

A New Skewed Distribution Function Applicable to Biology II.How to Calculate the Constant and its Application

メタデータ	言語: jpn 出版者: 公開日: 2017-10-04 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/8564

生物学に応用できる新しい非対称分布関数

II. 常数の求め方とその応用について

金沢大学医学部衛生学講座 (金沢大学衛生学協力研究員)

茶 谷 士 一

私は前回、本誌上に、非対称分布関数 II 型を発表したが、今回この関数に関して、その常数と、平均値、標準偏差、歪度、尖鋭度との関係を示す数値表を考案した。また一つの実例を挙げてこれらの表の応用方法を解説した。生物統計の解析に有用であると思う。

Key words skewed distribution function II, compound skewed distribution function

前回、本誌上に報告した非対称分布関数 II を、実際に適用するためには、この関数の有する常数を決定しなければならないが、この関数は従来の方法のように数理論的に、乗積能率から一般的な関係式を求めることは極めて困難である。したがって今の場合、乗積能率の算定可能な m が整数の場合を含め、近似的に算定し易い特定の数値に対応する、平均値、標準偏差、歪度、尖鋭度、を求めて、あらかじめ、これらの数値と常数との対照表を作製しておき、対照表から常数を読み取るようにしなければならない。以下、対照数表の作製方法と、事例について応用方法を解説する。

1. 数表の作製方法

1. m が整数のとき

非対象分布関数 II は、 $b=1$ のとき、

$$Z = y_0 \left[1 - \frac{m^2}{a^2} \left\{ (1+\chi)^{\frac{1}{m}-1} - 1 \right\}^2 \right]^m (1+\chi)^{\frac{1}{m}-1} \quad (1)$$

$$y_0 = \frac{1}{a} \frac{\Gamma(2m+2)}{2^{2m+1} \{\Gamma(m+1)\}^2} \quad (2)$$

と書けるが、この関数の平均値、並びに、原点の周りに於ける乗積能率は次のように表わされる。

$$Mi = y_0 \int_{\lambda_1}^{\lambda_2} \chi^i \left[1 - \frac{m^2}{a^2} \left\{ (1+\chi)^{\frac{1}{m}-1} - 1 \right\}^2 \right]^m (1+\chi)^{\frac{1}{m}-1} d\chi \quad (3)$$

$$i = 1, 2, 3, 4$$

ここに

$$\lambda_1 = \left(1 - \frac{a}{m}\right)^m - 1 \quad (4)$$

$$\lambda_2 = \left(1 + \frac{a}{m}\right)^m - 1 \quad (5)$$

である。

上の積分において

$$\chi = \left(\frac{t}{m} + 1\right)^m - 1$$

とおけば

$$d\chi = \left(\frac{t}{m} + 1\right)^{m-1} dt$$

$$t = m \left\{ \left(\chi + 1\right)^{\frac{1}{m}} - 1 \right\} \quad (6)$$

$\chi = \lambda_1, \lambda_2$ のとき、

(4), (5), (6) より t はそれぞれ $-a, a$ となる。したがって(3)は次のように変換される。

$$Mi = y_0 \int_{-a}^a \left\{ \left(\frac{t}{m} + 1\right)^m - 1 \right\}^i \left(1 - \frac{t^2}{a^2}\right)^m dt \quad (7)$$

ここで(7)の中括弧を展開すれば、 t に関して、 mi 次の多項式となるので、(7)は、これらの各項と次の関数 (pearson II)¹⁾ との積の積分

$$y = y_0 \left(1 - \frac{t^2}{a^2}\right)^m \quad (8)$$

の和となる。このとき、 t の奇数次の項は奇関数となるので積分値は 0 となり、偶数次の項の積分だけとなる。

しかるに偶数次の項の積分に関しては、次に示すような漸化式が成り立つ。

$$\begin{aligned} & y_0 \int_{-a}^a t^{2n} \left(1 - \frac{t^2}{a^2}\right)^m dt \\ &= y_0 \int_{-a}^a \left(-\frac{a^2}{2(m+1)} t^{2n-1} \right) \left\{ -(m+1) \frac{2}{a^2} t \right\} \left(1 - \frac{t^2}{a^2}\right)^m dt \\ &= y_0 \left[-\frac{a^2}{2(m+1)} t^{2n-1} \left(1 - \frac{t^2}{a^2}\right)^{m+1} \right]_{-a}^a \\ & \quad + \frac{a^2(2n-1)}{2(m+1)} y_0 \int_{-a}^a t^{2n-2} \left(1 - \frac{t^2}{a^2}\right)^{m+1} dt \end{aligned}$$

