数理統計学・期末試験問題 (2018.07.27)

- 電卓・スマートウォッチなどの計算機の使用禁止.
- 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.
- 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ので参考にされたい.

[1] 正規分布表を用いて, 次の問いに答えよ. (10 点 $\times 2 = 20$ 点)

(1) $X \sim N(4, 3^2)$ のとき, $P(X \geq 2.47)$ を求めよ.

(2) Y が $N(50, 10^2)$ に従う確率変数のとき, $P(Y \leq a) = 0.33$ を満たす a を求めよ.

[2] 3 つの事象 A, B, C が独立であるとき,

$$P(A|B^c \cup C) = P(A)$$

が成り立つことを条件付確率の定義に基づいて証明せよ. ただし, B^c は B の余事象である. (10 点)

[3] ある国では, 病気 A の感染者が $100q$ % 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. (10 点 $\times 2 = 20$ 点)

(1) $0 < q < 1$ として, 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 を求めよ. この P が q とともにどのように変化するか, その変化の特徴からこの確率 P を現実問題に適用する際の注意を述べよ.

(2) $q = 0.04$ とする. この検査を受けて陰性反応が出た人が非感染者である確率を求めよ. ただし, 答は既約分数または小数で表せ.

[4] 100 万世帯のうち 600 世帯を無作為抽出して番組 A の視聴率調査を行い, 信頼係数 95% で視聴率の信頼区間 $22.1 \pm 3.3\%$ を得た. (5 点 $\times 3 = 15$ 点)

(1) 一般に, 信頼区間はどのようにして導かれるか, 信頼係数の意味も含めて説明せよ.

(2) 信頼係数 90% の信頼区間を求めよ.

(3) 信頼係数 95% を保ったまま, 信頼区間の幅をより狭く $22.1\% \pm 1.1\%$ 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.

[5] ある工場の製造ラインで 1 万個の部品を製造した. 部品 1 個の重量は 20.00g になるように調整しているが, 製造ラインの特性によって, 製品の重量は正規分布に従って変動する. 製品から選んだ 9 個の標本の平均重量は 20.21g であった. (5 点 $\times 2 = 10$ 点)

(1) 製造ラインのゆらぎが, 標準偏差 0.24g の正規分布に従っているとき, 製造ラインに狂いが生じているかどうかを仮説検定で判定せよ.

(2) 標準偏差 0.24g があらかじめ知られていないときは, どのような検定をすればよいか. 計算の概略を述べよ.

[6] コインが公平かどうかを判断するために, コインを 400 回投げて表の回数を調べた. (5 点 $\times$ 2 $=$ 10 点)

(1) 有意水準 5% の仮説検定において生ずる第 2 種誤り確率とは何か説明せよ.

(2) (1) において, 第 2 種誤り確率が 10%以下になるのはどのような場合であるか答えよ.

[7] 辺の長さが L の正三角形の内部から, どの点も同等の確率で選ばれるようにランダムに選ばれた点を A とする. A から 3 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 から正三角形の周までの距離である. (5 点 $\times$ 3 $=$ 15 点)

(1) X の確率密度関数 $f(x)$ を求めよ.

(2) X の平均値を求めよ.

(3) X の分散を求めよ.

付録：標準正規分布表　　$P = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^z e^{-x^2/2} dx$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4773	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

数理統計学 (2018.07.27 実施) 期末試験解説

[1] (1) $Z \sim N(0,1)$ とする.

$$\begin{aligned}P(X \geq 2.47) &= P\Big(\frac{X-4}{3} \geq \frac{2.47-4}{3}\Big) = P(Z \geq -0.51)\\ &= 0.5 + P(0 \leq Z \leq 0.51) = 0.5 + 0.1950 = 0.6950\end{aligned}$$

(2) $Y \sim N(50, 10^2)$ を標準化する.

$$P(Y \leq a) = P\Big(\frac{Y-50}{10} \leq \frac{a-50}{10}\Big) = P\Big(Z \leq \frac{a-50}{10}\Big)$$

