統計数学後期試験問題

1997/01/07 服部哲弥

解答用紙は氏名と学籍番号を記入せよ．番号が 95CA$\cdots$ **以外** の者は名前を囲み，95CA$\cdots$ の者は名前に飾りをつけないこと．早く終わった者は答案を提出して退出してよい．また，２枚目の数表及びいくつかの数値は自由に用いてよい．

問 1． 次の文章は古典的な統計的推測の方法に関する基礎事項のいくつかを説明している．文章中の空欄 (1)–(20) に下の語群から適当なものを入れよ.（番号の等しい欄は同じ語が入り，番号の異なる欄は異なる語が入る．用いない語もある．解答は番号順に並べること.）

標準正規分布（平均 0 分散 1 の正規分布）の分布関数を F と書く． F は積分を用いて

$$F(x) = \frac{1}{\sqrt{\boxed{\text{(1)}}}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

と表される．ここで，$\boxed{\text{(1)}}$ は全体集合の確率が 1 であること，即ち，$\displaystyle\lim_{x\to\infty} F(x) = \boxed{\text{(2)}}$ から決まる．

確率空間で定義された実数値関数（正確には可測関数）を確率変数という．確率変数の列 $X_1, X_2, X_3, \cdots, X_n$ が，独立でどの確率変数も同じ分布に従うとき，この列を大きさ n の $\boxed{\text{(3)}}$ と呼び，確率変数が従う分布を母集団と呼ぶ．確率的な現象を n 回測定または観測して n 個の数値を得ることはこの $\boxed{\text{(3)}}$ の値 $X_1(\omega), X_2(\omega), X_3(\omega), \cdots, X_n(\omega)$ を得ることと考える．その場合，変数列の独立性は $\boxed{\text{(4)}}$ によって確保されると期待する．サンプル値の分布を経験分布と呼ぶ．度数分布表は経験分布を見やすくする一つの方法である．

$\boxed{\text{(5)}}$（母集団の平均）を μ $(= \mathrm{E}[\,X_1\,])$，$\boxed{\text{(6)}}$（母集団の分散）を σ^2 $(= \mathrm{V}[\,X_1\,])$ と書く．$\boxed{\text{(5)}}$ と $\boxed{\text{(6)}}$ を点推定する量としてよく用いられるものには，それぞれ，サンプル平均 $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ と不偏分散 $V_n = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ がある．$\bar{X}$ の右辺の分母が n なのに対して V_n の右辺の分母が $n-1$ である．このように定義する根拠として，例えば $\boxed{\text{(7)}}$ （$\mathrm{E}[\,\bar{X}_n\,] = \boxed{\text{(8)}}$, $\mathrm{E}[\,V_n\,] = \boxed{\text{(9)}}$ を満たすこと）がある．

サンプル平均は平均が $\mathrm{E}[\,\bar{X}_n\,] = \boxed{\text{(8)}}$, 分散が $\mathrm{V}[\,\bar{X}_n\,] = \mathrm{E}[\,(\bar{X}_n - \mu)^2\,] = \boxed{\text{(10)}}$ となる．前者は期待値の線形性から，後者は期待値の線形性と確率変数の独立性からわかる．従って，$\boxed{\text{(11)}}$ から，n が十分大きければ $\dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$ は標準正規分布 $N(0,1)$ にほぼ従う．このことを用いれば（$\boxed{\text{(3)}}$ が十分大きいとき）$\boxed{\text{(5)}}$ を次のようにして推定または検定することができる．

$F(x)$ の定義から

$$\mathrm{P}(\, z' < \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} < z\,) = F(z) - F(z')$$

となるので，$\boxed{\text{(12)}}$ α または信頼率 $1-\alpha$ が与えられたとき，$F(z) - F(z') = 1-\alpha$ となるように z, z' を決める．このとき，$z' < \dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} < z$ が $\boxed{\text{(13)}}$ を定める．一つの方程式から z, z' の二つの未知数を決めることはできない．通常は，正規分布の対称性から $F(z) + F(-z) = 1$ となることを用いて，$z' = -z$ と選ぶ．しかし，対立仮説のあり方によっては，$\boxed{\text{(13)}}$ の取り方が変わりうる．検定のときは α は $\boxed{\text{(14)}}$ が起きる確率（$\boxed{\text{(15)}}$ H が正しいのに棄却してしまう確率）に等しい．

