

2変数最大エントロピー分布の適用性

寒川典昭* 荒木正夫** 寺島 彰***

(昭和62年5月25日受理)

Applicability of the Two-Variate Maximum Entropy Distribution

Noriaki SOGAWA,* Masao ARAKI,**
and Akira TERASHIMA***

The applicability of the two-variate maximum entropy distribution (2MED) to the two-variate gamma type distribution (2GTD) and the observed hydrologic histogram (OHH), when several low-order moments are known, was examined by employing the iteration method to calculate relevant parameters. Real and estimated values of the moments were used to examine the possibility of improving the applicability of the 2MED to the 2GTD and OHH. After investigating the variation of the moments by using simulation data from the 2GTD, values of the moments were changed artificially and used to analyze the sensitivity of the 2MED. The 2M(4, 4, 2, 2), which stands for the 2MED given the first four moments with respect to variables x , y , and xy , was in good agreement with any types of 2GTDs and OHHs. The 2M(3, 3, 1, 1), which stands for the 2MED given the first three moments with respect to x and y and the first moment with respect to xy , was in good agreement with the two-variate exponential type distribution and the two-variate exponential type OHH.

1. ま え が き

2地点で観測された水文量を同時に頻度解析するとき、2変数の同時確率密度関数が必要である。従来水文統計における2変数分布として、指数分布^{1),2),3)}、正規分布⁴⁾、ガンマ分布^{2),3)}が用いられてきた。ところが、2変数指数分布、2変数正規分布は周辺分布が指数分布、正規分布をする場合であり、このような水文データは特殊な場合に限られている。一方、2変数ガンマ分布は広く非対象形を表現することができ、応用面において有効な分

* 土木工学科 助手

** 土木工学科 教授

*** 上田千曲高等学校

布とされている^{5),6)}が、変形ベッセル関数を含んでいるためその取り扱いに困難を生じる場合がある⁷⁾とともに、3変数以上に拡張したときパラメタ推定が難かしい。

このような現状に鑑み、1984年、我々⁸⁾は2変数最大エントロピー分布を提案した。これは、1変数最大エントロピー分布⁹⁾の性質を保持したままでの2変数への拡張であり、小標本を取り扱う水文統計にとって有効な分布であること、3変数以上の多変数¹⁰⁾及びm変数を条件とした(n-m)変数の条件つき分布¹¹⁾へと拡張可能であること、密度関数式が簡潔で利用しやすいこと、等の特徴を有し、水文量の多変数統計を可能ならしめた。しかし、母集団の再現性及び実測データの適用性の検討については、水文統計上重要な未解決課題として残されていた。

本稿は、我々が用いたパラメタ同定法について概説しながら、主として次の4項目を通じて上述の問題を解決したものである。

- 1) 典型的な5つの形状を示す2変数ガンマ分布を母集団と仮定して、2変数最大エントロピー分布で十分な精度まで再現するのに必要なモーメントの次数を把握する。
- 2) シミュレートしたデータから得られるモーメントの精度を明らかにし、シミュレーションデータからのモーメントを用いた最大エントロピー分布と真値のモーメントを用いた最大エントロピー分布の比較から必要なデータ数を推定するとともに、データ数を固定した場合の生起データの相違による最大エントロピー分布の変動性をとらえる。
- 3) モーメントを人為的に変動させ、最大エントロピー分布の感度を分析して、許容できるモーメントの変動域を明らかにする。
- 4) 実測の水文データを用いて最大エントロピー分布を求め、データの頻度確率と比較することにより、最大エントロピー分布の実データへの適用性、及び必要とされるモーメントの次数を検討する。

尚、3～6章において、適合度の改善及び適合度の良否の判断は、必要に応じて3次元立体図、等確率線図、断面図、メッシュ確率を用いながら、我々が頻度解析を通じて得た経験に基づいている。いずれ客観的な判断指標の導入を検討しなければならないと考えているが、現時点ではこのような方法に頼らざるを得ないことを付記しておく。

2. 理論式の概説****,8)

2.1 2変数最大エントロピー分布

確率変数 x, y の同時確率密度関数を $p(x, y)$ とするとエントロピーは次式で定義される。

$$H(X, Y) = - \iint p(x, y) \ln p(x, y) dx dy \quad (1)$$

いま、密度関数の具備すべき条件と任意関数 $g_r(x, y)$ の期待値を

$$\iint p(x, y) dx dy = 1, \quad (2)$$

**** 文献8)の記述は不十分であったため補足を加えて再記する。

$$\iint g_r(x, y) p(x, y) dx dy = E[g_r(x, y)], \quad r = 1, 2, \dots, M \quad (3)$$

と書くと, (2), (3) 式を制約条件として (1) 式を最大にする分布, すなわち最大エントロピー分布は次式で与えられる.

$$p(x, y) = \exp \left\{ -\lambda_0 - \sum_{r=1}^M \lambda_r g_r(x, y) \right\} \quad (4)$$

ここに, $E[\cdot]$ は期待値記号である. λ_r はラグランジュ乗数であり, (4) 式が (2), (3) 式を満足するように決定される. (4) 式は, 特定の $g_r(x, y)$ を与えることにより, 2変数の, 一様, 指数 (相関がない場合), 正規分布になり得る⁸⁾.