ここで, $P(Z \leq b) = 0.33$ を満たす b を求める必要がある. まず, $b < 0$ に注意して, $P(b \leq Z \leq 0) = 0.17$ である. 標準正規分布表から, $b = -0.44$ がわかる. そうすると,

$$a = 10 \times (-0.44) + 50 = 45.6$$

[2] 条件付確率の定義から

$$P(A|B^c \cup C) = \frac{P(A \cap (B^c \cup C))}{P(B^c \cup C)}$$

である. 分子について,

$$A \cap (B^c \cup C) = (A \cap B^c) \cup (A \cap C)$$

であるから,

$$\begin{aligned}P(A \cap (B^c \cup C)) &= P(A \cap B^c) + P(A \cap C) - P((A \cap B^c) \cap (A \cap C))\\ &= P(A \cap B^c) + P(A \cap C) - P(A \cap B^c \cap C)\\ &= a(1-b) + ac - a(1-b)c = a - ab + abc\end{aligned}$$

ここで, A, B が独立であれば, A, B^c も独立であることを用いた. 同様に, $P(B^c \cup C)$ については

$$P(B^c \cup C) = P(B^c) + P(C) - P(B^c \cap C) = (1-b) + c - (1-b)c = 1 - b + bc$$

が得られる. したがって,

$$P(A|B^c \cup C) = \frac{P(A \cap (B^c \cup C))}{P(B^c \cup C)} = \frac{a - ab + abc}{1 - b + bc} = a = P(A)$$

となり証明を終える.

※ A, B, C が独立であることから

$$P(A \cap (B^c \cup C)) = P(A)P(B^c \cup C)$$

が成り立つが, このことを証明せずに使ったものは減点.

※ 確率計算の経過は, その事象をどうとらえるかで様々ありうる. 上記は一例.

[3] 病気 A に感染している確率と感染していない確率は

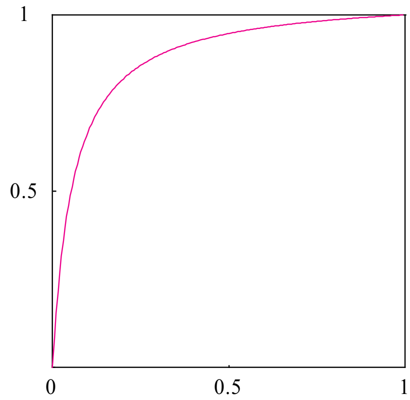
$$P(A) = q, \qquad P(A^c) = 1 - q.$$

検査 B に陽性反応を示す確率は, 条件付確率であって,

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

(1) ベイズの公式によって,

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



変化の特徴と現実問題に適用するときの注意 (2 例): (i) q が 0 から 1 に変化するとき, $P(A|B)$ も 0 から 1 に変化する. 一般に, q を知ることは困難 (全数調査が必要) であり, サンプリングなどによって推定することになる. したがって, q の値には幅を見ておく必要があり, 場合によっては P も大きな幅を見ておく必要がある.

(ii) q が小さいとき, たとえば $q \le 0.01$ のとき, $P \le 0.15$ となり, 陽性反応が出ても多くの場合は非感染者である. 検査の効率という点で注意を要する.

(2) ベイズの公式によって,

$$P(A^c|B^c) = \frac{P(A^c)P(B^c|A^c)}{P(A^c)P(B^c|A^c) + P(A)P(B^c|A)} = \frac{(1-q)\times 0.95}{(1-q)\times 0.95 + q\times 0.1} = \frac{19(1-q)}{19-17q}\,.$$

$q = 0.04$ とすれば,

$$P(A^c|B^c) = \frac{19\times 0.96}{19 - 17\times 0.04} = \frac{228}{229} \approx 0.9956...$$

※ $q=0$ のとき $P=0$, $q=1$ のとき $P=1$ であるから, グラフは $(0,0)$ と $(1,1)$ を結ぶ曲線である. 端点がきちんと把握されず, 曲線が浮いているようなものは不可.

