確率紙における推定パラメータの区間推定

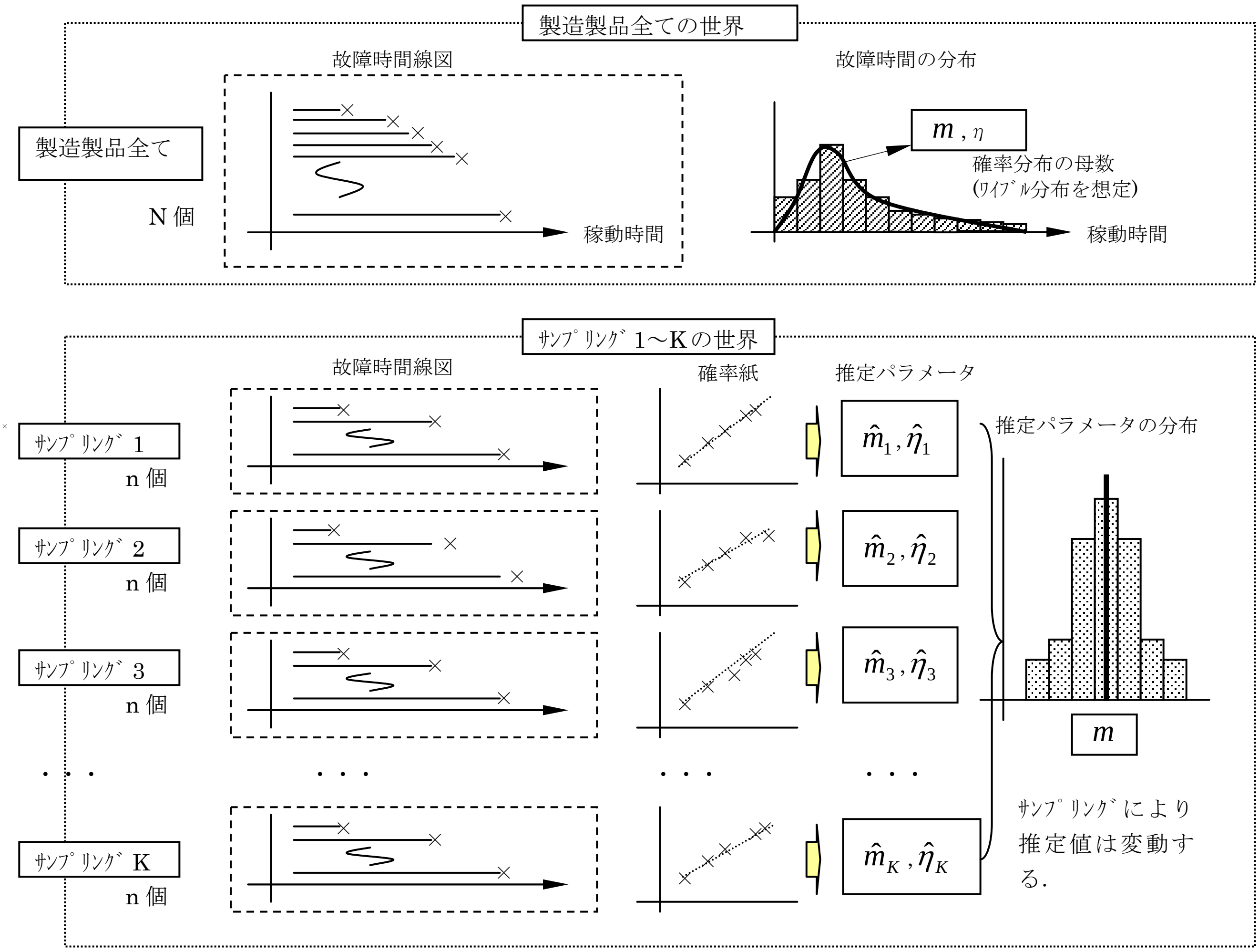
(株)日本科学技術研修所
長谷和彦

1. 確率紙を用いたパラメータ推定

製造した製品のスペックを保証することを考えようとすると，出荷前でも出荷後でも製品全体の寿命を知るために故障時間分布を知りたいというニーズがある．製造した製品の全てが故障するまで稼動させれば，その製品全体の故障時間分布を得ることができる．しかし，それは実質的に不可能である．出荷前であれば絶対に不可能である．そこで，いくつかサンプリングした製品の故障時間データから製造製品全体の故障分布を推定することになる．したがって選択されたサンプルの故障時間により，推定する故障時間分布も違ったものになる．

サンプルデータから製品全体の故障時間分布を推定する方法として，確率紙がよく利用されている．故障時間分布がなんらかの確率分布（主にワイブル分布）に従うと仮定し，対応する確率紙上にサンプルの故障時間データをプロットする．