昭和57年度（問　題）

1. 学生1,600人のうち，自宅通学者は755人，非自宅通学者は845人であった。 そのうちでアルバイトをするものは，自宅通学者425人，非自宅通学者405人であった。アルバイトをする学生の比率は自宅通学者と非自宅通学者とで差があると考えてよいか。ここに，$\chi_1^2(0.01)=6.63$，$\chi_2^2(0.01)=9.21$，$\chi_3^2(0.01)=11.34$，$\chi_4^2(0.01)=13.28$ である。

2. A，B 2グループの生徒に同一テストを行った結果，つぎのような結果を得た。

グループ	受験者数	平　均	分　散
A	$n_x=20$	$\overline{x}=72$	$s_x^2=83$
B	$n_y=25$	$\overline{y}=65$	$s_y^2=105$

A，B両グループの平均値間に差があるといえるか，有意水準0.05で検定せよ。

（注） $F_{19}^{24}(0.025)=2.45$，$t_{43}(0.025)=2.02$

3. 支持率の予想が全くつかないA候補につき，その支持率を信頼度99%で誤差が1%以内であるように推定したい。およそどのくらいの数の標本を抽出すればよいか。

また，A候補の支持率が約30%と予想される場合はどうか。

（注） $u(0.005)$　2.58

4. 同じ種類の新しい粒子を生み出すような粒子を考える。最初，このような粒子がただ1つだけあるとする。これを原始世代，すなわち第0世代とする。第n世代の粒子は互いに独立に第$(n+1)$世代を作るものとし，また，どの粒子もちょうどk個の新しい粒子を生み出す確率はp_k $(k=0, 1, 2;\cdots\cdots)$ であるとする。第n世代の粒子の数を確率変数X_nで表わし，X_1の母関数を$P(s)=\sum_{k=0}^{\infty} p_k s^k$とするとき，

(1) X_nの母関数$P_n(s)$は，次々と

$$P_1(s)=P(s),\ P_n(s)=P(P_{n-1}(s))$$

で求められることを示せ。

(2) $\mu=\sum_{k=0} kp_k$ を1つの粒子の生む次世代の粒子の数の期待値とするとき，X_nの期待値を求めよ。

5. 母集団分布が指数分布$f(x;\mu)=\frac{1}{\mu}e^{-\frac{x}{\mu}}$ $(x>0)$にしたがうとき，大きさnの標本$(x_1, x_2, \cdots\cdots, x_n)$により最尤推定量$\hat{\mu}$を求め，その不偏性，有効性，充足性，一致性を調べよ。

昭和57年度（解答例）

1. 1,600人の学生を，自宅通学かどうかとアルバイトをしているかどうかという，2種類の属性A，Bにより2×2分割表を作ってみると右表のようになる。

通学＼アルバイト	する	しない	合　計
自　宅	425人	330人	755人
非自宅	405	440	845
合　計	830	770	1,600

ここで，自宅通学者と非自宅通学者の間でアルバイトをする比率に差があるかどうかは，次の仮説 H_0 を検定すれば分る。

H_0 ：2つの属性AとBは独立である。

従って，分割表による独立性の検定を行うのだが，組度数が十分大きいので，自由度 $(2-1)\times(2-1)=1$ の χ^2 分布を用いて検定できる。そこで，2×2分割表の簡便計算法を用いると，

$$\chi^2 = \frac{(425\times440-330\times405)^2\times1600}{755\times845\times830\times770}$$

$$= 11.169 > \chi_1^2(0.01) = 6.63$$

であり，H_0 が棄却される。よって，有意水準1％で，アルバイトをする比率は自宅通学者と非自宅通学者の間に差があると言える。

〔1の別解〕

この問題を，百分率の差の検定問題として解いてみる。

自宅通学者のアルバイト比率を p_1

非自宅通学者のアルバイト比率を p_2

とおくと，それらの実現値 $\hat{p}_1$，$\hat{p}_2$ は，それぞれ $\hat{p}_1=\dfrac{425}{755}$，$\hat{p}_2=\dfrac{405}{845}$ である。

ここで，検定する仮説は H_0：$p_1=p_2$ であり，$n_1=755$，$n_2=845$ とおくとき，

$$|u| = \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} > u\left(\frac{\varepsilon}{2}\right)$$ のとき H_0 を棄却し，

