数理解析研究所講究録 1273 巻 2002 年 165-177

記録値に対する予測区間の構成について

筑波大・数学 飛田 英祐 (Eisuke Hida)
筑波大・数学 赤平 昌文 (Masafumi Akahira)

1. はじめに

通常の統計的推測理論では未知の母数をもつ母集団分布からの標本に基づいて, その母数の推測方式の最適性などについて論じる. それに対して統計的予測論では未観測の確率変数を観測データに基づいて予測する方式を考える ([Gu70], [T75], [Ak90]). その際, 観測データが従う分布は未知の母数をもつから, そのことも考慮しなければならない. そこで, 指数型分布族において十分統計量に基づいて予測区間を構成したり ([AkH00], [H00]), また, ベイズ的観点から予測方式を考え, その際の事前分布の選択の適切性に関していろいろ論じられている ([Ge93]).

本論では, まず, 推測理論の区間推定に対応する区間予測を, 1 母数指数分布の場合に, 記録値の区間予測として考える. 本論では, 指数分布の場合に Awad and Raqab [AwR00] が論じた観測可能な記録値 (record values) に基づいて未観測の記録値の予測区間について述べ, それらとは別のベイズ的方法および完備十分統計量を用いる方法によって予測区間を構成し, また, 数値的観点からもそれらの予測区間を比較検討する ([HAk01]). さらに, 一般の分布の場合の記録値の区間予測の方法についても提案し, 具体的には, 切断指数分布, ワイブル分布, ガンマ分布の場合について, 数値的検討も含めて, 考察する. なお, 記録値に関する分布論および推測については Arnold *et.al* [ArBN98] 参照.

2. 指数分布の場合の予測区間の構成

確率変数列 $X_1, X_2, \ldots$ は互いに独立にいずれも確率密度関数 (probability density function 略して p.d.f.)

$$f(x;\theta) = \begin{cases} \theta e^{-\theta x} & (x > 0), \\ 0 & (x \leq 0) \end{cases}$$

をもつ 1 母数指数分布 $Exp(1/\theta)$ に従うとする. ただし, θ は未知の正値母数とする. このとき, 記録時刻 (record time) を

$$N(1) := 1,$$
$$N(j) := \min\{k \mid k > N(j-1),\ X_k > X_{N(j-1)}\}, \qquad (j = 2, 3, \ldots)$$

とし, この記録時刻を用いて記録値を $Y_j := X_{N(j)}$ $(j \geq 1)$ と表わす. いま, この指数分布 $Exp(1/\theta)$ から観測された記録値ベクトル $\boldsymbol{Y} := (Y_1, \ldots, Y_m)$ に基づいて未観測の n 番目に更新される記録値 Y_n $(1 \leq m < n)$ の予測区間を未知母数 θ に無関係に構成する方法について考える.

指数分布 $Exp(1/\theta)$ から観測された記録値ベクトル $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ の同時確率密度関数 (joint(j). p.d.f.) は

$$f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_m; \theta) = \prod_{i=1}^{m-1} \left[\frac{f(y_i;\theta)}{1 - F(y_i;\theta)} \right] f(y_m;\theta)$$

$$= \begin{cases} \theta^m e^{-\theta y_m} & (0 < y_m < \infty), \\ 0 & (\text{その他}) \end{cases}$$

となる．ただし，$F(\cdot;\theta)$ は $Exp(1/\theta)$ の累積分布関数 (cumulative distribution function 略して c.d.f.) とする．これより，次のことが成り立つことが分かる（たとえば，Ahsanullah [A78], [A80] 参照).

(i) $Y_i = X_{N(i)}$はガンマ分布 $G(i, 1/\theta)$ に従う．

(ii) $Z_{n,m} := Y_n - Y_m$ $(1 \leq m < n)$ とすると，Y_mを与えたときの $Z_{n,m}$の条件付分布は $G(n-m, 1/\theta)$ となる．

(iii) (Y_m, Y_n) の j.p.d.f. は

$$f_{Y_m,Y_n}(y_m, y_n;\theta) = \begin{cases} \dfrac{1}{\Gamma(m)\Gamma(n-m)}\theta^n y_m^{m-1}(y_n - y_m)^{n-m-1}e^{-\theta y_n} & (0 < y_m < y_n < \infty), \\ 0 & (\text{その他}). \end{cases}$$

(iv) $\boldsymbol{Y}$ に基づくθの最尤推定量は$\hat{\theta}_{ML} := m/Y_m$.

そこで，(i) ~ (iv) に基づいて Awad and Raqab[AwR00] が論じた Y_n の区間予測 (A), (W), (P), (H) について述べる．また，ベイズ的観点と完備十分性の観点から Y_n の予測区間の構成法 (B), (C) についても考え，数値的に比較検討する．

(A) 近似法 (approximate method)

統計量 $W := 2\theta(Y_n - Y_m)$ は，自由度 $2n-2m$ のカイ二乗分布 χ^2_{2n-2m} 分布に従う．そこで，$0 < \alpha < 1$ とし，$\chi^2_\alpha(2n-2m)$ を χ^2_{2n-2m} 分布の下側 $100\alpha\%$ 点とすると，

$$\Pr\{\chi^2_{\alpha/2}(2n-2m) \leq W \leq \chi^2_{1-\alpha/2}(2n-2m)\} = 1-\alpha$$

となる．これより，Y_n の信頼係数 $1-\alpha$ の予測区間は，未知母数 θ の代わりに $\boldsymbol{Y}$ に基づく $\hat{\theta}_{ML}$ を用いて，

$$\Pr\left\{\left(1 + \frac{1}{2m}\chi^2_{\alpha/2}(2n-2m)\right)y_m \leq Y_n \leq \left(1 + \frac{1}{2m}\chi^2_{1-\alpha/2}(2n-2m)\right)y_m\right\} = 1-\alpha$$

