試験のためだけの損保数理
Bühlmann モデル

pseudomathematician

平成 27 年 5 月 24 日

1　はじめに

アクチュアリーの実務では，直近の実績に基づいて将来を推測することがある．例えば，なんらかのクレーム額を推定するとしよう．そのクレーム額の確率分布が完全にわかっていれば，すなわち平均・標準偏差などがわかっていれば推定は容易であろう．もしわかっていなければ，過去の実績を使うなどして合理的に推定するなどする．

Bühlmann モデルは，この推測方法のひとつを与えるものである．

2　Bühlmann モデルの具体例

Bühlmann モデルの問題例をみていく．

例 1

野球選手の年間ホームラン数 X はある分布に従っているとする．ある選手Aの年間ホームラン数は過去5年間で以下の通りであった．

x_1	x_2	x_3	x_4	x_5
10	20	0	30	20

しかし，各年のホームラン数は，互いに独立とは限らない，すなわち，X はある確率変数 Θ に依存すると考える．今，$\Theta = E(X|\Theta)$ であるとする．このもとで，翌年のAのホームラン数を推定したい．今，Θ の分布を調べるために，選手 200 人に対する年間ホームラン数実績を以下の通り抽出した．

年間ホームラン数	0	1	2	3
選手数	150	30	15	5

このとき，Aの翌年のホームラン数を推定せよ．

もし，Aの過去5年のホームラン数 $X_1, X_2, \cdots, X_5$ が互いに独立であれば，$\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_5}{5}$ を推定量として，$\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_5}{5}$ を推定値とする方法が考えられる．では，この X_i が互いに独立でないような場合，例えば，ホームラン王になった次の年はまったくホームランを打てなくなるような場合である．このような各年の相関関係は，ホームランを打つセンスによるもので，それは Θ が制御していると考えられる．この Θ の分布がわかれば，$E(X) = E(E(X|\Theta))$ が計算できるので，ホームランの推定に強力な材料となろう．例 1 の 200 人のホームラン数の情報は Θ の分布を調べるための情報である．このような問題例はのちほど過去問や参考書を例に学ぶこととする．

まず問題に慣れるため，もう少し簡単な例を見ていく．

例 2

ある人に長方形を n 個描いてもらいその i 個目の面積を X_i とおく．ただし，横の長さは最初に 1 回だけ描き，縦の長さは n 回描いてもらうものとする．縦・横の長さをそれぞれ A_i, Θ とおき，$X_i = A_i \times \Theta$ とする．このもとで，X_{n+1} を推定したい．

例 3

年間のクレーム件数 X が $X \sim \mathrm{Po}(\Theta)$, $\Theta \sim \Gamma(\alpha, \beta)$ とする．ある人を 1 人抽出して，n 年間のクレーム件数 X_i から X_{n+1} を推定したい．

以上の例からもわかるように，$\{X_i\}$ は互いに独立ではないため，実績値の単純平均を推定値とするわけにはいかない．しかし，理論値としての平均 $\mu = E(X)$ が計算できれば（特にこだわらない限り，X_i と同じ分布に従う確率変数を i を省いて X と書く），μ と $\overline{X}$ の加重平均を推定量と考えるのはひとつの方法と言えよう．今，重みを Z と書き，推定量を

$$C = Z\overline{X} + (1 - Z)\mu$$

とする．この形から，実績値が強く信頼されるような分布であれば $Z \approx 1$ となり，実績値の信頼が小さく，理論値が信頼されるようであれば $Z \approx 0$ となろう．こういった意味から Z を **信頼度** と呼ぶことにする．この Z の計算方法を与えるのが Bühlmann モデルである．

3　Bühlmann モデルにおける信頼度の定義

結論から述べる．

Bühlmann モデルによる信頼度

推定対象となるデータを確率変数 X で表し，過去 n 年間の実績を $X_1, X_2, \cdots, X_n$ とする．X はある確率分布に従うパラメータ Θ に依存し，条件付確率変数 $\{X_i|\Theta\}$ は互いに独立とする．このとき，X_{n+1} の推定量を

$$C = Z\overline{X} + (1-Z)\mu$$

と書いたとき，信頼度 Z は

$$Z = \frac{V(E(X|\Theta))}{V(\overline{X})} \underset{V(E(X|\Theta))\neq 0}{=} \frac{n}{n + \dfrac{E(V(X|\Theta))}{V(E(X|\Theta))}}$$

となる．

なぜ Z がこうなるかは教科書を参考にされたい．試験対策という意味では一切必要のない知識なので割愛する．上記の算式だけを覚えていればよい．

さて，$E(X|\Theta)$ がほぼ定数のような状況であれば（Θ にほとんど依存しない，すなわち $\{X_i\}$ はほぼ互いに独立），このとき $V(E(X|\Theta)) \approx 0,\ C \approx \mu$ となり理論値 μ が採用される．逆に，Θ に強く依存するような場合で，$V(E(X|\Theta))$ が非常に大きい値であれば $Z \approx 1,\ C \approx \overline{X}$ となり実績値 $\overline{X}$ のみが採用されることになる．以下で問題を見ていく．

4　Bühlmann モデルにおける信頼度の計算

信頼度の計算において，以下のパターンがある．

1. X の分布が完全に判明している場合．
2. X の分布が一部のみ判明している場合．
3. X の分布が一切わかっていない場合（ノンパラメトリック）．

まず簡便のため記号を

$$v = E(V(X|\Theta)),\ w = V(E(X|\Theta)),\ \mu = E(X)$$

と定義する．問題によっては若干記号の使い方が変わることがあるが文脈で判断できる場合はいちいちことわらないことにする．

パターン 1 については，Z の計算は容易だろう．パターン 2 については，$v, w, \mu, n, \overline{x}$ を問題で与えられた仮定から合理的に推測･計算しなくてはいけない．パターン 3 については，分布の情報が何もないので，$v, w, \mu, n, \overline{x}$ は全て実績値から推測する必要がある．まず，教科書の Bühlmann モデルの説明がおかしいのは，パターン 2 と 3 について，v, w, μ を推測するための実績値テーブルと，$n, \overline{x}$ を計算するための実績値テーブルについて，記号が混合するなど，しっかりと区別されていないことにある．このわかりにくさを解消すべく，実際の問題例を通じて整理していく．以下で，v, w, μ を推測するための実績値テーブルを「$v \cdot w \cdot \mu$ テーブル」，$n, \overline{x}$ を計算するための実績値テーブルを「$n \cdot \overline{x}$ テーブル」と呼ぶことにする．

