CRR DISCUSSION PAPER SERIES J

Discussion Paper No. J-7

# 死亡率統計のスムージングにおける数学的手法の提案

董　普・久保　英也

2010年12月

Center for Risk Research
Faculty of Economics
SHIGA UNIVERSITY

1-1-1 BANBA, HIKONE,
SHIGA 522-8522, JAPAN

滋賀大学経済学部附属リスク研究センター
〒522-8522滋賀県彦根市馬場1-1-1

# 死亡率統計のスムージングにおける数学的手法の提案

董　　　普（東北財経大学金融学院）
久保　英也（滋賀大学経済学研究科）

中国の経済発展に伴い、保険が国民日常生活に占める位置はますます重要となってきた。近代的中国保険制度の発足は遅いものの、近年には高い成長を示している。中国における生命保険市場は大きく変化し、今後人口の高齢化に伴う死亡保険市場の成熟化と企業年金など生存保障市場の成長が見込まれる。

他方、保険の供給者が増加し、市場における競争が一段と厳しくなった。こうした環境変化を背景に、保険会社は競争力を強化するために、より厳密な保険料算出方式の導入が必要となってきている。

死亡保障保険の保険料に影響する大きな要因として、（予定）死亡率と（予定）利率の 2 つがあげられる。とりわけ、死亡率については、リスク細分化保険などにおいて既存の標本ではアクチュアリーが想定する死亡率などのラインが標本数の制約からいびつになる可能性が高い。

本稿では、歪になった標本死亡率から数理計算に使用する死亡率をより正確に算定するための数学的手法のひとつを提案する。

## 1．問題提起

標本から得られた死亡率がいびつになる事態は高齢者ゾーンや死亡率が反転する幼児死亡率において時に観察される。ここでは観測データが歪になりがちな 2 つの例示を行う。

（1）例示 1

生命保険会社は保険の引き受けに際し、選択を行う。選択効果は時間と共に低減することから次の関係が成り立つ。

$x$ 歳かつ選択された人のその以降年度の死亡率は $q_{[x]+j}, j=0,1,\cdots$ とする。

その中、$x$ 歳は選択年齢であり、$q_{[x]+j}$ は $x$ 歳かつ選択された人の現在年齢が $x+j$ 歳で、$x+j+1$ 前での死亡確率である(たとえば、$q_{[38]+2}$ は 38 歳かつ選択された人の現在年齢が 40 歳で、41 歳前での死亡確率を表す)。

当然、同じ年齢の死亡率は選択時間の推移に伴って増大しつつある。つまり、

$$q_{[x-k+1]+k-1} \geq \cdots \geq q_{[x-1]+1} \geq q_{[x]} \tag{1.1}$$

(たとえば、 $q_{[35]+5} \geq \cdots \geq q_{[39]+1} \geq q_{[40]}$)

記号を簡素化するため、$x$-$i$+1 歳に選択した人で、$x$ 歳までに生きている $n_i$ 人について、

$\mu_i = q_{[x-i+1]+i-1}, i=1,2,\cdots k.$ とする。

(1.1)式からわかるように、

$$\mu_1 \le \mu_2 \le \cdots \le \mu_k \qquad (1.2)$$

その意味するところは、同じ $x$ 歳の人に対して、年齢が $x$-$s$ 歳時点に選択した人は $x$ 歳時選択した人より、時間が経過しているため、死亡率がより高いのである。この $\mu = (\mu_1, \mu_2, \cdots \mu_k)$ は、選択時期による死亡率の動きを表すパラメータである。

$x_{ij}$ は $x$-$i$+1 歳に選択し $x$ 歳時に生きている $n_i$ 人において、$j$ 人目の $x$+1 歳前の生死状況を表す。$j$=1,2,…$n_i$。その中

$$x_{ij} = \begin{cases} 1 & \text{死亡} \\ 0 & \text{生存} \end{cases}$$

かくして、$x_{ij}$ は密度関数が

$$\mu_i^x (1-\mu_i)^{1-x} = \exp\left\{x \ln \frac{\mu_i}{1-\mu_i} + \ln(1-\mu_i)\right\}$$

であるサンプルデータである。

通常、サンプルデータを用いてこの $\mu_i$ を推定する(この推定は制限条件なしの最大尤度推定)

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

しかしながら、多くの推定は $\mu_1 \le \mu_2 \le \cdots \le \mu_k$ を満たさず、それは理論とは異なる。

$n_i \bar{x}_i$ は二項分布に従うことは証明されうるので、尤度関数は以下のとおりである。

$$\prod_{i=1}^{k} \mu_i^{n_i \bar{x}_i} (1-\mu_i)^{n_i(1-\bar{x}_i)} \qquad (1.3)$$