すなわち結果として生じた 2 つの項のうち、始めの項の $\left[\right]_{-a}^a$ 中の関数は、 t の奇関数である。したがってその値は 0 となり、始めの被積分関数は、 t の次数が 2 だけ減った被積分関数に漸化される。よって、この操作を n 回繰り返すことにより、次のように漸化されその値を求めることが出来る。

$$\int_{-a}^a t^{2n} y dt = \frac{a^{2n}(2n-1)(2n-3)\cdots 1}{2^n (m+1)(m+2)\cdots(m+n)} y_0 \int_{-a}^a \left(1 - \frac{t^2}{a^2}\right)^{m+n} dt \quad (9)$$

$$y_0 = 1 / \sqrt{2\pi \log \frac{\epsilon - a}{M - a}} \quad a < M < \epsilon$$

M p(χ)の中央値
ε p(χ)の平均値

$$R = \sqrt{2 \log \frac{\epsilon - a}{M - a}} \quad t = \frac{\log \frac{\epsilon - a}{M - a}}{R}$$

において、λが負の値をとるとき、関数の両端は負となるが、(λ)の値が小さいときは、負領域を無視しても差し支えない場合が多く、確率論的な操作が容易なことから、歪度、尖鋭度の広い範囲に亘って、非対称分布関数Ⅱの代用となり得るが、(λ)の値が大きくなるにしたがって誤差は無視出来なくなる。この問題を詳細に論ずるには長編の論文を要するので、ここでは、その概略を窺い知る手段として、複関数のRが極小になったとき、すなわちその極限形としての次の関数について考察する。

$$y = \phi_0(\chi) + \lambda(\chi^4 - 6\chi^2 + 3)\phi_0(\chi) \quad \phi_0(\chi): \text{正規関数} \quad (12)$$

標本に(12)を当て嵌めたとき、λが負のときの先端の様子を図1で示す。標本を常套の手段のように階級(P₀, P₁, P_{m-1}, P_m)に分けて関数の値に応じて接分するとき、末端のP_mまではその大きさに応じて接分されるが、P_mより先∞にいたるまでの範囲で正の部分 Q_mP_mR と、負の部分 RS∞ を合わせて新たに設けた P_{m+1}との間に G'P_{m+1}H' としてあてられる。これは、Q_mP_mR: と等しい面積の GP_mP_{m+1}H から RS∞ の面積と等しい GG'H'H 差し引いた面積である。このような操作で出来た分布図では、面積は関数の面積に等しいが、乗積率が変わり、尖鋭度の値(複関数のときは歪度も含む)が変化する。その様子を関数の上下端がそれぞれ±3.6と±3.8の点で負領域に移る場合を検討してみた。

原点よりRまでを共に8区間に分けた。計算の結果を表4に示す。(今の場合 P_{m+1}とRは同一点なり) (A), (B)の場合における、この関数と同じ尖鋭度を有するピアソンⅡ型¹⁾の曲線の常数mは次の式から求められる。

$$\beta_2 = \frac{6m+9}{2m+5} \quad (13)$$

上の式に (A), (B) の β₂ の値を代入するとmの値はそれぞれ 9.15, 13.88と求められる。誤差率を見るに、(A) の場合は無視出来る程でないかも知れないが実用的には不可と云える程でもない。(B) の場合は、β₂の誤差率は1/300以下であり、これ位の誤差は標本変動を考えれば実際面では無視しても差し支えない場合が多い。この関数については次章で述べるように数表を利用すれば簡単に常数を決めることが出来るので、数表はmの出来るだけ大きいところまで作製すれば良いと思われるが、一方、mは無限大にまで及ぶので途中で打ち切らなければならない。しかしまた、上の例でも類推出来るように、複合非対称関数⁹⁾を当て嵌めた場合、mが大きくなると急速に誤差率が減りよく適分するようになるので2桁以上のmは特殊な目的のない限り実用的に意義は少ないと考えられる。本文では以上の事を考慮して、mの値12までの表を作製した。mの値は小さいが、この範囲内だけで、歪度が0に近づいた場合には、尖鋭度の値は2.33より2.793までを占め、非対称分布関数Ⅰ¹⁰⁾の尖鋭度3との間に0.207の間隔を残すだけとなり、この範囲内には複合非対称分布が充分にその機能を代行出来るのである(この場合の複合関数のλの値は負で-0.008625)。