4.1　分布が完全に判明している場合

[教 **3-45.** 練習問題 **3**] 契約者のクレーム件数は期待値 Θ のポアソン分布に従い，さらに，さまざまな契約者で Θ を観察したところ，Θ は確率密度関数が $f(\theta) = 3\theta^{-4}\ (\theta > 1)$ である分布に従うことがわかった．

このとき，1 年目，2 年目の 2 年間で計 20 件のクレームを起こしたある契約者について，Bühlmann モデルを用いて，3 年目のクレーム件数を推定せよ．

[解説] 分布が完全にわかっているため，$v \cdot w \cdot \mu$ テーブルは必要なく，以下の計算が可能となる．

$$v = E(V(X|\Theta)) = E(\Theta) = 1.5,$$
$$w = V(E(X|\Theta)) = V(\Theta) = 0.75,$$
$$\mu = E(X) = E(E(X|\Theta)) = E(\Theta) = 1.5.$$

次に $n \cdot \overline{x}$ テーブルであるが，問題より，$n = 2,\ \overline{x} = 10$ となり，以上より，

$$Z = \frac{2}{2 + \dfrac{1.5}{0.75}} = 0.5,\ C = 0.5 \times 10 + (1 - 0.5) \times 1.5 = 5.75.$$

■

[過去問 **H17**] 無作為に抽出した契約者の年間クレーム件数は期待値 Θ のポアソン分布に従い，Θ は契約者ごとにばらつきがある．今，Θ が確率密度関数 $f(\theta) = a\theta^{-\alpha-1}\ (\theta > 1,\ a > 2)$ で表される分布に従うものとして，過去 n 年間のデータに対して求めた信頼度は 0.4 であった．

この条件を変更して，Θ が確率密度関数 $f(\theta) = \left(\frac{2}{3}a\right)\left(\frac{2}{3}\theta\right)^{-a-1}\ (\theta > \frac{3}{2},\ a > 2)$ で表される分布に従うものとして，過去 n 年間のデータに対して信頼度を求めたとき，この信頼度を求めよ．

[解説]（割愛）■

[例題 **38**] 保険契約者Ａ社の労働災害保険のクレーム件数 N，クレーム額 X はそれぞれパラメータ $\Theta = (\Theta_1, \Theta_2)$ を持つ確率分布に従う，具体的には，

$$P(N = n|\Theta_1 = \theta_1) = e^{-\theta_1}\frac{\theta_1^n}{n!} \quad (n = 0, 1, 2, \cdots)$$
$$f_{X|\Theta_2}(x|\theta_2) = \frac{1}{\theta_2}e^{-\frac{1}{\theta_2}x} \quad (x \geq 0)$$

である．さて，以下のパラメータ

$$\mu_N = 0.2,\ \ v_N = E(V(N|\Theta_1)) = 0.2,$$
$$\mu_X = 200,\ \ v_X = E(V(X|\Theta_2)) = 50,000,$$
$$w_N = V(E(N|\Theta_1)) = 0.05$$

が判明しており，かつＡ社の 3 年間の実績クレーム総額が 180 であるとき，同社の年間クレーム総額 S を Bühlmann モデルにより推定せよ．

[解説] この問題はクレーム総額 $S = X_1 + X_2 + \cdots + X_N$ に対して Bühlmann モデルを適用する問題である．

まず，問題より，$n = 3,\ \overline{x} = 60$ である．次に v_S を計算する．

$$\begin{aligned} V(S|\Theta) &= E(V(S_{|N}|\Theta)) + V(E(S_{|N}|\Theta)) \\ &= E(N_{|\Theta_1} \times V(X|\Theta_2)) + V(N_{|\Theta_1} \times E(X|\Theta_2)) \\ &= E(N|\Theta_1) \times V(X|\Theta_2) + V(N|\Theta_1) \times E(X|\Theta_2)^2 \end{aligned}$$

より，

$$v_S = E(V(S|\Theta)) = \mu_N \cdot v_X + v_N \cdot E(E(X|\Theta_2)^2) \quad ①$$

次に，

$$w_X = V(E(X|\Theta_2)) = E(E(X|\Theta_2)^2) - \mu_X^2$$

より，①に代入して

$$v_S = \mu_N \cdot v_X + v_N \cdot (w_X + \mu_x^2) \quad ②$$

を得る．さらに，

$$\begin{aligned} w_X &= V(E(X|\Theta_2)) = V(\Theta_2) \\ &= E(\Theta_2^2) - E(\Theta_2)^2 = v_X - \mu_X^2 \end{aligned}$$

となるが，これを②に代入して

$$\begin{aligned} v_S &= \mu_N \cdot v_X + v_N \cdot v_X \\ &= 0.2 \times 50,000 + 0.2 \times 50,000 = 20,000 \end{aligned}$$

