(水曜 1 限, 尾畑)

# 数理統計学・期末試験問題 (2015.07.22)

- 電卓などの計算機の使用禁止. (Use of Calculators is prohibited.)
- 判読不能な文字 (薄い, 小さい, 汚いなど) や論理不明瞭な文章は読みません. (Illegible answers are excluded from marking.)
- 試験終了後, 問題の解説を担当者のホームページに掲載する. (Summary of this examination will be posted in Obata's website.)

**[1]** 辺の長さが $L$ の正方形の内部から, どの点も同等の確率で選ばれるようにランダムに 1 点を選び, $A$ とする. $A$ から 4 辺に下ろした垂線のうち最短なものの長さを $X$ とする. 一言で言えば, $X$ は 点 $A$ から正方形の周までの距離である. (5 点 $\times 4 = 20$ 点) (Let $A$ be a point randomly chosen from the interir of a square with side length $L$. Let $X$ be the shortest distance from $A$ to the edges.)

(1) $X$ の分布関数 $F(x) = P(X \le x)$ を求め, そのグラフを図示せよ. ただし, $x \in \mathbf{R}$ とする. (Find the distribution function $F(x) = P(X \le x)$ and draw its graph, where $x \in \mathbf{R}$.)

(2) $X$ の確率密度関数 $f(x)$ を求めよ. (Find the probability density function $f(x)$ of $X$.)

(3) $X$ の平均値を求めよ. (Calculate the mean value of $X$.)

(4) $X$ の分散を求めよ. (Calculate the variance of $X$.)

**[2]** 確率変数 $T$ がパラメータ $\lambda > 0$ の指数分布に従うとは,

$$P(T \le x) = \int_0^x \lambda e^{-\lambda t} dt, \quad x > 0 \text{ のとき}; \qquad P(T \le x) = 0, \quad x \le 0 \text{ のとき},$$

が成り立つときにいう. (5 点 $\times$ 2 $=$ 10 点) (Let $T$ be a random variable obeying the exponential distribution with parameter $\lambda > 0$, i.e., the above relations hold.)

(1) $a \ge 0$ とするとき, $P(T \ge a)$ を求めよ. (For $a \ge 0$ find $P(T \ge a)$.)

(2) $a \ge 0$, $b \ge 0$ とするとき, $P(T \ge a + b | T \ge a) = P(T \ge b)$ が成り立つことを示せ. ただし, $P(A|B)$ は条件 $B$ の下での $A$ の条件付確率を表す. (For $a \ge 0$ and $b \ge 0$, prove the relation $P(T \ge a+b | T \ge a) = P(T \ge b)$. Here $P(A|B)$ stands for the conditional probability of $A$ under the condition $B$.)

**[3]** ある国では, 病気 A の感染者は $100q$ % であるという ($0 < q < 1$). 検査 B は, 感染者の 90%に陽性反応を示すが, 非感染者の 5% にも陽性反応が出てしまう. (5 点 $\times 2 = 10$ 点) (There is a certain disease A randomly found in $100q$ % of the total population in a country ($0 < q < 1$). A certain clinical test B is 90% effective in detecting the presence of this disease, but it also yields false-positive results in 5% of the cases where the disease is not present.)

(1) $q = 0.04$ とする. この検査を受けて陽性反応が出た人が感染者である確率を求めよ. (Suppose $q = 0.04$. Calculate the probability that a positive test result will be a true positive.)

(2) この検査を受けて陽性反応が出た人が感染者である確率 $P$ は, $0 < q < 1$ とともにどのように変化するか? グラフを示して, その変化の特徴を述べよ. (Let $P$ be the probability that a positive test result will be a true positive. How does $P$ behave as $q$ varies over $0 < q < 1$. Draw the graph and state its characteristics.)

**[4]** 100 万世帯に対して番組 A の視聴率を調べるために, 600 世帯を無作為抽出して視聴率 22.1% を得た. これをもと母比率の区間推定をして, 信頼係数 95% の信頼区間 $22.1 \pm 3.3$% が導かれた. このことについて次の問いに答えよ. (10 点 $\times 2 = 20$ 点) (The ratings of the TV program A was 22.1%, obtained by 600 random samples among 1,000,000 population. Applying the interval estimation, one obtains the confidence interval $22.1 \pm 3.3$% with confidence coefficient 95%. Answer the following questions.)

(1) 信頼区間の導出の仕方について, その基本的な考え方も合わせて説明せよ. (Explain derivation of the confidence interval together with fundamental ideas.)