プロット点から回帰線を推定し，その回帰係数から確率分布のパラメータを知るという仕組みである．したがって，推定する故障時間分布の違いは回帰係数の違いとなって現れる．

確率紙で得られた回帰線の傾き(回帰係数)は，ワイブル分布であれば m，正規分布であれば $1/\sigma$ に対応する．本稿では，確率紙上の回帰線の傾きがサンプリングによりばらつくことをコンピュータシミュレーションにより観察し，その傾きの区間推定方法を考察する．



現実にはｻﾝﾌﾟﾘﾝｸﾞ 1 回の結果で判断することになる．確率紙から得られたパラメータがどの程度のばらつきの中の値なのかを理解しておくことは判断するうえで参考になると思われる．

2. 確率紙における回帰線の傾きのばらつき　-シミュレーションでの考察

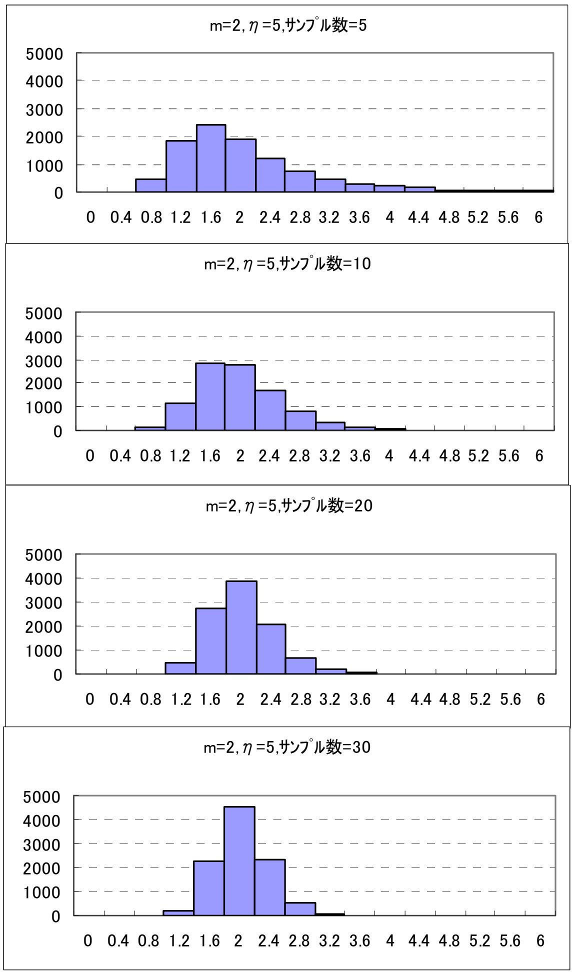
製品全体（母集団）の確率分布を設定し，該当する確率分布の乱数を生成し，確率紙上にプロットした際の傾きの値を計算した.

・確率分布はワイブル分布と正規分布のそれぞれで実行した.