を得る．一方，

$$\begin{aligned} w_S &= V(E(S|\Theta)) = E(E(S|\Theta)^2) - E(E(S|\Theta))^2 \\ &= E(E(N|\Theta_1)^2 \times E(X|\Theta_2)^2) \\ &\qquad - E(E(N|\Theta_1) \times E(X|\Theta_2))^2 \\ &= E(E(N|\Theta_1)^2) \times E(E(X|\Theta_2)^2) \\ &\qquad - E(E(N|\Theta_1))^2 \times E(E(X|\Theta_2))^2 \\ &= \{V(E(N|\Theta_1)) + E(E(N|\Theta_1))^2\} \\ &\qquad \times \{V(E(X|\Theta_2)) + E(E(X|\Theta_2))^2\} \\ &\qquad - E(E(N|\Theta_1))^2 \times E(E(X|\Theta_2))^2 \\ &= (w_N + \mu_N^2) \times v_X - \mu_N^2 \times \mu_X^2 \\ &= (0.05 + 0.2^2) \times 50,000 - 0.2^2 \times 200^2 = 2,900 \end{aligned}$$

となる．また，$\mu = \mu_N \times \mu_X = 40$ となり，以上より，

$$Z = \frac{3}{3 + \dfrac{20,000}{2,900}} = 0.303,$$
$$C = 0.303 \times 60 + (1 - 0.303) \times 40 = 46.06.$$

■

[過去問 **H23**] 無作為に抽出した契約者のクレーム件数は幾何分布 $P(N = n) = p(1-p)^n\ (n = 0, 1, 2, \cdots)$ に従い，さらに，幾何分布のパラメータ p は契約者ごとにばらつきがあり，$0.1 \leq p \leq 0.5$ の一様分布に従うものとする．なお，契約者単位の時系列で観察した場合，同一の契約者の p が同一であるとする．このとき，過去 1 年間のクレーム件数が 10 件の契約者の翌年度のクレーム件数を，Bühlmann モデルを用いて推定せよ．

[解説] 同様．パラメータ p を Θ と思えばいい． ■

[参考書 例題 **40**] 保険会社Ｐ社の地域別マーケットシェアは，下表のとおりであることがわかっている．ある特定地区をランダムに選択（地区の選択は均等になされた），100 件の保険契約について調査したところ，5 件の契約がＰ社の契約であることが判明した．このとき，再度，同地区で 100 件の契約調査を行った場合に，確認されるであろうＰ社の契約数を Bühlmann モデルで推定せよ．

	Ａ地区	Ｂ地区	Ｃ地区
マーケットシェア	5 %	10 %	1 %

[解説] 今までの問題とやや傾向が異なる．この状況での Θ はＡ地区・Ｂ地区・Ｃ地区であり，問題の設定から $P(\Theta = \text{Ａ地区}) = P(\Theta = \text{Ｂ地区}) = P(\Theta = \text{Ｃ地区}) = \dfrac{1}{3}$ を満たす確率分布になる．

次に，ある 1 件が契約していれば 1，契約していなければ 0 となる確率変数を N としたとき，N は Θ に依存し，

$$P(N = k|\Theta = \text{Ａ地区}) = \begin{cases} 0.95 & k = 0 \\ 0.05 & k = 1 \end{cases}$$
$$P(N = k|\Theta = \text{Ｂ地区}) = \begin{cases} 0.90 & k = 0 \\ 0.10 & k = 1 \end{cases}$$
$$P(N = k|\Theta = \text{Ｃ地区}) = \begin{cases} 0.99 & k = 0 \\ 0.01 & k = 1 \end{cases}$$

を満たす確率分布になる．これで N の分布は完全にわかったことになる．この分布から $N_1, N_2, \cdots, N_{100}$ を抽出し，このうち契約数が何件かを Bühlmann モデルで推定するということなので，契約数を確率変数 S とすれば，$S = N_1 + N_2 + \cdots + N_{100}$ とでき，これについて Bühlmann モデルを適用する．まず．$E(S|\Theta) = 100 \times E(N|\Theta),\ V(S|\Theta) = 100 \times V(N|\Theta)$ に気をつけて，

$$E(N|\Theta) = \begin{cases} 0 \times 0.95 + 1 \times 0.05 = 0.05 & \Theta = \text{Ａ地区} \\ 0 \times 0.90 + 1 \times 0.10 = 0.10 & \Theta = \text{Ｂ地区} \\ 0 \times 0.99 + 1 \times 0.01 = 0.01 & \Theta = \text{Ｃ地区} \end{cases}$$

$$E(N^2|\Theta) = \begin{cases} 0^2 \times 0.95 + 1^2 \times 0.05 = 0.05 & \Theta = \text{Ａ地区} \\ 0^2 \times 0.90 + 1^2 \times 0.10 = 0.10 & \Theta = \text{Ｂ地区} \\ 0^2 \times 0.99 + 1^2 \times 0.01 = 0.01 & \Theta = \text{Ｃ地区} \end{cases}$$

となるから，以下の表を得る．

Θ	A地区	B地区	C地区
$E(S\|\Theta)$	5	10	1
$V(S\|\Theta)$	4.75	9.00	0.99

以上より，

$$\begin{aligned}
\mu &= E(E(S|\Theta)) = \frac{5+10+1}{3} = 5.333,\\
v &= E(V(S|\Theta)) = \frac{4.75+9.00+0.99}{3} = 4.913,\\
w &= V(E(S|\Theta))\\
&= \frac{(5-5.333)^2+(10-5.333)^2+(1-5.333)^2}{3}\\
&= 13.556
\end{aligned}$$

となる．また，問題より，$n=1$, $\overline{x}=5.333$ であるため，

$$Z = \frac{1}{1+\dfrac{4.913}{13.556}} = 0.734,$$
$$C = 0.734\times 5 + (1-0.734)\times 5.333 = 5.089.$$

■

[過去問 **H18**] ある保険会社の地域別ロス状況は下表の通りである．

	事故の有無	
	無	有
A地域	0.8	0.2
B地域	0.7	0.3
C地域	0.5	0.5
D地域	0.2	0.8

今，ある特定の地域を無作為に（確率 1/4 で）選択し，5件の保険契約を抽出したところ，3件の事故があった．再度同地域から5件を抽出したとき，何件の事故件数になるかを，Bühlmann モデルによって推定せよ．

