数理統計学・期末試験問題 (2018.07.25)

- 電卓・スマートウォッチなどの計算機の使用禁止.
- 提出する解答用紙には学籍番号と氏名を記入せよ.
- 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] 3 つの事象 A, B, C は独立であって, $P(A) = a$, $P(B) = b$, $P(C) = c$ がわかっている. (5 点 $\times 2 = 10$ 点)

(1) $P(A^c \cap (B \cup C))$ を a, b, c で表せ. ただし, A^c は A の余事象である.

(2) 条件付確率 $P(A \cap B | A \cup C)$ を a, b, c で表せ.

[2] $0 < a < b$ とする. 隣り合う辺の長さが a, b である長方形の内部から, どの点も同等の確率で選ばれるようにランダムに選ばれた点を 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 から長方形の周までの距離である. (5 点 $\times 4 = 20$ 点)

(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 は実数である.

(2) X の確率密度関数 $f(x)$ を求めよ.

(3) X の平均値を求めよ.

(4) X の分散を求めよ.

[3] 大規模な選抜試験が実施され, 上位 5%が合格となる. 試験の結果, 平均点は 68 点, 標準偏差は 8 点であった. 受験者全体の得点分布は正規分布であると仮定して, 合格するための最低点を求めよ. (10 点)

[4] ある地域では, 病気 A の感染者は 1000 人に 4 人の割合であるという. 検査 B は, 感染者の 90%に陽性反応を示すが, 非感染者の $100p\%$ $(0 < p < 1)$ にも陽性反応が出るという. (5 点 $\times 2 = 10$ 点)

(1) 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 p で表せ.　この P が p とともにどのように変化するか, その変化の特徴を述べよ.

(2) $p = 0.05$ のとき, 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. 答は既約分数または小数で表せ.

[5] 100 万世帯から 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% の信頼区間 22.1±3.3% が得られた. この例に即して, 信頼区間とは何か, その導出も含めて 500 文字を目安に説明せよ. (500 文字には数式を含まない. この問題文の 1 行が 50 文字程度である) (15 点)

[6] 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, 重量は標準偏差 0.24g の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 20.19g であった．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．また, 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．(15 点)

[7] コインが公平かどうかを有意水準 5% の仮説検定で判断したい．そのためにコインを 400 回投げて表の回数を調べた．(10 点 $\times$ 2 $=$ 20 点)

(1) この問題に即して, 第 1 種誤り確率・第 2 種誤り確率とは何か説明せよ．

(2) 第 2 種誤り確率が 25%以上になるのはどのような場合であるか答えよ．

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

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 (2018.07.25 実施) 期末試験解説

[1]　(1) まず,

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

を用いて,

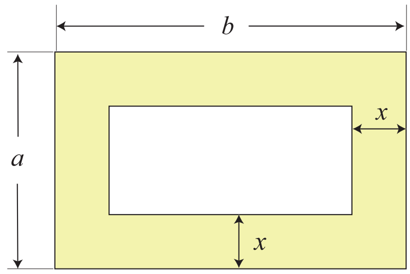
$$\begin{aligned} P(A^c \cap (B \cup C) &= P(A^c \cap B) + P(A^c \cap C) - P((A^c \cap B) \cap (A^c \cap C)) \\ &= P(A^c \cap B) + P(A^c \cap C) - P(A^c \cap B \cap C) \\ &= (1-a)b + (1-a)c - (1-a)bc = (1-a)(b+c-bc) \end{aligned}$$

簡単には, A, B, C が独立だから A^c と $B \cup C$ は独立であることを用いてもよい.

(2) $A \cap B \subset A \cup C$ であるから, $(A \cap B) \cap (A \cup C) = A \cap B$ である. したがって,

$$P(A \cap B | A \cup C) = \frac{P(A \cap B)}{P(A \cup C)} = \frac{P(A)P(B)}{P(A) + P(C) - P(A)P(C)} = \frac{ab}{a + c - ac}$$

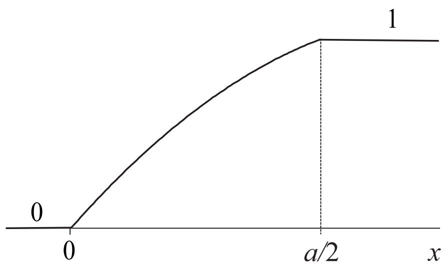
[2] $0 \le x \le a/2$ とする.



確率は面積比で与えるのが適当であるから,

$$F(x) = P(X \le x) = \frac{ab - (a-2x)(b-2x)}{ab} = 1 - (1 - \frac{2x}{a})(1 - \frac{2x}{b}) = (\frac{2}{a} + \frac{2}{b})x - \frac{4x^2}{ab}$$

$x < 0$ では $F(x) = 0$, $x > a/2$ では $F(x) = 1$.



(2) $0 \le x \le a/2$ では

$$f(x) = F'(x) = (\frac{2}{a} + \frac{2}{b}) - \frac{8x}{ab}.$$

それ以外では, $f(x) = 0$.

