数理統計学　期末試験

問 1，問 2，問 3 をそれぞれ別の解答用紙に解答せよ．

2004/07/20 服部哲弥

問 1 (40)．　下記表は，標準正規分布 $N(0,1)$ の数値表の抜粋で，標準正規分布の下で実数 x に対して x 以上が生じる確率 $\mathrm{P}[\,[x,\infty)\,]$ を有効数字 4 桁以内で示す．（スペースの節約上一部のみ掲げ，書式も標準的な教科書と異なる．）以下に答えよ．

x	0.00	1.00	1.645	1.96	2.00	2.326	2.576	3.00
$\mathrm{P}[\,[x,\infty)\,]$	0.5000	0.1587	0.0500	0.0250	0.0228	0.0100	0.00500	0.00135

(i) $N(0,1)$ の密度関数 $\rho(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のおおよその形を図示し，表の中の 0.0228 という数字がその図のどの部分の面積を表すか図示せよ．ρ の図は一見しておかしくなければかまわないが，面積は必要な数値を明示し斜線を施すなど明確にすること．

(ii) 母集団が標準正規分布に従うとき，取り出したデータが 99% の確率で区間 $[-a, a]$ に入るような a を表から求めよ．

(iii) 平均 m 分散 v の正規分布 $N(m, v)$ に従う確率変数 X に対して $Y = a(X - b)$ が標準正規分布に従うような定数 a, b を m, v を用いて表せ．$a > 0$ とする．答だけでよい．

(iv) 母集団が分散 1 の正規分布 $N(p, 1)$ とする．平均 p は未知である．p を推定するために標本を無作為に取り出したところ，その値が x であった．信頼係数 99% で p の値を区間推定せよ．

問 2 (40)．　某テレビ番組によれば，さいころは目の数によって材料を削り取る量が異なるため目の出る確率が正確に $p = 1/6$ ずつではない，という．これを統計学的に検証するために，さいころを $n = 10000$ 回投げて 6 の出た回数 m を調べ，比 $X = m/n$ をデータとして帰無仮説　H: $p = 1/6$　を危険率（有意水準）1% で検定することにした．以下の問に答えよ．

(i) 帰無仮説 H の下で 6 の目は 1 万回中平均（期待値で）何回出るか？答は四捨五入して整数で．

(ii) H が棄却されるのは m が平均（期待値）からいくつ以上ずれたときか?「約 2100 以上ずれたとき」など有効数字 2 桁程度で答えよ．なお，H の下で X が従う分布は平均 p，分散 $v_n = \dfrac{p(1-p)}{n}$ の正規分布 $N(p, v_n)$ とし，必要ならば問 1 の表と $\sqrt{5} = 2.236$ を用いよ．

問 3 (40)．　地震調査研究推進本部 web page にある宮城県沖地震の過去の発生間隔をデータ（データ数 5）として標本平均 37.06 年と不偏分散の平方根 6.61 年を算出し，平均と標準偏差の点推定値とすることで発生間隔の分布を求めた．これに基づいて，諸君の大半が学部，大学院博士前期課程（学部卒業後東北大学大学院に進学した場合），博士後期課程を卒業・修了する予定時期までに地震の発生する確率を算出したものが右表である．以下に答えよ．

期限	期限までの発生確率
2007/03/31	6.5%
2009/03/31	15%
2012/03/31	32%

(i) データを X_1, X_2, X_3, X_4, X_5 とおくとき，標本平均 $\hat{X}$ をこれらを用いて表せ．答だけでよい．

(ii) 不偏分散 V を X_k, $k = 1, 2, 3, 4, 5$, と $\hat{X}$ を用いて表せ．答だけでよい．

(iii) 諸君が東北大学大学院への進学を考える際に上表の結果はどの程度の障害となるかならないかを簡単に説明せよ．（意志決定に統計学的データをどう用いるかという視点の明確さを採点基準とする．書き方は自由で，他の判断材料の列挙や重要性の比較，進学に対する考えの中への諸要素の位置づけなど．表の数字の価値を論じてもよいが，題意は意志決定であって確率・統計に関する議論ではない．現時点で判断できない場合は意志決定のために必要などういう情報が不足しているかが明確であることを基準とする．

参考までに，1978 年の地震では死者 13 人でうち 9 人はブロック塀の倒壊によった．ただし，死者がほとんどいなかったのは幸運が重なって火災が 8 件しかなかったためでもある．）

数理統計学　期末試験　**略解**　　2004/07/21 服部哲弥

問 1 (10*4)．　$N(0,1)$ の表について答えよ．

(i) 標準正規分布の密度関数の図に表の数値 0.0228 を示せ．

図略【釣り鐘型で左右対称，遠方で 0 に近づく．横軸の値 2 から右の範囲に斜線．値 2, 0.0228, 斜線の明記を採点基準．遠方で 0 に近づかないなどあまりにひどいグラフは減点.】

(ii) $N(0,1)$ からとったデータが 99% の確率で $[-a,a]$ に入る a を求めよ．

左右対称性から $\mathrm{P}[\,[a,\infty)\,] = 0.001/2 = 0.005$ となる a を求めればよいから，$a = 2.576$.