(1.2)に制限される(1.3)の最大値は、(1.2)に制限される $\mu = (\mu_1, \mu_2, \cdots \mu_k)$ の最大尤度推定である。つまり、以下の数学モデルである。

$$\max \prod_{i=1}^{k} \mu_i^{n_i \bar{x}_i} (1-\mu_i)^{n_i(1-\bar{x}_i)}$$

$$\text{st } \mu_1 \le \cdots \le \mu_k \qquad (1.4)$$

ここまでの論述からわかるように、$x$ 歳かつ選択された人のその以降年度の死亡率を予測することは、この数学モデルの解を求める作業となる。

（2）例示 2

$\mu_x$ を $x$ 歳人の 1 年以内の死亡率とする。かくして、 $\mu_x \le \mu_{x+1} \le \cdots \le \mu_{x+k-1}$ . である。

たとえば、$\mu_{70} \le \mu_{71} \le \cdots \le \mu_{84}$ は 70 歳の人の今後 1 年以内の死亡率は 71 歳の人の同死亡率より小さいことを意味する。つまり、

$x_{xj}$ は、$x$ 歳の $n_x$ 人において、$j$ 人目の 1 年以内の生死状況を表す。$j$=1,2,… $n_x$。その中、

$$x_{xj} = \begin{cases} 1 & \text{死亡} \\ 0 & \text{生存} \end{cases}$$

かくして、 $x_{ij}$ は密度関数が

$$\mu_x^y (1-\mu_x)^{1-y} = \exp\left\{ y \ln \frac{\mu_x}{1-\mu_x} + \ln(1-\mu_x) \right\}$$

であるサンプルデータである。

問題提起 1 の検討と同様に、死亡率 $\mu_x, \mu_{x+1}, \cdots, \mu_{x+k-1}$ への推定は数学モデルの解を求める作業となる。

$$\max \prod_{i=x}^{x+k-1} \mu_i^{n_i \bar{x}_i} (1-\mu_i)^{n_i(1-\bar{x}_i)}$$

$$\text{st}\ \mu_1 \le \cdots \le \mu_{x+k-1}$$

ここでは、 $\bar{x}_y = \frac{1}{n_y} \sum_{j=1}^{n_y} x_{yj}$

**2. 問題解決**

これらの問題の解決に向け、Isotonic regression を用いる。

（1）Isotonic regression の定義

$X = (x_1, x_2, \cdots, x_k)$ は既定関数とする。もし $X^* = (x_1^*,\ x_2^*,\ \cdots, x_k^*) \in G$ 、かつ

$$\sum_{i=1}^{k} (x_i - x_i^*)^2 w_i = \min_{Y \in G} \sum_{i=1}^{k} (x_i - y_i)^2 w_i$$

であれば、 $X^*$ は simple semi-order の下での($X$, $W$)における Isotonic regression である。ここ

で、$G=\{y_1,\cdots,y_k)\,|\,y_1\le\cdots\le y_k\}$、$W=(w_1,w_2,\cdots,w_k)$、$w_i>0$ は既定の weighted functions である。

（２）Isotonic regression の解

Isotonic regression の解を求める最適な方法に Ayer et al(1955)の提案したた PAVA 計算方法がある。

$B$ を $K=(1, 2, \cdots, k)$の部分集合とする。$A_V(B)=\sum_{i\in B}x_i w_i/\sum_{i\in B}w_i$。

① もし $x\in G,$ であれば、$x^*=x$。

② もし $x_j>x_{j+1},$ を成立させる j が存在すれば、B={j,j+1}, $x_B=A_V(B),\omega_B=\omega_j+\omega_{j+1}.$ として、

$\tilde{x}=(x_1,\cdots,x_{j-1},x_B,x_{j+1},\cdots,x_k),\tilde{\omega}=(\omega_1,\cdots,\omega_{j-1},\omega_B,\omega_{j+1},\cdots,\omega_k),$ とする。

③ ②の繰り返しを通して、K を $l$ 個の $B_1,\cdots B_l,$ に分解し、$A_V(B_1)<\cdots<A_V(B_l).$ を満たす。かくして、

$x_i^*=A_V(B_t),i\in B_t,t=1,\cdots,l$。

この方法から得られる解が、Isotonic regression の解である。その証明は以下の通り、先行研究によりなされている。

ある$\sigma$有限測度は以下の密度関数の指数分布族を有する。

$$f(y;\theta,\tau)=\exp\{f_1(\theta)f_2(\tau)H(y,\tau)+s(y,\tau)+q(\theta,\tau)\} \qquad (2.1)$$

