2016年 1級 統計数理 係数

問2

指数分布からガンマ分布を帰納法で導出する問がポイントとなる問題。

それ以外の問いは比較的簡単なので、手早く処理したい。

(1) 略

(2)

$Q(c) = P(X > c)$ を積分により愚直に計算する。

$$Q(c) = \int_c^{\infty} \lambda e^{-\lambda x}\, dx = \langle 略 \rangle = e^{-\lambda c}$$

$u(\alpha)$ を求めるには、$c = u(\alpha)$, $Q(c) = \alpha$ を代入して整理する。

$$u(\alpha) = -\frac{1}{\lambda}\log\alpha$$

(3)

対数尤度を計算し、λ の停留点を最尤推定量 $\hat{\lambda}$ とする。

$$0 = \frac{d}{d\lambda}\log L(\lambda)\Big|_{\lambda=\hat{\lambda}}$$

$$= \frac{d}{d\lambda}\left\{\sum_{i=1}^{n}(\log\lambda - \lambda X_i)\right\}\Big|_{\lambda=\hat{\lambda}}$$

$$= \langle 略 \rangle$$

$$= \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} X_i$$

$$\Longleftrightarrow \hat{\lambda} = \frac{1}{\bar{X}} \qquad (ここで、\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i)$$

$Q(c)$、$u(\alpha)$ に $\lambda = \hat{\lambda}$ を代入することで最尤推定量を求める。

$$\hat{Q}(c) = e^{-\frac{c}{\bar{x}}}, \quad \hat{u}(\alpha) = -\bar{x}\log\alpha$$

また、$E(\hat{u}(\alpha)) = -E[\bar{x}]\log\alpha = -\frac{1}{\lambda}\log\alpha$ より、$\hat{u}(\alpha)$ は不偏である。

[4]

帰納法を使って解く。もし帰納法を使わずに解こうとすると、X_1 から X_n までの和が Y となるようなあらゆる組み合わせで $dx_1 \cdots dx_n$ の畳み込み積分をすることとなり、直接解くのが難しくなる。帰納法を使おうとするモチベーションは次の2点より湧いてくる。

■ Y の確率密度関数が既にわかっている。

■ ガンマ関数の性質 $\Gamma(n+1) = (n+1)\Gamma(n)$ を使えそう。

$Y_n = X_1 + X_2 + \cdots + X_n$ とおく。$n=1$ のとき、$Y_1 = X_1$ より Y_1 の p.d.f は、$f(y)$ であり、

$$f(y) = \frac{\lambda^1}{\Gamma(1)} y^0 \cdot e^{-\lambda y} = g_1(y)$$

が成り立つ。Y_n の p.d.f が $g_n(y)$ と仮定すると、Y_{n+1} の p.d.f. は次の式になる。

$$(\text{p.d.f of } Y_{n+1}) = \int_0^y \underbrace{g_n(y')}_{X_1,\cdots,X_n\text{の和が}y'\text{となる確率密度}} \cdot \underbrace{f(y-y')}_{X_{n+1}=y-y'\text{となる密度}} dy'$$

$$= \int_0^y \frac{\lambda^n}{\Gamma(n)} y'^{n-1} e^{-\lambda y'} \cdot \lambda e^{-\lambda(y-y')} dy'$$

$$= \frac{\lambda^{n+1}}{\Gamma(n)} \cdot e^{-\lambda y} \int_0^y y'^{n-1} \underbrace{e^{-\lambda y'} \cdot e^{\lambda y'}}_{=1} dy'$$

$$= \frac{\lambda^{n+1}}{\Gamma(n)} e^{-\lambda y} \left[\frac{y'^n}{n} \right]_0^y$$

$$= \frac{\lambda^{n+1}}{\Gamma(n)} e^{-\lambda y} \cdot \frac{y^n}{n}$$

$$= \frac{\lambda^{n+1}}{\Gamma(n+1)} y^{(n+1)-1} e^{-\lambda y}$$

$$= g_{n+1}(y)$$

よって帰納法より、Y_n の p.d.f. が $g_n(y)$ となることを示せた。

$E(\tilde{Q}(c)) = Q(c)$ を示す。

$$E(\tilde{Q}(c)) = \int_0^c 0 \cdot g_n(y)\,dy + \int_c^\infty \left(1-\frac{c}{y}\right)^{n-1} \cdot g_n(y)\,dy$$

$$= \int_c^\infty \left(1-\frac{c}{y}\right)^{n-1} \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}\,dy$$

