応用統計学 2017 第 5 回 確率論の復習，推測統計の基礎

2017 年 10 月 25 日 (水)
清 智也 sei@mist.i.u-tokyo.ac.jp
講義用ページ http://ur0.pw/yTzt

データを確率変数の実現値と見なし，背後にある確率構造に関する知見を得ることを，一般に推測統計 (inferential statistics) という。

- そのようなことを考えたくなる動機 → 回帰モデルで説明。
- 用語：標本，母集団，データ生成過程，統計モデル，パラメータ，統計量[1]。
- 2 つの考え方：頻度論的 or ベイズ的[2]。
  - 統計モデルを用いるという点は共通。
  - "Parameters are fixed" or "Data are fixed"
    `http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf`
- 確率論の復習については別紙参照。

演習問題

問題 5-1. $X_1, \ldots, X_n$ は平均 μ，分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ に独立に従う確率変数とする。ただし μ, σ^2 はパラメータとする。このとき，標本平均

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

の分布を求めよ。

問題 5-2 (単回帰モデル)**.** $x_1, \ldots, x_n$ は実数とし，$Y_1, \ldots, Y_n$ は

$$Y_t = a + bx_t + \varepsilon_t, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2),$$

により定まる確率変数とする。ただし a, b, σ^2 はパラメータとする。Y_t の確率分布を求めよ。また，$\bar{x} = n^{-1}\sum_{t=1}^{n} x_t = 0$ のとき，最小二乗法により求められる回帰係数 $(\hat{a}, \hat{b})'$ の確率分布を求めよ。

---

[1] sample, population, data generating process, statistical model, parameter, statistic.
[2] frequentist approach, Bayesian approach.

問題 5-3 (判別モデル). 確率変数ベクトル $(\boldsymbol{X}, Y)$ の分布を以下で定める：

$$\mathrm{P}(Y=1)=\mathrm{P}(Y=-1)=\frac{1}{2}, \quad \boldsymbol{X}|\{Y=1\} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \boldsymbol{X}|\{Y=-1\} \sim \mathrm{N}(\boldsymbol{\nu}, \boldsymbol{\Sigma}).$$

ただし $\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\Sigma}$ はパラメータとする。このとき，与えられた実数ベクトル $\boldsymbol{x}$ に対して条件付き確率 $P(Y|\boldsymbol{X}=\boldsymbol{x})$ を求めよ。

問題 5-4 (頻度論的推測とベイズ的推測). ある野球選手が，n 回の打席でヒットを打つ回数 X は二項分布 $\mathrm{Bin}(n, p)$ に従うという。

(i) 統計量 $\hat{p}=X/n$ の分布を求めよ。また $n=10$ とし，$p=0.3$ および $p=0$ のときの $\hat{p}$ の分布を図示せよ。

(ii) p が $[0,1]$ 上の一様分布に従うと仮定する。$X=x$ を与えたもとでの p の条件付き分布（事後分布）を求めよ。また $n=10$ とし，$x=3$ および $x=0$ のときの事後分布を図示せよ。

宿題 5

問題 5-5. $X_1, \ldots, X_n$ は独立に平均 μ の指数分布に従うと仮定する。このとき標本平均 $\bar{X}$ の従う分布を求めよ。