H27 海事統計学 期末試験（１／３）

答案には導出過程を記述すること。

問題 1

確率 p で $x = 1$、確率 $1 - p$ で $x = -1$ となる確率変数 x の期待値と分散を求めよ。（5 点）

問題 2

確率変数 θ が $0 \leq \theta \leq 2\pi$ の範囲において確率密度関数 $f(\theta) = \frac{1}{2\pi}$ で一様に分布するとき、この確率変数 θ の期待値および分散を求めよ。（5 点）

問題 3

K-S 検定では、データと確率分布モデルとの間の累積確率分布の差の最大値、あるいは 2 つのデータ間の累積確率分布の差の最大値を K-S 統計量として計算し検定を行う。確率分布モデルが以下の指数分布 $f(x)$ の場合の <u>累積確率分布関数</u>$F(x)$ を示せ。（5 点）

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{, where } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

問題 4

期待値 4，分散 9 の正規分布に従う確率変数 x が $7 \leq x \leq 10$ の範囲の値となる確率を正規分布表から求めよ。（5 点）

問題 5

造船所の安全対策の調査のため、無作為に選んだ従業員 2500 人にアンケート調査を行ったら、1 か月間のヒヤリハットの経験者は 500 人だった。このとき、造船所内でのヒヤリハットの割合 p の 95 %信頼区間を求めよ。（10 点）

問題６

3 人の子供を持つ 200 組の船員の家庭について子供の性別を調べたところ、男児の数が以下の表のようになった。

男児数	0	1	2	3
家庭数	30	85	70	15

これについて以下の問いに答えよ。

【問６–１】(5 点)
上記の表で、男女の出生比率が１：１であると仮定した場合の、男児数 0,1,2,3 名に対応する家庭数の期待度数をそれぞれ求めよ。

【問６–２】(10 点)
このデータが、男女の出生比率１：１という仮説に適合しているかどうかを有意水準 0.05（５％）で検定せよ。

問題７

故障や事故などの偶発的な事象が発生する時間間隔は、指数分布になることが知られており、以下の確率密度関数で表される：

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \le x \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

ただし e は自然対数の底を表す。このとき以下の問に答えよ。

【問７–１】(10 点)
実際に観測された事象の時間間隔の標本を $x_1, x_2, \cdots, x_n$ と表す。
このとき、確率密度関数を式 (1) で表す場合の標本の対数ユウ度はどのようになるか $x_1, x_2, \cdots, x_n$ を用いて示せ。

【問７–２】(10 点)
事象の時間間隔の標本が【問７–１】のように与えられる場合、式 (1) の指数分布のパラメータ λ の最尤推定値を導出し、上記の標本 $x_1, x_2, \cdots, x_n$ を使って表せ。導出過程も明記すること。

【問７–３】(5 点)
上記の式 (1) で表される発生時間間隔が、単位時間以下（つまり $x \le 1$）となる確率を e と λ を用いて式で表せ。

【問７–４】(10 点)
事象の発生する時間間隔が指数分布に従うとき、単位時間あたりに事象が発生する回数はポアソン分布に従うことが知られている。この事実を利用して、事象の発生時間間隔が式 (1) で表される事象が、単位時間あたり１回以上起きる確率を計算し、式 (1) 中で用いられているパラメータ λ や e を用いて表せ。

問題８

以下はある造船所で建造された肥大船の主要目と、あるフルード数におけるその船の流体抵抗値についての水槽試験データである。

要目＼船の番号	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	計画中
L/B	5.94	7.59	6.02	6.02	6.06	6.40	5.80	5.50	5.59
B/d	2.49	2.37	2.47	2.44	2.72	2.60	2.29	2.88	2.78
C_b	0.777	0.819	0.771	0.795	0.800	0.821	0.781	0.811	0.800
C_m	0.991	0.994	0.993	0.994	0.996	0.996	0.990	0.998	0.995
lcb	-2.40	-2.60	-2.28	-2.75	-3.10	-3.01	-2.35	-3.48	-3.11
流体抵抗値	2.80	3.03	3.16	3.11	3.39	4.26	3.13	4.11	?

