I-330　統計データと安全性指標について

京都大学工学部　正員　古田均
京都大学工学部　正員　白石成人

1．<u>まえがき</u>　近年の信頼性解析における２次モーメント法の発展は著しく、種々の問題への適用が試みられている。また、安全性指標 β の種々の問題点も明らかとなり、その物理的意味も明確になりつつある[1]。しかしながら、この安全性指標は観測・実験等により得られた標本値に統計的処理を施すことにより求められるものであり、もし確率変数の平均値・標準偏差等の推定が正しく行なわれなければ、得られた β の値は信用の置けないものとなる。本報告では、フィディシアル統計学の考え方を用いて、変数が正規分布に従う場合の母集団の平均値・標準偏差と標本平均・標本標準偏差との関係を導く。つぎに、この関係を安全性指標の計算に適用し、安全性指標とデータ数との関連を調べ、β の期待値を求めるための修正係数の近似式を提案する。最後に、ノンパラメトリック統計学の信頼性解析への応用について若干の考察を加える。

2．<u>パラメトリック統計学による平均値・標準偏差の推定[2]</u>　いま、確率変数 X が正規分布に従い、n 個の標本 X_i $(i=1\sim n)$ が得られたとすると、母集団の平均 μ_x，標準偏差 σ_x の結合確率密度関数 $f(\mu_x, \sigma_x)$ はフィディシアル統計学の考え方を用いることにより次式のように求められる。

$$f(\mu_x, \sigma_x) = C\,\frac{1}{\sigma_x^2}\left(\frac{s_x}{\sigma_x}\right)^{n-1}\exp\left[-\frac{n}{2\sigma_x^2}\left\{(\bar{x}-\mu_x)^2+s_x^2\right\}\right],\quad C=\frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \qquad (1)$$

ここで、$\Gamma(\cdot)$ はガンマ関数、$\bar{x}$・s_x は標本平均・標本標準偏差をそれぞれ表わす。
このとき、μ_x, σ_x の期待値は式(1)を積分することにより以下のように求められる。

$$E[\mu_x] = \int_{-\infty}^{\infty}\mu_x\left[\int_0^{\infty} f(\mu_x, \sigma_x)\,d\sigma_x\right]d\mu_x = \bar{x} \qquad (2)$$

$$E[\sigma_x] = \int_0^{\infty}\sigma_x\left[\int_{-\infty}^{\infty} f(\mu_x, \sigma_x)\,d\mu_x\right]d\sigma_x = \sqrt{\frac{n}{2}}\,\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\,s_x \qquad (3)$$

3．<u>データ数の安全性指標への影響</u>　いま、簡単のために、Z (safety margin) に対する標本 Z_i $(i=1\sim n)$ が得られたとすると、安全性指標 $\beta\,(=\mu_z/\sigma_z)$ の確率密度関数は、式(1)に変数変換および周辺積分を行なうことにより以下のように計算できる。

$$f_\beta(\beta) = 2C\int_0^{\infty}\frac{1}{\sigma_z}\left(\frac{s_z}{\sigma_z}\right)^{n-1}\exp\left[-\frac{n}{2\sigma_z^2}\left\{(\bar{z}-\beta\sigma_z)^2+s_z^2\right\}\right]d\sigma_z \qquad (4)$$

式(4)から β の平均値・標準偏差を求めることは可能であるが、一般にその積分を解析的に行なうことは困難であるので、ここでは一次近似の考え方を用いて安全性指標とデータの個数との関係を求める。

$$E[\beta] = E[\mu_z]\,/\,E[\sigma_z] \qquad (5)$$

いま、式(2),(3)の添字 X を Z に置きかえると、$E[\mu_z]$，$E[\sigma_z]$ は容易に求まり、$E[\beta]$ は修正係数 γ_β を導入することにより、標本平均・標本標準偏差の値を用いて以下のように求めることができる。

$$E[\beta] = \gamma_\beta\cdot\hat{\beta}\quad,\quad \hat{\beta}=\frac{\bar{z}}{s_z}\quad,\ \gamma_\beta=\sqrt{\frac{2}{n}}\,\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \qquad (6)$$

つぎに、γ_β をもっと簡単に求めるための近似式を誘導することを試みる。いま、$\gamma_\beta=\sqrt{\theta/n}$ とすると、θ の対数は以下のように表わされる[3]。

$$\log\theta = \log 2 + 2\log\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \qquad (7)$$

ここで、$\log[\Gamma(x+h)/\Gamma(x)]$ はつぎのように展開できるので、

$$\log\frac{\Gamma(x+h)}{\Gamma(x)} = h\log x + \frac{1}{x}\frac{h(h-1)}{2} - \frac{1}{x^2}\frac{h(h-1)(2h-1)}{12} + \frac{1}{x^3}\frac{h^2(h-1)^2}{12}\cdots \quad (8)$$

θ, γ_β は以下のように近似的に求められる。(ただし、γ_β の計算では θ の3項目以下は無視されている)

$$\theta = (n-1) - 3/2 + 1/[8(n-2)] \quad (9)$$

$$\gamma_\beta = \sqrt{(n-2.5)/n} \quad (10)$$

表1 γ_β の近似解

n	Exact	Appr.	Error(%)
3	0.4607	0.4082	-11.40
4	0.6267	0.6124	-2.28
5	0.7137	0.7071	-0.92
6	0.7675	0.7638	-0.48
7	0.8042	0.8018	-0.30
8	0.8308	0.8292	-0.19
9	0.8511	0.8498	-0.15
10	0.8670	0.8660	-0.12
20	0.9356	0.9354	-0.02
30	0.9575	0.9574	-0.01
40	0.9683	0.9682	-0.01
50	0.9747	0.9747	0.