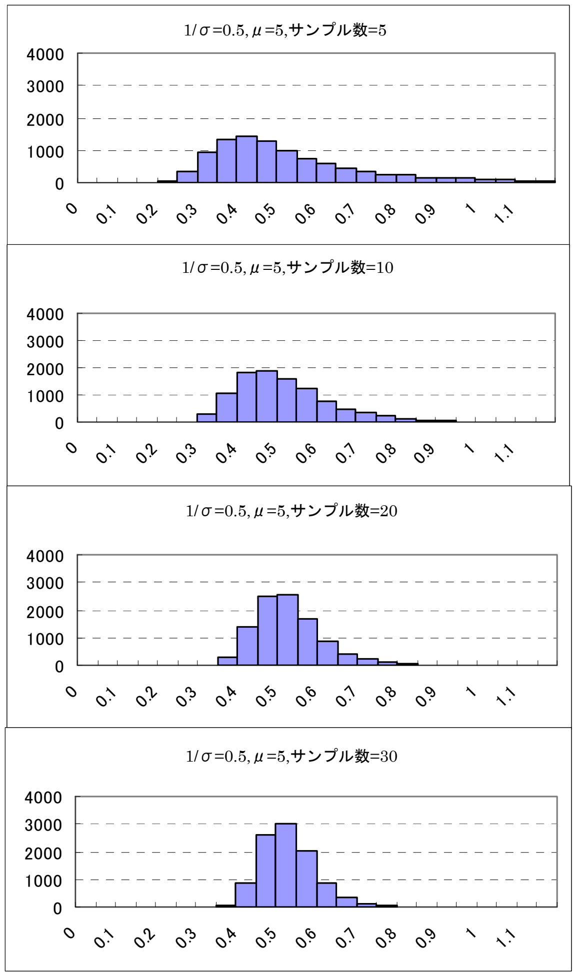
・サンプリング数(上図 n)は 5 個，10 個，20 個，30 個とし，それぞれ 10000 回(上図 K)繰り返した.

なお，ソフトウェアは Microsoft Excel 2003 を利用した.

ワイブル分布（m=2, η=5）の m の推定値の分布



正規分布（σ=2, μ=5）の 1/σの推定値の分布



サンプリングにより傾きの推定値は，ばらつくことがわかる．分布は対数正規分布のような形状をしている.

ワイブル確率紙や正規確率紙の場合,傾きは正の値しかとらないので,このような形状になると思われる.

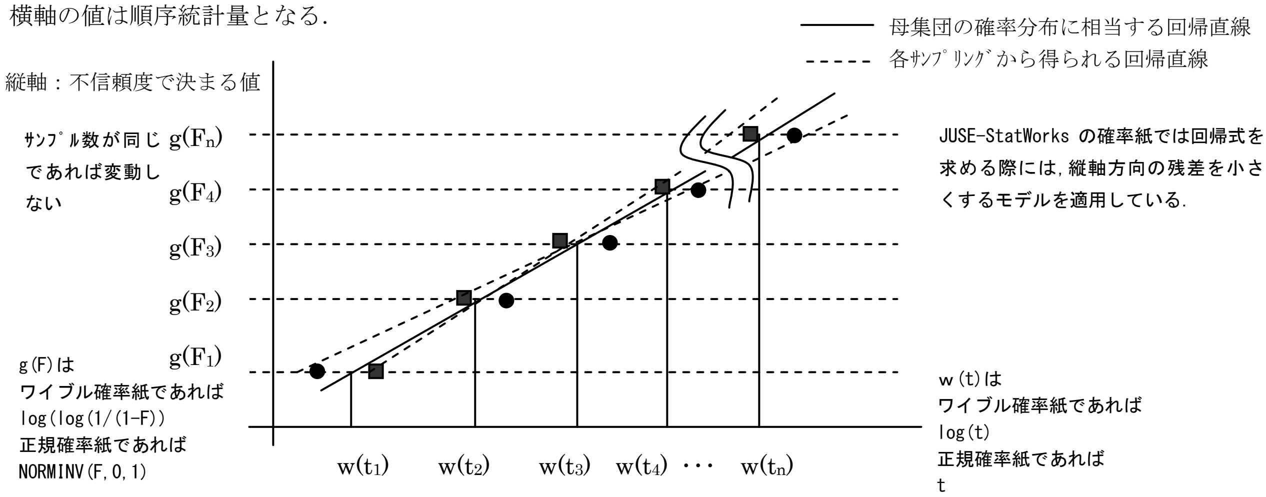
サンプリング数が増えるごとにばらつきが減少し，対象な分布に近づいていくように見える．ただし，完全にはひずみがとれることはないように見える.