[解説] Θ は以下を満たす確率分布に従う：

$$\begin{aligned}
P(\Theta=\text{A地域}) &= P(\Theta=\text{B地域})\\
&= P(\Theta=\text{C地域})\\
&= P(\Theta=\text{D地域}) = \frac{1}{4}.
\end{aligned}$$

次に，N を，事故が無ければ 0，事故が有れば 1 となる確率変数としたとき，以下を満たす確率分布に従う：

$$P(N=k|\Theta=\text{A地域}) = \begin{cases} 0.80 & k=0\\ 0.20 & k=1 \end{cases}$$

$$P(N=k|\Theta=\text{B地区}) = \begin{cases} 0.70 & k=0\\ 0.30 & k=1 \end{cases}$$

$$P(N=k|\Theta=\text{C地区}) = \begin{cases} 0.50 & k=0\\ 0.50 & k=1 \end{cases}$$

$$P(N=k|\Theta=\text{D地区}) = \begin{cases} 0.20 & k=0\\ 0.80 & k=1. \end{cases}$$

これで N の分布は完全にわかったことになる．この分布から $N_1, N_2, \cdots, N_5$ を抽出し，このうち事故件数が何件かを Bühlmann モデルで推定するということなので，事故件数を確率変数 S とすれば，$S=N_1+N_2+\cdots+N_5$ とでき，これについて Bühlmann モデルを適用する．まず．$E(S|\Theta)=5\times E(N|\Theta)$, $V(S|\Theta)=5\times V(N|\Theta)$ に気をつけて，

$$E(N|\Theta) = \begin{cases} 0\times 0.80+1\times 0.20 = 0.20 & \Theta=\text{A地域}\\ 0\times 0.70+1\times 0.30 = 0.30 & \Theta=\text{B地域}\\ 0\times 0.50+1\times 0.50 = 0.50 & \Theta=\text{C地域}\\ 0\times 0.20+1\times 0.80 = 0.80 & \Theta=\text{D地域} \end{cases}$$

$$E(N^2|\Theta) = \begin{cases} 0^2\times 0.80+1^2\times 0.20 = 0.20 & \Theta=\text{A地域}\\ 0^2\times 0.70+1^2\times 0.30 = 0.30 & \Theta=\text{B地域}\\ 0^2\times 0.50+1^2\times 0.50 = 0.50 & \Theta=\text{C地域}\\ 0^2\times 0.20+1^2\times 0.80 = 0.80 & \Theta=\text{D地域} \end{cases}$$

となるから，以下の表を得る．

Θ	A地区	B地区	C地区	D地区
$E(S\|\Theta)$	1.00	1.50	2.50	4.00
$V(S\|\Theta)$	0.80	1.05	1.25	0.80

以上より，

$$\begin{aligned}
\mu &= E(E(S|\Theta)) = \frac{1.00+1.50+2.50+4.00}{4} = 2.25,\\
v &= E(V(S|\Theta)) = \frac{0.80+1.05+1.25+0.80}{4} = 0.975,\\
w &= V(E(S|\Theta))\\
&= \frac{1}{4}((1.00-0.975)^2+(1.50-0.975)^2\\
&\qquad +(2.50-0.975)^2+(4.00-0.975)^2) = 1.3125
\end{aligned}$$

となる．また，問題より，$n=1$, $\overline{x}=3$ であるため，

$$Z = \frac{1}{1+\dfrac{0.975}{1.3125}} = 0.57377,$$
$$C = 0.57377\times 3 + (1-0.57377)\times 2.25 = 2.68.$$

■

[過去問 **H14**] それぞれ独立したリスクの集団がある．リスクAとBは二つのクラスに分類され，かつクラスAとクラスBのリスクの数は同一である．

ここで，次の条件が与えられたとする．

① それぞれのリスクは，1年間に1回だけ事故が発生する確率が 20 %で，1回も事故が発生しない確率が 80 %．

② クラスAのクレーム額は，すべて 2 である．

③ クラスBのクレーム額は，すべて c（定数）である．

無作為に一つのリスクを抽出して 1 年間のクレームを観察し，翌年のクレームコストの期待値を算出したい．このとき次の問に答えよ．
(1) 信頼度 Z を c であらわせ．
(2) $c > 10$ であることがわかっているとするとき，Z の取りうる範囲を求めよ．
[解説] Θ は以下を満たす確率分布に従う：
$$P(\Theta = \text{クラスA}) = P(\Theta = \text{クラスB}) = \frac{1}{2}.$$
次に，クレーム額を X としたとき，以下を満たす確率分布に従う：
$$P(X = k|\Theta = \text{クラスA}) = \begin{cases} 0.80 & k = 0 \\ 0.20 & k = 2 \end{cases}$$
$$P(X = k|\Theta = \text{クラスB}) = \begin{cases} 0.80 & k = 0 \\ 0.20 & k = c \end{cases}$$
を満たす確率分布になる．これで X の分布は完全にわかったことになる．以上より以下の表を得る．

Θ	クラスA	クラスB
$E(X\|\Theta)$	$0.2 \times 2 = 0.4$	$0.2 \times c = 0.2c$
$E(X^2\|\Theta)$	$0.2 \times 2^2 = 0.8$	$0.2 \times c^2 = 0.2c^2$
$V(X\|\Theta)$	$0.8 - 0.4^2 = 0.64$	$0.2c^2 - (0.2c)^2 = 0.16c^2$

以上より，
$$v = 0.08(c^2 - 4),\ w = 0.01(c-2)^2$$
となるので，
$$Z = \frac{(c-2)^2}{9c^2 - 4c + 36}$$
を得る．これの取りうる範囲を求めればよい． ■