このように，$\dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$ が標準正規分布に従うことを利用することによって，$\boxed{\text{(6)}}$ の値を推定する方法があれば（正規母集団でなくても）母集団の $\boxed{\text{(5)}}$ の推定および検定を行うことができる．$\boxed{\text{(6)}}$

は，例えば，n が十分大きいときは $\boxed{(16)}$ で近似できる．正規分布に従う確率変数の $\boxed{(17)}$ も正規分布に従うことを用いれば，２つの系列（例えば二個所でおこなった独立な測定結果）の $\boxed{(5)}$ の $\boxed{(17)}$ について検定することができるので，例えば二つの工場で作られた製品の品質が統計的に等しいかどうかを検査できる．他の分布を用いることで他の母集団特性値の推定および検定を行うこともできる．例えばカイ平方分布を用いれば正規母集団の $\boxed{(6)}$ の推定や検定，および，分布の $\boxed{(18)}$（母集団の分布が，与えられた分布と一致しているかどうか）や（因子の）$\boxed{(19)}$ などの検定を行うことができるし，F分布を用いれば（$\boxed{(6)}$ についての仮定なしに，小さな n でも）正規母集団の $\boxed{(5)}$ の推定や検定を行うことができる．カイ平方分布に基づいて正規母集団の $\boxed{(6)}$ の推定ができるのは，正規分布 $N(\mu,\sigma^2)$ に従う $\boxed{(3)}$ $X_1,\ X_2,\ \cdots,\ X_n$ に対して，$\chi^2 = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 = \dfrac{1}{\sigma^2}(n-1)V_n$ が，自由度 $\boxed{(20)}$ のカイ平方分布に従うことによる．

語群

-1	2π	σ^2	$\mathrm{E}[\,\bar{X}_n\,]$	和	適合度	帰無仮説
0	π^2	σ^2/n	$\mathrm{V}[\,\bar{X}_n\,]$	差	不偏性	無作為抽出
1	$2\pi^2$	X_1	$\mathrm{E}[\,V_n\,]$	積	独立性	第１種の過誤
2	μ	$\bar{X}_n$	$\mathrm{V}[\,V_n\,]$	商	信頼区間	第２種の過誤
n	$n-1$	V_n	$1-\alpha$	危険率	サンプル	大数の法則
F	$F(0)$	α	π	母平均	母分散	中心極限定理

問 2.

問 2–1. 硬貨を続けて 200 回投げたところ，117 回表が出た．この硬貨は正常か．危険率 1% で検定せよ．ここで正常な硬貨とは表の出る確率が 0.5 ということであるとし，試行が独立であることは疑わないものとする．

この硬貨が正常であるか表が出やすいかどちらかである，という信頼できる情報があった場合はどうか．

問 2–2. あるスーパーマーケットで卵 10 個パックを買って卵の重さを測ったところ 62, 62, 64, 64, 64, 65, 66, 67, 67, 70, であった（単位はグラム）．母集団は正規分布とするとき，重さの母分散 σ^2（の信頼区間）を信頼係数 90% で求めよ．

別のパックの重さは 58, 61, 62, 63, 64, 64, 64, 66, 67, 68, であった．σ^2 を信頼係数 90% で求めよ．

この二つのパックは平均的な重さが等しい母集団から得られたといえるか．

問 3.