(2) 信頼係数 95% を保ったまま, 信頼区間の幅をより狭く $22.1\% \pm 0.33\%$ のように精度を高めるためには, どうすればよいか? 理由も合わせて述べよ. (What can you do in order to increase the precision so as to have the confidence interval $22.1\% \pm 0.33\%$ with the same confidence coefficient?)

**[5]** 化学物質の精製実験を同一条件下で 9 回行い, 生成物の重量を測定して, 平均値 $\bar{x} = 32.75$, 不偏分散 $u^2 = 0.25 = 0.5^2$ を得た. この方法で得られる生成物の平均重量の 90%信頼区間を求めよ. (10 点) (The weight of refined chemical is measured. After 9 experiments under the same condition, one obtained the mean value $\bar{x} = 32.75$, and the unbiased variance $u^2 = 0.25 = 0.5^2$. Find the confidence interval for the mean value with confidence coefficient 90%.)

**[6]** コインの公平さを判断するために, コインを 400 回投げた. ($10$ 点 $\times 2 = 20$ 点) (A coin is tossed 400 times for judgement of the fairness.)

(1) 表が 222 回出た. このコインは公平であるといえるだろうか? 有意水準 5% で仮説検定せよ. 有意水準 1% ではどうか? 結果を比較し, 知るところを述べよ. (The heads occurs 222 times. Can you claim this coin is fair? Perform hypothesis testing with significance level 5% and 1%. Then compare these results.)

(2) 表が 215 回出た. 有意水準 5% の仮説検定によってコインは公平といえるかどうかを判定するときに生ずる第 2 種誤り確率とは何か説明せよ. 次に, 第 2 種誤り確率が 5%以下になるような状況はどのような場合であるか答えよ. (The heads occurs 215 times. What is the type II error in the testing hypothesis with significance level 5%? Explain when the probability of type II error is less than 5%.)

**[7]** 日本の国民全体の平均年齢は 44.5 歳, 標準偏差は 23.5 歳である. あるサークルのメンバー 25 名の平均年齢は 34 歳である. このサークルは日本人の無作為標本といえるだろうか? (10 点) (The average age of the total Japanese population is 44.5 (years), and the standard deviation is 23.5 (years). There is a group consisting of 25 persons of which average age is 34. Can you claim that this group is a random sample of the Japanese population?)

標準正規分布 $N(0,1)$ の両側 $\alpha$ 点 $P(|Z| \geq z) = \alpha$)

| $\alpha$ | 0.317 | 0.100 | 0.050 | 0.045 | 0.010 | 0.003 | 0.001 |
|---|---|---|---|---|---|---|---|
| $z$ | 1.00 | 1.64 | 1.96 | 2.00 | 2.58 | 3.00 | 3.29 |

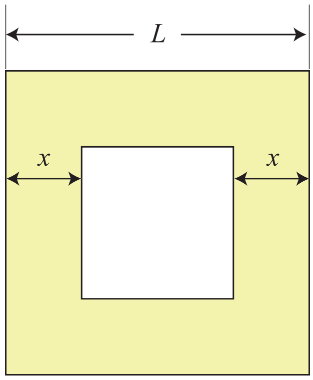
自由度 $n$ の $t$ 分布 $t_n$ の両側 $\alpha$ 点 $P(|T| \geq t) = \alpha$)

| $n\backslash\alpha$ | 0.100 | 0.050 | 0.020 | 0.010 |
|---|---|---|---|---|
| 1 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.812 | 2.228 | 2.764 | 3.169 |
| $\infty$ | 1.645 | 1.960 | 2.326 | 2.576 |

# 数理統計学 (2015.07.22 実施) 期末試験解説

**[1]** (1) $x \leq 0$ のとき $F(x) = 0$, $x \geq L/2$ のとき $F(x) = 1$ は明らか. $0 < x < L/2$ とする.

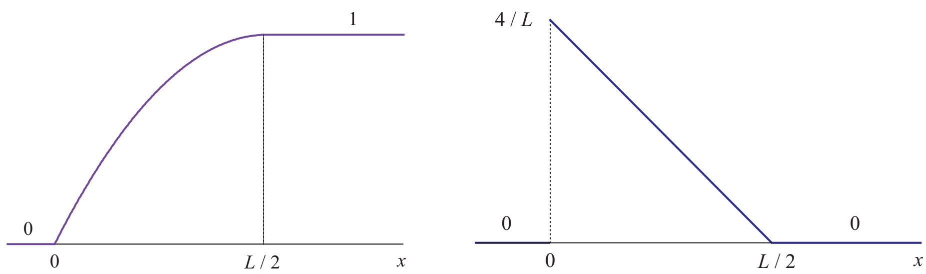
$$F(x) = P(X \leq x) = \frac{L^2 - (L - 2x)^2}{L^2} = \frac{1}{L^2}(-4x^2 + 4Lx)$$




(2) $0 < x < L/2$ のとき,

$$f(x) = F'(x) = \frac{1}{L^2}(-8x + 4L)$$

それ以外では, $f(x) = 0$.