4.2 分布が一部のみ判明している場合

[**教 3-41. 例 3-11**] ある野球選手の年間被死球数を調べたい．年間の被死球数 $X \sim \mathrm{Po}(\Theta)$ とし，Θ の分布は不明とする．去年 1 年間のデータは以下のようになった．

被死球数（個）	0	1	2	3	4	5
人数（人）	123	97	49	21	8	2

このとき，過去 1 年に 2 個の死球を受けた選手は，来年に何個の死球を受けるか，Bühlmann モデルで推定せよ．
[解説] 被死球数 X がポアソン分布に従うという情報はあるものの，そのパラメータ Θ の分布は不明であるので，与えられた情報から Θ がどういう確率分布に従うか推定しなければいけない．そのため，表を $v \cdot w \cdot \mu$ テーブルとみて $v,\ w,\ \mu$ を推定する．また，問題より $n = 1,\ \overline{x} = 2$ である．ポアソン分布の性質より，$V(X|\Theta) = E(X|\Theta)$ となるので，
$$v = E(E(X|\Theta)) = \mu.$$
また，
$$w = E(V(X|\Theta)) = V(X) - V(E(X|\Theta)) = \sigma^2 - \mu$$
となるが，$v \cdot w \cdot \mu$ テーブルから平均と不偏分散を計算すれば $\mu = 1,\ \sigma^2 = 1.2$ と推定でき，以下を得る，
$$v = 1,\ w = 0.2,\ \mu = 1.$$
以上より，
$$Z = \frac{1}{1 + \frac{1}{0.2}} = \frac{1}{6},\ C = \frac{1}{6} \times 2 + \left(1 - \frac{1}{6}\right) \times 1 = \frac{7}{6}.$$
■

[**参考書 例題 36**] 下表は，昨年度の P 地区の車両保険契約者の車両の年間盗難クレーム件数データである．盗難実績 1 件の契約者と 0 件の契約者との間で，今年度どれくらい盗難クレーム頻度較差をつけるべきか，Bühlmann モデルにより推定せよ．なお，各契約者クレーム件数はポアソン分布に従い，また，契約者ごとにポアソン分布のパラメータは異なるものとする．

盗難クレーム件数	0	1	2	3	合計
契約者数	300	40	15	5	360

[解説] 問題は，過去実績が 1 件と 0 件であった契約者の今年度のクレーム件数をそれぞれ求め，その比を求めることである．クレーム件数 N はポアソン分布に従うという情報はあるものの，そのパラメータ Θ の分布は不明であるので，与えられた情報から Θ がどういう確率分布に従うか推定しなければいけない．そのため，表を $v \cdot w \cdot \mu$ テーブルとみて $v,\ w,\ \mu$ を推定する．ポアソン分布の性質より，$V(N|\Theta) = E(N|\Theta)$ となるので，
$$v = E(E(N|\Theta)) = \mu.$$
また，
$$w = E(V(N|\Theta)) = V(N) - V(E(N|\Theta)) = \sigma^2 - \mu$$
となるが，$v \cdot w \cdot \mu$ テーブルから標本平均と標本不偏分散を計算すれば $\mu = 0.2361,\ \sigma^2 = 0.3480$ と推定でき，以下を得る．
$$v = 0.2361,\ w = 0.1119,\ \mu = 0.2361.$$
さらに，$n \cdot \overline{x}$ テーブルから，$n = 1$ で，実績 0 件の契約者Aについては $\overline{x}_A = 0$，実績 1 件の契約者Bについては $\overline{x}_B = 1$ となる．以上より，以下を得る．
$$Z = \frac{1}{1 + \frac{0.2361}{0.1119}} = 0.3216,$$
$$C_A = 0.3216 \times 0\ + (1 - 0.3216) \times 0.2361 = 0.1602$$
$$C_B = 0.3216 \times 1\ + (1 - 0.3216) \times 0.2361 = 0.4818\ .$$
よって，実績 1 件の契約者の盗難クレーム件数は実績 0 件の契約者の $0.4818/0.1602 = 3.007$ 倍程度となる． ■

[参考書 例題 **37**] 下表は，昨年度の車両保険契約者のクレーム件数データである．各契約者のクレーム件数は幾何分布に従い，また，契約者ごとに幾何分布のパラメータは異なるものとする．

クレーム件数	0	1	2	3	4	合計
契約者数	3,000	210	20	8	2	3,240

(1) 契約者ごとのクレーム件数 X の条件付期待値および分散を求めよ．

(2) Bühlmann モデルによって，昨年度の契約者個別の実績クレーム件数に対する信頼度を求めよ．

[解説] この問題は $n\cdot\overline{x}$ テーブルが不明，特に n が具体的に与えられていない状況で信頼度を求める問題である．こういうときは，$v\cdot w\cdot\mu$ テーブルを流用するが，実績データは過去 1 年分だけなので，$n=1$ と考える．

X はパラメータが Θ の幾何分布に従うので，条件付確率変数 $X|\Theta$ の確率密度関数は $f_{X|\Theta}(x|\theta)=\theta(1-\theta)^x$ $(x=0,1,2,\cdots)$ ということがわかっている．まず，

$$E(X|\Theta)=\frac{1-\Theta}{\Theta},\ V(X|\Theta)=\frac{1-\Theta}{\Theta^2}$$

より

$$\begin{aligned}
v&=E(V(X|\Theta))=E\left(\frac{1-\Theta}{\Theta^2}\right)\\
&=E\left[\left(\frac{1-\Theta}{\Theta}\right)^2+\frac{1-\Theta}{\Theta}\right]\\
&=E\left[\left(\frac{1-\Theta}{\Theta}\right)^2\right]+E\left(\frac{1-\Theta}{\Theta}\right)\\
&=E(E(X|\Theta)^2)+E(E(X|\Theta))\\
&=E(E(X|\Theta)^2)+\mu
\end{aligned}$$

