海事統計学 定期試験問題（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 $x=0$ となる確率変数 x の確率分布は、ベルヌーイ分布と呼ばれている。
このとき以下の問に答えよ。

【問１–１】（5 点）
上記のベルヌーイ分布に従う確率変数 x の期待値、および分散を p を用いて示せ。

【問１–２】（5 点）
ベルヌーイ分布に従って独立な試行を n 回行い、その結果を $x_1, x_2, \cdots, x_n$ と表す。
ここで、この確率変数を全て合計した新たな確率変数 $y = x_1 + x_2 + \cdots + x_n$ を定義すると、
y は二項分布 $B(n,p)$ に従うことが知られている。二項分布 $B(n,p)$ の期待値、および分散を p と n を用いて示せ。

【問１–３】（5 点）
二項分布 $B(n,p)$ の確率分布を式で表せ。

【問１–４】（5 点）
上記の確率変数 y の確率分布は、合計するベルヌーイ分布の個数 n を大きくすると正規分布へ近づく。
この現象について説明した法則または定理の名称を答えよ。

【問１–５】（10 点）
確率変数 y が二項分布 $B(10000, 0.5)$ に従うとき、y が 4900 以上 5050 以下の値をとる確率 $P(4900 \leq y \leq 5050)$ を求めよ。ただし二項分布 $B(n,p)$ の n が大きい場合に正規分布で近似できる性質を利用し、正規分布表から近似的に求めよ。

問題２

ある確率変数 x が以下の確率密度関数に従うとき、x の期待値および分散を求めよ。ただしパラメータ $\lambda > 0$ である。（5 点）

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (0 \leq x) \\ 0 & (x < 0) \end{cases}$$

問題３

あるメーカーのロープについて、破断にいたる荷重を調べたところ、64 回の試行における破断荷重の平均 4.8 トン，不偏分散 0.25 であった。破断荷重が正規分布に従うと仮定すると、平均値 μ のとりうる値を信頼係数 99 %で区間推定せよ。分布表からの読み取りの有効数字は２桁でよい。（10 点）

海事統計学 定期試験問題（２／３）

問題４

以下の表は、ある類似船型の貨物船 12 隻において 1 年間にバルブの故障が発生した件数を表している。
このとき以下の問いに答えよ。

船の番号	1	2	3	4	5	6	7	8	9	10	11	12
年間故障発生件数	0	1	0	0	4	1	0	2	2	0	1	1

【問４–１】（5 点）
１隻あたり１年間で発生するバルブ故障の平均件数を計算せよ。

【問４–２】（5 点）
年間に発生する故障の発生件数 x の確率 $P(x)$ をポアソン分布 $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ のモデルで説明したい。
船の番号 i における年間故障発生件数を x_i と表す。このときポアソン分布によるモデルの対数尤度を x_i を用いて式で示せ。

【問４–３】（5 点）
問４–２の対数尤度を最大化する λ はどのような値か？　x_i を用いて式で示せ。

【問４–４】（5 点）
表に示されたデータを最も尤もらしく説明するポアソン分布モデルの λ はどのような値か？
表のデータから計算して示せ。

【問４–５】（10 点）
表に示されたデータがポアソン分布に従っているといえるかどうか検定したい。どのような検定方法が考えられるか検討せよ。また検定方法が使えないなら、その理由を述べよ。

【問４–６】（5 点）
ポアソン分布と指数分布の関係について説明せよ。

海事統計学 定期試験問題（３／３）

問題５

以下は 1980 年代の自動車の要目と燃費についてのデータである。

車種＼要目	x_1	x_2	x_3	x_4	x_5	x_6	10 モード走行 (km/L)
クラウン	1.360	4.778	2.4251	125	17.5	8.8	8.7
マーク II	1.245	4.100	2.4082	125	17.5	8.8	9.5
カムリ	1.070	3.214	2.3575	120	17.6	8.7	10.6
ソアラ	1.235	4.100	2.3052	125	17.5	8.8	9.2
セドリック	1.420	4.625	2.4251	130	17.5	9.5	8.9
ローレル	1.175	3.889	2.3660	125	17.0	9.1	9.2
スカイライン	1.175	4.111	2.3198	125	17.0	9.1	9.2
レパード	1.220	3.900	2.2899	125	17.0	9.1	9.4
カペラ	1.030	3.450	2.3829	120	17.0	8.6	10.2
ギャラン Σ	1.180	3.665	2.3645	110	16.7	8.5	10.6
ルーチェ	1.150	3.909	2.3829	120	17.0	8.6	?

ただし要目 x_i については、x_1:車体重量 (トン)，x_2：減速比，x_3：幅×高さ (m^2)，x_4：最大出力 (ps)，x_5：最大トルク (kgm)，x_6：圧縮比

【問５－１】（15 点）上記のデータのうち、ルーチェの 10 モード走行の値が不明であるので、他の車のデータを利用し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、ルーチェの燃費推定の計算手順を説明せよ。