(3)

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{L^2} \int_0^{L/2} x(-8x + 4L) dx = \frac{L}{6}$$

(4)

$$\mathbf{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{1}{L^2} \int_0^{L/2} x^2(-8x + 4L) dx = \frac{L^2}{24}$$

したがって,

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{L^2}{24} - \left(\frac{L}{6}\right)^2 = \frac{L^2}{72}.$$

**[2]** (1)

$$P(T \geq a) = \int_a^{+\infty} \lambda e^{-\lambda x} dx = \left[ -e^{-\lambda x} \right]_a^{+\infty} = e^{-\lambda a}.$$

(2) 条件付確率の定義から,

$$P(T \geq a+b | T \geq a) = \frac{P(T \geq a+b, T \geq a)}{P(T \geq a)}$$

であるが, $T \geq a+b$ ならば $T \geq a$ であるから, $\{T \geq a+b, T \geq a\} = \{T \geq a+b\}$ である. そうすると, (1) の結果も用いて,

$$P(T \geq a+b | T \geq a) = \frac{P(T \geq a+b)}{P(T \geq a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P(T \geq b).$$

この性質を指数分布の無記憶性という.

**[3]** 病気 A に感染している確率と感染していない確率は

$$P(A) = q, \qquad P(A^c) = 1 - q.$$

検査 B に陽性反応を示す確率は, 条件付確率であって,

$$P(B|A) = 0.9 \qquad P(B|A^c) = 0.05$$

ベイズの公式によって,

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\ &= \frac{q \times 0.9}{q \times 0.9 + (1-q) \times 0.05} = \frac{18q}{17q+1}. \end{aligned}$$

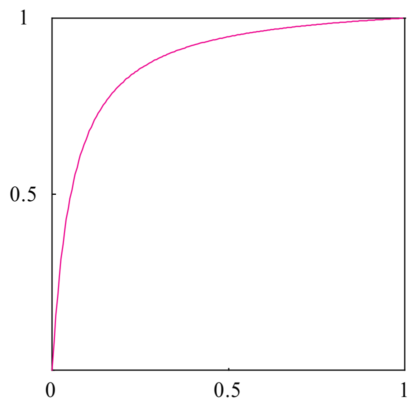
(1) $q = 0.04$ とすれば,

$$P(A|B) = \frac{18 \times 0.04}{17 \times 0.04 + 1} = \frac{3}{7} \approx 0.428$$

(2)

$$P(A|B) = \frac{18q}{17q+1} = \frac{18}{17}\left(1 - \frac{1}{17q+1}\right)$$

$q = 0$ から $q = 1$ に変化するとき, $P(A|B)$ は 0 から 1 まで単調に増加する. しかも, $q = 0$ の近くではかなり急激に増加しているといえる.



[**4**]　(1) 内容は教科書を参照のこと.

(2) 一般に, 95%信頼区間は

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

で与えられる. 本問では母比率を問うているので, 標本比率 $\hat{p}$ を用いて,

$$\hat{p} \pm 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

となる. いずれにせよ, 信頼係数を保ったまま, 信頼区間の幅を 1/10 にするためには標本数は 100 倍必要になる.

※ 試験問題として出題しているのだから、数理的な解答が必要。「精度を上げるためには標本数を増やせばよい」などという情緒的な解答の人は、何のための数理統計学だったのか、反省せよ。

[**5**]　標本平均を $\bar{X}$ とすれば,

$$\frac{\bar{X}-m}{U/\sqrt{n}}$$

が自由度 $n-1$ の $t$ 分布にしたがう. 90% 信頼区間は $t_8$ の両側 10% 点が 1.86 であることから,

$$\bar{X} \pm 1.86 \times \frac{U}{\sqrt{9}} = 32.75 \pm 1.86 \times \frac{0.5}{3} = 32.75 \pm 0.31$$

[**6**] (1) $H_0 : p = \dfrac{1}{2}$, $H_1 : p \neq \dfrac{1}{2}$, $\alpha = 0.05$ として検定を行う. コインを 400 回投げるときの表の回数を $X$ とすると, $X \sim B(400, 1/2) \approx N(200, 10^2)$. $H_1$ の形から両側検定である. 棄却域は

$$\left|\frac{x-200}{10}\right| \geq 1.96$$

であり, 実現値 $\bar{x} = 222$ は棄却域に落ちる. したがって, 有意水準 5% で $H_0$ は棄却され, コインは公平ではないと判定される.

