H25 海事統計学 定期試験問題（１／３）

以下、必要ならば $\sqrt{2} = 1.41$, $\sqrt{3} = 1.73$, $\sqrt{5} = 2.24$, $\sqrt{10} = 3.16$ として計算せよ。
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)

H25 海事統計学 定期試験問題（２／３）

問題３

スクラップタイヤを貨物として積載した貨物船があり、その貨物であるタイヤは大・中・小の 3 種類の大きさに分類される。「貨物についての説明」の記述によると、全体の数の $\frac{1}{4}$ が小タイヤ、$\frac{1}{2}$ が中タイヤ、残りの $\frac{1}{4}$ が大タイヤとなっている。貨物の積込み時においてタイヤ 40 本をサンプル調査したところ、以下の表のような結果となった。
この調査結果は、上記の「貨物についての説明」がウソであることを指摘する証拠として十分か？有意水準 0.05 で検定せよ。(15 点)

タイヤの大きさ	大	中	小	合計
本数	11	15	14	40

問題４

故障や事故などの偶発的な事象が単位時間あたりに発生する回数は、ポアソン分布になることが知られている。ポアソン分布では、単位時間に事象が x 回発生する確率 $P(x)$ は、以下の確率で与えられる：

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{1}$$

ただし e は自然対数の底、λ はパラメータを表す。このとき以下の問に答えよ。

【問４–１】(5 点)
単位時間あたりに観測された事象の発生回数の標本を $x_1, x_2, \cdots, x_n$ と表す。
このとき、事象が単位時間あたりに x 回発生する確率を式 (1) で表す場合、標本の対数ユウ度はどのように表されるか $x_1, x_2, \cdots, x_n$ を用いて式を示せ。

【問４–２】(10 点)
単位時間あたりに観測された事象の発生回数の標本が【問４–１】のように与えられる場合、式 (1) 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４–３】(10 点)
ポアソン分布と指数分布の関係について説明せよ。

問題５

以下はある造船所で建造されたタンカーの主要目とその工事に伴う溶接長の実績である。

船の番号	L	B	D	Cb	延べ溶接長
No.1	168	32.2	17.0	0.80	11.3
No.2	192	32.2	15.2	0.79	12.4
No.3	116	20.0	13.7	0.78	3.32
No.4	175	23.1	15.2	0.82	11.7
No.5	137	23.1	14.5	0.78	8.72
No.6	172	32.2	18.1	0.80	12.3
No.7	97	15.5	10.5	0.70	2.1

【問５–１】（10 点）新しい工事予定の船の要目 L, B, D, Cb が与えられたとき、上記のデータを利用して、この工事予定の船における延べ溶接長を多重回帰を用いて推定したい。
<u>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。</u>（数値計算はしなくて良い）

【問５–２】（10 点）問５–１の推定をより精度良く行うため、以下のような項目 1,2,3,4 のような工夫を検討した。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, D, Cb に加えて、船殻の体積を反映する新しい特徴量 $x_5 = L \cdot B \cdot D \cdot C_b$ を導入し、問５–１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, D, Cb に加えて、船殻の長さを反映する新しい特徴量 $x_5 = L + B + D$ を導入し、問５–１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、Cb 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５–１と同様の推定方法で計算をやり直す。
4. 新しく番号 No.8 についてのデータをテーブルに追加し、問５–１と同様の推定方法で計算をやり直す。このとき、この No.8 は No.6 と同型船であるため、L, B, D, Cb の値は No.6 と完全に同一であるが、延べ溶接長は若干異なっている。

海事統計学 平成25年度定期試験問題（2013年7月29日実施）解答

問題１

【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、$\frac{4900-5000}{\sqrt{2500}} \le z \le \frac{5050-5000}{\sqrt{2500}}$ 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、$-2 \le z < 0$ の面積（確率）は 0.4772, $0 \le z \le 1$ 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

正規分布の期待値の区間推定なので、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。（10点）

問題３

タイヤの大きさ（クラス）k	大（$k=1$）	中（$k=2$）	小（$k=3$）	合計
本数（観測度数）f_k	11	15	14	40
期待度数 f_k^*	10	20	10	40