から得られる．また，Y_n の信頼係数 $1-\alpha$ の下側，上側予測限界はそれぞれ

$$\left\{1 + \frac{1}{2m}\chi^2_\alpha(2n-2m)\right\}y_m, \quad \left\{1 + \frac{1}{2m}\chi^2_{1-\alpha}(2n-2m)\right\}y_m$$

になる．

(W) Wald 法

Y_n の p.d.f. は

$$f_{Y_n}(y_n) = \begin{cases} \dfrac{\theta^n}{\Gamma(n)}y_n^{n-1}e^{\theta y_n} & (y_n > 0), \\ 0 & (y_n \leq 0) \end{cases}$$

になる. この p.d.f. において未知母数 θ を $\boldsymbol{Y}$ に基づく $\hat{\theta}_{ML}$ で置き換えたものを $g(\cdot)$ とすると

$$g(y_n) = \begin{cases} \dfrac{1}{\Gamma(n)}\left(\dfrac{m}{y_m}\right)^n y_n^{n-1} e^{-my_n/y_m} & (0 < y_m < y_n < \infty), \\ 0 & (\text{その他}) \end{cases}$$

となる. ここで, $Z := 2mY_n/Y_m$ とすると, Z は χ^2_{2n} 分布に従うので, Y_n の信頼係数 $1-\alpha$ の予測区間は

$$\Pr\left\{\frac{y_m}{2m}\chi^2_{\alpha/2}(2n) \leq Y_n \leq \frac{y_m}{2m}\chi^2_{1-\alpha/2}(2n)\right\} = 1-\alpha$$

から得られる. また, Y_n の信頼係数 $1-\alpha$ の下側, 上側予測限界はそれぞれ $\dfrac{y_m}{2m}\chi^2_\alpha(2n)$, $\dfrac{y_m}{2m}\chi^2_{1-\alpha}(2n)$ になる.

(P) 枢軸法 (pivot method)

$W = 2\theta(Y_n - Y_m)$ と $T := 2\theta Y_m$ は互いに独立に, それぞれ χ^2_{2n-2m} 分布, χ^2_{2m} 分布に従うことから,

$$F := \frac{W}{2n-2m}\Big/\frac{T}{2m} = \frac{m(Y_n - Y_m)}{(n-m)Y_m} \sim \mathrm{F}_{2n-2m,2m}\text{分布 (自由度 } 2n-2m,\ 2m \text{ の F 分布)}$$

となる. $\mathrm{F}_\alpha(a,b)$ を $\mathrm{F}_{a,b}$ 分布の下側 $100\alpha\%$ 点とすると, Y_n の信頼係数 $1-\alpha$ の予測区間は

$$\Pr\left\{\left(1+\frac{n-m}{m}\mathrm{F}_{\alpha/2}(2n-2m,2m)\right)y_m \leq Y_n \leq \left(1+\frac{n-m}{m}\mathrm{F}_{1-\alpha/2}(2n-2m,2m)\right)y_m\right\} = 1-\alpha$$

から得られる. また, Y_n の信頼係数 $1-\alpha$ の下側, 上側予測限界はそれぞれ $\left(1+\dfrac{n-m}{m}\mathrm{F}_\alpha(2n-2m,2m)\right)y_m$, $\left(1+\dfrac{n-m}{m}\mathrm{F}_{1-\alpha}(2n-2m,2m)\right)y_m$ になる.

(H) 最大条件付密度法 (highest conditional density method)

$\boldsymbol{Y} = (Y_1,\ldots,Y_m)$ を与えたときの Y_n の条件付密度は

$$f_{Y_n|Y_m}(y_n|y_m;\theta) = \begin{cases} \dfrac{1}{\Gamma(n-m)}\theta^{n-m}(y_n-y_m)^{n-m-1}e^{-\theta(y_n-y_m)} & (0 < y_m < y_n < \infty), \\ 0 & (\text{その他}) \end{cases}$$

になり, ここで, 未知母数 θ を $\boldsymbol{Y}$ に基づく $\hat{\theta}_{ML}$ で置き換えたものを $\hat{f}$ とすると

$$\hat{f}_{Y_n|Y_m}(y_n|y_m) = \begin{cases} \dfrac{1}{\Gamma(n-m)y_m}m^{n-m}\left(\dfrac{y_n-y_m}{y_m}\right)^{n-m-1} e^{-m(y_n-y_m)/y_m} & (0 < y_m < y_n < \infty), \\ 0 & (\text{その他}) \end{cases}$$

となる. $\hat{f}$ は未知母数 θ に無関係であるので, Y_n の信頼係数 $1-\alpha$ の最大条件密度法 (HCD) 予測区間は

$$\Pr\{(1+w_1)\,y_m \leq Y_n \leq (1+w_2)\,y_m\} = 1-\alpha$$

から得られる. ただし, (w_1, w_2) は

$$\begin{cases} \displaystyle\int_{(1+w_1)y_m}^{(1+w_2)y_m} \hat{f}_{Y_n|Y_m}(y_n|y_m)dy_n = 1-\alpha, \\ \\ \hat{f}_{Y_n|Y_m}((1+w_1)y_m|y_m) = \hat{f}_{Y_n|Y_m}((1+w_2)y_m|y_m) \end{cases}$$

を同時に満たすものとする.

次に, [AR00] とは別の観点から, Y_n の予測区間を構成する方法 (B), (C) を考える.