問 3–1. ２つの数を区間 $(0,1)$ の中で，ランダムに（一様分布に従って）独立に選ぶ．小さいほうの数が 0.5 より小さいという条件の下で，大きいほうの数が 0.5 より大きい確率を求めよ．

問 3–2. ある番組の視聴率を信頼係数（信頼度）95% で誤差が 2% 以内になるように推定したいとき，何人以上のモニターを調査すれば十分か求めよ．p が実数ならば $p(1-p)\le 1/4$ が成り立つことを用いてもよい．（注：この問題では信頼区間が $[c-a, c+a]$ という形ならば，a のことを誤差と呼ぶ.）

問 4. 統計数学の後期の講義で最も印象に残った内容を書け．もし万が一，印象に残る内容を得るほど出席しなかった場合は，出席しなかった時間をどのように有効に利用したのか，なぜ，それがこの講義への出席より有効な時間利用の方法と考えたのかを書け．説得力および具体性（正直に書いているように見えるか，具体的な経験に基づかなければ書けない内容にふみこんでいるか）を採点基準とする．

標準正規分布の分布関数

$$F(x) = \int_{-\infty}^{x} f_{0,1}(y)\,dy \propto \int_{-\infty}^{x} e^{-y^2/2}\,dy\,.$$

x	$F(x)$
0	1/2
1	0.8413
1.6448	0.9500
1.9600	0.9750
2	0.9772
2.326	0.9900
2.5758	0.9950
3	0.9986
4	0.99997

X が自由度 n のカイ平方分布に従うとき，$\alpha = \mathrm{P}(\ X > c\ ) = \displaystyle\int_c^{\infty} f_n(x)\,dx$ となる $c = c(n,\alpha)$ の表.

$n \setminus \alpha$	0.99	0.95	0.05	0.01
1	0.00	0.00	3.84	6.63
2	0.02	0.10	5.99	9.21
3	0.12	0.35	7.81	11.34
4	0.30	0.71	9.49	13.28
5	0.55	1.15	11.07	15.09
6	0.87	1.64	12.59	16.81
7	1.24	2.17	14.07	18.48
8	1.65	2.73	15.51	20.09
9	2.09	3.33	16.92	21.67
10	2.56	3.94	18.31	23.21

計算にあたって，次の数値を用いてもよい.

$\sqrt{2} = 1.4142,$
$1.7\sqrt{2} = 2.404,$
$1.4\sqrt{\dfrac{10}{7.87}} = 1.578,$
$3.1^2 \times 2 + 1.1^2 \times 3 + 0.1^2 + 0.9^2 + 1.9^2 \times 2 + 4.9^2 = 54.9,$
$5.7^2 + 2.7^2 + 1.7^2 + 0.7^2 + 0.3^2 \times 3 + 2.3^2 + 3.3^2 + 4.3^2 = 78.1\,.$

統計数学後期試験解答

1997/01/07 服部哲弥

問 1 (40).

(1) 2π	(2) 1	(3) サンプル	(4) 無作為抽出	(5) 母平均
(6) 母分散	(7) 不偏性	(8) μ	(9) σ^2	(10) σ^2/n
(11) 中心極限定理	(12) 危険率	(13) 信頼区間	(14) 第 1 種の過誤	(15) 帰無仮説
(16) V_n	(17) 差	(18) 適合度	(19) 独立性	(20) $n-1$

問 2–1 (20). （前半小針 p.200 例題 2，後半講義）

帰無仮説 H: $p=0.5$ を検定する．N 回投げたときの表の出る割合 s は平均 p, 分散 $\dfrac{p(1-p)}{N}$ の正規分布にほぼ従うから，$\dfrac{s-p}{\sqrt{p(1-p)/N}}$ は標準正規分布にほぼ従う．危険率 1% で信頼区間は

$$-z < \frac{s-p}{\sqrt{p(1-p)/N}} < z$$

となる．ここで z は $F(z)=1-0.01/2=0.995$ を満たす数なので， $z=2.5758$．実際に表の出た割合は $s=117/200=0.585$. $N=200$, $p=0.5$, も代入すると，$\dfrac{s-p}{\sqrt{p(1-p)/N}}=1.7\sqrt{2}=2.404$. $-2.5758<2.404<2.5758$ だから仮説は棄却できない．即ち，正常でないとは言えない．

表が出やすいという情報があれば，対立仮説は H': $p>0.5$ となるので，信頼区間は

$$\frac{s-p}{\sqrt{p(1-p)/N}} < z$$

ととるべきである．z は $F(z)=1-0.01=0.99$ を満たす数なので， $z=2.326$．$2.404>2.326$ だから仮説 H は棄却される．即ち，情報の追加によって，同じ危険率でこの硬貨はいかさまと判断する．