であるから，

$$E(E(X|\Theta)^2)=v-\mu$$

を得る．これを利用すると，

$$\begin{aligned}
w&=V(E(X|\Theta))=E(E(X|\Theta)^2)-E(E(X|\Theta))^2\\
&=E(E(X|\Theta)^2)-\mu^2=v-\mu-\mu^2
\end{aligned}$$

を得る．さらに，

$$w=V(E(X|\Theta))=V(X)-E(V(X|\Theta))=\sigma^2-v$$

でもあるので，以上より，以下を得る．

$$v=\frac{1}{2}(\sigma^2+\mu+\mu^2),\ w=\frac{1}{2}(\sigma^2-\mu-\mu^2).$$

これらは $v\cdot w\cdot\mu$ テーブルから計算できるが，μ を標本平均，σ^2 を標本不偏分散で推定すれば，$\mu=0.0870$, $\sigma^2=0.1141$ となり，以上より，

$$v=0.1043,\ w=0.0098.$$

よって，

$$Z=\frac{1}{1+\dfrac{0.1043}{0.0098}}=0.0859.$$

■

[過去問 **H22**] 保険事故が発生した場合には保険金 1,000 を支払う保険契約について，過去 1 年間のクレーム件数を調べたところ，次のデータが得られた．ここで，1 契約者あたりのクレーム件数はポアソン分布に従い，契約者ごとにポアソン分布のパラメータは異なるものとする．このとき，次の (1)，(2) の各問に答えなさい．

過去 1 年間のクレーム件数	0	1	2	3
契約者数	500	160	70	35

(1) 信頼度を求めよ．

(2) 過去 1 年間のクレーム件数が 2 件の契約者の純保険料を求めよ．

[解説] 前問と同様． ■

4.3 分布が一切わかっていない場合（ノンパラメトリック）

今までは X や Θ の分布に情報が与えられていたので，それを頼りに計算を行ってきた．以下では，分布の情報がなく実績データのみで推測する方法を考えることになる．今，測定期間が s，対象数 r の $v\cdot w\cdot\mu$ テーブルが以下のように与えられたとする．

i \ j	1	2	$\cdots$	s
1	X_{11}	X_{12}	$\cdots$	X_{1s}
2	X_{21}	X_{22}	$\cdots$	X_{2s}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
r	X_{r1}	X_{r2}	$\cdots$	X_{rs}

この表から v,w,μ の推定量を作る．

i	平均	不偏分散
1	$\overline{X}_{1\bullet}=\frac{\sum_j X_{1j}}{s}$	$v_1=\frac{\sum_j(X_{1j}-\overline{X}_{1\bullet})^2}{n-1}$
2	$\overline{X}_{2\bullet}=\frac{\sum_j X_{2j}}{s}$	$v_2=\frac{\sum_j(X_{2j}-\overline{X}_{2\bullet})^2}{n-1}$
$\cdots$	$\cdots$	$\cdots$
r	$\overline{X}_{r\bullet}=\frac{\sum_j X_{rj}}{s}$	$v_r=\frac{\sum_j(X_{rj}-\overline{X}_{r\bullet})^2}{n-1}$
平均	$\mu=\overline{X}_{\bullet\bullet}=\frac{\sum_{i,j}X_{ij}}{rs}$	$v=\frac{\sum_i v_i}{r}$
不偏分散	$w=\frac{\sum_i(\overline{X}_{i\bullet}-\overline{X}_{\bullet\bullet})^2}{r-1}-\frac{v}{s}$	-

これで v,w,μ の推定量ができた．この推定量だけ記憶しておけば問題はない．念のため，w について補足する．

契約者 i のデータ X_i の平均の推定量は，$\overline{X}_{i\bullet}$ であるが，この分散を計算してみる．ここでは，契約者 i のパ

ラメータは Θ_i と書く．

$$\begin{aligned} V(\overline{X}_{i\bullet}) &= V(E(\overline{X}_{i\bullet}|\Theta_i)) + E(V(\overline{X}_{i\bullet}|\Theta_i)) \\ &= V(E(X_i|\Theta_i)) + E\left(\frac{V(X_i|\Theta_i)}{s}\right) \\ &= V(E(X|\Theta)) + E\left(\frac{V(X|\Theta)}{s}\right) \\ &= w + \frac{v}{s} \end{aligned}$$

となり，契約者に依存しない値が得られる．一方，左辺は不偏分散

$$\frac{1}{r-1}\sum_{i=1}^{r}\left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right)^2$$

が推定量となるので，

$$w = \frac{1}{r-1}\sum_{i=1}^{r}\left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right)^2 - \frac{v}{s}$$

が得られる．

実際の問題例は以下で見ていくが，上記の推定量をただ計算するだけである．このパターンが，Bühlmann モデルで一番簡単な問題といえよう．

[例題 **39**] 3 名の契約者の過去 3 年のクレーム額実績データとして下表を得ている．

	1 年目	2 年目	3 年目
契約者A	200	250	300
契約者B	600	500	400
契約者C	800	600	900

(1) 信頼度を Bühlmann モデルで求めよ．

(2) 契約者Aの 4 年目のクレーム額を推定せよ．

[解説] 表より，$s=3,\ r=3$ がわかる．次に，$n\cdot\overline{x}$ テーブルであるが，$v\cdot w\cdot\mu$ テーブルである表にすでに推定対象者Aの実績があるので，これを $n\cdot\overline{x}$ テーブルとみて，$n=3,\ \overline{x}=250$ を得る．

	平均	不偏分散
契約者A	250.0	2,500.0
契約者B	500.0	10,000.0
契約者C	766.7	23,333.3
平均	505.6	11,944
不偏分散	62,787	—

以上より，

$$Z = \frac{3}{3+\dfrac{11,944}{92,787}} = 0.940,$$

$$C = 0.940\times 250 + (1-0.940)\times 505.6 = 265.3.$$

■

5　Bühlmann-Straub モデル

前節にて，実績テーブルは 1 件単位でデータが与えられている状態であった．しかし実務の際，1 件単位のデータが得られにくいことがある．そのようなときは，合計データと，対象者数だけから推測するなどしなくてはいけない．その場合，Bühlmann モデルを適用することは不可能で，改良を加えた Bühlmann-Straub モデルというものが必要になる．以下それを解説する．