(B) ベイズ法 (Bayesian method)

ベイズの観点から, 未知母数 θ の事前密度として, Hartigan の漸近局所不変事前密度 $p_H(\theta) = c$, $(0 < c < \infty)$ をとる. このとき, (Y_n, Y_m) の同時事後密度と Y_m の事後密度はそれぞれ

$$\begin{aligned} f(y_n, y_m) &= \int_0^\infty f_{Y_n,Y_m}(y_n, y_m; \theta)p_H(\theta)d\theta, \\ &\propto y_m^{m-1}(y_n - y_m)^{n-m-1}/y_n^{n+1}, \end{aligned}$$

$$\begin{aligned} f(y_m) &= \int_0^\infty f_{Y_m}(y_m; \theta)p_H(\theta)d\theta \\ &\propto y_m^{-2} \end{aligned}$$

となることから, Y_n の予測密度は

$$f(y_n|y_m) = \frac{f(y_n, y_m)}{f(y_m)} = \begin{cases} K\left(\dfrac{y_m}{y_n}\right)^{m+1}\left(1 - \dfrac{y_m}{y_n}\right)^{n-m-1}\dfrac{1}{y_n} & (0 < y_m < y_n < \infty), \\ 0 & (\text{その他}) \end{cases}$$

となる. ただし, $K := 1/B(m+1, n-m)$ とする. この予測密度より, Y_n の信頼係数 $1-\alpha$ の予測区間は

$$\int_{\underline{b}}^{\overline{b}} f(y_n|y_m)dy_n = 1 - \alpha$$

を満たす $[\underline{b}, \overline{b}]$ で与えられる. また,

$$\int_0^{\underline{b}^*} f(y_n|y_m)dy_n = \alpha, \quad \int_0^{\overline{b}^*} f(y_n|y_m)dy_n = 1 - \alpha$$

となる $\underline{b}^*$, $\overline{b}^*$ について, Y_n の信頼係数 $1-\alpha$ の下側, 上側予測限界はそれぞれ $\underline{b}^*$, $\overline{b}^*$ になる.

(C) 完備十分統計 (量による方) 法 (complete sufficient statistic method)

(A), (W), (P), (H) の方法では未知母数 θ の代わりに $\hat{\theta}_{ML}$ にするのが特徴であり, (B) では θ の事前密度として Hartigan のものを用いるのが特徴である. しかし, この場合には完備十分統計量が存在するから, それに基づいて区間予測を行う方が自然であり, 妥当に思われる. そこで, 一般的に

X を観測された確率変数, Y を未観測の確率変数とし, (X,Y) の j.p.d.f. を $f_{X,Y}(x,y;\theta)$ $(\theta\in\Theta)$ とする. このとき, $S(x)$ を $X=x$ に依存する Y の値域の部分集合として, 予測関数 $\phi(x,y)$ を

$$\phi(x,y) := \begin{cases} 1 & (y \in S(x) \text{ のとき}), \\ 0 & (y \notin S(x) \text{ のとき}) \end{cases}$$

によって定義する. そして, $\mathcal{Y}\subset\mathbf{R}^1$ とし, $0<\alpha<1$ について

$$\begin{aligned} 1-\alpha = \mathrm{P}_\theta\{Y\in S(X)\} &= \mathrm{E}_\theta[\phi(X,Y)] \\ &= \iint \phi(x,y) f_{X,Y}(x,y;\theta)dxdy, \qquad \theta\in\Theta \end{aligned} \tag{2.1}$$

となるように区間 $S(X)(\subset\mathcal{Y})$ を定めるとき, これを Y の信頼係数 $1-\alpha$ の予測区間という. 次に, Y が θ に対する完備十分統計量とする. (2.1) より

$$\iint \phi(x,y) f_{X|Y}(x|y) f_Y^\theta(y)dxdy = 1-\alpha, \qquad \theta\in\Theta$$

であるから

$$\int\left\{\int \phi(x,y) f_{X|Y}(x|y)dx - (1-\alpha)\right\} f_Y^\theta(y)dy = 0, \qquad \theta\in\Theta$$

となる. よって, Y が θ の完備であるから

$$\int \phi(x,y) f_{X|Y}(x|y)dx = 1-\alpha, \qquad a.a.\ y \tag{2.2}$$

となる. いま, $U:=T(X,Y)$ とし, $X=g(U,Y)$ と表わせたとすると, (2.2) は

$$\int \phi(g(u,y),y) f_{U|Y}(u|y)du = 1-\alpha, \qquad a.a.\ y \tag{2.3}$$

となる. ここで,

$$\begin{cases} f_{U|Y}(u|y) = f_U(u) \\ \phi(g(u,y),y) = \psi(u) \end{cases}$$

であると仮定すると, (2.3) より

$$1-\alpha = \int \psi(u) f_U(u)du = \mathrm{P_U}\{\underline{u}\le U\le \overline{u}\}$$

となる $\underline{u}$, $\overline{u}$ が存在する. よって

$$1-\alpha = \mathrm{P_{X,Y}}\{\underline{u}\le T(X,Y)\le\overline{u}\} = \mathrm{P}_{X,Y}\{a(X)\le Y\le b(X)\}$$

となれば, 区間 $[a(X),\ b(X)]$ は Y の信頼係数 $1-\alpha$ の予測区間となる.

