統計　宿題第２回解答例

問1　尤度関数の理解を深めるための問題である。

(1)　尤度関数は確率密度関数と同一の関数である。授業で配布したスライド資料より、パラメータβの指数分布の確率密度関数は、 $f(x \mid \beta) = \dfrac{1}{\beta}\exp\left(-\dfrac{x}{\beta}\right)$ **であるから、** X_1 **に対する尤度関数は、**

$L(\beta \mid X_1) = \dfrac{1}{\beta}\exp\left(-\dfrac{X_1}{\beta}\right)$ **となる。**

(2)　標本データ $X_1, \cdots, X_n$ **に対する尤度関数は、独立なランダム標本であるという仮定より**

$L(\beta \mid X_1, \cdots, X_n) = \prod_{i=1}^{n} \dfrac{1}{\beta}\exp\left(-\dfrac{X_i}{\beta}\right)$ **。よって、対数尤度関数** $\log L(\beta \mid X_1, \cdots, X_n)$ **は**

$$\log L(\beta \mid X_1, \cdots, X_n) = -n \log \beta - \frac{1}{\beta}\sum_{i=1}^{n} X_i$$

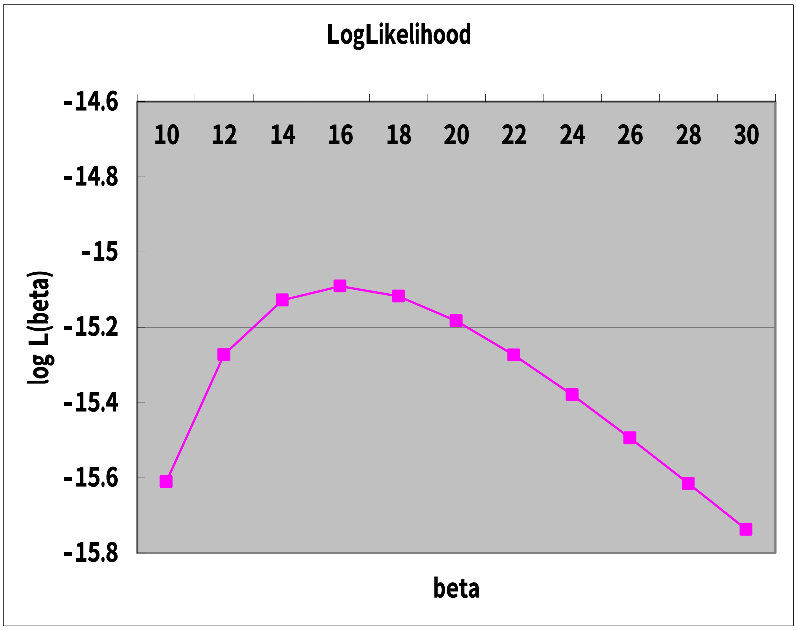
となる。

(3)　最尤推定量は対数尤度関数が最大となる1階の条件、

$\dfrac{\partial \log L(\beta)}{\partial \beta} = -n \times \dfrac{1}{\beta} - (-1) \times \dfrac{1}{\beta^2}\sum_{i=1}^{n} X_i = 0$ **を満たすβになる。よって、** $\hat{\beta} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ **。**

(4)　⑵で求めた対数尤度関数に $X_1 = 12,\ X_2 = 28,\ X_3 = 20,\ X_4 = 4$ **を代入すると、下図のグラフが得られる。**

beta	LogLikelihood
10	-15.6103404
12	-15.2729599
14	-15.1276579
16	-15.0903549
18	-15.1170426
20	-15.1829291
22	-15.2732607
24	-15.378882
26	-15.4939246
28	-15.6145323
30	-15.7381229



これよりβ=16 のとき、対数尤度が最大となることが分かる。

問2 推定量の分布から、未知の母数（パラメータ）の区間推定および仮説検定に持ち込む流れを理解するための問題である。設問の仮定より、機械 i の不良品の個数が正規分布にしたがう、すなわち $X_i \sim \mathrm{N}(\mu, \sigma^2)$ であることに注意。