(iii) $X \sim N(m,v)$ のとき $Y = a(X-b) \sim N(0,1)$ となる a,b を求めよ．

【積分変数変換 $z = a(x-b)$ によって

$$\mathrm{P}[\,Y > y\,] = \mathrm{P}[\,X > \frac{y}{a} + b\,] = \int_{\frac{y}{a}+b}^{\infty} e^{-(x-m)^2/(2v)}\,\frac{dx}{\sqrt{2\pi v}} = \int_y^{\infty} e^{-(z+(b-m)a)^2/(2va^2)}\,\frac{dz}{\sqrt{2\pi v a^2}}.$$

これが $\displaystyle\int_y^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$ に恒等的に等しいためには,】$a = \dfrac{1}{\sqrt{v}}$, $b = m$.

(iv) $N(p,1)$ に従う母集団から取り出した標本 x を元に信頼係数 99% で p の値を区間推定せよ．

$N(p,1)$ の分布を P_p と書くとき，P_p の下で x より大きい値も小さい値も確率 $0.01/2 = 0.005$ より頻繁に起きるような p の範囲，つまり，$\mathrm{P}_p[\,(x,\infty)\,] \geqq 0.005$ かつ $\mathrm{P}_p[\,(-\infty,x)\,] \geqq 0.005$ なる p の範囲を求めればよい．(iii) から $\mathrm{P}_p[\,[x,\infty)\,] = \mathrm{P}_0[\,[x-p,\infty)\,]$ かつ
$\mathrm{P}_p[\,(-\infty,x]\,] = \mathrm{P}_0[\,(-\infty,x-p]\,] = \mathrm{P}_0[\,[p-x,\infty)\,]$ の一方，(ii) から $\mathrm{P}_0[\,[2.576,\infty)\,] = 0.005$ なので $|x-p| \leqq 2.576$ 即ち，$p \in [x-2.576, x+2.576]$ となるから，これが求める区間である．

問 2 (10+30)．　さいころを $n = 10000$ 回投げて 6 の出た比率 $X = m/n$ をデータとして帰無仮説 H: $p = 1/6$ を危険率 1% で検定する．

(i) H の下で 6 の目は 1 万回中平均何回出るか？

$np = 1667$ 回.

(ii) H が棄却される $|m-np|$ の範囲を求めよ．

問 1(iii) より $Y = \dfrac{X-p}{\sqrt{v_n}}$ は $N(0,1)$ に従う．よって問 1(ii) から危険率 1% で H を $|Y| > 2.576$ のとき棄却し $|Y| < 2.576$ のとき採択する．すなわち
$|X-p| \geqq 2.576\sqrt{v_n} = 2.576\sqrt{\frac{p(1-p)}{n}} = 2.576\sqrt{\frac{5}{6^2\times 10000}} = 0.0096$，より，$|m-np| \geqq 0.0096n = 96$．約 100 回よぶんにまたは少なく出れば H を棄却する．

（この問題を示唆して下さった須藤砂織さんに感謝します.）

問 3 (5+5+30)．　宮城県沖地震の過去の発生間隔データ（サイズ 5）に基づいてある分布族のパラメータを点推定して得られた分布を用いて，2007, 2009, 2012 年度末までに地震の発生する確率を計算した．

(i) データを X_1, X_2, X_3, X_4, X_5 とおくとき，標本平均 $\hat{X}$ をこれらを用いて表せ．答だけでよい．

$$\hat{X} = \frac{1}{5}\sum_{k=1}^{5} X_k.$$

(ii) 不偏分散 V を X_k, $k = 1,2,3,4,5$, と $\hat{X}$ を用いて表せ．答だけでよい．

$$V = \frac{1}{4}\sum_{k=1}^{5} (X_k - \hat{X})^2.$$

(iii) 各期限までの地震の発生する確率の計算結果が東北大学大学院への進学希望にどの程度障害となるか．

採点基準：【意志決定支援の手段としての統計学という観点】から，意志決定における客観的根拠としての統計学データと希望・主観・人生観との区別の明確さ．意志決定できない場合は意志決定に必要なデータを認識して入手しようという視点.

参考：地震調査研究推進本部仙台沖地震の URL：

http://www.jishin.go.jp/main/chousa/00nov4/miyagi.htm

なお，この URL が採用する分布は不適当なので問題文の数字は出題者の研究による．

仙台市による 1978 年地震の被害の概要説明の URL：

http://www.city.sendai.jp/syoubou/bousai/sairai/

当時の仙台市域で死者 13 人，うち 9 人はブロック塀の倒壊による．その後ブロック塀は減った．ただし，幸運が重なって火災が 8 件．条件が悪ければ被害は大きくなる．住家全半壊が 4000 戸強．電気電話は 1,2 日でほぼ全面復旧したが，水道は 1 週間強（ガスは約 1 月）とのこと．