そこで, 前記の記録値の場合について $X:=Y_m,\ Y:=Y_n$ とおくと, $(X,\ Y)$ の同時密度は

$$f_{X,\ Y}^\theta(x,y) = \begin{cases} \dfrac{\theta^n}{(m-1)!(n-m-1)!}x^{m-1}(y-x)^{n-m-1}e^{-\theta y} & (0<x<y<\infty), \\ 0 & (\text{その他}) \end{cases}$$

より, Y は未知母数 θ に対する完備十分統計量になる. したがって, $Y = y$ を与えたときの X の条件付密度は

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{1}{B(m, n-m)} \left(\dfrac{x}{y}\right)^{m-1} \left(1 - \dfrac{x}{y}\right)^{n-m-1} \dfrac{1}{y} & (0 < x < y < \infty), \\ 0 & (\text{その他}) \end{cases}$$

となり, θ に無関係になる. この条件付密度を用いて, $U := X/Y$ とおけば, U はベータ分布 $\mathrm{Be}(m, n-m)$ に従うから, Y の信頼係数 $1-\alpha$ の予測区間は

$$1 - \alpha = P_U\{\underline{u} \le U \le \overline{u}\} = P_{X,Y}\left\{\underline{u} \le \frac{X}{Y} \le \overline{u}\right\} = P_{X,Y}\left\{\frac{X}{\overline{u}} \le Y \le \frac{X}{\underline{u}}\right\}$$

より得られる. また,

$$P_U\{U \le \underline{u}^*\} = \alpha, \quad P_U\{U \le \overline{u}^*\} = 1 - \alpha$$

となる $\underline{u}^*$, $\overline{u}^*$ について, Y の信頼係数 $1-\alpha$ の下側, 上側予測限界はそれぞれ $X/\overline{u}^*$, $X/\underline{u}^*$ になる.

3. 数値的検討

前節において論じた近似法 (A), Wald 法 (W), 枢軸法 (P), 最大条件密度法 (H), ベイズ法 (B), 完備十分統計法 (C) について $\theta = 0.1, 0.2, 0.5$; $m = 3$; $n = 5$ の場合に, 10000 回繰り返しシミュレーションを行い, それぞれの予測区間の被覆確率, 区間の幅, 予測区間を得た (表 1 ~ 3 参照).

表 1. $m = 3$, $n = 5$ の場合の予測区間の被覆確率

$1-\alpha$	θ ＼ 方法	(A)	(W)	(H)	(P)	(B)	(C)
95%	0.1	0.9190	0.9192	0.8355	0.9497	0.9098	0.9491
	0.2	0.9219	0.9220	0.8338	0.9477	0.9130	0.9483
	0.5	0.9187	0.9188	0.8348	0.9472	0.9086	0.9482
90%	0.1	0.8944	0.8944	0.7740	0.8997	0.8496	0.8988
	0.2	0.8867	0.8868	0.7693	0.8961	0.8084	0.8941
	0.5	0.8954	0.8956	0.7758	0.8947	0.8179	0.8954
80%	0.1	0.8473	0.8474	0.6793	0.7935	0.6964	0.7951
	0.2	0.8473	0.8475	0.6795	0.7960	0.6956	0.7976
	0.5	0.8433	0.8434	0.6779	0.7958	0.6931	0.7951

表 2. $m = 3$, $n = 5$ の場合の 95% 予測区間の区間幅

θ ＼ 方法	(A)	(W)	(H)	(P)	(B)	(C)
0.1	52.9384	85.6055	46.9126	121.5517	59.7468	99.7117
0.2	26.6703	43.1278	23.6345	61.2374	30.1003	50.2347
0.5	10.5707	17.0936	9.3674	24.2712	11.9302	19.9103

表 3. 指数分布 $Exp(1/\theta)$, $\theta = 0.5$ から導かれた記録値 2.0248, 4.0466, 6.0333, 8.0338, 10.018 の場合に, $y_3 = 6.0333$ まで観測した場合の $y_5 = 10.018$ の予測区間

$1-\alpha$	(A)	(W)	(H)	(P)	(B)	(C)
95%	[6.520, 17.238]	[3.265, 20.597]	[6.119, 15.617]	[6.471, 31.080]	[6.195, 18.291]	[6.310, 26.498]
90%	[6.748, 15.574]	[3.962, 18.409]	[6.202, 13.941]	[6.686, 24.269]	[6.301, 15.267]	[6.473, 20.963]
80%	[7.103, 13.856]	[4.892, 16.076]	[6.370, 12.228]	[7.036, 18.827]	[6.487, 12.706]	[6.749, 16.510]
70%	[7.407, 12.816]	[5.601, 14.615]	[6.542, 11.205]	[7.352, 16.155]	[6.831, 11.367]	[7.004, 14.310]
60%	[7.691, 12.055]	[6.213, 13.517]	[6.722, 10.469]	[7.660, 14.449]	[6.831, 10.520]	[7.255, 12.900]
50%	[7.967, 11.449]	[6.775, 12.619]	[6.909, 9.890]	[7.970, 13.222]	[7.005, 9.881]	[7.510, 11.881]

4. 応用例

1997 年から現在までの関東地方 (北緯 34.5～37.2 度, 東経 138.1～141.0 度の地域) で起こった地震のマグニチュードの大きさに関して次のような記録値がある.

$$3.7, 3.9, 4.4, 4.5, 4.7, 4.9, 5.0, 5.1, 5.2, 5.4, 5.8, 6.1 \ (M).$$

このとき, 10 番目までの記録値 $y_{10} = 5.4$ から, y_{12} を予測したときの信頼係数 $1-\alpha$ の予測区間は次の表のようになる.