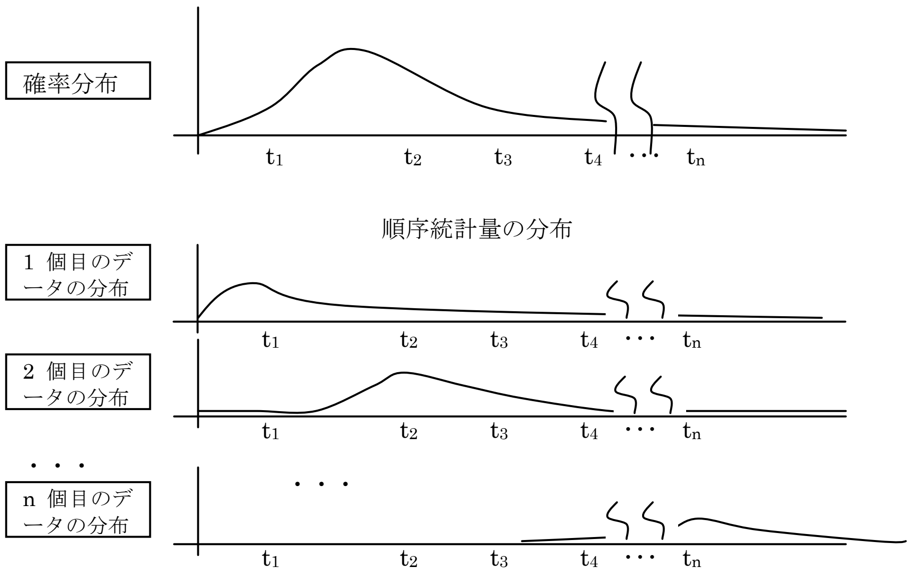
一般には傾きは正規分布に漸近するといわれている.

3. 確率紙における回帰線の傾きのばらつき -推定方法の考察

３－１．確率紙の仕組み

確率紙の縦軸値は不信頼度の推定値で決まる値となる．不信頼度の推定値は，特性値型でタイデータが無いデータをプロットする際には，メジアンランク法でも累積ハザード法でもサンプル数と順位のみで決まる値である．
サンプリングにより横軸の値が変動することにより，回帰線の傾きが変動する．
横軸の値は順序統計量となる．





３－２．回帰線の傾き（回帰係数）の分散と順序統計量の分散共分散の関係

ここでは確率紙での回帰線の傾きの分散を導出する．傾きは漸近的に正規分布に従うと仮定する．
Y 方向の残差を小さくする最小二乗法による回帰線の傾きの推定値 a は以下の式で表現される．

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - x)^2}$$

ワイブル確率紙の場合，$a = m$，$x_i = \log(t_i)$，$y_i = \log(\log(1/(1 - F_i)))$ である．

正規確率紙の場合，$a = 1/\sigma$，$x_i = t_i$，$y_i = NORMINV(F_i, 0, 1)$ である．

(NORMINV は正規分布の累積分布関数の逆関数である．)
確率紙の性質から縦軸の値は定数とみなすことができる．
そこで

$$a=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-x)^2}=\sum_{i=1}^{n}c_iq_i$$

とする．ただし$c_i=y_i-\bar{y}$，$q_i=\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$　とした．