例 4

ある保険会社にて，翌期の契約者一人当たりのクレーム額を推測する．契約者一人当たりのデータはすぐに得られなかったため，以下のような合計データから推測したい．

	1 年目	2 年目	3 年目
クレーム額計	—	12,000	15,000
契約者数	—	50	60

このとき，この保険会社の実績値の信頼度を求めよ．

もし，$X,\ \Theta$ の確率分布がわかっていれば，今までと同様に v,w,μ を計算すればよい．わかっていなければ，表を $v\cdot w\cdot\mu$ テーブルとみて計算を行う必要がある．さらに，問題によっては，表を $n\cdot\overline{x}$ テーブルとみて計算を行う必要がある．この例については，$n=50+60=110$ としなくてはならない．

Bühlmann-Straub モデルについては，具体例で学ぶこととする．

5.1　分布が完全に判明している場合

分布が完全に判明している場合は，Bühlmann モデルと同様に v,w,μ を同様に計算すればよいが，n についてだけさきほど述べたように注意すればよい．

Bühlmann モデルでは n は実績抽出期間としていたが，Bühlmann-Straub モデルでは，エクスポージャ数（データ抽出対象の数）を用いるようにすればよい．

[過去問 **H24**] ある 1 年契約の保険商品について, 今年度（X 年度）は 17 件の契約を引き受けている．このうち次年度に契約を更改する件数（更改件数）を推定する．各契約者が契約を更改するか否かは，パラメータ Θ の二項分布 $B(1,\Theta)$,（$f_{Y|\Theta}(y|\theta) = \begin{pmatrix}1\\y\end{pmatrix}\theta^y(1-\theta)^{1-y}$，確率変数 Y は，契約を更改する場合は 1 を，契約を更改しない場合は 0 をとるものとする）に従い，Θ は，確率密度関数が，$f_\Theta(\theta) = 72\theta^7(1-\theta)\ (0<\theta<1)$ のベータ分布に従うことが分かっている．また，過去 3 年間の契約件

数と更改件数は下表の通りとなっている．このとき，次の (1)，(2) の各問に答えなさい．

	X-3 年度	X-2 年度	X-1 年度	X 年度
契約件数	12	13	15	17
更改件数	6	7	7	①

(1) 実績データに対する信頼度を Bühlmann-Straub モデルにより求めよ．
(2) Bühlmann-Straub モデルにより①を推定せよ．
[解説] 与えられた表が $n\cdot\overline{x}$ テーブルであり，これから $n=12+13+15=40$, $\overline{x}=0.5$ がわかる．あとは確率分布の仮定から v,w,μ を計算すれば，契約件数 1 件が契約を更改する確率が推定できる．それに X 年度の契約件数を乗じればよい． ■

[参考書 例題 41] 1 契約当たりのクレーム件数は，パラメータ Λ のポアソン分布に従い，また，Θ は確率密度関数が $f_\Theta(\theta)=\dfrac{100}{\Gamma(5)}e^{-100\theta}(100\theta)^4$ のガンマ分布に従うことが判明している．ある契約集団の過去 2 年間のクレームを調べたところ下表のデータが得られた．

	1 年目	2 年目
クレーム件数	60	40
契約件数	100	50

(1) このクレームに対する信頼度を Bühlmann モデルで求めよ．
(2) 3 年目の契約件数は 120 件と判明している．3 年目のクレーム件数を Bühlmann モデルで推定せよ．
[解説] 前問と同様に，与えられた表が $n\cdot\overline{x}$ テーブルなので，$n=150$, $\overline{x}=(60+40)/(100+50)=100/150$ がわかる．あとは確率分布の仮定から v,w,μ を計算すれば，契約件数 1 件あたりのクレーム件数が推定できる．それに 120 を乗じればよい． ■

[例題 42] 下表は，ある契約集団の過去 2 年間の成績である．加えて，下表以外に同契約のデータとして以下のことが判明している．
(a) 業界全体ベースでの平均クレームコストは $\mu=90$
(b) $v=1,200 \quad w=15$
このとき Bühlmann-Straub モデルでのこの契約集団の純保険料を推定せよ．

	実績クレームコスト	契約者数
1 年目	120	300
2 年目	100	200

[解説] すでに v,w,μ は与えられている．次に $n=300+200=500$ かつ
$$\overline{x}=\frac{120\times300+100\times200}{300+200}=112$$
より
$$Z=\frac{500}{500+\frac{1,200}{15}}=0.862$$
となる．よって，
$$C=0.862\times112+(1-0.862)\times90=108.96.$$
■

5.2 分布が一切わかっていない場合（ノンパラメトリック）

分布の情報がない場合は，Bühlmann モデルと同様に推定量を作る必要がある．以下で定義する推定量は，証明が困難なことから，結論のみ述べることとする．

今，グループ i $(1\le i\le r)$ に対して，過去 n_i 年間の実績データ値合計を得たとする．ここで，j $(1\le j\le n_i)$ 年目のグループのエクスポージャ数を m_{ij}，実績データ値合計を $S_{ij}=X_{i,j,1}+X_{i,j,2}+X_{i,j,m_{ij}}$，平均を $X_{ij}=\dfrac{S_{ij}}{m_{ij}}$ とする．また，
$$m_i=\sum_{j=1}^{n_i}m_{ij},\ m=\sum_{i=1}^{r}m_i,$$
$$\overline{X}_{i\bullet}=\frac{1}{m_i}\sum_{j=1}^{n_i}m_{ij}X_{ij},\ \overline{X}=\frac{1}{m}\sum_{i=1}^{r}m_i\overline{X}_{j\bullet}$$
と定義する．ただし，$X_{i,j,k}$ はグループ i の過去 j 年目の k 人目の対象者のデータをあらわす確率変数である．