有意水準 $\alpha = 0.01$ とすると, 両側検定の棄却域は

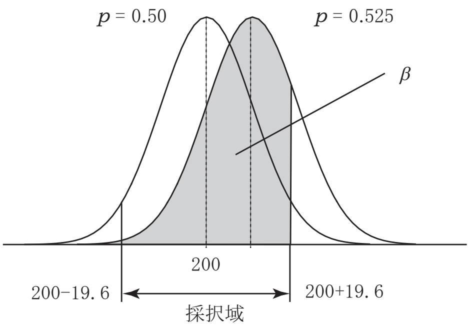
$$\left|\frac{x-200}{10}\right| \geq 2.58$$

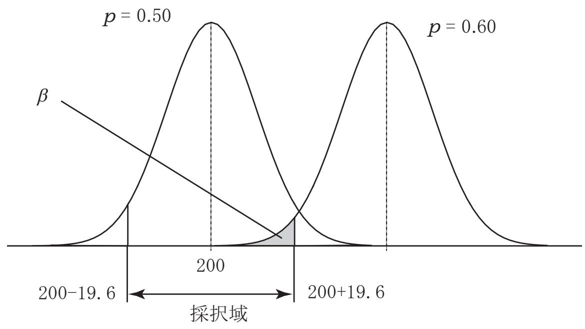
となり, 実現値 $\bar{x} = 222$ は棄却域に落ちない. したがって, 有意水準 1% で $H_0$ は棄却されない. したがって, 有意水準 5% では有意差を見出したが, 高度に有意ではない.

(2) (1) と同様に, $H_0 : p = \dfrac{1}{2}$, $H_1 : p \neq \dfrac{1}{2}$, $\alpha = 0.05$ として検定を行う. 実現値 $\bar{x} = 215$ は棄却域に落ちないので, 有意水準 5% で $H_0$ は棄却されず, コインは公平であると判定される. この判定が間違っている確率, すなわち, コインは公平ではないのにもかかわらず公平であると判定してまう確率が第 2 種誤り確率である.

コインが公正ではない場合, 可能な $p$ は無限にあり, 第 2 種誤り確率を評価することができない. たとえば, $p = 0.525$ (左図) と $p = 0.6$ (右図) の場合を考えよう. $p = 0.525$ の場合は, $p = 0.5$ の分布と近いため, 採択域の範囲にかぶさってくる範囲がかなり大きく, 図からわかるように $\beta \geq 0.5$ である. つまり, 公正なコインに近い偽物に対しては, 公正であるという判定をしがちであり, その間違い確率 (第 2 種誤り確率) は極めて大きい.

$p = 0.6$ の場合は, 2 つの分布がかなり離れており, $\beta \leq 0.025$ となることが見て取れる. したがって, $\beta = 0.05$ となるのは $p$ が 0.6 より少し小さい時である. このとき, 分散はほぼ $10^2$ と考えてよいので, $219.6 + 1.64 \times 10 = 236$ から $p = 0.59$ がわかる. つまり, 調べているコインが公平 ($p = 0.5$) であるか, あるいは表が出やすく $p \geq 0.59$ であるか, いずれかであれば, 第 2 種誤り確率が 0.05 に抑え込まれる.




**[7]** 日本人全体の平均年齢を $m$ とする. $H_0 : m = 44.5$, $H_1 : m \neq 44.5$, $\alpha = 0.05$ として検定を行う. 大きさ $n = 25$ の無作為標本の標本平均 $\bar{X}$ に対して,

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} = \frac{\bar{X} - 44.5}{23.5/\sqrt{25}} = \frac{\bar{X} - 44.5}{4.7}$$

が $N(0, 1)$ に従う. 実現値 $\bar{x} = 34$ は,

$$z = \frac{34 - 44.5}{4.7} = -2.23.$$

有意水準 5% の棄却域は $|z| > 1.96$ であるから, 実現値 $z = -2.23$ は棄却域に落ち, $H_0$ が棄却される. つまり, このグループは日本人全体からの無作為標本とは言えない.

有意水準 1% なら $H_0$ は棄却されない. したがって, このグループは日本人全体からの無作為標本と言える.