$1-\alpha$	(A)	(W)	(H)	(P)	(B)	(C)
95%	[5.531, 8.409]	[3.348, 10.628]	[5.423, 7.973]	[5.526, 9.196]	[5.430, 8.199]	[5.434, 8.537]
90%	[5.592, 7.962]	[3.739, 9.832]	[5.445, 7.523]	[5.586, 8.495]	[5.455, 7.634]	[5.463, 7.895]
80%	[5.687, 7.500]	[4.228, 8.406]	[5.490, 7.063]	[5.681, 7.829]	[5.503, 7.098]	[5.516, 7.291]
70%	[5.769, 7.221]	[4.877, 7.979]	[5.537, 6.789]	[5.764, 7.451]	[5.550, 6.797]	[5.569, 6.954]
60%	[5.845, 7.017]	[4.877, 7.979]	[5.585, 6.591]	[5.842, 7.187]	[5.598, 6.588]	[5.621, 6.720]
50%	[5.919, 6.854]	[5.140, 7.625]	[5.635, 6.436]	[5.919, 6.982]	[5.647, 6.429]	[5.676, 6.543]

5. 一般の分布の場合の予測区間の構成

一般に, 確率変数 X が p.d.f. $p(x)$, c.d.f. $F(x)$ をもつときに, $U := F(X)$ とすれば, U は区間 $[0,1]$ 上の一様分布 $U(0,1)$ に従う. 従って

$$Y := -\log(1 - F(X)) \tag{5.1}$$

とおくと, Y は p.d.f.

$$f_Y(y) = \begin{cases} e^{-y} & (y > 0), \\ 0 & (y \leq 0) \end{cases}$$

をもつ指数分布 $Exp(1)$ に従うことが分かる. ここで, (5.1) から Y は X の非減少関数であることに注意. そこで, いま, $X_1, \ldots, X_n$ がたがいに独立にいずれも p.d.f. $p(x, \theta)$ をもつ分布に従うとする. ただし, $\theta \in \Theta$ とする. まず, $\boldsymbol{X} := (X_1, \ldots, X_n)$ として, 母数 θ の適当な推定量 $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$ について $\boldsymbol{X} = \boldsymbol{x} = (x_1, \ldots, x_n)$ のときの推定値 $\hat{\theta}(\boldsymbol{x})$ を用いて, $\hat{p}(\cdot) = p(\cdot, \hat{\theta}(\boldsymbol{x}))$ をつくる. そして, $X_1, \ldots, X_n$ がたがいに独立にいずれも p.d.f. $\hat{p}$ に従うと見なして変換 (5.1) によって $Y_i = -\log(1 - \hat{F}(X_i))$ $(i = 1, \ldots, n)$ とすれば, $Y_1, \ldots, Y_n$ はたがいに独立にいずれも指数分布 $Exp(1)$ に従うことになる. ただし, $\hat{F}$ は $\hat{p}$ を p.d.f. としてもつ c.d.f. とする. 従って, 後は第 2 節の指数分布の場合の予測区間の構成に帰着できる.

5.1. 切断指数分布の場合

まず, $X_1, \dots, X_n$ をたがいに独立に, いずれも p.d.f.

$$p(x;\alpha,\beta) = \begin{cases} c_{\alpha,\beta}e^{-x/\beta} & (0 < x < \alpha), \\ 0 & (\text{その他}) \end{cases}$$

をもつ切断指数分布に従うとする. ただし, α, β は未知とし,

$$c_{\alpha,\beta} = \frac{1}{\beta(1-e^{-\alpha/\beta})}$$

とする. このとき, 統計量 $(\max_{1\leq i\leq n} X_i, \bar{X})$ が (α,β) に対する十分統計量になる. いま, $\boldsymbol{X} := (X_1, \dots, X_n)$ の実現値を $\boldsymbol{x} := (x_1, \dots, x_n)$ とすると, 母数 (α,β) の尤度関数は

$$L(\alpha,\beta;\boldsymbol{x}) := c^n_{\alpha,\beta}\exp\left(-\sum_{i=1}^n x_i/\beta\right) \quad (0 < x_{(1)} \leq x_{(n)} < \alpha)$$

になり, 尤度方程式 $\partial \log L/\partial\beta = 0$ より

$$\beta = \bar{x} + \frac{\alpha e^{-\alpha/\beta}}{1-e^{-\alpha/\beta}} =: \bar{x} + \alpha\gamma \tag{5.2}$$

になる. ただし, $x_{(1)} = \min_{1\leq i\leq n} x_i$, $x_{(n)} = \max_{1\leq i\leq n} x_i$, $\bar{x} = (1/n)\sum_{i=1}^n x_i$, $\gamma = e^{-\alpha/\beta}\big/\left(1-e^{-\alpha/\beta}\right)$ とする.

このとき, α, β のそれぞれの推定量を $\hat{\alpha}$, $\hat{\beta}$ とすれば, (5.2) から, γ の推定量として

$$\hat{\gamma} = e^{-\hat{\alpha}/\hat{\beta}}\Big/\left(1-e^{-\hat{\alpha}/\hat{\beta}}\right) \tag{5.3}$$

を考える. そこで, 実際には, $\hat{\alpha}_1 := X_{(n)} = \max_{1\leq i\leq n} X_i$ とし, (5.2) より $\hat{\beta} = \bar{X} + \hat{\alpha}_1\hat{\gamma}$ を (5.3) に代入して

$$\hat{\gamma} = \frac{e^{-\hat{\alpha}_1/(\bar{X}+\hat{\alpha}_1\hat{\gamma})}}{1-e^{-\hat{\alpha}_1/(\bar{X}+\hat{\alpha}_1\hat{\gamma})}}$$

から $\hat{\gamma} = \hat{\gamma}_1$ を求めて, 再び (5.2) より $\hat{\beta}_1 = \bar{X} + \hat{\alpha}_1\hat{\gamma}_1$ を得る. 次に, $\hat{\beta}_1$, $\hat{\gamma}_1$ を用いて

$$\hat{\alpha}_2 := X_{(n)} + \frac{\hat{\beta}_1}{n\hat{\gamma}_1}$$

とし, (5.2) より $\hat{\beta} = \bar{X} + \hat{\alpha}_2\hat{\gamma}$ を (5.3) に代入して

$$\hat{\gamma} = \frac{e^{-\hat{\alpha}_2/(\bar{X}+\hat{\alpha}_2\hat{\gamma})}}{1-e^{-\hat{\alpha}_2/(\bar{X}+\hat{\alpha}_2\hat{\gamma})}}$$