このとき，v,w,μ の推定量は，
$$\mu=\overline{X},$$
$$v=\frac{\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{n_i}m_{ij}(X_{ij}-\overline{X}_i)^2}{\sum_{i=1}^{r}(n_i-1)},$$
$$w=\frac{\displaystyle\sum_{i=1}^{r}m_i(\overline{X}_i-\overline{X})^2-v(r-1)}{m-\dfrac{1}{m}\displaystyle\sum_{i=1}^{r}m_i^2}$$
となる．

[教 **3-45** 練習問題 **4**] ある保険会社は，次のような実績を持つ団体契約を 4 年目に引き受けることになった．

		1 年目	2 年目	3 年目
団体Ａ	クレーム額計	—	12,000	15,000
	加入者数	—	50	60
団体Ｂ	クレーム額計	19,000	23,000	16,000
	加入者数	100	150	160

4 年目の団体Ａ，団体Ｂの加入者がそれぞれ 80 人，180 人であるとする．このとき，保険会社が 4 年目に領収すべき総純保険料を Bühlmann-Straub モデルにより推定せよ．ただし，各計算過程において，小数点 3 位で四捨五入するものとする．

[解説] 団体Ａ・団体Ｂの一人当たりの推定クレーム額をそれぞれ C_A, C_B としたとき，求める値は，$80 \times C_A + 180 \times C_B$ である．まず，団体Ａの信頼度，団体Ｂの信頼度をそれぞれ計算する．

問題の表が $v{\cdot}w{\cdot}\mu$ テーブルである．団体Ａ・団体Ｂの $n{\cdot}\overline{x}$ テーブルは与えられていないので，$v{\cdot}w{\cdot}\mu$ テーブルを流用すると，$n_A = 110$, $\overline{x}_A = 245.45$, $n_B = 410$, $\overline{x}_B = 141.46$ がわかる．次に，各数値を計算していく．まず，

$$\mu = \frac{12,000 + 15,000 + 19,000 + 23,000 + 16,000}{50 + 60 + 100 + 150 + 160} = \frac{85,000}{520} = 163.46$$

である．次に，$\sum_{i=1}^{r} m_i(\overline{X}_i - \overline{X})^2$ は以下の通り計算される．

i	m_i	$\overline{X}_i$	$m_i(\overline{X}_i - \overline{X})^2$
1	110	245.45	739,459.61
2	410	141.46	198,440.00
合計	520		937,899.61

次に，$\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \overline{X}_i)^2$ は以下の通り計算される．

	1 年目	2 年目	3 年目	合計
団体Ａ	—	1,485.12	1,242.15	2,727.27
団体Ｂ	235,613.16	21,134.54	275,029.06	296,163.60
合計				534,504.03

ただし，

$$X_{11} = X_{\text{団体A,2 年目}} = 12,000/50 = 240,$$
$$X_{12} = X_{\text{団体A,3 年目}} = 15,000/60 = 250,$$
$$X_{21} = X_{\text{団体B,1 年目}} = 19,000/100 = 190,$$
$$X_{22} = X_{\text{団体B,2 年目}} = 23,000/150 = 153.33,$$
$$X_{23} = X_{\text{団体B,3 年目}} = 16,000/160 = 100$$

である．これらより，

$$v = \frac{534,504.03}{(2-1)+(3-1)} = 178,168.01,$$
$$w = \frac{937,899.61 - 178,168.01 \times (2-1)}{520 - \dfrac{110^2 + 410^2}{520}} = 4,379.83$$

となる．これで v, w, μ がすべて求まった．以上より，

$$Z_A = \frac{110}{110 + \dfrac{178.168.01}{4,379.83}} = 0.73,$$
$$Z_B = \frac{410}{410 + \dfrac{178.168.01}{4,379.83}} = 0.91,$$

$$C_A = 0.73 \times 245.45 + (1 - 0.73) \times 163.46 = 223,$$
$$C_B = 0.91 \times 141.46 + (1 - 0.91) \times 163.46 = 143$$

よって，求める保険料は

$$80 \times 223 + 180 \times 143 = 43,580.$$

■

[例題 **43**] 下表は，2 つの団体の傷害保険（実損払）の過去 3 年分の契約年度別クレーム総額実績（S_{ij}）および（m_{ij}）である．現在判明している 4 年度目の契約者数を用いて，4 年度目のクレーム総額を推定せよ．

		1 年度	2 年度	3 年度	4 年度
団体Ａ	クレーム額計	—	2,000	2,100	①
	契約者数	—	50	70	60
団体Ｂ	クレーム額計	1,900	3,600	1,210	②
	契約者数	100	120	110	90

[解説] 前問と同様に計算すればよい． ■

[例題 **44**] 年齢グループ別（グループ数 : $r = 3$）に過去 3 年間（$n = 3$）の実績クレームコスト（X_{ij}）の保険契約成績として下表を得た．

i	グループ	m_i	$\overline{X}_i$	v_i	$m_i(\overline{X}_i - \overline{X})^2$
1	20 歳未満	200	1.50	0.90	1.35
2	21〜29 歳	250	1.60	0.80	1.20
3	30 歳以上	350	1.40	0.75	1.10

Bühlmann-Straub モデルを用いて，各年齢グループの純保険料（クレームコストの理論値）を推定せよ．ただし，推定される純保険料の合計が実績のクレーム合計と同一になるようにせよ（すなわち総ファンドを一定とせよ）．

[解説] 前問と同様に計算すればよいが，実績クレーム額と，推定した総純保険料が等しくなるようにするためには，特別な計算が必要となり，これは教科書の内容を逸脱するようであるため，ここでは紹介しない．詳しくは参考書を参考にされたい． ■

以上