(3)

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{a/2} \Big\{ (\frac{2}{a} + \frac{2}{b})x - \frac{8x^2}{ab} \Big\} dx = \frac{a}{4} - \frac{a^2}{12b} = \frac{a^2}{12}\Big(\frac{3}{a} - \frac{1}{b}\Big).$$

(4)

$$\mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{a/2} \Big\{ (\frac{2}{a} + \frac{2}{b})x^2 - \frac{8x^3}{ab} \Big\} dx = \frac{a^3}{24}\Big(\frac{2}{a} - \frac{1}{b}\Big).$$

したがって,

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{a^3}{24}\Big(\frac{2}{a} - \frac{1}{b}\Big) - \bigg(\frac{a^2}{12}\Big(\frac{3}{a} - \frac{1}{b}\Big)\bigg)^2 = \frac{a^4}{144}\Big(\frac{3}{a^2} - \frac{1}{b^2}\Big).$$

[3] 受験者の得点を X とすれば, $X \sim N(68, 8^2)$. 合格するための最低点を a とすれば,

$$P(X \geq a) = 0.05.$$

標準正規分布において, $P(Z \geq 1.64) = 0.05$ がわかるから,

$$\frac{a-68}{8} = 1.64 \qquad \Longleftrightarrow \qquad a = 68 + 1.64 \times 8 = 81.12$$

[4]　病気 A に感染している確率と感染していない確率は

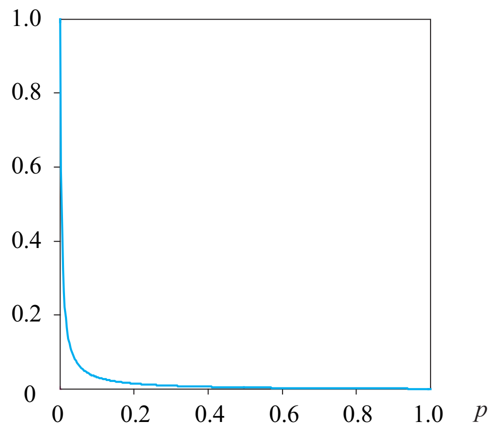
$$P(A) = \frac{4}{1000}, \qquad P(A^c) = \frac{996}{1000}.$$

検査 B に陽性反応を示す確率は, 条件付確率であって,

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

ベイズの公式によって,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{4 \times 0.9}{4 \times 0.9 + 996 \times p} = \frac{3}{3 + 830p}.$$



変化の特徴 (例示): p が 0 から 1 に変化するとき, $P(A|B)$ は 1 から 3/833 という小さな値に単調減少する. この減少は $p = 0$ の近くで急激であって, $P(A|B) \geq 0.5$ となるためには $p \leq 3/830 \approx 0.0036$ が必要である. 陽性から感染者を効率よく発見するためには, この検査の誤差ともいえる p を相当に小さくする必要がある.

※ $p = 0$ のとき, $P(A|B) = 1$ であって, 「反比例のグラフ」ではない.

(2) ベイズの公式によって,

$$P(A^c|B^c) = \frac{P(A^c)P(B^c|A^c)}{P(A^c)P(B^c|A^c) + P(A)P(B^c|A)} = \frac{996(1-p)}{996(1-p) + 4 \times 0.1}.$$

$p = 0.05$ とすれば,

$$P(A^c|B^c) = \frac{996 \times 0.95}{996 \times 0.95 + 4 \times 0.1} = \frac{946.2}{946.6} = \frac{4731}{4733}.$$

[5] 省略

[6] 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 の標本平均は $\bar{X} \sim N\left(20, \dfrac{0.24^2}{9}\right) = N(20, 0.08^2)$. 標準化して

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

実現値 $\bar{x} = 20.19$ を代入して,

$$z = \frac{20.19 - 20}{0.08} = 2.375$$

H_1 から両側検定となる. 有意水準 $\alpha = 0.05$ (5%) に対する棄却域は $|z| \geq 1.96$. 実現値 $z = 2.375$ 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, 製造ラインに狂いが生じていると判定される.

有意水準 $\alpha = 0.01$ とすると, 両側検定の 1%棄却域は $|z| \geq 2.58$. 実現値 $z = 2.375$ は棄却域に落ちない. したがって, 有意水準 1% では H_0 は棄却されない. （高度に有意ではない）

標準偏差があらかじめ知られていないときは, t 検定を行う. 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.

[7] (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. 公平なコインであっても, 偶然に実現値が棄却域に落ちることがある. このとき, H_0 は棄却されるので, 判定を誤ることになる. この誤りを第 1 種誤りといい, その確率を第 1 種誤り確率という. 第 1 種誤り確率は有意水準に一致する.

$p \neq 0.5$ 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 $\beta = 0.25$ となるような p を求めればよい. 標準正規分布表から $P(0 \leq Z \leq 0.675) = 0.25$ である. したがって, $\beta = 0.25$ となる p は

$$400p = 200 + 19.6 + 6.75 = 226.35$$

を満たす．つまり $p = 0.566$. したがって, 実際の p が $0.5 < p \leq 0.566$ であれば $\beta \geq 0.25$ となる．対称性から, $p < 0.5$ の場合も考えて, 表の出る確率 p が $0.434 \leq p \leq 0.566$ のように 0.5 に近い場合, 仮説検定の結果, 採択する誤り (第 2 種誤り) を犯す確率が 0.25 以上になる．