【問８－１】（10 点）
新しく計画中の船の要目 $L/B = 5.59$, $B/d = 2.78$, $C_b = 0.800$, $C_m = 0.995$, $lcb = -3.11$ が与えられたとき、上記のデータを利用して、この計画中の船における流体抵抗値について多重回帰を用いて推定したい。
上記の表データの数字を用いて式を作り、計算手順を説明せよ。

【問８－２】（10 点）
問８－１の推定をより精度良く行うため、以下のような 5 通りの工夫を検討した。
このとき、明らかに推定計算上不都合が生じると考えられるものを全て挙げ、その理由を説明せよ。ただし単なる計算量の増加は、計算上の不都合とは考えないものとする。

1. 各船の特徴量として L/B, B/d, C_b, C_m, lcb に加えて、新しい特徴量 $x_6 = (L/B) \cdot (B/d) \cdot C_b \cdot C_m$ を導入し、問８－１と同様の推定方法で計算をやり直す。
2. 各船の特徴量として L/B, B/d, C_b, C_m, lcb に加えて、新しい特徴量 $x_6 = (L/B) + (B/D)$ を導入し、問８－１と同様の推定方法で計算をやり直す。
3. 上記の特徴量のうち、C_m の値はどの船もほとんど変わらないので、特徴量から除外し、問８－１と同様の推定方法で計算をやり直す。
4. 新しく No.9 の船についてのデータをテーブルに追加し、問８－１と同様の推定方法で計算をやり直す。このとき、この No.9 は No.2 と同型船であるため、L/B, B/d, C_b, C_m, lcb の値は No.2 と完全に同一であるが、異なる水槽で実験したので流体抵抗値は若干異なっている。

海事統計学 平成27年度期末試験（2015年8月4日実施）解答

問題１（５点）

期待値：

$$1 \times p + (-1) \times (1-p) = 2p - 1$$

分散：

$$p(1-(2p-1))^2 + (1-p)(-1-(2p-1))^2 = 4p - 4p^2$$

問題２（5点）

期待値：

$$E\{\theta\} = \int_0^{2\pi} \theta \frac{1}{2\pi} d\theta = \pi$$

分散：

$$V\{\theta\} = \int_0^{2\pi} (\theta - \pi)^2 \frac{1}{2\pi} d\theta = \frac{\pi^2}{3}$$

問題３（5点）

$$F(x) = \int_{-\infty}^{x} f(z)dz = \begin{cases} \int_0^x \lambda e^{-\lambda z} dz & \text{, where } x \geq 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \left[-e^{-\lambda z}\right]_0^x & \text{, where } x \geq 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 - e^{-\lambda x} & \text{, where } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

問題４（5点）

期待値 μ，分散 σ^2 の正規分布に従う確率変数 x を標準化した確率変数

$$z = \frac{x-\mu}{\sigma}$$

は期待値 0、分散 1 の標準正規分布 $N(0,1)$ に従う。よって $7 \leq x \leq 10$ の範囲をとる確率は、$1 \leq z \leq 2$ の範囲をとる確率と等価である。標準正規分布表より、$0 \leq z \leq 2$ の範囲の面積は 0.4772、$0 \leq z \leq 1$ の範囲の面積は 0.3413 なので、$0.4772 - 0.3413 = 0.1359$ が答えとなる。

問題５（10点）

母比率の大標本区間推定：計算には標準正規分布表を利用
$\bar{x} = \frac{500}{2500} = \frac{1}{5}$ 正規分布表より $1-\alpha = 0.95$ となるような $z(\alpha/2) = 1.96$ と読みとれることから、母比率 p の 95 %信頼区間は、

$$\left[\bar{x} - 1.96 \times \sqrt{\frac{\bar{x}(1-\bar{x})}{2500}} \quad , \quad \bar{x} + 1.96 \times \sqrt{\frac{\bar{x}(1-\bar{x})}{2500}}\right] = [0.184, 0.216]$$

問題６

【問６–１】（5 点）
男女の出生比率が１：１であると仮定した場合の男児数の期待度数は 2 項分布に従うので、

男児数	0	1	2	3
家庭数	30	85	70	15
期待度数	25	75	75	25

【問６–２】（10 点）
これは 4 個のクラスによる適合度検定だから、帰無仮説（男女の出生比率が１：１）が正しければ以下の統計量は自由度 $4-1=3$ のカイ 2 乗分布に従うはずである：

$$\chi^2 = \frac{(30-25)^2}{25} + \frac{(85-75)^2}{75} + \frac{(70-75)^2}{75} + \frac{(15-25)^2}{25} = 5 + \frac{5}{3} = 6.667$$

自由度３のカイ 2 乗分布の５%棄却域（片側）は $\chi^2 > 7.81$ であるので、棄却域に入らず、帰無仮説は棄却できない。よって男女の出産比率が１：１であるという仮説に適合していないとはいえない。

問題７

【問７–１】（10 点）

$$\ln L(\lambda) = \ln f(x_1) + \ln f(x_2) + \cdots + \ln f(x_n) = \sum_{i=1}^{n} \ln \lambda \exp(-\lambda x_i)$$