その中、$y$ はある集合 $A$ から取得し、$\theta$ はある区間 $(\underline{\theta},\ \overline{\theta})$に属する。$\tau\in T$ は嫌なパラメータである。$\theta$ の推計問題を考慮するため、以下の条件が成立すると仮定する。

1、$f_1$ と $q(\cdot,\tau)$は、$(\underline{\theta},\ \overline{\theta})$ での 2 次連続導関数を有する。

2、$f_1'(\theta)>0,\forall\theta\in(\underline{\theta},\ \overline{\theta})$, $f_2(\tau)>0,\ \forall\tau\in T$

3、$q'(\theta,\tau)=-\theta f_1'(\theta)f_2(\tau),\forall\theta\in(\underline{\theta},\ \overline{\theta}),\ \forall\tau\in T$

上記の条件から、

$E(H(Y,\tau))=\theta$、 $\mathrm{Var}(H(Y,\tau))=(f_1'(\theta)f_2(\tau))^{-1}$ が得られる。

今、仮に $k$ グループがある。$y_{ij}$、$j=1, \cdots, n_i$ を密度関数が $f(y_i;\theta_i,\tau_i)$であるサンプルデータとする。$i=1, \cdots, k$。条件 1、2、3 から、$\theta_i$ の最大尤度推定は以下のようになる。

$$\hat{\theta}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}H(y_{ij},\tau_i) \qquad (2.2)$$

$\theta=(\theta_1,\theta_2,\cdots,\theta_k)'$ とする。もし $\theta$ の分量が semi-order に制限されれば、$\theta$ の最大尤度推定は以下の結果に示される。

定理：semi-order 制限がある場合、$\theta$ の最大尤度推定は $\hat{\theta}$ の Isotonic regression である。その中、$\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)'$、$\theta_i$ は(2.2)から算出される。

weighted functions は　$\omega=(n_1 f_2(\tau_1),\cdots,n_k f_2(\tau_k))'$ である。

これを、以下のモデルに適用すると

$$\max\prod_{i=1}^{k}\mu_i^{n_i\bar{x}_i}(1-\mu_i)^{n_i(1-\bar{x}_i)}$$

$$st\ \mu_1\leq\cdots\leq\mu_k$$

この解を求める問題は、$\mu_1\leq\mu_2\leq\cdots\leq\mu_k$ 制限の下で、$\sum_{i=1}^{k}n_i(\bar{x}-\mu_i)^2$ の最小値を求める問題と同じである。それは $X=(\bar{x}_1,\cdots\bar{x}_k)$ の Isotonic regression と同じであり、その中の weighted functions は $N=(n_1,n_2,\ldots n_k)$ である。

証明：$X=\begin{cases}1 & \text{死亡}\\ 0 & \text{生存}\end{cases}$ とする。その中、死亡の確率は $\mu$ である。密度関数は

$$\mu^x(1-\mu_x)^{1-x}=\exp\left\{x\ln\frac{\mu}{1-\mu}+\ln(1-\mu)\right\}$$

となる。

(2．1)式に対応して、ここの $\mu$ を $\theta$ にみなし、$x$ を $y$ にみなしている。$s(x,\tau)=0$、$q(\mu,\tau)=\ln(1-\mu)$、$H(x,\tau)=x$、$f_2(\tau)=1$、$f_1(\mu)=\ln(\mu/(1-\mu))$。

1、$f_1(\mu)=\ln(\mu/(1-\mu))$ と $q(\mu,\tau)=\ln(1-\mu)$ は、(０，１) での 2 次連続導関数を有する。

2、$f_1'(\mu)=\dfrac{1}{\mu(1-\mu)}>0,\forall\mu\in$ (０，１), $f_2(\tau)=1>0,\ \forall\tau\in T$

3、$q'(\mu,\tau)=-\dfrac{1}{(1-\mu)}$ であるため、$q'(\mu,\tau)=-\mu f_1'(\mu)f_2(\tau),\forall\mu\in$ (０，１) $\forall\tau\in T$ である。

公準の条件を満たしている。(2.2)式に対応して、$\hat{\theta}_i=\dfrac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}=\bar{x}_i,\quad i=1,\cdots,k.$ である。

公準からわかるように、

$$\max\prod_{i=1}^{k}\mu_i^{n_i\bar{x}_i}(1-\mu_i)^{n_i(1-\bar{x}_i)}$$

$$st\ \mu_1 \le \cdots \le \mu_k$$

　である。

したがって、前述の通り、この解を求める問題は、 $\mu_1 \le \mu_2 \le \cdots \le \mu_k$ 制限の下で、 $\sum_{i=1}^{k} n_i(\bar{x}-\mu_i)^2$ の最小値を求める問題と同じである。それは $X=(\bar{x}_1,\cdots\bar{x}_k)$ の Isotonic regression とは同じであり、その中の weighted functions は $N=(n_1,n_2,\ldots n_k)$である。

