II-295　二つの対数正規分布の合成標本の統計処理

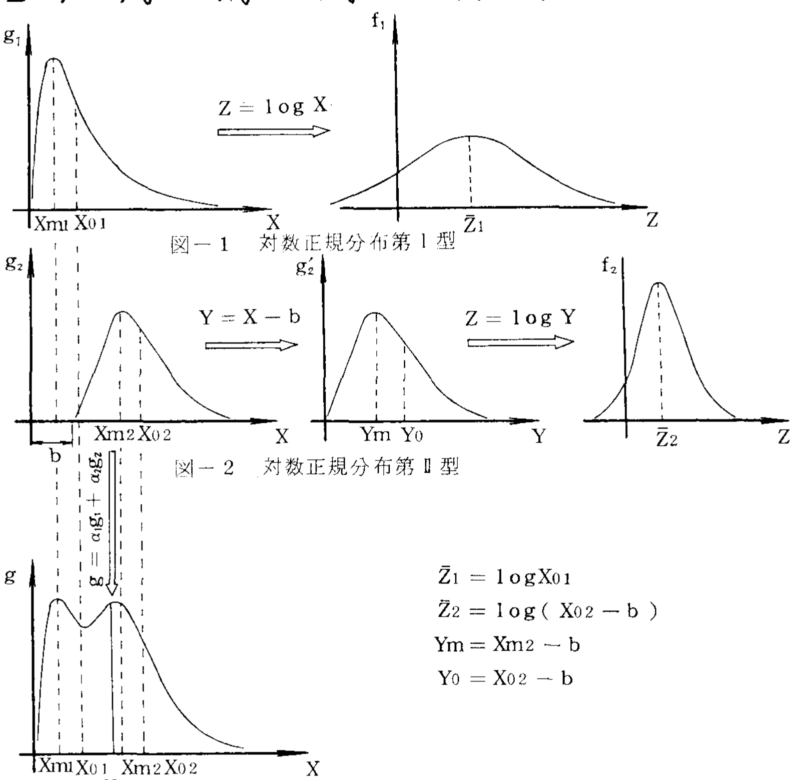
日本大学工学部　正員　安田禎輔

まえがき　二つのピークを持つ度数分布の統計処理のうち，この度数分布が二つの正規分布の合成，b=0 の二つの対数正規分布の合成とみなせる場合，および標本の個数のはなはだ少ない場合の処理については本講演会においてすでに報告した。

対数正規分布のうちで最も一般的な形はbを持つ場合である。本報においては，bを有する二つの対数正規分布の合成標本とみなせる場合の統計的処理方法を論ずる。

1. 第I,II型の対数正規分布とこれらの合成分布

最も一般的な対数正規分布は図-2 に示したような分布であり，b=0 の場合が図-1となる。図-1を第I型の対数正規分布とし，図-2を第II型分布とする。図-3は，I型とII型の混合した分布であり，α_1およびα_2をそれぞれI型，II型の加重とすれば，$g(x)=\alpha_1 \cdot g_1(x)+\alpha_2 \cdot g_2(x)$ で示される。



図－1　対数正規分布第I型

図－2　対数正規分布第II型

図－3　第I，II型混合分布

2. 第II型の対数正規分布

第II型の対数正規分布は，確率変数xを$z=\log_{10}(x-b)$によりzに変換すれば正規分布となる分布である。

$$g(x)=\frac{\log_{10}e}{\sigma_0\sqrt{2\pi}}\cdot\frac{1}{x-b}\,e^{-\frac{1}{2\sigma_0^2}\left(\log_{10}\frac{x-b}{x_0-b}\right)^2} \qquad (1)$$

は，置換$z=\log_{10}(x-b)$により

$$g(x)dx=\frac{1}{\sigma_0\sqrt{2\pi}}\,e^{-\frac{(z-\bar{z})^2}{2\sigma_0^2}}dz=f(z)dz$$

となる。

$$\therefore \quad f(z)=\frac{1}{\sigma_0\sqrt{2\pi}}e^{-\frac{(z-\bar{z})^2}{2\sigma_0^2}}$$