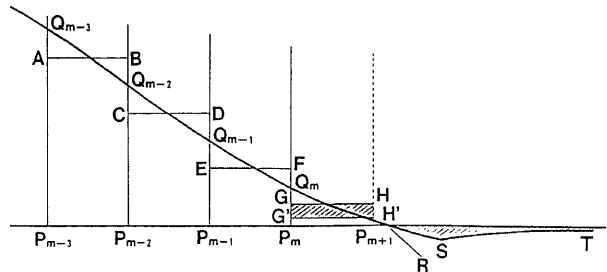


Fig.1 The upper end of the histogram made from the sample applied in the function. The area of the rectangle GG'H'H is equal to the area of the indefinite form RST. R is the point at which the curve of the function intersect the X-axis. In this case, the moments of the histogram is always larger than moments of the applied function.

Table 4. The error ratio of the moments and the peakness of the histogram to which the function was applied

(A) λ = -0.01073							
f(x)							
Class	-3.6	3.6	-3.6	3.6	Total range	Applied distribution	Error ratio
U ₂	1.0026		-0.0026		1.0	1.00086	0.086%
U ₄	2.7876		-0.0457		2.7419	2.7678	0.94%
β ₂					2.7419	2.7630	0.77%
(B) λ = -0.00801							
f(x)							
Class	-3.6	3.6	-3.6	3.6	Total range	Applied distribution	Error ratio
U ₂	1.0012		-0.0012		1.0	1.00042	0.042%
U ₄	2.8287		-0.0209		2.8078	2.8193	0.410%
β ₂					2.8078	2.8169	0.324%

f(x) = Φ₀(x) + λ(x⁴ - 6x² + 3)Φ₀(x) Φ₀(x): normal distribution
The curve of function intersect the X-axis at the point P. In the case A, P is 3.6, and in the case B, P is 3.8.

Table 5. Distribution of blood contents of VB₁ in spring

Blood contents of VB ₁ (γ/dl)	4.1-5.0	5.1-6.0	6.1-7.0	7.1-8.0	8.1-9.0	9.1-10.0	Total number of persons
Number of persons	11	14	13	8	2	1	49

One of the 50 blood contents of VB₁ revealed 16.0 γ/dl which is especially high and thought to be in a special condition. This was eliminated from the calculation.

II. 標本への適用

実際例について、この曲線を当て嵌める仕方の概要を説明する。例として挙げたのは、当教室の業報第39号(昭和34年10月)に記載されているものでその論文中の春季血中総 VB₁ 量分布⁹⁾ についてである。

1. 常数の決定

表5に示した成績の(5.1~6.0)の級の midpoint 5.55を原点とし、級間を単位として4次までの乗積能率を計算すると

$$M_1=0.5714, M_2=1.8367, M_3=3.7551, M_4=11.6327$$

したがって、平均値 $\epsilon=5.55+0.5714=6.0714$

次に ϵ を原点とする2次以上の乗積能率を計算すると

$$\mu_2 = M_2 - M_1^2 = 1.5102$$

$$\mu_3 = M_3 - 3M_2M_1 + 2M_1^3 = 0.9798$$

$$\mu_4 = M_4 - 4M_3M_1 + 6M_2M_1^2 - 3M_1^4 = 6.3283$$

したがって

$$\sqrt{\mu_2} = 1.2289 \quad \beta_1 = \frac{\mu_3}{\mu_2^2} = 0.2787 \quad \sqrt{\beta_1} = 0.5279$$