問 2–2 (20). （小針 p.225 (p.281) 練習問題 4）

データの大きさ $n=10$, 標本平均 $\bar{X}_n=65.1$，カイ平方 $\chi^2=(n-1)V_n/\sigma^2=(3.1^2\times 2+1.1^2\times 3+0.1^2+0.9^2+1.9^2\times 2+4.9^2)/\sigma^2=54.9/\sigma^2$．自由度 $n=10-1=9$．両側に 5% ずつの危険域をとることにして，$\alpha=95\%$, 5% に対応するカイ平方の値は表より，それぞれ 3.33, 16.92 だから，信頼区間は $3.33<54.9/\sigma^2<16.92$．即ち，$3.24<\sigma^2<16.5$.

後半は，データの大きさ $n=10$, 標本平均 $\bar{X}'_n=63.7$，カイ平方 $\chi'^2=(n-1)V'_n/\sigma^2=(5.7^2+2.7^2+1.7^2+0.7^2+0.3^2\times 3+2.3^2+3.3^2+4.3^2)/\sigma^2=78.1/\sigma^2$．前半と同様に，信頼区間は $4.62<\sigma'^2<23.5$.

以下，次のように考えれば講義の範囲で検定ができる．二つのパックの卵の重さの母平均をそれぞれ μ, μ' とする．帰無仮説 H: $\mu=\mu'$．$\bar{X}_n$, $\bar{X}'_n$ はそれぞれ $N(\mu,\sigma^2/n)$, $N(\mu',\sigma'^2/n)$ に従うので，$\bar{X}_n-\bar{X}'_n$ は $N(\mu-\mu',(\sigma^2+\sigma'^2)/n)$ に従う．よって仮説 H の下で，$Z=(\bar{X}_n-\bar{X}'_n)\sqrt{\dfrac{10}{\sigma^2+\sigma'^2}}$ は $N(0,1)$ に従う．$\alpha=0.05$ に対して，$F(z)-F(-z)=1-\alpha$ となる $z>0$ を選ぶと，$z=1.960$. 即ち，Z の信頼区間 $-1.960<Z<1.960$．データからは

$$Z(\omega)=(65.1-63.7)\sqrt{\frac{10}{\sigma^2+\sigma'^2}}\le 1.4\sqrt{\frac{10}{3.25+4.62}}=1.58\,.$$

よって，H は棄却できない．二つのパックが同じ母集団から得られたとして重さに関して矛盾はない．($\alpha=0.1$ としても $-1.645<Z<1.645$ なので結論は変わらない．)

問 3–1 (15). （アクチュアリ H7-数学 1 問 1(2)）

X, Y を区間 $(0,1)$ で一様分布する確率変数とする．求める条件付き確率は

$$\frac{1}{1-P(X>0.5,\ Y>0.5)}(P(X>0.5,\ Y<0.5)+P(X<0.5,\ Y>0.5))=\frac{0.5^2+0.5^2}{1-0.5^2}=\frac{2}{3}\,.$$

問 3–2 (15). （アクチュアリ H7-数学２問 1(1)）

視聴率を p とすると一人が番組を見るか見ないかを表す確率変数 X（見たとき $X = 1$, 見ないとき $X = 0$）は平均 p 分散 $p(1-p)$ なので，n 人の視聴率のサンプル平均 $\bar{X}_n$ は平均 p 分散 $p(1-p)/n$ の正規分布に近い．よって，正数 z に対して，

$$P(p - z\sqrt{\frac{p(1-p)}{n}} < \bar{X}_n < p + z\sqrt{\frac{p(1-p)}{n}}) = F(z) - F(-z) = 2F(z) - 1\,.$$

これが信頼率 0.95 に等しくなるためには $F(z) = 0.975$ だから数表より $z = 1.96$. $p(1-p) \le 1/4$ を用いると，誤差は

$$z\sqrt{\frac{p(1-p)}{n}} \le \frac{0.98}{\sqrt{n}}\,.$$

これが 0.02 以下ならばよい．これを解くと $n \ge 2401$ を得る．

問 4 (10). （原理的に解答例等は存在しません.）