【問５－２】（5 点）多重回帰による推定において問題となる多重共線性とは何か、どのようなデータを扱うと生じるかについて説明せよ。

海事統計学 平成 29 年度定期試験問題（2017 年 6 月 6 日実施）解答

問題１解答

【問１–１】（5 点）　期待値 p, 分散 $p(1-p)$

【問１–２】（5 点）　期待値 np, 分散 $np(1-p)$

【問１–３】（5 点）

$$P(x) = B(n,p) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

【問１–４】（5 点）　中心値極限定理

【問１–５】（10 点）二項分布 $B(10000, 0.5)$ の期待値 $np = 5000$, 分散 $np(1-p) = 2500$ より、$z = \frac{x-5000}{\sqrt{2500}}$ は標準正規分布 $N(0,1)$ に従う。よって、$\frac{4900-5000}{\sqrt{2500}} \leq z \leq \frac{5050-5000}{\sqrt{2500}}$ となる標準正規分布 $N(0,1)$ に従う z の確率を正規分布表から読み取ると、$-2 \leq z < 0$ の面積（確率）は 0.4772, $0 \leq z \leq 1$ の面積（確率）は 0.3413 より、求める確率は $0.4772 + 0.3413 = 0.8185$

問題２ 解答

【指数分布の期待値と分散】（5 点）　指数分布の期待値は $1/\lambda$、分散は $1/\lambda^2$（導出は講義資料を参照）

問題３ 解答

【正規分布の区間推定】（10 点）　データ数 $n = 64$ より自由度 63 の t 分布表から面積が 0.99 となる区間を読み取らなければならない。しかしデータ数が 40 以上であれば、t 分布は正規分布で近似できるので、正規分布表から読み取ると、2.57〜2.58 である。t 分布表の自由度 60 で 2.66, 自由度 120 で 2.62 なので、おおよそ 2.6 とすると、信頼区間は $4.8 \pm 2.6(s/\sqrt{n})$ すなわち 4.8 ± 0.1625 となる。

問題４解答

【問４–１】（5 点）　$(0+1+0+0+4+1+0+2+2+0+1+1)/12 = 1$　よって １件／年

【問４–２】（5 点）

$$\sum_{i=1}^{12} \ln P(x_i) = \sum_{i=1}^{12} \ln \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

【問４–３】（5 点）

$$\frac{\partial}{\partial \lambda} \sum_{i=1}^{12} \ln \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = 0 \quad \text{この方程式を } \lambda \text{ について解くと} \quad \lambda = \frac{\sum_{i=1}^{12} x_i}{12}$$

【問４–４】（5 点）　問４－３より、標本平均に等しいので　$\lambda = 1$

【問４–５】（10 点）　正規分布でも比率の分布でもないので、それらの平均値や分散を検定する方法は使えない。
カイ２乗分布を使ったあてはまりの検定は、期待度数を５以上にしなければならないが、それができないので使えない。
また K-S 検定はデータ数が 20 より大きくないと使えない。よって講義で扱った検定方法はどれも適用できない。

【問４−６】（5 点）　ポアソン分布における発生時間間隔は、指数分布に従う。事象の発生時間間隔を x としたとき、指数分布の確率密度関数は $f(x) = \lambda e^{-\lambda x}$ で表される。このときの指数分布のパラメータ λ の値はポアソン分布のパラメータ λ と同一になる。

問題５解答

【問５−１】（15 点）　10 モード走行 (km/L) を y とし、各要目の値 $x_1, x_2, x_3, x_4, x_5, x_6$ を変数（説明変数）として以下の多重回帰によって説明する：

$$y \;=\; b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_6 + e$$

ただし $b_0, b_1, b_2, b_3, b_4, b_5, b_6$ 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 $b_0, b_1, \cdots, b_6$ 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：

$$\boldsymbol{y} = \begin{bmatrix} 8.7 \\ 9.5 \\ 10.6 \\ 9.2 \\ 8.9 \\ 9.2 \\ 9.2 \\ 9.4 \\ 10.2 \\ 10.6 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1.360 & 4.778 & 2.4251 & 125 & 17.5 & 8.8 \\ 1 & 1.245 & 4.100 & 2.4082 & 125 & 17.5 & 8.8 \\ 1 & 1.070 & 3.214 & 2.3575 & 120 & 17.6 & 8.7 \\ 1 & 1.235 & 4.100 & 2.3052 & 125 & 17.5 & 8.8 \\ 1 & 1.420 & 4.625 & 2.4251 & 130 & 17.5 & 9.5 \\ 1 & 1.175 & 3.889 & 2.3660 & 125 & 17.0 & 9.1 \\ 1 & 1.175 & 4.111 & 2.3198 & 125 & 17.0 & 9.1 \\ 1 & 1.220 & 3.900 & 2.2899 & 125 & 17.0 & 9.1 \\ 1 & 1.030 & 3.450 & 2.3829 & 120 & 17.0 & 8.6 \\ 1 & 1.180 & 3.665 & 2.3645 & 110 & 16.7 & 8.5 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$

【注意１】行列にルーチェのデータを含めないこと
【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：

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

ルーチェの要目を $x_1 = 1.150, x_2 = 3.909, x_3 = 2.3829, x_4 = 120, x_5 = 17.0, x_6 = 8.6$ へ代入し、上で求めた回帰係数 $\hat{\boldsymbol{b}}$ を使って $y \;=\; \hat{b_0} + \hat{b_1} x_1 + \hat{b_2} x_2 + \hat{b_3} x_3 + \hat{b_4} x_4 + \hat{b_5} x_5 + \hat{b_6} x_6$ よりルーチェの 10 モード走行 (km/L) の推定値 $\hat{y}$ を得る。
【注意３】この推定時の式には誤差 e は入れないこと

【問５−２】（5 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。

本試験問題および解答に疑問がある場合は，6 月 13 日（火）午後 4 時までに，
W２号館６３４号室の木村まで申し出ること。

2017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。