2010 年度後期　情報統計学　第１０回

総合演習　問題

※過去の試験問題から抜粋しました．

１．さいころを１回投げたとき，ある１つの目が出る確率は 1/6 であるといわれている．では，未使用のさいころが１つあるとする．このさいころを１回投げたとき１の目が出る確率が，1/6 であることを証明できるか．できるならばそのやり方を記せ．できないならば，なぜできないのかを説明せよ．

２．ある店には，１時間あたりにやってくる男性客の数の期待値は２人，女性客のそれは１人であるという．各々の客は独立に一定の確率で店を訪れているとするとき，１時間の間に男女を問わず客がちょうど２人やってくる確率を求めよ．

３．「航空機事故は続けて起こりやすい」と，一般によく言われる．

1. ある瞬間に地球上のどこかで航空機事故が起きる危険の度合いは，常に一定であると仮定する．このとき，ある時刻から次に航空機事故が起きるまでの経過時間は，どのような確率分布にしたがうと考えられるか．この講義で説明した知識にもとづいて論じよ．
2. 上の仮定が正しいとするとき，「航空機事故は続けて起こりやすい」という言葉が妥当かどうかについて論じよ．

４．第２次世界大戦中，ドイツ軍では，暗号をイギリス軍に解読されるのを防ぐため，通信を暗号化する鍵となるコード（いくつかの数字の列）を毎日変更していた．その際，可能な数字の組み合わせの中からランダムに数字を選び，さらにその月に一度使ったコードは月内には２度と使わないことにしていた．この方法は，コードをランダムにして暗号を見破られにくくする方法として正しいか．

５．ある区間推定の問題で，母平均 μ の 95%信頼区間を計算すると，$[40, 50]$ であった．この答を，「$\Pr(40 \leq \mu \leq 50) = 0.95$」と書くと，それは誤りである．それはなぜか．
（Pr( ) は，「カッコ内の事象が起きる確率」を意味する）

６．ある機械には２つの重要な部品 A, B があって，２つの部品の両方が正常でないと，この機械は正常に稼働しないという．また，部品 A, B は，どの時刻においても常に，その時刻から単位時間の間に，一定の確率で独立に故障するという．新品の部品 A の故障までの時間の期待値は２ヶ月，B のそれは１ヶ月である．この機械の A, B 以外の部品は故障しないとするとき，新品の機械が稼働し始めてから１ヶ月後に，この機械が正常に稼働している確率はいくらか．

解答例

１．証明することはできない．
さいころを１回投げたとき，ある１つの目が出る確率が 1/6 であるといわれているのは，さいころのどの目も同じ確率で出る，と仮定できるときに言えることである．しかし，さいころのある目が出る確率とは，これから未来にわたって十分多くの回数さいころをふったときに，その目が出る回数の割合であるから，実験によって調べることはできず，したがって上の仮定も実験によって確かめることはできない．

２．各々の客は独立に一定の確率で店を訪れるので，１時間あたりにやって来る客の数は男女それぞれポアソン分布にしたがうと考えられる．ポアソン分布にしたがう確率変数を X とすると，その確率分布は

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

と表され，このとき X の期待値は λ となる．よって，１時間あたりにやってくる男性客の数を M，女性客の数を F とすると，各々の期待値が 2 および 1 であるから，各々の確率分布は

$$P(M = m) = \frac{e^{-2}2^{m}}{m!}, P(F = f) = \frac{e^{-1}1^{f}}{f!}, \tag{2}$$