[4] (1) 教科書参照.

(2) 信頼係数 95% の信頼区間と信頼係数 90%　の信頼区間は

$$\hat{p} \pm 1.96\frac{\sigma}{\sqrt{n}}\,, \qquad \hat{p} \pm 1.64\frac{\sigma}{\sqrt{n}}\,, \qquad \sigma = \sqrt{\hat{p}(1-\hat{p})}\,,$$

で与えられる. 信頼係数 95% の信頼区間が $22.1 \pm 3.3\%$ なので,

$$1.96\frac{\sigma}{\sqrt{n}} = 0.033$$

である. そうすると,

$$1.64\frac{\sigma}{\sqrt{n}} = \frac{1.64}{1.96}\times 0.033 = 0.028.$$

したがって, 信頼係数 90% の信頼区間は

$$22.1 \pm 2.8$$

(3) 信頼係数 95% の信頼区間は

$$\hat{p} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

で与えられるので, 幅を ± 3.3 から ± 1.1 のように 1/3 にするためには標本数を 9 倍, つまり, $600\times 9 = 5400$ 世帯を無作為抽出する必要がある.

※ 単に「標本数を増やす」という解答は 0 点.

[5] (1) 帰無仮説と対立仮説を

$$H_0 : m = 20 \qquad H_1 : m \neq 20$$

とする. 大きさ 9 の標本平均 $\bar{X}$ は

$$\bar{X} \sim N\left(20, \frac{0.24^2}{9}\right) = N(20, 0.08^2)$$

であうから, 標準化して

$$Z = \frac{\bar{X} - 20}{0.08} \sim N(0, 1).$$

実現値 $\bar{x} = 20.21$ を代入して,

$$z = \frac{20.21 - 20}{0.08} = 2.625$$

を得る. 対立仮説 H_1 から両側検定とする.

有意水準 $\alpha = 0.05$ (5%) に対する棄却域は $|z| \geq 1.96$ であり, 実現値 $z = 2.625$ は棄却域に落ちる. したがって, H_0 は棄却される.

さらに, 有意水準 $\alpha = 0.01$ では両側検定の 1%棄却域は $|z| \geq 2.58$ である. 実現値 $z = 2.625$ は棄却域に落ちるから, H_0 は棄却される. つまり, 高度に有意である.

(2) t 検定を用いる. 内容は教科書参照.

※ 不偏分散 U^2, 検定統計量 T, 自由度 $n-1$ の t 分布の用語あるいは式が含まれていないものは不可.

[6] (1) 仮説検定に際して帰無仮説と対立仮説を設定する. この問題においては, 表の出る確率を p として

$$H_0 : p = \frac{1}{2} \qquad H_1 : p \neq \frac{1}{2}$$

とする. コインを 400 回投げるときの表の回数を X とすると, $X \sim B(400, 1/2) \approx N(200, 10^2)$ である. 有意水準 $\alpha = 0.05$ ということは, X の実現値がその平均値 200 から外れて, 確率 0.05 の範囲を棄却域, 逆に平均値を中心に 0.95 の範囲を採択域とすることを意味する. 標準正規分布表から $P(|Z| \geq 1.96) = 0.05$ が知られているので, 本問題において採択域は

$$|x - 200| < 1.96 \times 10 = 19.6$$

となる. $p \neq 0.5$ であるコインであっても実現値がこの採択域に落ちれば, H_0 を採択し, その結果, $p = 0.5$ であると判定する. これは判定として誤っているのだが, この誤りを第 2 種誤りといい, その確率を第 2 種誤り確率という. 第 2 種誤り確率は p によって定まるので, 事前にこの値を知ることは不可能である.