いま, 1変数における4次までの統計モーメントは, それをもつ確率分布の形状と深くかかわっているため, ここでも統計モーメントで情報を与えることを考える. $x[0, \infty)$, $y[0, \infty)$ とすると, (3) 式は次のように書き換えられる.

$$\int_0^{\infty} \int_0^{\infty} x^a p(x, y) dx dy = x^a \mu_a, \quad a = 1, 2, \dots, N_a, \quad (5)$$

$$\int_0^{\infty} \int_0^{\infty} y^b p(x, y) dx dy = y^b \mu_b, \quad b = 1, 2, \dots, N_b, \quad (6)$$

$$\int_0^{\infty} \int_0^{\infty} x^c y^d p(x, y) dx dy = x^c y^d \mu_{cd},$$

$$c = 1, 2, \dots, N_c, \quad d = 1, 2, \dots, N_d \quad (7)$$

このとき, (4) 式の最大エントロピー分布は次式となる.

$$p(x, y) = \exp \left(-\alpha - \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) \quad (8)$$

ここに, $\alpha, \beta_a, \gamma_b, \delta_{cd}$ はラグランジュ乗数 (母数あるいはパラメタと呼ぶこともある) であり, (8) 式が (2), (5)~(7) 式を満足するように決定される. このようにして求めたラグランジュ乗数をもつ (8) 式を $2M(N_a, N_b, N_c, N_d)$ と略記する.

2.2 パラメタ同定法

(8) 式を (5)~(7) 式に代入して整理すると次式が得られる.

$$\int_0^{\infty} \int_0^{\infty} x^i \exp \left(-\sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy$$

$$= x^i \mu_i \int_0^{\infty} \int_0^{\infty} \exp \left(-\sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy,$$

$$i = 1, 2, \dots, N_a, \quad (9)$$

$$\begin{aligned}
& \int_0^\infty \int_0^\infty y^j \exp \left(- \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy \\
&= y^j \int_0^\infty \int_0^\infty \exp \left(- \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy, \\
& \qquad \qquad \qquad j = 1, 2, \dots, N_b, \tag{10}
\end{aligned}$$

$$\begin{aligned}
& \int_0^\infty \int_0^\infty x^k y^l \exp \left(- \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy \\
&= x y^l \int_0^\infty \int_0^\infty \exp \left(- \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy, \\
& \qquad \qquad \qquad k = 1, 2, \dots, N_c, \quad l = 1, 2, \dots, N_d \tag{11}
\end{aligned}$$

(9)~(11)式は、 β_a , γ_b , δ_{cd} に関する非線形連立方程式であるため、このままの形では解は求まらない。いま、 β_a , γ_b , δ_{cd} の近似値を β_a^0 , γ_b^0 , δ_{cd}^0 , 残差を $\beta \varepsilon_a$, $r \varepsilon_b$, $\delta \varepsilon_{cd}$ とおくと

$$\beta_a = \beta_a^0 + \beta \varepsilon_a, \quad \gamma_b = \gamma_b^0 + r \varepsilon_b, \quad \delta_{cd} = \delta_{cd}^0 + \delta \varepsilon_{cd} \tag{12}$$

となる。これらを(9)~(11)式に代入して、 $\beta \varepsilon_a$, $r \varepsilon_b$, $\delta \varepsilon_{cd}$ に関してテーラー展開し、 $(\beta \varepsilon_a)^2$, $(r \varepsilon_b)^2$, $(\delta \varepsilon_{cd})^2$ 以上の微小量の項を無視して整理すると次式が得られる。

$$\begin{aligned}
& \sum_{a=1}^{N_a} (Z_{i+a,0} - x \mu_i Z_{a,0}) \beta \varepsilon_a + \sum_{b=1}^{N_b} (Z_{i,b} - x \mu_i Z_{0,b}) r \varepsilon_b + \\
& \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} (Z_{i+c,d} - x \mu_i Z_{c,d}) \delta \varepsilon_{cd} = Z_{i,0} - x \mu_i Z_{0,0}, \quad i = 1, 2, \dots, N_a, \tag{13}
\end{aligned}$$

$$\begin{aligned}
& \sum_{a=1}^{N_a} (Z_{a,j} - y \mu_j Z_{a,0}) \beta \varepsilon_a + \sum_{b=1}^{N_b} (Z_{0,j+b} - y \mu_j Z_{0,b}) r \varepsilon_b + \\
& \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} (Z_{c,j+d} - y \mu_j Z_{c,d}) \delta \varepsilon_{cd} = Z_{0,j} - y \mu_j Z_{0,0}, \quad j = 1, 2, \dots, N_b, \tag{14}
\end{aligned}$$

$$\begin{aligned}
& \sum_{a=1}^{N_a} (Z_{k+a,l} - x y^l \mu_{kl} Z_{a,0}) \beta \varepsilon_a + \sum_{b=1}^{N_b} (Z_{k,l+b} - x y^l \mu_{kl} Z_{0,b}) r \varepsilon_b + \\
& \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} (Z_{k+c,l+d} - x y^l \mu_{kl} Z_{c,d}) \delta \varepsilon_{cd} = Z_{k,l} - x y^l \mu_{kl} Z_{0,0}, \\
& \qquad \qquad \qquad k = 1, 2, \dots, N_c, \quad l = 1, 2, \dots, N_d \tag{15}
\end{aligned}$$