帰無仮説を「大・中・小のタイヤ個数比率が $\frac{1}{4} : \frac{1}{2} : \frac{1}{4}$ の度数分布に従う」として度数分布の適合度検定を行う。

$$\chi^2 = \sum_{k=1}^{3} \frac{(f_k - f_k^*)^2}{f_k^*} = \frac{(11-10)^2}{10} + \frac{(15-20)^2}{20} + \frac{(14-10)^2}{10} = 2.95$$

カイ2乗分布表より、自由度 $3-1=2$ で有意水準 0.05 の片側検定における棄却域を読み取ると、棄却域は $\chi_2^2(0.05) = 5.99$ より大きい領域である。ここで上記の $2.95 < 5.99$ で観測データは棄却域に入らないので帰無仮説は棄却できない。よって「貨物についての説明」がウソであることを指摘する証拠としては不十分である。■（15点）

問題４

【問4－1】（5点）

$$\ln L(\lambda) = \ln P(x_1) + \ln P(x_2) + \cdots + \ln P(x_n) = \sum_{i=1}^{n} \ln \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

【問4－2】（10 点）対数尤度が最大になる λ を求めるため、対数尤度をパラメータ λ で微分した式＝０の方程式を解く

$$\frac{\partial}{\partial\lambda}\sum_{i=1}^{n}\ln\frac{e^{-\lambda}\lambda_i^x}{x_i!}=\sum_{i=1}^{n}\frac{1}{\frac{e^{-\lambda}\lambda_i^x}{x_i!}}\frac{-e^{-\lambda}\lambda_i^x+e^{-\lambda}x_i\lambda_i^x}{x_i!}=\sum_{i=1}^{n}-1+\frac{x_i}{\lambda}=0$$

この式を λ について解くと

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

【問4－3】（10 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

【問５－１】（10 点）各船の特徴量 L, B, D, Cb を以下の変数（説明変数）x_1,x_2,x_3,x_4 で表し、延べ溶接長 y を以下の多重回帰によって説明する：

$$y\ =\ b_0+b_1x_1+b_2x_2+b_3x_3+b_4x_4+e$$

ただし b_0,b_1,b_2,b_3,b_4 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 $b_0,b_1,\cdots,b_4$ を求める。ここで、

$$\boldsymbol{y}=\begin{bmatrix}11.3\\12.4\\3.32\\11.7\\8.72\\12.3\\2.1\end{bmatrix}\quad \boldsymbol{X}=\begin{bmatrix}1&168&32.2&17.0&0.80\\1&192&32.2&15.2&0.79\\1&116&20.0&13.7&0.78\\1&175&23.1&15.2&0.82\\1&137&23.1&14.5&0.78\\1&172&32.2&18.1&0.80\\1&97&15.5&10.5&0.70\end{bmatrix}\quad \boldsymbol{b}=\begin{bmatrix}b_0\\b_1\\b_2\\b_3\\b_4\end{bmatrix}\quad \boldsymbol{e}=\begin{bmatrix}e_1\\e_2\\e_3\\e_4\\e_5\\e_6\\e_7\end{bmatrix}$$

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：

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

新しい工事予定の船の要目 L, B, D, Cb を $x_1=L,x_2=B,x_3=D,x_4=C_b$ へ代入し、上で求めた回帰係数 $\hat{\boldsymbol{b}}$ を使って $y\ =\ \hat{b_0}+\hat{b_1}x_1+\hat{b_2}x_2+\hat{b_3}x_3+\hat{b_4}x_4$ より新しい工事予定の船の延べ溶接長の推定値 y を得る。

【問５－２】（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 $a_0+a_1x_1+a_2x_2+\cdots a_kx_k=0$ 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２番目の方法は明らかに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の数を減らすだけなので何ら問題は生じない。４番目の方法におけるデータの重複については、多重回帰は最小２乗法であるため、問題は無い。

本試験問題および解答に疑問がある場合は，８月 26 日（月）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25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。