$$= \frac{\lambda^n}{\Gamma(n)} \int_c^\infty (y-c)^{n-1} e^{-\lambda y}\,dy$$

$t = y - c$ と置換すると、

$$E(\tilde{Q}(c)) = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty t^{n-1} e^{-\lambda t} \cdot e^{-\lambda c} \cdot dt$$

$$= e^{-\lambda c} \int_0^\infty \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}\,dt$$

$$= e^{-\lambda c} = Q(c)$$

□

今回扱った確率変数Yの正体は、「故障率λの電球が壊れたら付け替えるを繰り返した場合にn個壊れるまでの時間」に相当する。これはガンマ分布に従い、n=1の時は指数分布に従い、また、時間 w の間に故障する個数は期待値 wλ のポアソン分布に従うことは知識として入れておくのが良い。

問4

モンテカルロ法を用いたパラメータ推定の問題。

分散を効率よく計算する能力や、 e^{-u^2} の積分計算に標準正規分布の付表を使う閃きが試される。

(1)

X は 二項分布 $Bin(n, \theta)$ に従う. よって.

$$V(\hat{\theta}_1) = \frac{1}{n^2} V[X] = \frac{1}{n}\theta(1-\theta)$$

(2)

$P(|Z| \leq 1) = 2 \cdot P(0 \leq Z \leq 1) = 2\theta$ より.

Y は $Bin(n, 2\theta)$ に従う. よって.

$$V[\hat{\theta}_2] = \frac{1}{4n^2} V[Y] = \frac{1}{4n^2} n \cdot 2\theta(1-2\theta) = \frac{\theta}{2n}(1-2\theta)$$

(3)

$$V[\hat{\theta}_3] = V\left[\frac{1}{n} \cdot \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left[-\frac{U_i^2}{2}\right]\right]$$ を計算する。

ここで. $V[\hat{\theta}_3] = E[(\hat{\theta}_3 - \theta)^2]$ or $E(\hat{\theta}_3^2) - (E(\hat{\theta}_3))^2$ に飛びつくと失敗する。というのも. $(\bigcirc \Sigma \triangle)^2$ の形となり交差項が表れ、複雑さがぐんと増す. $V[\bigcirc \Sigma \triangle]$ の形は、先に Σ を V の外に出すのがポイントとなる.

$$V[\hat{\theta}_3] = \frac{1}{n^2}\sum_{i=1}^{n} V\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{U_i^2}{2}\right)\right]$$

あえて中に入れる。

$E\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{U_i^2}{2}\right)\right) = \theta$ を使うため。

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[E\left\{\frac{1}{2\pi}e^{-U_i^2}\right\} - \left(E\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{U_i^2}{2}\right)\right\}\right)^2\right]$$

$$= \frac{1}{n}\left\{\frac{1}{2\pi}\int_0^1 e^{-u^2}\cdot 1\,du - \theta^2\right\}$$

(ポイント)
積分を陽に展開できないので、標準正規分布の分布表の値を使う。

$\downarrow u = \frac{s}{\sqrt{2}}$ で変換。

$$= \frac{1}{n}\left\{\frac{1}{2\pi}\int_0^{\sqrt{2}} e^{-\frac{s^2}{2}}\cdot\frac{1}{\sqrt{2}}ds - \theta^2\right\}$$

$$= \frac{1}{n}\left\{\frac{1}{2\pi}\frac{\sqrt{2\pi}}{\sqrt{2}}\int_0^{\sqrt{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds - \theta^2\right\}$$

$$= \frac{1}{n}\left\{\frac{1}{2\sqrt{\pi}}P(0 \leq z \leq \sqrt{2}) - \theta^2\right\}$$

$$\fallingdotseq \frac{0.0023}{n}$$

(4)

i) $V(\hat{\theta}_1)_{n=10000} \geq V(\hat{\theta}_2)_{n=n_2}$ となる n_2 と、

ii) $V(\hat{\theta}_1)_{n=10000} \geq V(\hat{\theta}_3)_{n=n_3}$ となる n_3 を求める。

i) $V(\hat{\theta}_1)_{n=10000} \geq V(\hat{\theta}_2)_{n=n_2}$

$\Leftrightarrow \frac{0.2248}{10000} \geq \frac{0.0542}{n_2}$

$\Leftrightarrow n_2 \geq 2411$

よって $\hat{\theta}_2$ では 2411 個

(ii) $\dfrac{0.2248}{10000} \geq \dfrac{0.0023}{n_3}$

$\Leftrightarrow \quad n_3 \geq 102.3$

よって $\hat{\theta}_3$ では 103個