で表される．

さて，「男女を問わず客がちょうど２人やってくる」という事象は，「男性客が２人やって来て，かつ女性客が０人やってくる」「男性客が１人かつ女性客が１人」「男性客が０人かつ女性客が２人」のいずれかである．これらの事象は排反で，また男性客や女性客は独立にやって来るので，求める確率は

$$\begin{aligned}
&「男２女０」 \quad \rightarrow \quad P(M = 2) \times P(F = 0) = \frac{e^{-2}2^{2}}{2!} \times \frac{e^{-1}1^{0}}{0!} = 0.0996 \\
&「男１女１」 \quad \rightarrow \quad P(M = 1) \times P(F = 1) = \frac{e^{-2}2^{1}}{1!} \times \frac{e^{-1}1^{1}}{1!} = 0.0996 \\
&「男０女２」 \quad \rightarrow \quad P(M = 0) \times P(F = 2) = \frac{e^{-2}2^{0}}{0!} \times \frac{e^{-1}1^{2}}{2!} = 0.0249
\end{aligned}$$

の和で，0.224 となる．

注：　この問題では男女の客は独立に店にやって来るので，男女を問わず１時間あたりにやって来る客の数もポアソン分布にしたがう．また，２つの確率変数の和の期待値はそれぞれの期待値の和になるので，その期待値は３人となる．これからただちに上と同じ解が得られる．

３．

1. 「ある瞬間に事故が起きる危険の度合い」は，講義第７回のハザード関数に相当する．問題の条件では，ハザード関数は時刻によらず一定という条件なので，ある時刻から次に事故が起きるまでの待ち時間は，指数分布にしたがう．
2. 生存関数のある時刻での値は，「その時刻までに事故が起きていない確率」と考えることができる．指数分布では，その関数は時刻を t とするとき $e^{-\lambda t}$ $(\lambda > 0)$ であるから，時間が経つにつれ急激に減少する．すなわち，次の事故が起きるまでの時間は，短いほどその可能性は高く，「事故は続け

ておきやすい」というのは間違ってはいない．

（東京大学教養学部統計学教室編「基礎統計学 I　統計学入門」125 ページの記述によると，実際に過去の「事故間隔」を横軸，その頻度を縦軸にグラフにプロットしてみると，確かに指数分布が当てはまっているそうである）

４．正しくない．その月に一度使ったコードを月内には２度と使わないことにすると，その時点で，そのコードが選ばれる確率は 0 となり，他の各コードが選ばれる確率が少しずつ増えることになるので，暗号を見破るヒントを与えてしまう．無作為抽出にするためには，すでに一度使ったコードも，次の抽出の機会には平等なチャンスで選ばれるようにしなければならない．

（ドイツ軍は，実際にこの誤りを犯し，そのためにイギリス軍に暗号を見破りやすくさせてしまった，といわれている）

５．母平均 μ は，人が知らないだけで，実際にはひとつに決まっている．だから，「40 以上 50 以下」という具体的区間を求めた段階で，母平均が「40 以上 50 以下」であるかどうかはすでに決まっているのであり，μ は確率変数ではない．だから，「40 以上 50 以下である確率」には意味がない．「95% 信頼区間」とは，この方法で信頼区間を何度も求めると，そのうち 95% は母平均を本当に含んでいる区間である，という意味である．

６．問題の条件から，各部品が故障するまでの時間は指数分布にしたがう．時間を t（月），部品が故障するまでの時間を T とすると，

$$P(T \leq t) = 1 - e^{-\lambda t} \tag{3}$$

であり，その期待値 $E(T)$ は $1/\lambda$ である．部品 A,B が故障するまでの時間をそれぞれ T_A, T_B とすると，部品 A については $E(T_A) = 2$ であるから $\lambda = 1/2$ で，$P(T_A \leq 1) = 1 - e^{-1/2}$ となり，$P(T_A \geq 1) = e^{-1/2}$ である．一方，部品 B については $E(T_B) = 1$ であるから $\lambda = 1$ で，$P(T_B \leq 1) = 1 - e^{-1}$ となり，$P(T_B \geq 1) = e^{-1}$ である．

問題の確率は，「１ヶ月後に A も B も故障していない確率」であり，A,B はそれぞれ独立に故障するから，$P(T_A \geq 1)P(T_B \geq 1)$ で表される．この値は $e^{-3/2} = 0.223$ である．