4. ノンパラメトリック統計学による推定　前節までは確率変数に正規分布を仮定していたが、ここでは分布形を仮定せずに、μ_x, σ_x^2 を推定することを考える。Ferguson[4] によると、Dirichlet Process を用いることにより、n個の標本 X_i が得られたときの μ_x, σ_x^2 の推定値 $\hat{\mu}_n$, $\hat{\sigma}_n^2$ は以下のように表わされる。

$$\hat{\mu}_n = p_n\mu_0 + (1-p_n)\bar{X} \quad (11)$$

$$\hat{\sigma}_n^2 = \frac{\alpha(R)+n}{\alpha(R)+n+1}\left[p_n\sigma_0^2 + (1-p_n)S_X^2 + p_n(1-p_n)(\mu_0-\bar{X})^2\right] \quad (12)$$

ここで、$p_n = \dfrac{\alpha(R)}{\alpha(R)+n}$, $\mu_0 = \dfrac{1}{\alpha(R)}\int x\,d\alpha(x)$, $\sigma_0^2 = \dfrac{1}{\alpha(R)}\int x^2 d\alpha(x) - \mu_0^2$

また、$\alpha(\cdot)$ はディレクレ分布のパラメーターであり、Rは可測度空間における一つの実数軸を表わす。

5. 数値計算例　表1に γ_β の近似解の計算結果を示す。厳密には式(10)はnの値がかなり大きくないと成立しないが、計算結果からは $n=6$ 以上で非常に精度の良い解が得られている。図1に $\bar{Z} = 800$ kg/cm², $S_Z = 210$ kg/cm² が与えられた場合の安全性指標のデータ数による変化を示す。つぎに、$\mu_X = 800$ kg/cm², $\sigma_X = 210$ kg/cm² としてモンテカルロ法で標本を発生させた場合の $\hat{\sigma}_n$ の計算結果を表2に示す。標本は10個づつ発生させている。なお、この場合、$\hat{\mu}_n$ は $\bar{X}$ と等しくなる。標本標準偏差 S_X と比較すると、ノンパラメトリック統計学による標準偏差の推定値は常に小さめの値をとっている。これは $\alpha(R)$ として前に発生させた標本数をとり、$\mu_0 = \sum_{i=1}^{n} X_i/n$, $\sigma_0^2 = \sum_{i=1}^{n}(X_i-\mu_0)^2/(n+1)$ (ガウス分布の場合のミニマックス推定値)と仮定したことによる影響とも考えられる。(詳細は当日発表)

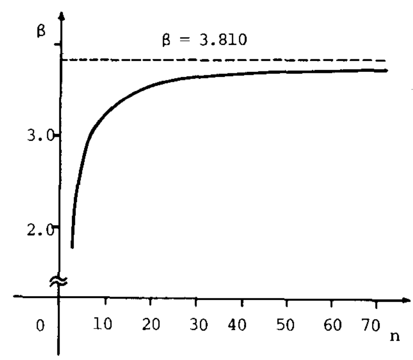

図1 安全性指標とデータ数との関係

n	$\bar{x}$	s_x	$\hat{\sigma}_n$
20	765.2	250.2	234.6
30	779.2	233.8	225.8
40	799.0	224.9	219.3
50	804.3	244.1	239.7
60	830.0	247.7	243.7
70	821.7	234.5	231.0
80	812.0	227.9	225.0
90	810.0	223.4	220.9
100	807.0	220.4	218.2

表2 $\hat{\sigma}_n$ の計算結果

6. あとがき　フィディシアル統計学の考え方を用い安全性指標とデータ数との関係を導き、修正係数を簡単に求めるための近似式を提案した。この式はnが小さくてもかなりの精度をもつ解を与えることが判明した。ノンパラメトリック統計学は未だ信頼性解析への適用には多くの問題を含んでいるものの、種々の理論上あるいは計算上の利点を有していると思われ、今後その可能性について十分検討されることが望まれる。

"参考文献" 1) 例えば、A. Hasofer & N. Lind, "Exact and Invariant Second Moment Code Format", ASCE, EM1, 1974, 2) 白石・古田・中野, "安全性指標の信頼性解析への適用に関する2・3の考察", 土木論文集投稿中, 3) J. Gurland & R. Tripathi, "A Simple Approximation for Unbiased Estimation of the Standard Deviation", Amer. Statistician, Oct. 1971 4) T. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems", Annals of Statistics, Vol. 1, 1973