(1) 母平均μの不偏推定量として、標本データの算術平均 $\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ を用いる。計算すると、

$\hat{\mu} = \overline{X} = 33$。

(2) 設問の仮定より $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$。いま $n = 20,\ \sigma^2 = 45$ であるから、$\overline{X} \sim \mathrm{N}\left(\mu, \frac{9}{4}\right)$。

(3) 母分散の不偏推定量は $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ で与えられるから、計算すると $\hat{\sigma}^2 = \frac{1102}{19} = 58$ となる。

(4) 自由度 n-1 の t 分布にしたがう。

(5) 信頼係数αは 0.95 であるから、p=0.025。自由度 19（=20-1）のt分布の上側 2.5%点はt分布表より $G^{-1}(0.025; 19)=2.093$。よって、

$$
\begin{aligned}
&\mathrm{P}(\overline{X} - G^{-1}(0.025;19)\sqrt{\hat{\sigma}^2/n} < \mu \le \overline{X} + G^{-1}(0.025;19)\sqrt{\hat{\sigma}^2/n}) \\
&= \mathrm{P}(33 - 2.093\times\sqrt{58/20} < \mu \le 33 + 2.093\times\sqrt{58/20}) \\
&= \mathrm{P}(33 - 3.564 < \mu \le 33 + 3.564) = \mathrm{P}(29.436 < \mu \le 36.564) = 0.95
\end{aligned}
$$
。

ゆえに母平均μは 95%の確率で [29.436, 35.564] の中に入っている。

(6) 帰無仮説が μ=36、対立仮説が μ < 36 であるから、片側検定になる。検定統計量τは

$\tau = \frac{\overline{X} - \mu}{\sqrt{\hat{\sigma}^2/n}} = \frac{33-36}{\sqrt{58/20}} = \frac{-3}{1.703} = -1.762$。有意水準は 5%であるから、自由度 19 のt分布の下側 5%はt分布表より $-G^{-1}(0.05; 19)$ （$= -t_{0.05}(19)$）$= -1.729$。これより、$\tau < -G^{-1}(0.05; 19)$であるから、帰無仮説を棄却する。すなわち母平均が 36 であるという帰無仮説は有意水準 5%で棄却された。

問3 ベルヌイ分布にしたがう独立な確率変数の算術平均が、中心極限定理によって正規分布にしたがうという性質を用いて、仮説検定に持ち込む問題である。

調査対象の個々の世帯は番組Aを観た($X_i = 1$)か、観なかった($X_i = 0$)かのどちらかであるから、ベルヌイ分布にしたがっていることが分かる。ベルヌイ分布にしたがう確率変数の期待値は p、分散は p (1-p) であるから、調査世帯の視聴率 $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ の平均（期待値）は、

$$\mathrm{E}\left(\overline{X}_n\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(X_i\right) = \frac{np}{n} = p$$
、

となる。また分散は

$$\mathrm{Var}\left(\overline{X}_n\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right)=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$$

となるので、n → ∞ のとき、中心極限定理より $Z=\dfrac{\overline{X}_n-p}{\sqrt{\dfrac{p(1-p)}{n}}}\sim \mathrm{N}(0,1)$ **が成立する。**

いま、帰無仮説は p = 0.2、対立仮説は編成局長の発言から p < 0.2 となり、片側検定になる。帰無仮説の下で、検定統計量 Z を計算すると、 $Z=\dfrac{\overline{X}_n-p}{\sqrt{\dfrac{p(1-p)}{n}}}=\dfrac{0.175-0.2}{\sqrt{\dfrac{0.2\times 0.8}{600}}}=\dfrac{-0.025}{0.0163}=-1.53$ 。

有意水準を 5%に設定すると、標準正規分布の下側 5%点は、統計表より-1.64。Z > -1.64 であるので、帰無仮説は棄却できない。ゆえに真の視聴率が 20%であるという主張は、有意水準 5%の統計的仮説検定から支持される。