$$\beta_2 = \frac{\mu_4}{\mu_2^3} = 2.7743 \quad \sqrt{\beta_2} = 1.6657$$

このようにして標本の $\sqrt{\beta_1}$, $\sqrt{\beta_2}$ が求められたら次はこの $\sqrt{\beta_1}$, $\sqrt{\beta_2}$ の値を持つ関数を数表から読み取らねばならない。それには先ず表3の2, の $m=12$ の行で、上記の値を持つ $\sqrt{\beta_1}$ が、 $\sqrt{\beta_2}$ の右側にあることを確かめる。すなわち $\sqrt{\beta_1}$ に対応する k 値が、 $\sqrt{\beta_2}$ に対応する k 値より大きい場合である。もしこの順序が反対ならば、 m は12より大きく、2の表を利用することは出来ない。今の場合、 $\sqrt{\beta_1}$ に対応する k 値は、0.08と0.1との間にあり、 $\sqrt{\beta_2}$ に対応する k 値は0.02より小さい。すなわち前述の条件を満たしている。このような場合には、 $\sqrt{\beta_1}$, $\sqrt{\beta_2}$ の存在する大凡の位置を、 m の小さい方、すなわち上の方へ順次なぞって行くと、両者の間は漸次狭くなり、ついに順序が逆転する場合がある。この途中で両者が一致する場所が求める m, k の値である。今の場合、 m の3の行線上で両者は共に k 値が0.35と0.4との間にあるので先ずこの線上で標本の $\sqrt{\beta_1}$, $\sqrt{\beta_2}$ の値を持つ二点 P_1, Q_1 を比例部分の理で求める。次にこの行線の上下何れかで順序が逆になっている行線上で同じ値を持つ二点 P_2, Q_2 を求める。線分 $\overline{P_1P_2}$, $\overline{Q_1Q_2}$ の交点の座標が求める m, k の値である。

m を x 軸, k を y 軸に取れば

$x=3.0$ のとき

$$P_1 \text{ の } y \text{ 座標} = 0.35 + \frac{0.5279 - 0.501}{0.570 - 0.501} \times 0.05 = 0.3695$$

$$Q_1 \text{ の } y \text{ 座標} = 0.35 + \frac{1.6657 - 1.664}{1.691 - 1.664} \times 0.05 = 0.3531$$

すなわち $m=3$ の行線上では $\sqrt{\beta_1}$, $\sqrt{\beta_2}$ の大きさの順序が $m=12$ のときと変わらないから次は、上の $m=2.5$ の行線上で、 P_2, Q_2 を定める。

$x=2.5$ のとき

$$P_2 \text{ の } y \text{ 座標} = 0.45 + \frac{0.5279 - 0.492}{0.544 - 0.492} \times 0.05 = 0.4845$$

$$Q_2 \text{ の } y \text{ 座標} = 0.50 + \frac{1.6657 - 1.657}{1.678 - 1.657} \times 0.05 = 0.5207$$

二点 $P_1(3, 0.3695)$, $P_2(2.5, 0.4845)$ 並びに、二点 $Q_1(3, 0.3531)$, $Q_2(2.5, 0.5207)$ を通る直線の方程式は、次に示す二点 (x_1, y_1) (x_2, y_2) 通る直線の方程式

$$(y_1 - y_2)x + (x_2 - x_1)y + x_1y_2 - x_2y_1 = 0$$

に各点の座標を代入すれば、

$$(0.3695 - 0.4845)x + (2.5 - 3)y + 3 \times 0.4845 - 2.5 \times 0.3695 = 0$$

$$(0.3531 - 0.5207)x + (2.5 - 3)y + 3 \times 0.5207 - 2.5 \times 0.3531 = 0$$

となる。これを解いて

$$\begin{aligned} x &= 2.84 & \text{従って} & m = 2.844 \\ y &= 0.405 & & k = 0.405 \\ & & & a = 1.1518 \end{aligned}$$

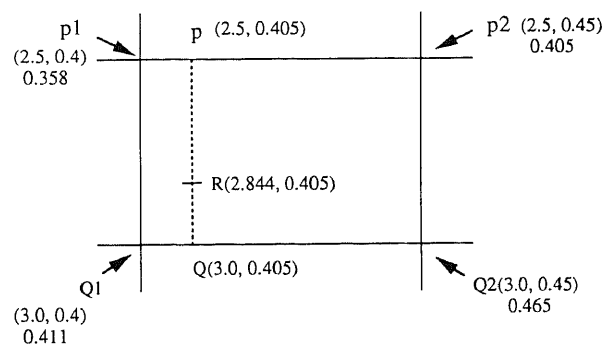


Fig. 2 The rectangle which consists of four points (P_1, P_2, Q_1, Q_2) on the Table 2. Four points (P_1, P_2, Q_1, Q_2) circle the point R which has the coordinates (m, k) that are equal to the numerical value of the constants (m, k) calculated from Table 3. The numerical values in the parenthesis shows the coordinates (m, k) of the point. The numerical values under the parenthesis shows the standard deviation of the point. The m -coordinate of the point P is equal to m -coordinates of points (P_1, P_2) and the k -coordinate of the point P is equal to the k -coordinate of the point R. The m -coordinates of the point Q is equal to m -coordinate of the point Q is equal to m -coordinates of the points (Q_1, Q_2) and the k -coordinate of the point Q is equal to the k -coordinate of the point R.