から $\hat{\gamma} = \hat{\gamma}_2$ を求めて, (5.2) より $\hat{\beta}_2 = \bar{X} + \hat{\alpha}_2\hat{\gamma}_2$ を得る. さらに

$$\hat{\alpha}_3 := X_{(n)} + \frac{\hat{\beta}_2}{n\hat{\gamma}_2}$$

として, 以下, 同様の手続きを繰り返して, $\hat{\alpha}_k$, $\hat{\beta}_k$ の $k \to \infty$ のときの極限をそれぞれ $\hat{\alpha}^*$, $\hat{\beta}^*$ を得る. そこで, $p(\cdot\,;\alpha,\beta)$ を $p(\cdot\,;\hat{\alpha}^*,\hat{\beta}^*)$ で推定する.

例 5.1 ある p.d.f. $p(\cdot;\alpha,\beta)$ をもつ切断指数分布から大きさ 30 の無作為標本を求めると, 次のようになった.

$$70.4878, 75.8511, 72.3118, 73.4242, 71.492, 74.6821, 72.0552, 80.2314, 72.4681, 74.3722,$$
$$80.4015, 73.0246, 73.7204, 80.4591, 77.78, 80.9469, 78.1324, 75.6065, 77.5339, 72.6765,$$
$$73.7422, 70.1697, 76.6218, 76.9422, 78.997, 73.7404, 79.3767, 77.5787, 71.5385, 77.7493.$$

このデータから第 2 節の意味での記録値を求めると

$$70.4878, 75.8511, 80.2314, 80.4015, 80.4591, 80.9469$$

となる. このとき, まず, この切断指数分布の p.d.f. の母数 α, β の推定値を上記の方法で求めると, それぞれ $\hat{\alpha}^* = 81.136$, $\hat{\beta}^* = -5.665$ となる.

次に, p.d.f. $p(\cdot;\hat{\alpha}^*,\hat{\beta}^*)$ に従う確率変数の変換 (5.1) を用いると, 上記の記録値は

$$0.165657, 0.499961, 1.91368, 2.10735, 2.18409, 3.41782$$

となる. この記録値において, 4 番目の記録値 $Y_4 = 2.10735$ から 6 番目の記録値 $Y_6(= 3.41782)$ の予測区間を第 2 節の各方法で構成し, さらに (5.1) の逆変換で元に戻すと, $X_4 = 80.4015$ に基づく信頼係数 $1-\alpha$ の $X_6(= 80.9469)$ の予測区間は次の表のようになる. いずれの場合にも, X_6 の値が予測区間に含まれていることが分かる.

$1-\alpha$	(A)	(W)	(H)	(P)	(B)	(C)
95%	[80.495, 81.099]	[79.005, 81.124]	[80.488, 81.132]	[80.419, 81.080]	[80.431, 81.103]	[80.444, 81.126]
90%	[80.534, 81.079]	[79.488, 81.113]	[80.526, 81.124]	[80.435, 81.048]	[80.451, 81.069]	[80.471, 81.108]
80%	[80.590, 81.046]	[79.942, 81.093]	[80.582, 81.100]	[80.467, 80.998]	[80.485, 81.010]	[80.515, 81.068]
70%	[80.634, 81.018]	[80.192, 81.071]	[80.628, 81.073]	[80.499, 80.955]	[80.516, 80.961]	[80.554, 81.029]

例 5.2 日本陸上競技において, ハンマー投げ競技で日本記録を保持している室伏広治選手の 1998 年から 2001 年までの新記録の成績が

$$76.65, 76.67, 77.35, 78.41, 78.57, 79.17, 80.23, 80.56, 81.08, 82.23, 82.60, 83.47 \text{ (m)}$$

であった. この記録値は, ある切断指数分布に従っていると考えて, 上記の方法から母数 α, β の推定値は, それぞれ $\hat{\alpha}^* = 91.058$, $\hat{\beta}^* = 1.82 \times 10^8$ となる. さらに, 変換 (5.1) を用いると, 上記の記録値は

$$1.844, 1.845, 1.894, 1.974, 1.987, 2.036, 2.129, 2.160, 2.211, 2.334, 2.376, 2.485$$

となる. この記録値において, 10 番目の記録値 $Y_{10} = 2.334$ から 12 番目の記録値 $Y_{12}(= 2.485)$ の予測区間を第 2 節の各方法で構成し, さらに (5.1) の逆変換で元に戻すと, $X_{10} = 82.23$ に基づく信頼係数 $1-\alpha$ の $X_{12}(= 83.47)$ の予測区間は次の表のようになる. Wald 法による予測区間については, 区間の長さは他のものより長いが, いずれの予測区間にも X_{12} の値が含まれる.