（３）Miller の事例

　ここでは、有名な Miller の事例を紹介し、上記の意味を具体的に考えることとする。表 １ に見るような年齢別の被保険者の観測データが存在すると仮定する。そこから観測死亡率を算出する。

表１　観測死亡率表の事例

| 年齢（歳）<br>$x$ | 観測データ数<br>$n_x$ | 観測死亡数<br>$h_x$ | 観測死亡率<br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71 | 143 | 12 | 0.084 |
| 72 | 140 | 10 | 0.071 |
| 73 | 144 | 11 | 0.076 |
| 74 | 149 | 6 | 0.040 |
| 75 | 154 | 16 | 0.104 |
| 76 | 150 | 24 | 0.160 |
| 77 | 139 | 8 | 0.058 |
| 78 | 145 | 16 | 0.110 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

　先験的に、死亡率の序列は、①年齢を重ねるにつれて増加、② 高齢者ゾーンにおいて、死亡率曲線は著しく上昇、とされる。しかしながらこの表に示した観測死亡率は 72 歳、74 歳、77 歳、79 歳で想定死亡率のラインとは異なるものとなっている。

　そこで、本来、単調増加（もしくは、単調減少）であるデータ系列に序列が逆転するよ

うな事象が発生した場合には、これを均す作業を行う。これをスムージングと呼ぶ。
PAVA 法によりスムージングを行うと表 2 の結果が得られる。具体的には 71 歳と 72 歳のデータの平均を算出したものである（ウェートは 1：1)。

表 2　観測死亡率の修正（１）

| 年齢（歳）<br>$X$ | 観測データ数<br>$n_x$ | 観測死亡数<br>$h_x$ | 観測死亡率<br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~72 | 283 | 22 | 0.078 |
| 73 | 144 | 11 | 0.076 |
| 74 | 149 | 6 | 0.040 |
| 75 | 154 | 16 | 0.104 |
| 76 | 150 | 24 | 0.160 |
| 77 | 139 | 8 | 0.058 |
| 78 | 145 | 16 | 0.110 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

しかしながら、71 歳～72 歳、73 歳、74 歳の間で依然、逆転現象が起こっているため、さらにスムージングを進める。その結果を表 3 に示した。

表 3　観測死亡率の修正（2）

| 年齢（歳）<br>$x$ | 観測データ数<br>$n_x$ | 観測死亡数<br>$h_x$ | 観測死亡率<br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~73 | 427 | 33 | 0.077 |
| 74 | 149 | 6 | 0.040 |

| | | | |
|---|---|---|---|
| 75 | 154 | 16 | 0.104 |
| 76 | 150 | 24 | 0.160 |
| 77 | 139 | 8 | 0.058 |
| 78 | 145 | 16 | 0.110 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

まだ、71～73 歳と 74 歳の間で逆転現象が見られるため、さらにスムージングを進めた結果が表 4 である。

表 4　観測死亡率の修正（3）

| 年齢（歳） $x$ | 観測データ数 $n_x$ | 観測死亡数 $h_x$ | 観測死亡率 $u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~74 | 576 | 39 | 0.068 |
| 75 | 154 | 16 | 0.104 |
| 76 | 150 | 24 | 0.160 |
| 77 | 139 | 8 | 0.058 |
| 78 | 145 | 16 | 0.110 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

これにより、70 歳前半の死亡率の逆転現象は解消したが、未だ、77 歳において逆転現象が見られる。そこでこのランクについてスムージングを行う。その結果を表 5 に示した。

表 5　観測死亡率の修正（4）

| 年齢（歳）<br><br>$x$ | 観測データ数<br><br>$n_x$ | 観測死亡数<br><br>$h_x$ | 観測死亡率<br><br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~74 | 576 | 39 | 0.068 |
| 75 | 154 | 16 | 0.104 |
| 76~77 | 289 | 32 | 0.111 |
| 78 | 145 | 16 | 0.110 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

同様に、76～77 歳、78 歳、79 歳における逆転現象についても、表 6 の通りスムージングを行う。

表 6　観測死亡率の修正（5）

| 年齢（歳）<br><br>$x$ | 観測データ数<br><br>$n_x$ | 観測死亡数<br><br>$h_x$ | 観測死亡率<br><br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~74 | 576 | 39 | 0.068 |
| 75 | 154 | 16 | 0.104 |
| 76~78 | 434 | 48 | 0.111 |
| 79 | 140 | 13 | 0.093 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