$|u| < u\left(\dfrac{\varepsilon}{2}\right)$ のとき H_0 を採択すればよい。

そこで $|u|$ を計算すると，

$$|u| = \frac{0.5629-0.4793}{\sqrt{0.0003258+0.0002953}}$$

$$\fallingdotseq 3.355 > u\left(\frac{0.01}{2}\right) = 2.58 \qquad \left(\begin{array}{l} u(0.005)\text{の値は3番の} \\ \text{問題で与えられている。} \end{array}\right)$$

であるので，H_0 は有意水準１％で棄却される。

2. まず等分散性の検定を行う。ここで，A，B両グループの分散を σ_x^2 ，σ_y^2 とするとき，帰無仮説は H_0：$\sigma_x^2 = \sigma_y^2$ ，対立仮説は H_1：$\sigma_x^2 \fallingdotseq \sigma_y^2$ である。不偏分散比を求めると，

$$F_0 = \frac{n_y s_y^2 / (n_y-1)}{n_x s_x^2 / (n_x-1)} = \frac{25\times105 / 24}{20\times83 / 19} = 1.252$$

である。ここで，$F_{n_x-1}^{n_y-1}(0.025) = F_{19}^{24}(0.025) = 2.45$ であるから，

$$F_0 < F_{19}^{24}(0.025)$$

となり，H_0 は採択される。

そこで，$\sigma^2 = \sigma_x^2 = \sigma_y^2$ の推定値は，

$$\hat{\sigma}^2 = \frac{1}{n_x+n_y-2}(n_x s_x^2 + n_y s_y^2) = \frac{1}{43}(20\times83+25\times105) = 99.7$$

とおける。

次に，平均値の差について t 検定を行う。A，B両グループの平均値を μ_x，μ_y とするとき，帰無仮説は H_0：$\mu_x = \mu_y$ ，対立仮説は H_1：$\mu_x \fallingdotseq \mu_y$ である。統計量 t を計算すると，

$$t = \frac{|\overline{x}-\overline{y}|}{\sqrt{\dfrac{n_x s_x^2 + n_y s_y^2}{n_x+n_y-2}\left(\dfrac{1}{n_x}+\dfrac{1}{n_y}\right)}} = \frac{72-65}{\hat{\sigma}\sqrt{\dfrac{1}{n_x}+\dfrac{1}{n_y}}} = 2.337$$

であり，$t_{n_x+n_y-2}(0.025) = t_{43}(0.025) = 2.02$ より，

$$t > t_{43}(0.025)$$

であるので，有意水準５％で，A，B両グループの平均値に差があると言える。

3.　A候補の支持率を p とする。

大きさ n の標本をとった場合の標本支持率を $\hat{p}$ とすると，n は十分大きいと考えられるので，$\hat{p}$ は近似的に正規分布 $N\left(p, \dfrac{p(1-p)}{n}\right)$ に従う。よって，統計量

$$U = \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

は，標準正規分布 $N(0,1)$ に従う。

$$\therefore\ P_r\left\{-u\left(\frac{\varepsilon}{2}\right) < \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} < u\left(\frac{\varepsilon}{2}\right)\right\} = 1-\varepsilon$$

ここで，$\varepsilon = 0.01$　とすると，$u\left(\dfrac{0.01}{2}\right) = 2.58$　であるから，

$$P_r\left\{-2.58 < \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} < 2.58\right\}$$

$$= P_r\left\{\hat{p}-2.58\sqrt{\frac{p(1-p)}{n}} < p < \hat{p}+2.58\sqrt{\frac{p(1-p)}{n}}\right\} = 0.99$$

従って，信頼度99％で，

$$\hat{p}-2.58\sqrt{\frac{p(1-p)}{n}} < p < \hat{p}+2.58\sqrt{\frac{p(1-p)}{n}}$$

である。ここで，簡便法を用い上式の両辺の p を推定値 $\hat{p}$ でおきかえると，

$$\hat{p}-2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p}+2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

