(A_i \cap B_j) = P(A_i)P(B_j) = p_{i.}p_{.j}$$

となる．独立であるとの仮定のもとで，各セルの生起確率の推定量は $\hat{p}_{i.} = n_{i.}/n_{..}$ のように計算できるので，期待度数は，

$$n_{..}\hat{p}_{ij} = n_{..}\hat{p}_{i.}\hat{p}_{.j} = \frac{n_{i.}n_{.j}}{n_{..}}$$

と計算できる．独立性の検定は，検定統計量

$$Q_{(r-1)(c-1)} := \sum_i \sum_j \frac{(n_{ij} - n_{..}\hat{p}_{ij})^2}{n_{..}\hat{p}_{ij}} \sim \chi^2((r-1)(c-1))$$

を利用して検定を行う．

¡note¿