回帰線の傾きの分散は

$$V(a)=V(\sum_{i=1}^{n}c_iq_i)$$

$$=\left(\frac{1}{K-1}\right)\sum_{l=1}^{K}(\sum_{i=1}^{n}c_iq_{ik}-\sum_{i=1}^{n}c_i\bar{q}_i)^2$$

$$=\left(\frac{1}{K-1}\right)\sum_{l=1}^{K}\left\{\sum_{i=1}^{n}c_i(q_{ik}-\bar{q}_i\ )\right\}^2$$

$$=\left(\frac{1}{K-1}\right)\sum_{l=1}^{K}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j(q_{ik}-\bar{q}_i)(q_{jk}-\bar{q}_j)\right\}$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\left(\frac{1}{K-1}\right)\sum_{l=1}^{K}(q_{ik}-\bar{q}_i)(q_{jk}-\bar{q}_j)$$

$$\cong\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\,\mathrm{cov}(q_i,q_j)\quad\cdots(1)$$

となる．
順序統計量の分散共分散の極限形式は参考文献[1]によれば

p_1：順序統計量s_1の確率，p_2：順序統計量s_2の確率，$f(s)$：確率密度関数，n：サンプル数　とすると

$$p_2\le p_1\quad\text{の場合，}\ \mathrm{cov}(s_1,s_2)=\frac{p_2\cdot(1-p_1)}{n\cdot f(s_1)\cdot f(s_2)}$$

$$p_2>p_1\quad\text{の場合，}\ \mathrm{cov}(s_1,s_2)=\mathrm{cov}(s_2,s_1)\quad\cdots(2)$$

である．

ワイブル確率紙の場合，$q=\frac{x-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$，$x=\log(t)$ より，$f(q)=f^*(t)\cdot\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\cdot t$　・・・(3)

正規確率紙の場合，$q=\frac{x-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$，$x=t$ より，$f(q)=f^*(t)\cdot\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$

t 軸における確率密度関数を $f^*(t)$ と表現している．

となる．順序統計量の期待値において,(1)，(2)，(3)式を用いることにより，確率紙の回帰線の傾きの分布が漸近

的に正規分布をすると仮定した場合の分散を推定することができる．正規分布を仮定し，両側 95%点（2.5%点と 97.5%点）の推定は以下のとおり．

$$a_{2.5\%} = a + NORMINV(0.025,0,1)\sqrt{V(a)}$$

$$a_{97.5\%} = a + NORMINV(0.975,0,1)\sqrt{V(a)}$$

4. ワイブル確率紙, 正規確率紙での結果　-シミュレーション結果との比較

上式の結果とコンピューターシミュレーションの結果を比較した．

ワイブル確率紙（いずれも η=5）

		シミュレーション結果			上式を用いた計算値		
m	ｻﾝﾌﾟﾙ数	分散	2.5%点	97.5%点	分散	2.5%点	97.5%点
0.1	5	0.00288669	0.035759246	0.227511109	0.00205841	0.011077015	0.188922985
	10	0.000888783	0.045038105	0.161349239	0.000903743	0.041078938	0.158921062
	20	0.000437281	0.055103411	0.13786498	0.000443193	0.058738545	0.141261455
	30	0.000309307	0.060773792	0.130295843	0.000298729	0.066124417	0.133875583
0.5	5	0.072167264	0.17879623	1.137555545	0.051460254	0.055385058	0.944614942
	10	0.022219587	0.225190525	0.806746194	0.022593579	0.205394665	0.794605335
	20	0.010932029	0.275517054	0.6893249	0.01107982	0.293692772	0.706307228
	30	0.007732681	0.303868961	0.651479213	0.007468237	0.330621947	0.669378053
1.0	5	0.288669056	0.35759246	2.275111091	0.205841015	0.110770118	1.889229882
	10	0.088878346	0.450381051	1.613492388	0.090374317	0.410789327	1.589210673
	20	0.043728114	0.551034109	1.3786498	0.044319282	0.587385535	1.412614465
	30	0.030930724	0.607737922	1.302958426	0.029872946	0.661243905	1.338756095
2.0	5	1.154676222	0.715184919	4.550222181	0.823364058	0.221540239	3.778459761
	10	0.355513386	0.900762102	3.226984775	0.361497269	0.821578652	3.178421348
	20	0.174912458	1.102068218	2.7572996	0.177277128	1.17477107	2.82522893
	30	0.123722898	1.215475845	2.605916852	0.119491784	1.32248781	2.67751219
5.0	5	7.216726389	1.787962298	11.37555545	5.146025365	0.553850595	9.446149405
	10	2.22195866	2.251905255	8.067461939	2.259357933	2.05394663	7.94605337
	20	1.09320286	2.755170545	6.893249	1.10798205	2.936927675	7.063072325
	30	0.773268112	3.038689612	6.514792129	0.746823653	3.306219522	6.693780478