となる。そこで，誤差を１％以内とするのであるから，

$$2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leqq 0.01$$

でなくてはならない。つまり，

$$n \geqq 258^2\,\hat{p}(1-\hat{p}) \qquad \cdots\cdots\cdots\cdots\cdots(1)$$

である。

a．支持率の予想が全くつかない場合

$\hat{p}(1-\hat{p})$ は $\hat{p}=\dfrac{1}{2}$ で最大値をとり，そのときの(1)式の右辺は，

$$258^2\times0.5\times0.5 = 16641$$

よって，約16,600以上の標本をとればよい。

b．支持率が0.30と予想される場合

(1)式の右辺に $\hat{p}=0.30$ を代入して

$$258^2\times0.30\times0.70 = 13978.44$$

となるので，約14,000の標本をとればよい。

4. (1) 確率変数 X で，ある粒子が生み出す粒子の個数を表わすものとすると，題意により $P_r(X=k)=p_k\ (k=0,1,2,\cdots\cdots)$ である。また，確率変数 $Y_n\ (n=0,1,2,\cdots\cdots)$ を次のように定義する。

$$Y_n = \underbrace{X+X+\cdots\cdots+X}_{n\text{ 個}} \qquad \text{ただし } Y_0=0 \text{ とする。}$$

まず，$P_1(s)$ を求める。母関数の定義により，

$$P_1(s)=\sum_{k=0}^{\infty}P_r(X_1=k)s^k$$

である。ここで，X_1 は第0世代（粒子1個）が生み出す第1世代の粒子数を表わしているから，その粒子を生み出す確率分布は X と同じである。

よって，

$$P_1(s)=\sum_{k=0}^{\infty}P_r(X=k)s^k=\sum_{k=0}^{\infty}p_k s^k=P(s)$$

次に，$P_2(s)$ を求める。定義により，

$$P_2(s)=\sum_{k=0}^{\infty}P_r(X_2=k)s^k$$

であるが，ここで，

$$P_r(X_2 = k) = \sum_{j=0}^{\infty} P_r(X_2 = k, X_1 = j)$$

$$= \sum_{j=0}^{\infty} P_r(X_2 = k) P_r(X_1 = j) \quad (\because\ X_1 \text{と} X_2 \text{は独立})$$

$$= \sum_{j=0}^{\infty} P_r(Y_j = k) P_r(X = j)$$

$$\therefore\ P_2(s) = \sum_{k=0}^{\infty} \left\{ \sum_{j=0}^{\infty} P_r(Y_j = k) P_r(X = j) \right\} s^k$$

$$= \sum_{j=0}^{\infty} \left\{ \sum_{k=0}^{\infty} P_r(Y_j = k) s^k \right\} P_r(X = j)$$

$$= \sum_{j=0}^{\infty} E(s^{Y_j}) P_r(X = j)$$

$$= \sum_{j=0}^{\infty} \left\{ \underbrace{E(s^X) E(s^X) \cdots E(s^X)}_{j\text{個}} \right\} P_r(X = j) \quad (\because\ \text{各} X \text{は独立})$$

$$= \sum_{j=0}^{\infty} \left\{ E(s^X) \right\}^j P_r(X = j)$$

$$= \sum_{j=0}^{\infty} \left\{ P(s) \right\}^j P_r(X = j) = P(P(s)) = P(P_1(s))$$

となり，$P_2(s) = P(P_1(s))$ が言えた。

次に，$k = n$ まで $P_k(s) = P(P_{k-1}(s))$ が成り立つとして，$k = n+1$ の場合を考える。$P_2(s)$を求めた時と同じ考えで，

$$P_{n+1}(s) = \sum_{k=0}^{\infty} P_r(X_{n+1} = k) s^k$$

$$= \sum_{k=0}^{\infty} \left\{ \sum_{j=0}^{\infty} P_r(Y_j = k) P_r(X_n = j) \right\} s^k$$

$$= \sum_{j=0}^{\infty} \left\{ \sum_{k=0}^{\infty} P_r(Y_j = k) s^k \right\} P_r(X_n = j)$$

$$= \sum_{j=0}^{\infty} \left\{ P(s) \right\}^j P_r(X_n = j)$$

$$= P_n(P(s)) = P(P_{n-1}(P(s))) = \cdots = \underbrace{P(P(\cdots(P(}_{n\text{ 個}}P(s))\cdots)$$

$$= \underbrace{P(P(\cdots(P(}_{n\text{ 個}}P_1(s))\cdots) = \underbrace{P(P(\cdots(P(}_{n-1\text{個}}P_2(s))\cdots) = \cdots\cdots$$