同様に、76～78 歳ゾーンと 79 歳の逆転現象にスムージングを施す。これによるすべての年齢区分において逆転現象は表 7 の通り、解消することになる。

表 7　観測死亡率の修正（6）

| 年齢（歳）<br><br>$x$ | 観測データ数<br><br>$n_x$ | 観測死亡数<br><br>$h_x$ | 観測死亡率<br><br>$u_x$ |
|---|---|---|---|
| 70 | 135 | 6 | 0.044 |
| 71~74 | 576 | 39 | 0.068 |
| 75 | 154 | 16 | 0.104 |
| 76~79 | 574 | 61 | 0.106 |
| 80 | 137 | 19 | 0.139 |
| 81 | 136 | 21 | 0.154 |
| 82 | 126 | 23 | 0.183 |
| 83 | 126 | 26 | 0.206 |
| 84 | 109 | 26 | 0.239 |

これにより、観測死亡率は理論や経験に沿った系列となる。しかしながら、このようなウェートを所与とした人為的なスムージングは、観測データの持つ意味そのものを消失させる可能性はないのであろうかという疑問が生じる。この点は冷静に判断する必要がある。

## 3．死亡率における新たなスムージング手法の提案

まず、先行研究において、スムージングの定義を見てみよう。

（1）Andrews および Nesbitt の定義：法則性を有する自然現象の若干の観測データから、規則的な修正を通して、その現象を代表する系列を求めること。

（2）Miller の定義：ある不規則な序列を有する連続変数から、規則的な修正序列が得られるように連続変数を修正し、得られた系列は観測データの基本的性格を維持していること。

これらの定義には次の基本的な考え方がある。すなわち、観測データは常に不規則な序列であり、かつ、これらのデータは修正されるべきと考えられるが、修正後のデータ系列は観測データから大きくは乖離していないことが重要である。この状態を、①「Fit」という。

これに加え、アクチュアリーは死亡率の修正に際し、②常に先験情報を使い、③そのデータ系列は円滑な序列である、と信じている。

これらの 3 要素を数学的に表現すると以下の通りとなる。

① Fit：

$$\frac{\frac{(\bar{X}_0-p_0)^2}{p_0(1-p_0)}}{n_0}+\frac{\frac{(\bar{X}_1-p_1)^2}{p_1(1-p_1)}}{n_1}+\frac{\frac{(\bar{X}_2-p_2)^2}{p_2(1-p_2)}}{n_2}+\cdots\frac{\frac{(\bar{X}_{104}-p_{104})^2}{p_{104}(1-p_{104})}}{n_{104}} \quad (1.1)$$

② 先験情報：　$p_0 \geq p_1 \geq p_2 \geq \cdots \geq p_{11} \leq p_{12} \leq \cdots \leq p_{104} < 1$　（1.2）

③ 円滑性：　$S=\sum_i (\Delta^4 p_i)^2$　（1.3）

すなわち、(1.2)の制限条件の下で、(1.1)の最小値を求めると共に、(1.3)が比較的に小さいことを検証する必要がある。詳しくは、以下の数学モデル A を解くことになる。

$$\min\{\frac{\frac{(\bar{X}_0-p_0)^2}{p_0(1-p_0)}}{n_0}+\frac{\frac{(\bar{X}_1-p_1)^2}{p_1(1-p_1)}}{n_1}+\frac{\frac{(\bar{X}_2-p_2)^2}{p_2(1-p_2)}}{n_2}+\cdots\frac{\frac{(\bar{X}_{104}-p_{104})^2}{p_{104}(1-p_{104})}}{n_{104}}\} \quad \text{（モデル A）}$$

St: $0< p_0 \geq p_1 \geq p_2 \geq \cdots \geq p_{11} \leq p_{12} \leq \cdots \leq p_{104} < 1$

ただし、$\bar{X}_i$ は $i$ 母体目の粗死亡率をあらわし、$p_i$ は $i$ 母体目の死亡率である。

この計算を簡素化するために、以下の計算方法を提案する。

すなわち、

$$\min\{\frac{\frac{(\bar{X}_1-p_1)^2}{p_1(1-p_1)}}{n_1}+\frac{\frac{(\bar{X}_2-p_2)^2}{p_2(1-p_2)}}{n_2}+\cdots\frac{\frac{(\bar{X}_k-p_k)^2}{p_k(1-p_k)}}{n_k}\} \quad (1.4)$$