$1-\alpha$	(A)	(W)	(H)	(P)	(B)	(C)
95%	[82.72, 88.65]	[69.63, 90.14]	[82.70, 89.35]	[82.32, 88.15]	[82.34, 88.43]	[82.36, 88.78]
90%	[82.93, 88.14]	[72.96, 89.76]	[82.91, 88.74]	[82.40, 87.53]	[82.44, 87.70]	[82.47, 88.05]
80%	[83.26, 87.50]	[76.41, 89.17]	[83.24, 87.97]	[82.57, 86.76]	[82.62, 86.82]	[82.66, 87.16]
70%	[83.53, 87.04]	[78.48, 88.65]	[83.51, 87.42]	[82.73, 86.21]	[82.79, 86.23]	[82.85, 86.55]

5.2. ワイブル分布の場合

$X_1, X_2, \dots, X_n, \dots$ をたがいに独立に，いずれも p.d.f.

$$p_\alpha(x,\theta) = \begin{cases} \dfrac{\alpha x^{\alpha-1}}{\theta^\alpha} e^{-(x/\theta)^\alpha} & (x>0), \\ 0 & (x \le 0) \end{cases} \tag{5.4}$$

をもつワイブル (Weibull) 分布 $W(\alpha,\theta)$ に従う確率変数列とする．ただし，$\theta > 0$, $\alpha > 0$ とする．このとき，巾母数 α について，Menon [M63] は α^{-1} の推定量として

$$\tilde{\alpha}^{-1} := \frac{\sqrt{6}}{\pi}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\log X_i - \overline{\log X}\right)^2}$$

を提案した．ただし，$\overline{\log X} = (1/n)\sum_{i=1}^{n} \log X_i$ とする．推定量 $\tilde{\alpha}^{-1}$ は α^{-1} の漸近正規性かつ漸近不偏性を持ち，その分散は

$$\mathrm{Var}\left(\tilde{\alpha}^{-1}\right) = \left\{1.1 + O\left(\frac{1}{n}\right)\right\}\frac{\alpha^{-2}}{n}$$

となり，その Cramér-Rao の下界に対する漸近効率は約 55%であることが知られている (Johnson *et.al* [JKB94])．

いま，確率変数 X が p.d.f. (5.4) をもつワイブル分布 $W(\alpha,\theta)$ に従うとき，$Y := X^\alpha$, $\eta := \theta^\alpha$ とおくと，Y は指数分布 $Exp(\eta)$ に従うことが分かる．そこで，$X_i = x_i$ $(i = 1,2,\dots,n,\dots)$ であるとき，$\tilde{\alpha}^{-1}$ の実現値を計算して

$$Y_i := X_i^{1/\tilde{\alpha}^{-1}} \qquad (i = 1,2,\dots,n,\dots) \tag{5.5}$$

という変数変換を行って，$Y_1, Y_2, \dots, Y_n, \dots$ がたがいに独立に，いずれも指数分布 $Exp\left(\theta^{1/\tilde{\alpha}^{-1}}\right)$ に従うと見なして，第 2 節の議論に帰着できる．

例 5.3 1993 年から 1997 年までに，神通川の神通大橋観測所で観測された日平均流量データから，得られた第 2 節の意味での記録値が

$$137.52, 156.4, 312.32, 339.97, 414.2, 418.69, 512.12, 839.59, 1183.8, 1315.86, 1410.07\ (\mathrm{m}^3/\mathrm{s})$$

であった．この記録値は，あるワイブル分布に従っていると考えて，上記の方法から母数 α^{-1} の推定値は，$\tilde{\alpha}^{-1} = 0.48443$ となる．さらに，変換 (5.5) を用いると，上記の記録値は，それぞれ

$$25953.1, 33847.1, 141109, 168115, 252731, 258419, 391656,$$
$$1.08667\times10^6, 2.20857\times10^6, 2.74743\times10^6, 3.16897\times10^6$$

となる．この記録値において，9 番目の記録値 $Y_9 = 2.20857 \times 10^6$ から 11 番目の記録値 $Y_{11}(= 3.16897 \times 10^6)$ の予測区間を第 2 節の各方法で構成し，さらに (5.5) の逆変換で元に戻すと，$X_9 = 1183.8$ に基づく信頼係数 $1-\alpha$ の $X_{11}(= 1410.07)$ の予測区間は次の表のようになる．

$1-\alpha$	(A)	(W)	(H)	(P)	(B)	(C)
95%	[1199.1, 1495.0]	[931.8, 1673.5]	[1198.5, 1574.6]	[1186.5, 1454.4]	[1187.4, 1477.9]	[1188.0, 1513.5]
90%	[1206.2, 1453.3]	[985.9, 1609.2]	[1205.5, 1509.1]	[1189.1, 1411.1]	[1190.4, 1423.0]	[1191.4, 1451.5]
80%	[1217.2, 1408.8]	[1049.6, 1536.0]	[1216.4, 1444.4]	[1194.4, 1365.2]	[1196.1, 1369.1]	[1197.8, 1391.0]
70%	[1226.5, 1381.1]	[1093.4, 1487.1]	[1225.8, 1406.6]	[1199.8, 1337.1]	[1201.6, 1338.1]	[1204.0, 1356.2]

5.3. ガンマ分布の場合

$X_1, X_2, \ldots, X_n, \ldots$ をたがいに独立に，いずれも p.d.f.

$$p(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} & (x>0), \\ 0 & (x \le 0) \end{cases} \tag{5.6}$$

をもつガンマ (Gamma) 分布 $G(\alpha,\beta)$ に従う確率変数列とする．ただし，$\alpha, \beta > 0$ とする．このとき，$\boldsymbol{X} = (X_1, \ldots, X_n)$ の実現値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ について，(α,β) の尤度関数は

$$L(\alpha,\beta|\boldsymbol{x}) = \prod_{i=1}^n p(x_i;\alpha,\beta) = \frac{\left(\prod_{i=1}^n x_i\right)^{\alpha-1} e^{-(1/\beta)\sum_{i=1}^n x_i}}{\beta^{n\alpha}\Gamma(\alpha)^n}$$

となるから，尤度方程式 $(\partial/\partial\alpha)\log L = 0$, $(\partial/\partial\beta)\log L = 0$ より

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^n \log X_i = \log\beta + \psi(\alpha), & (5.7) \\ \bar{X} = \alpha\beta & (5.8) \end{cases}$$