$$= P(P_n(s))$$

となり，帰納的に漸化式が示された。

(2) 母関数の性質を用いて

$$E(X_n) = P_n'(1) = P'(P_{n-1}(1))P_{n-1}'(1)$$
$$= P'(1)P_{n-1}'(1)$$

であるが，$P'(1) = \sum_{k=0}^{\infty} kp_k = \mu$ であるから，

$$E(X_n) = \mu \cdot E(X_{n-1})$$

がわかる。よって帰納的に，

$$E(X_n) = \mu^n$$

である。

5. 尤度関数 $l(\mu)$ を求めると，

$$l(\mu) = \prod_{i=1}^{n} f(x_i; \mu) = \prod_{i=1}^{n}\left(\frac{1}{\mu}e^{-\frac{x_i}{\mu}}\right) = \frac{1}{\mu^n}e^{-\frac{1}{\mu}\sum_{i=1}^{n}x_i}$$

$$\therefore \log l(\mu) = -n\log\mu - \frac{1}{\mu}\sum_{i=1}^{n}x_i$$

これより，

$$\frac{\partial}{\partial\mu}\log l(\mu) = -\frac{n}{\mu}+\frac{1}{\mu^2}\sum_{i=1}^{n}x_i = 0$$

を解くと，

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

よって，μの最尤推定量$\hat{\mu}$は，$X_i(i=1,2,\cdots\cdots)$ をそれぞれ指数分布

$f(x;\mu)=\dfrac{1}{\mu}e^{-\frac{x}{\mu}}$ $(x>0)$ に従う確率変数としたとき，

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$$

である。

次に，$\hat{\mu}$に関して問題の 4 つの性質につき調べる。

a．不偏性

$$E(\hat{\mu})=E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\cdot n\mu=\mu$$

よって不偏性を持つ。

b．有効性

これを見るために，次のクラメール・ラオの不等式の両辺を計算してみる。

$$V(\hat{\mu})\geqq\frac{1}{nE\left[\left\{\frac{\partial}{\partial\mu}\log f(X;\mu)\right\}^2\right]}$$

左辺は，

$$V(\hat{\mu})=V\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}V(X_i)=\frac{n\mu^2}{n^2}=\frac{\mu^2}{n}$$

であり，一方

$$E\left[\left\{\frac{\partial}{\partial\mu}\log f(X;\mu)\right\}^2\right]=E\left[\left\{\frac{\partial}{\partial\mu}\left(-\log\mu-\frac{X}{\mu}\right)\right\}^2\right]$$

$$=\frac{1}{\mu^4}E\left[\left\{X-\mu\right\}^2\right]=\frac{\mu^2}{\mu^4}=\frac{1}{\mu^2}$$

であるから，右辺も$\dfrac{\mu^2}{n}$となり，不等式の等号が成り立つことが示され，有効推定量であることがわかる。

c．充足性

これについては，

$$\prod_{i=1}^{n}f(x_i;\mu)=\frac{1}{\mu^n}e^{-\frac{1}{\mu}\sum_{i=1}^{n}x_i}=\frac{1}{\mu^n}e^{-\frac{1}{\mu}n\bar{x}}\times 1$$

ここで，$g(\overline{x};\mu)=\dfrac{1}{\mu^n}e^{-\frac{1}{\mu}n\overline{x}}$, $h(x_1,x_2,\cdots,x_n)=1$　とおくと，

$$\prod_{i=1}^{n}f(x_i;\mu)=g(\overline{x};\mu)\times h(x_1,x_2,\cdots,x_n)$$

と表わされ，$h(x_1,x_2,\cdots,x_n)$ は μ に無関係となるので，充足性は満たされている。

d．一致性

チェビシェフの不等式により，

$$P_r(\,|\hat{\mu}-E(\hat{\mu})|>\varepsilon)<\frac{V(\hat{\mu})}{\varepsilon^2}$$

である。

$$\therefore\ P_r(\,|\hat{\mu}-E(\overline{X})|>\varepsilon)<\frac{V(\overline{X})}{\varepsilon^2}$$

$$P_r(\,|\hat{\mu}-\mu|>\varepsilon)<\frac{\mu^2}{n\varepsilon^2}$$

ここで $n\to\infty$ とすると，右辺は 0 となるので，$\hat{\mu}$ が一致性を持つことが分る。