(2) $p \neq 0.5$ のコインを 400 回振ったとき, 表の回数を Y とすれば,

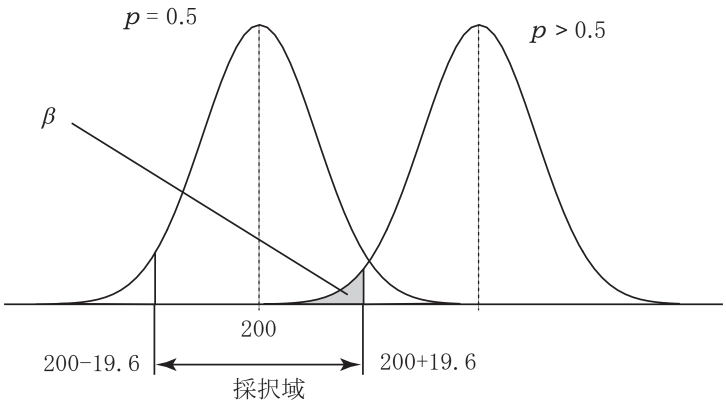
$$Y \sim B(400, p) \approx N(400p, 400p(1-p))$$

となる. $p \approx 0.6$ であれば $400p(1-p) = 96 = 9.8^2$ なので, 10 として近似しておく.

図は, $p > 0.5$ として Y の分布を X の分布に重ねて描いたものであり, Y の分布において, 採択域に値をとる確率が第 2 種誤り確率 β である. この $\beta = 0.1$ となるような p を求めればよい. 標準正規分布表から $P(Z \geq 1.28) = 0.40$ であるから $P(Z \leq -1.28) = 0.10$ となる. したがって, $\beta = 0.10$ となる p は

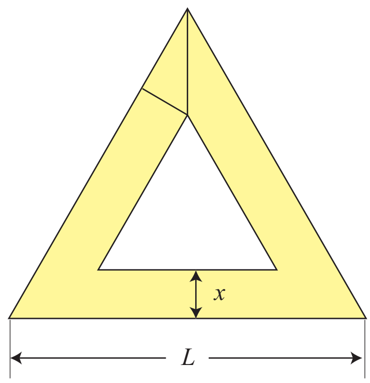
$$400p = 200 + 19.6 + 12.8 = 232.4$$

を満たす. つまり $p = 0.581$. したがって, 実際の p が $p \geq 0.58$ であれば $\beta \leq 0.1$ となる. 対称性から, $p < 0.5$ の場合は $p \leq 0.42$ ならばよい.



こうして, 表の出る確率 p が $p \geq 0.58$ または $p \leq 0.42$ のように 0.5 から相当にずれていれば, 仮説検定の結果, 採択する誤り (第 2 種誤り) を犯す確率が 0.1 以下になる.

[7]



(1) $0 \leq x \leq L/2\sqrt{3}$ で考えればよい. 辺の長さが L の正三角形の周から内側に x 離れている正三角形の辺の長さは $L - 2\sqrt{3}\,x$ である. したがって,

$$F(x) = P(X \leq x) = \frac{\dfrac{\sqrt{3}}{4}L^2 - \dfrac{\sqrt{3}}{4}(L - 2\sqrt{3}\,x)^2}{\dfrac{\sqrt{3}}{4}L^2} = 1 - \left(1 - \frac{2\sqrt{3}}{L}\,x\right)^2$$

これを微分して,

$$f(x) = F'(x) = \frac{4}{L^2}\,(-6x + \sqrt{3}\,L), \qquad 0 \leq x \leq \frac{L}{2\sqrt{3}}\,,$$

それ以外では, $f(x) = 0$.

(2)

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx = \frac{4}{L^2}\int_0^{L/2\sqrt{3}} x(-6x + \sqrt{3}\,L)dx = \frac{\sqrt{3}\,L}{18}$$

(3)

$$\mathbf{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{4}{L^2}\int_0^{L/2\sqrt{3}} x^2(-6x + \sqrt{3}\,L)dx = \frac{L^2}{72}$$

したがって,

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{L^2}{72} - \left(\frac{\sqrt{3}\,L}{18}\right)^2 = \frac{L^2}{216}\,.$$