を得る。

次に表2を利用して今求めた m, k に対応する $\sqrt{\mu_2}$ の値を求め、表2よりこの m, k の値を取り囲む4点を取り出し図示すれば図2の如くなる。点の座標の下の数はその点の m, k の座標値と同じ値の m, k をもつ関数の標準偏差値である。

線分 $\overline{P_1P_2}$ 上に点 $P(2.5, 0.405)$ をとると、

P に於ける標準偏差は

$$0.358 + \frac{0.405 - 0.4}{0.45 - 0.4} (0.405 - 0.358) = 0.3628$$

線分 $\overline{Q_1Q_2}$ 上に点 $Q(3.0, 0.405)$ をとると

Q における標準偏差は

$$0.411 + \frac{0.405 - 0.4}{0.45 - 0.4} (0.465 - 0.411) = 0.4164$$

線分 \overline{PQ} 上に点 $R(2.844, 0.405)$ をとると、

R における標準偏差は

$$0.3628 + \frac{2.844 - 2.5}{3.0 - 2.5} (0.4164 - 0.3628) = 0.3997$$

次に、標本の標準偏差を今得られた標準偏差と比較して、 b を決定する。

$$b = \frac{1.2289}{0.3997} = 3.075$$

以上で常数が全て決定したので関数は次のように書ける。

$$Z = \frac{y_0}{3.075} \left[1 - 6.097 \left\{ \left(1 + \frac{\chi}{3.075} \right)^{\frac{1}{2.844}} - 1 \right\}^2 \right] \left(1 + \frac{\chi}{3.075} \right)^{-0.6484} \quad (10)$$

$$y = \frac{1}{a} \cdot \frac{\Gamma(2m+2)}{2^{2m+1} \{\Gamma(m+1)\}^2} \quad \left\{ \begin{array}{l} a=1.1518 \\ m=2.844 \end{array} \right\} \text{ 原点は中央値}$$

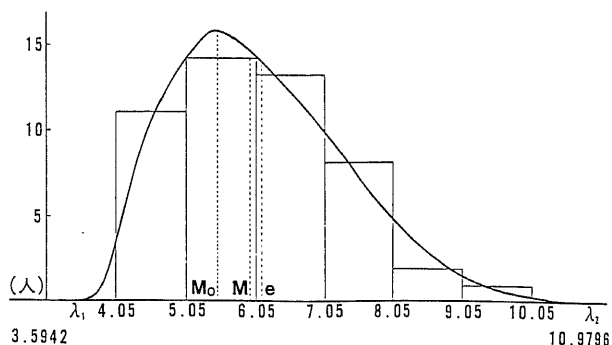


Fig. 3 The histogram of the sample and the curve of the applied function. The curved line shows the figure of skewed distribution function II, by the calculated value of Table 5. The histogram shows the distribution by classes of the sample. Both, the curve and the histogram parallel with each other.

Table 6. Theoretical frequency of applied function and the number of person in the classes

Class	Middle	χ	$f(\chi)$	area	ratio	X49	Sample
λ_1		2.3756	0.0				
		-2.1477	0.0575	0.0711	2.13	1.04	0
4.05		-1.9198	0.2380				
	4.55	-1.4198	0.7014	0.6707	20.07	9.84	11
5.05		-0.9198	0.9807				
MO		-0.4585	1.0527				
	5.55	-0.4198	1.0523	1.0287	30.74	15.09	14
M		0	1.0				
6.05		0.0802	0.9820				
	6.55	0.5802	0.8307	0.8270	24.75	12.13	13
7.05		1.0802	0.6570				
	7.55	1.5802	0.4810	0.4844	14.50	7.10	8
8.05		2.0802	0.3252				
	8.55	2.5802	0.1996	0.2052	6.14	3.01	2
9.05		3.0802	0.1073				
	9.55	3.5802	0.0470	0.0512	1.53	0.75	1
10.05		4.0802	0.0119				
		4.5450	0.0018	0.0030	0.09	0.04	0
λ_2		5.0098	0.0				
Sum				3.3413	100	49.00	49
				χ^2 test		0.0513	
				degree of freedom		3	
				1% point		11.341	