を得る．ただし，$\psi(\alpha) := (\partial/\partial\alpha)\log\Gamma(\alpha)$ とする．このとき，(5.8) より $\beta = \bar{X}/\alpha$ となるから，これを (5.7) に代入して

$$\frac{1}{n}\sum_{i=1}^n \log X_i - \log\bar{X} = \psi(\alpha) - \log\alpha,$$

すなわち

$$R_n := \log\frac{\bar{X}}{\left(\prod_{i=1}^n X_i\right)^{1/n}} = \log\alpha - \psi(\alpha) \tag{5.9}$$

になり，(5.9) の α の解を $\hat{\alpha} = \hat{\alpha}(\boldsymbol{X})$ とする．ここで，R_n は $X_1, \ldots, X_n$ の算術平均と幾何平均の比の対数であることに注意．そして，これを (5.8) に代入して $\hat{\beta} := \hat{\beta}(\boldsymbol{X}) = \bar{X}/\hat{\alpha}$ とすれば，$(\hat{\alpha}, \hat{\beta})$ は (α,β) の最尤推定量になる ([JKB94])．実際には，$X_i = x_i$ $(i = 1, \ldots, n)$ のとき $\hat{\alpha}(\boldsymbol{X})$, $\hat{\beta}(\boldsymbol{X})$ の実現値を $\hat{\alpha}^*$, $\hat{\beta}^*$ として，$X_1, \ldots, X_n$ をたがいに独立に，いずれも (5.6) の p.d.f. $p(x;\hat{\alpha},\hat{\beta})$ をもつガンマ分布 $G(\hat{\alpha},\hat{\beta})$ に従うと見なし，その c.d.f. を $F(x;\hat{\alpha}^*,\hat{\beta}^*)$ として

$$Y_i := -\log\left(1 - F(X_i;\hat{\alpha}^*,\hat{\beta}^*)\right) \quad (i = 1, \ldots, n) \tag{5.10}$$

と変数変換して，第 2 節の指数分布の場合に帰着させることができる．

例 5.4 ある p.d.f. $p(\cdot;\alpha,\beta)$ をもつガンマ分布から大きさ 100 の無作為標本を ら記録値を求めると

$$1.01357, 6.38898, 7.54277, 8.27365, 9.26006, 10.4708, 11.3563, 12.9$$

となった．このとき，まず，このガンマ分布の p.d.f. の母数 α, β の推定値を上記 それぞれ $\hat{\alpha}^* = 3.30488$, $\hat{\beta}^* = 1.43078$ となる．

次に，p.d.f. $p(\cdot;\hat{\alpha}^*,\hat{\beta}^*)$ に従う確率変数の変換 (5.10) を用いると，上記の記録

$$0.0212516, 1.49216, 1.99346, 2.32937, 2.80147, 3.40568, 3.8621, 4.7$$

となる．この記録値において，6 番目の記録値 $Y_6 = 3.40568$ から 8 番目の記録値 予測する予測区間を第 2 節の各方法で構成し，さらに (5.10) の逆変換で元に戻す に基づく信頼係数 $1-\alpha$ の $X_8(=12.9817)$ の予測区間は次の表のようになる．

$1-\alpha$	(A)	(W)	(H)	(P)	(B)
95%	[10.740, 16.291]	[7.469, 19.086]	[10.725, 18.912]	[10.518, 15.484]	[10.542, 16
90%	[10.865, 15.463]	[8.124, 17.846]	[10.846, 17.229]	[10.564, 14.640]	[10.597, 14
80%	[11.059, 14.597]	[8.933, 16.489]	[11.039, 15.680]	[10.657, 13.766]	[10.697, 13
70%	[11.225, 14.068]	[9.513, 15.616]	[11.208, 14.825]	[10.752, 13.238]	[10.794, 13

以上では，切断指数分布，ワイブル分布，ガンマ分布の場合に記録値の区間予測 的な検討の結果，Wald 法による予測区間が他の方法によるものと比べて，やや区 る傾向にある．他の方法による予測区間は，それほど大差はないと言える．

参考文献

[Ah78] Ahsanullah, M. (1978). Record values of exponentially distributed ra ables. *Ann. Inst. Statist. Math.*, **30**, 429–433.

[Ah80] Ahsanullah, M. (1980). Linear prediction of record values for the two exponential distribution. *Ann. Inst. Statist. Math.*, **32**, 363–368.

[Ak90] Akahira, M. (1990). *Theory of Statistical Prediction*. Lecture Note at East Technical University, Ankara.

[AkH00] Akahira, M. and Hida, E. (2000). Prediction intervals for a discrete family of distributions and its applications. *Istatistik* **3**, 58–82

[ArBN98] Arnold, B., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*. New York.

[AwR00] Awad, A. M. and Raqab, M. Z. (2000). Prediction intervals for the fu values from exponential distribution: comparative study. *J. Statist. Comp* **65**, 325–340.

[Ge93] Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman & York.

[Gu70] Guttman, I. (1970). *Statistical Tolerance Regions: Classical and Bayesian.* Griffin, London.

[H00] Hida, E. (2000). Prediction intervals for a chi distribution with a scale parameter. *Istatistik* **3**, 1–12.

[HAk01] 飛田 英祐, 赤平 昌文. (2001). Prediction intervals for unobserved record values. 科研費シンポジウム「統計的領域推定とそれに関連する手法の開発とその応用」講演予稿集 1–7.

[JKB94] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1 (2nd ed.), Wiley, New York.

[M63] Menon, M. V. (1963). Estimation of the shape and scale parameters of the Weibull distribution. *Technometrics* **5**, 175–182.

[T75] 竹内 啓 (1975). 統計的予測論. 培風館.