St:$0< p_1 \leq p_2 \leq \cdots \leq p_k < 1$ の時、(1.4)式は

$$\frac{n_1(\bar{X}_1-p_1)^2}{p_1(1-p_1)}+\frac{n_2(\bar{X}_2-p_2)^2}{p_2(1-p_2)}+\cdots\frac{n_k(\bar{X}_k-p_k)^2}{p_k(1-p_k)}$$ となる。　(1.5)

ここで、公準 1：$0 \leq x \leq 1$ のとき、方程式 $(1-2x)u^2+2x^2u-x^2=0$ は、2 つの根を持つ。

すなわち、$U_1=x,$, $U_2=\frac{x}{2x-1}$　$g(u)=(1-2x)u^2+2x^2u-x^2$ とする。

かくして、$0<u<x$ の時、$g(u)<0$

$x<u<1$ の時、$g(u)>0.$ である。これは、二次関数を用いて証明できる。

次に、公準 2： 関数 $f(u)=\frac{n(x-u)^2}{u(1-u)}$ において、$0<x<1$ の時、$f(u)$ は $(0,x)$ において

単調減少し、$x<u<1$の時に単調増加する。そして $u=x$ のときには最小値となる。

証明：$f'(u)=\dfrac{n}{u^2(1-u)^2}\{(u-x)[(1-2x)u+x]\}$

もし、$0\le x\le 1/2$ であれば、$(1-2x)u+x>0$。

$0<u<x$ の場合、$f'(u)<0$、 $x<u<1$ の場合、$f'(u)>0$。したがって、$f(u)$ は$(0,x)$においてて単調減少し、$x<u<1$の時に単調増加する。そして $u=x$ のときには最小値となる。

もし $1/2<x\le 1$ であれば、$\{\dfrac{x}{2x-1}\}'=-\dfrac{1}{(2x-1)^2}<0$、$\dfrac{x}{2x-1}$は単調減少する。

したがって、$\dfrac{x}{2x-1}>1>x$。

公準 1 が示すように、$0<u<x$ のとき、$(u-x)<0,[(1-2x)u+x]>0$。$f(u)$ は$(0,x)$において単調減少である。$x<u<1$のとき、同様に公準 1 から、$(u-x)>0,[(1-2x)u+x]>0$。$f(u)$ は$x<u<1$の時、単調増加である。$u=x$ のときには最小値となる。

定理 1：仮に関数$f(p_1,p_2,\cdots,p_k)=\dfrac{(\bar{X}_1-p_1)^2}{\frac{p_1(1-p_1)}{n_1}}+\dfrac{(\bar{X}_2-p_2)^2}{\frac{p_2(1-p_2)}{n_2}}+\cdots\dfrac{(\bar{X}_k-p_k)^2}{\frac{p_k(1-p_k)}{n_k}}$は、

$0<p_1\le p_2\le\cdots\le p_k<1$ の制限の下での最小値点$p_1^*\le p^*_2\le\cdots\le p^*_k$を満たす。もし$\bar{X}_j>\bar{X}_{j+1}$であれば、$p^*_j=p^*_{j+1}$である。

仮に$p^*_j<p^*_{j+1}$　を仮定すると、以下の 3 つの可能性がある。

1° $p^*_j<p^*_{j+1}\le\bar{X}_j$

2° $p^*_j\le\bar{X}_{j+1}<\bar{X}_j<p^*_{j+1}$

3° $\bar{X}_{j+1}<p^*_j<p^*_{j+1}$

まず、1° $p^*_j<p^*_{j+1}\le\bar{X}_j$の場合、

$y^*_1=p_1^*,\ y^*_2=p^*_2,\cdots,y^*_{j-1}=p^*_{j-1},\ y^*_{j+1}=p^*_{j+1},\cdots,y^*_k=p^*_k$とする。

$y_j^*$ は $p_j^*<y_j^*<p_{j+1}^*$

を満たす任意値であり、公準 1 から、

$$f(y^*_1,y^*_2,\cdots,y^*_k)-f(p^*_1,p^*_2,\cdots,p^*_k)<0,$$

$p_1^* \le p^*_2 \le \cdots \le p^*_k$ である。

$$f(p_1, p_2, \cdots, p_k) = \frac{(\bar{X}_1 - p_1)^2}{\frac{p_1(1-p_1)}{n_1}} + \frac{(\bar{X}_2 - p_2)^2}{\frac{p_2(1-p_2)}{n_2}} + \cdots \frac{(\bar{X}_k - p_k)^2}{\frac{p_k(1-p_k)}{n_k}}$$ は、