2. 曲線の当て嵌め方

関数曲線を標本に当て嵌めるには、先ず確率論的に最も安定している平均値を一致させねばならない。関数の原点は中央値であるから、平均値を求めねばならないが、これには前節で標準偏差を求めたときと同じ要領で、(表1)より次のように求められる。

$$P \text{ の } \sqrt{\text{平均値}} = 0.1935 + (0.2177 - 0.1935) \times \frac{0.405 - 0.4}{0.45 - 0.4} = 0.19592$$

$$Q \text{ の } \sqrt{\text{平均値}} = 0.2309 + (0.2598 - 0.2309) \times \frac{0.405 - 0.4}{0.45 - 0.4} = 0.23379$$

$$R \text{ の } \sqrt{\text{平均値}} = 0.19592 + (0.23379 - 0.19592) \times \frac{0.2844 - 0.25}{0.3 - 0.25} = 0.2220$$

したがって Rの平均値 = $0.222^2 = 0.0493$

標本の平均値 = $0.0493b = 0.0493 \times 3.075 = 0.1516$

標本の平均値は区間間隔が $1 \frac{\gamma}{dl}$ で単位間隔と数値は等しく、平

均値が假の平均値5.55より0.5714大である。したがって標本の中央値(関数の原点に当たる)は

$$5.55 + 0.5714 - 0.1516 = 5.9698$$

となる。従って関数を標本に当て嵌める計算をする場合、関数の変数 x には標本の境界値より、5.9698減じた数値を当てねばならない。上記の操作によって標本と関数の平均値が一致するのである。また、関数の値は、変数に対応する相対的な値を求めるだけで、以後の解析は充分であるから、関数 Z の y_0/b を外して計算した。このようにして標本の各級の境界点、級の中央

点に対応する関数の値を求め、次に Simpson の方法²⁾によって各級内の面積を求め、これに対応する人数を計算した。以上を表6、並びに図3にまとめて示す。

結 論

以上私は、非対称分布関数IIの適合範囲のうち、複合非対称分布関数では充分正確にその機能を代行し得ない範囲、すなわち m の小さい範囲で(本文では $m=2$ より12まで)。関数の有する常数と、その平均値、標準偏差、歪度、尖鋭度との関係を対照表としてまとめ、実例を挙げてその応用方法を解説した。これによって、非対象分布関数Iに比べ尖鋭度の小さい範囲の分布全般に対して、複合非対称分布関数で代行出来ない部分にも広く適用の範囲を広めることが出来た。

謝 辞

稿を終えるに当たり、御校閲の労をお願いしました橋本和夫名誉教授に深く感謝の意を表します。また数表の作製に当り膨大な計算を分担して戴いた、高松病院小野江信介先生並びに、いろいろ御世話して戴いた教室の皆さんに厚く御礼申し上げます。

文 献

- 1) 佐藤良一郎：数理統計学，増補11版，107-108頁，151-155頁，倍風管，東京，1954.
- 2) 矢野健太郎：微分積分学，第16版，133頁，裳華房，東京，1963.
- 3) 茶谷士一：生物学に適用できる一種の複合非対称分布関数と其の理論。十全医会誌，76，254-266(1968).
- 4) 茶谷士一：生物学に適用できる一種の非対称分布関数と其の理論。十全医会誌，66，383-392(1960).
- 5) 松井一信：北陸地方紡績女子工員の血中V.B1量並びにその季節的変動について。金沢大学医学部衛生学教室業報，39，25-29(1959).

A New Skewed Distribution Function Applicable to Biology. (II) How to Calculate the Constant and its Application Shiichi Chadani, Department of Hygiene, School of Medicine, Kanazawa University, Kanazawa 920-J. Juzen Med Soc., 103, 656-663 (1994)

Key words skewed distribution function II, compound skewed distribution function

Abstract

I have already presented a paper on the function "Skewed Distribution II" in this journal. Concerning this function, this time I have devised three tables which indicate the relation between its constant and mean, standard deviation, distortion and sharpness. I have also explained how to apply these tables with an example. I think that these tables are useful for analyzing biological statistics.