となり，$f(z)$ は正規分布関数となるから，$g(x)$ は第Ⅱ型の対数正規分布関数である。

ただし，
$$\bar{z}=\overline{\log_{10}(x-b)}=\log_{10}(x_0-b) \tag{2}$$
$$\sigma_0=\sqrt{\frac{1}{n}\sum\left(\log_{10}\frac{x-b}{x_0-b}\right)} \tag{3}$$

で，x_0は分布$g(x)$の中央値であり，σ_0は正規分布$f(z)$の標準偏差である。

つぎに，分布$g(x)$の最頻値x_mを求めると
$$(x_m-b)=(x_0-b)\cdot 10^{-\sigma_0^2/\log_{10}e} \tag{4}$$

となり，
$$g(x_m)=\frac{\log_{10}e}{\sigma_0\sqrt{2\pi}}\cdot\frac{10^{(\sigma_0^2/\log_{10}e)}}{x_0-b}\cdot e^{-\frac{1}{2}(\sigma_0/\log_{10}e)^2} \tag{5}$$

となる。bは(4)式より
$$b=\frac{x_m-x_0\cdot 10^{-(\sigma_0^2/\log_{10}e)}}{1-10^{-(\sigma_0^2/\log_{10}e)}} \tag{6}$$

となる。

3. 対数正規分布第Ⅰ，Ⅱ型混合標本の分離法

混合標本nを生起原因別に二つの標本n_1およびn_2とに分離する手順を箇条書きに示す。

① 度数分布曲線または，確率密度曲線図-3を描き，近似的に最頻値x_{m1}を読み取る。

② 図-3のx座標を$z=\log_{10}x$によって変換した曲線より，x_{01}を読み取る。

③ x_{m1}およびx_{01}の読み取り値を(4)式（b=0と置き）に代入し，σ_{01}を求める。

$$\sigma_{01}=\sqrt{\log_{10}e\cdot\log_{10}\frac{x_{01}}{x_{m1}}}$$

④ b=0とし，(5)式より，$g_1(x_{m1})$を計算する。

$$g_1(x_{m1})=\frac{\log_{10}e}{\sigma_{01}\sqrt{2\pi}}\cdot\frac{10^{(\sigma_{01}^2/\log_{10}e)}}{x_{01}}\cdot e^{-\frac{1}{2}(\sigma_{01}/\log_{10}e)^2}$$

⑤ 混合標本にピークが二つ現われることおよびb≠0などから，$g(x_{m1})\fallingdotseq\alpha_1\cdot g_1(x_{m1})$，$b\geqq x_{m1}$と考えられる。したがって
$$\alpha_1=\frac{g(x_{m1})}{g_1(x_{m1})} \quad \text{および} \quad \alpha_1+\alpha_2=1$$

より，α_1およびα_2が求められる。

⑥ 次式により，曲線$g_2(x)$を描き，bとx_{m2}を読み取る。

$$g_2(x)=\frac{1}{\alpha_2}\{g(x)-\alpha_1\cdot g_1(x)\}$$

⑦ $g_2(x)$曲線を対数変換$z=\log_{10}(x-b)$してx_{02}を読み取る。ただし，この変換は次式による。

$$f_2(z)=\frac{x-b}{\log_{10}e}\cdot g_2(x)$$

⑧ x_{m2}，x_{02}，bを(4)式に代入してσ_{02}を求める。

$$\sigma_{02}=\sqrt{\log_{10}e\cdot\log_{10}\frac{x_{02}-b}{x_{m2}-b}}$$

⑨ 以上で求めたσ_{01}，σ_{02}，x_{01}，x_{02}，bを用いた曲線

$$g(x)=\alpha_1\cdot g_1(x)+\alpha_2\cdot g_2(x)$$

と原曲線$g(x)$とを比較し，一致するまで①～⑨を修正しながら繰り返す。

以上のようにして，標本nを生起原因別にn_1，n_2に分離して，確率，超過確率および非超過確率などを計算すれば分離しない場合より精度は高まる。

<参考文献>
1) 安田禎輔：二つのピークを持つ度数分布の統計的処理，第36回土木学会年次学術講演会講演集 1981.10.
2) 安田禎輔：二つの対数正規分布の合成標本の統計処理，日本大学理工学部学術講演会論文集，1981.11.