【問７–２】（10 点）

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

導出は最尤推定の講義資料参照

【問７–３】（5 点）

$$\int_0^1 \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^1 = 1 - e^{-\lambda}$$

【問７–４】（10 点）
ポアソン分布は、単位時間あたりの事象の発生回数が平均で λ 回のとき $P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ より、単位時間あたり１回も事象が発生しない確率は $P(0) = e^{-\lambda}$ である。よって「単位時間あたり１回以上起きる」というのは「単位時間あたり１回も事象が発生しない」の余事象だから答えは $1 - e^{-\lambda}$ である。

問題８

【問８－１】（10 点）各船の特徴量 L/B, B/d, C_b, C_m, lcb を以下の変数（説明変数）x_1, x_2, x_3, x_4, x_5 で表し、推進抵抗 y を以下の多重回帰によって説明する：

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + e$$

ただし $b_0, b_1, b_2, b_3, b_4, b_5$ は回帰係数、e は確率変動（誤差）である。
ここで、表のデータを用いて、回帰係数 $b_0, b_1, \cdots, b_5$ を求める。行と列を間違えないよう注意しながら表のデータを行列に変換すると、以下のようになる：

$$\boldsymbol{y} = \begin{bmatrix} 2.80 \\ 3.03 \\ 3.16 \\ 3.11 \\ 3.39 \\ 4.26 \\ 3.13 \\ 4.11 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 5.94 & 2.49 & 0.777 & 0.991 & -2.40 \\ 1 & 7.59 & 2.37 & 0.819 & 0.994 & -2.60 \\ 1 & 6.02 & 2.47 & 0.771 & 0.993 & -2.28 \\ 1 & 6.02 & 2.44 & 0.795 & 0.994 & -2.75 \\ 1 & 6.06 & 2.72 & 0.800 & 0.996 & -3.10 \\ 1 & 6.40 & 2.60 & 0.821 & 0.996 & -3.01 \\ 1 & 5.80 & 2.29 & 0.781 & 0.990 & -2.35 \\ 1 & 5.50 & 2.88 & 0.811 & 0.998 & -3.48 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

【注意１】行列 X には、推定の対象である「計画中の船」の要目データを含めない

【注意２】行列 X の１列目の要素はすべて 1
と表すと、$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$ とした場合の誤差ベクトル $\boldsymbol{e}$ の平方和を最小にする回帰係数 $\hat{\boldsymbol{b}}$ は以下の式で与えられる：

$$\hat{\boldsymbol{b}} = \left(\boldsymbol{X}^{Trans}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{Trans}\boldsymbol{y}$$

推定の対象である「計画中の船」の要目 $L/B = 5.59$, $B/d = 2.78$, $C_b = 0.800$, $C_m = 0.995$, $lcb = -3.11$ を $x_1 = 5.59, x_2 = 2.78, x_3 = 0.800, x_4 = 0.995, x_5 = -3.11$ へ代入し、上で求めた回帰係数 $\hat{\boldsymbol{b}}$ を使って $y = \hat{b_0} + \hat{b_1} x_1 + \hat{b_2} x_2 + \hat{b_3} x_3 + \hat{b_4} x_4 + \hat{b_5} x_5$ より計画中の船の推進抵抗の推定値 y を得る。
【注意３】推定時には誤差 e の項をゼロで計算する。

【問８－２】（10 点）
説明変数 $x_1, x_2, \cdots x_k$ について、任意の定数 $a_0, a_1, \cdots a_k$ を用いて関係式 $a_0 + a_1 x_1 + a_2 x_2 + \cdots a_k x_k = 0$ に近い関係が成り立つとき、データは強い多重共線関係（多重共線性）があるという。このとき回帰係数ベクトルを求める逆行列計算が不安定になり、無意味な解が出やすい。
ここで、問題文のうち、明らかに 2 番目の方法は上記の関係式を生み出してしまい、回帰係数ベクトルを求める逆行列計算ができないという不都合が生じる。

1 番目の方法は、新しい説明変数を既存の説明変数から生成しているが、非線形関数になっているので多重共線性は生じず、問題は無い。
3 番目の方法は説明変数を減らすだけなので何ら問題は生じない。
4 番目の方法は、多重回帰は最小２乗法であり、データの重複は問題ではない。

本試験問題および解答に疑問がある場合は，8 月 21 日（金）午後５時までに，
W２号館６３４号室の木村まで申し出ること。

平成 27 年度 海事統計学 解答用紙　　学籍番号＿＿＿＿＿＿＿＿氏名＿＿＿＿＿＿＿＿＿＿

紙面が足りない場合は裏面も使用すること。