$0 < p_1 \le p_2 \le \cdots \le p_k < 1$ の制限の下での最小値点を満たされず、仮定と矛盾する。

次に、公準 3： $0 < \mathrm{a_i} < 1, \mathrm{i} = 1, \cdots, \mathrm{m}$ のとき、方程式

$$((\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i} - 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i})u^2 + 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2 u - \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i} = 0$$ は 2 つの根がある。

$$u_1 = \frac{-\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2 - \sqrt{(\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2)\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i (a_i - 1)^2}}}{\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i} - 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}},$$

$$u_2 = \frac{-\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2 + \sqrt{(\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2)\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i (a_i - 1)^2}}}{\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i} - 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}}$$ 。

$\mathrm{g(u)} = ((\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i} - 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i})u^2 + 2\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}^2 u - \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{n_i a_i}$ とする。かくして、 $0 < \mathrm{u} < \mathrm{u_2}$ の時、 $g(u) < 0$。$u_2 < u < 1$の時、 $g(u) > 0.$　これも二次関数を用いて、証明できる。

定理 2：関数 $f\,(\mathrm{u}) = \dfrac{\sum_{\mathrm{i=1}}^{\mathrm{m}} n_{\mathrm{i}}(\mathrm{a_i} - \mathrm{u})^2}{\mathrm{u(1-u)}}$ について、 $0 < \mathrm{a_i} < 1, \mathrm{i} = 1, \cdots, \mathrm{m}$　とした場合、

$f\,(\mathrm{u})$ は $0 < \mathrm{u} < \mathrm{u_2}$ において単調減少し、 $u_2 < u < 1$ において単調増加する。そして、 $u$= $\mathrm{u_2}$ のとき、最小値となる。

公準 3 と公準 2 の証明と同様に、以下の式が得られる。

$$\min\{\frac{(\bar{X}_1 - p_1)^2}{\frac{p_1(1-p_1)}{n_1}} + \frac{(\bar{X}_2 - p_2)^2}{\frac{p_2(1-p_2)}{n_2}} + \cdots \frac{(\bar{X}_k - p_k)^2}{\frac{p_k(1-p_k)}{n_k}}\}$$

$$\mathrm{St:}0 < p_1 \le p_2 \le \cdots \le p_k < 1$$

この式は以下のように書く変えることができる。これをモデル B とする。

$$\frac{n_1(\overline{X}_1-p_1)^2}{p_1(1-p_1)}+\frac{n_2(\overline{X}_2-p_2)^2}{p_2(1-p_2)}+\cdots\frac{n_k(\overline{X}_k-p_k)^2}{p_k(1-p_k)} \qquad \text{（モデル B）}$$

$$\text{St:}0< p_1 \le p_2 \le \ \cdots \ \le p_k < 1$$

このモデル B の解を求める計算方法をここで提案する。

1°　もし $\overline{X}_1 \le \overline{X}_2 \le \cdots \le \overline{X}_k$ であれば、$(p_1^*, p_2^* \cdots, p_k^*) = (\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_k)$ である。成立しなければ、次に転入する。

2° もし $\overline{X}_j > \overline{X}_{j+1}$ が存在すれば、かつ

$$\overline{X}_1 \le \cdots \le \overline{X}_{j-1} \le \frac{-\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2} + \sqrt{(\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2}) \sum_{i=j}^{j+1} n_i (\overline{x}_i - 1)^2}}{\sum_{i=j}^{j+1} n_i - 2\sum_{i=j}^{j+1} n_i \overline{x}_i} \le \overline{X}_{j+2} \le \cdots \le \overline{X}_k \text{ であれば、}$$

$$(p_1^*, p_2^* \cdots, p_{j-1}^*) = (\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_{j-1}),\ p_{j+1}^* = p_j^* = \frac{-\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2} + \sqrt{(\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2}) \sum_{i=j}^{j+1} n_i (\overline{x}_i - 1)^2}}{\sum_{i=j}^{j+1} n_i - 2\sum_{i=j}^{j+1} n_i \overline{x}_i} \text{ である。}$$

$(p_{j+2}^*, \cdots, p_k^*) = (\overline{X}_{j+2}, \cdots, \overline{X}_k)$ が成立しなければ、

$$B = \{j, j+1\},\ \overline{X}_B = \frac{-\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2} + \sqrt{(\sum_{i=j}^{j+1} n_i \overline{x}_i^{\,2}) \sum_{i=j}^{j+1} n_i (\overline{x}_i - 1)^2}}{\sum_{i=j}^{j+1} n_i - 2\sum_{i=j}^{j+1} n_i \overline{x}_i} \text{ である。}$$