正規確率紙（いずれも μ=5）

		シミュレーション結果			上式を用いた計算値		
1/σ	ｻﾝﾌﾟﾙ数	分散	2.5%点	97.5%点	分散	2.5%点	97.5%点
1/0.1	5	26.48162788	4.738016813	23.17783183	17.08558365	1.898545	18.10145
	10	6.608113034	5.934754908	15.79734415	6.610530663	4.960746	15.03925
	20	2.679861002	6.960152136	13.38852297	2.892627276	6.666547	13.33345
	30	1.771054331	7.443409168	12.58562232	1.841240133	7.34048	12.65952
1/0.5	5	1.059265115	0.947603363	4.635566366	0.683423346	0.379709	3.620291
	10	0.264324521	1.186950982	3.159468829	0.264421227	0.992149	3.007851
	20	0.10719444	1.392030427	2.677704594	0.115705091	1.333309	2.666691
	30	0.070842173	1.488681834	2.517124463	0.073649605	1.468096	2.531904
1/1.0	5	0.264816279	0.473801681	2.317783183	0.170855836	0.189855	1.810145
	10	0.06608113	0.593475491	1.579734415	0.066105307	0.496075	1.503925
	20	0.02679861	0.696015214	1.338852297	0.028926273	0.666655	1.333345
	30	0.017710543	0.744340917	1.258562232	0.018412401	0.734048	1.265952
1/2.0	5	0.06620407	0.236900841	1.158891592	0.042713959	0.094927	0.905073
	10	0.016520283	0.296737745	0.789867207	0.016526327	0.248037	0.751963
	20	0.006699653	0.348007607	0.669426149	0.007231568	0.333327	0.666673
	30	0.004427636	0.372170458	0.629281116	0.0046031	0.367024	0.632976
1/5.0	5	0.010592651	0.094760336	0.463556637	0.006834233	0.037971	0.362029
	10	0.002643245	0.118695098	0.315946883	0.002644212	0.099215	0.300785
	20	0.001071944	0.139203043	0.267770459	0.001157051	0.133331	0.266669
	30	0.000708422	0.148868183	0.251712446	0.000736496	0.14681	0.25319

サンプル数 10 以上でよく似た値になっていることがわかる．

5. 今後

サンプル数が少ない場合の推定精度を上げることと，各確率分布のもう一つのパラメータにおいても区間推定の方法を考察していきたいと考えている．

６．参考文献

[1]Stuart and Ord(1994),”Kendall’s Advanced Theory of Statistics,Vol1:Distribution Theory 6th Edition”,John Wiley & Sons Ltd,P356-360.

所属および役職等は，公開当時のものです．

■公開資料ページ
弊社ウェブページで各種資料をご覧いただけます http://www.i-juse.co.jp/statistics/jirei/

■お問い合わせ先
(株)日科技研 数理事業部 パッケージサポート係 http://www.i-juse.co.jp/statistics/support/contact.html