$(\overline{X}_1, \overline{X}_2, \cdots \overline{X}_{j-1}, \overline{X}_B, \overline{X}_{j+2}, \cdots, \overline{X}_k)$ に対して、次に転入する。

3° もし $\overline{X}_{B_s} > \overline{X}_{B_{s+1}}$ であれば、

$$B = B_S \cup B_{S+1},\ \overline{X}_B = \frac{-\sum_{i \in B_S \cup B_{S+1}} n_i \overline{x}_i^{\,2} + \sqrt{(\sum_{i \in B_S \cup B_{S+1}} n_i \overline{x}_i^{\,2}) \sum_{i \in B_S \cup B_{S+1}} n_i (\overline{x}_i - 1)^2}}{\sum_{i \in B_S \cup B_{S+1}} n_i - 2\sum_{i \in B_S \cup B_{S+1}} n_i \overline{x}_i} \text{ である。}$$

3 の繰り返しを通して、集合を m 個の $B_1$，$B_2$,⋯，$B_m$ に分解し、$\overline{X}_{B_1} \le \overline{X}_{B_2} \le \cdots \le \overline{X}_{B_m}$ を満たさせる。かくして、$p_i^* = \overline{X}_{B_t}, i \in B_t, i = 1, \cdots, k$ となる。

これは、定理 1 と定理 2 から証明でき、k 回の計算により、最終解を得ることができる。

このように、スムージングを数学的手法により行えば、原系列の特質を壊すことなく、先見情報と齟齬の無い、円滑なデータ系列を得ることができる。

## 4. 結語

観察データの少ない中国の生命保険業界において、正確な死亡率の算出が重要であることは論をまたず、今回提案したスムージングの手法は貢献度も大きい。

一方、日本においても生命表作成時に裁断とスムージングが用いられている。具体的には、収集した観察データに若年齢層を補整し粗死亡率を決める。これに数学的危険論による第一次補整、Greville による第 2 次補整、そして Gomperlz-Mkeham による第 3 次補整を行い、生命保険標準生命表が作成されている。主に、標本の収集が相対的に少ない乳幼児のゾーンと高齢者ゾーンに適用され、第 2 次補整と第 3 次補整は、偶然変動を除去するために所与のパラメータを使用したスムージングを行っている。ただ、前述した Fit、先験情報との調和、少ない乖離の 3 要件は導入当初満たされていたとしても現在も満たしているかなど検証の情報が少ない中で使用されている。今回提案した数学的なスムージングの方策はその情報を補完する 1 つの手段候補となる可能性がある。

生命保険の市場構造や研究者の研究視点の異なる中国の研究者との共同研究は、日本の研究者では気づきにくい新たな課題を明確にしてくれる。今後とも異質な視点を取り込み日本の保険研究の水準の向上に貢献できれば幸いである。

以　上

参考文献（中国語）


1. 盧玉貞・董普(2004)、「多維保序回帰及最大似然估計」、『大連海事大学学報』12 号、Vol.102。
2. 董普・盧玉貞(2003)、「正態分布的均値被簡単樹半序約束方差為未知的最大似然估計」、『数学的実践和認識』11 号、Vol.33。
3. 董普(2003)、「多維正態分布均値在序約束下的仮説検験」、『数学進展』ISSN 1000-0917、Vol.32、No.1。
4. 董普・盧玉貞(2003)、「正態分布的均値和方差別分別被簡単樹半序和簡単半序約束下的保序最大似然估計」、『応用概率統計』ISSN 1001-4268、Vol.19、No.1。
5. 盧玉貞・董普(2002)、「選択死亡率的一種估計方法」、『中国高教論衆』

ISBNT-5437-3063-4/G.2788、Vol.22。

6. 董普・盧玉貞(2001) 、「多維保序回帰和正態分布均値和方差的估計」、『中国高教論衆』ISBN7-5437-3003-O/G·2751、Vol.22、No.2。

参考文献（その他）

[1] SHI, N.Z. [1993] , " Isotonic regression and maximum likelihood estimate " , *Chinese Journal of Applied Probability and statistics* Vol.9, NO.2, May (1993) pp.203-215.

[2] SHI, N.Z. [1993] , " Maximum likelihood of isotonic normal means with unknown variances " *The Journal of Multivariate Analysis* (1993), Vol.64, pp.183-195.

[3] Barlow, R.E, Bartholoremew, D.J., Bremner, J.M. and Brunk, H.D.[1972] , " Statistical Inference under Order Restrictions " The Theory and Application of Isotonic Regression. New York: Wiley,.

[4] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. [1955] , " An Empirical Distribution Function for Sampling with Incomplete Information " *The Annals of Mathematical Statistics.*,Vol.26, pp.641-647.