ここに,

$$Z_{i,j} = \int_0^\infty \int_0^\infty x^i y^j \exp \left(- \sum_{a=1}^{N_a} \beta_a^0 x^a - \sum_{b=1}^{N_b} \gamma_b^0 y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd}^0 x^c y^d \right) dx dy \quad (16)$$

である。(13)~(15)式は未知数 ($\beta_a, \gamma_b, \delta_{cd}$) の数と方程式の数が一致した線形連立1次方程式であり, 解くことが可能である。

具体的には図1のフローチャートに従って $\beta_a, \gamma_b, \delta_{cd}$ を算定する。ここで, 初期値 $\beta_a^0 = 0.0, \gamma_b^0 = 0.0, \delta_{cd}^0 = 0.0$ としているのは, この場合が経験的に最も収束性がよいためである。

$\beta_a, \gamma_b, \delta_{cd}$ が求まると, (8)式を(2)式に代入することにより α は次式で算定される。

$$\alpha = \ln \left\{ \int_0^\infty \int_0^\infty \exp \left(- \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d \right) dx dy \right\} \quad (17)$$

尚, (4)式あるいは(8)式で示された最大エントロピー分布を 2MED (Two-Variate Maximum Entropy Distribution) と略記することもある。

3. 真のモーメントを用いた場合の適用性

3.1 仮定した母集団

周辺分布が形状母数 $\nu_1 = n + m, \nu_2 = n (m, n \geq 0)$ のガンマ分布である2変数 x, y についての密度関数は井沢³⁾ により次のように書かれている。

$$p(x, y) = \frac{(xy)^{\frac{n-1}{2}} x^m}{\Gamma(n) \Gamma(m) (\sigma_1 \sigma_2)^{\frac{n+1}{2}} \sigma_1^m (1-\rho) \rho^{\frac{n-1}{2}}} \exp \left\{ - \frac{x}{\sigma_1(1-\rho)} - \frac{y}{\sigma_2(1-\rho)} \right\} \\ \times \int_0^1 (1-t)^{\frac{n-1}{2}} t^{m-1} \exp \left\{ \frac{\rho x t}{\sigma_1(1-\rho)} \right\} I_{n-1} \left(\frac{2\sqrt{\rho}}{1-\rho} \sqrt{\frac{xy}{\sigma_1 \sigma_2}} (1-t) \right) dt \quad (18)$$

ここに, $\Gamma(\cdot)$ はガンマ関数, $I_\nu(\cdot)$ は ν 次の変形ベッセル関数であり, σ_1, σ_2, ρ は定数である。

一方, 2変数 x, y についての原点まわりの p, q 次のモーメントを ν_{pq} と書くとき

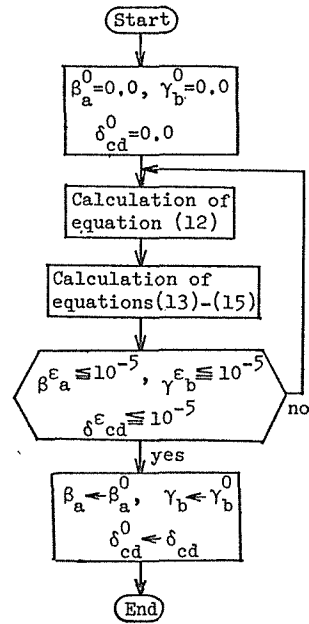


Fig. 1 Flow chart of the estimation procedure for parameters of 2MED.

$$m_{pq} = \frac{\nu_{pq}}{(\nu_{10})^p (\nu_{01})^q} \quad (19)$$

で定義されたモーメントを基準化されたモーメントと呼ぶことにする。

以下では形状母数 ν_1, ν_2 が $\nu_1 \neq \nu_2$ のとき, (18)式を $2G(\nu_1, \nu_2, \rho)$, $\nu_1 = \nu_2 = \nu$ のとき, $2G(\nu, \rho)$ と記し, 母集団と仮定する. このとき, 尺度母数 σ_1, σ_2 は $\nu_{10} = \nu_{01} = 1$ となるように決定する.

3.2 $\nu_1 = \nu_2 = \nu$ の母集団に対する適用性

ここでは, 指数分布 $2G(1, 0.2)$, 緩やかな分布 $2G(2, 0.5)$, 鋭い分布 $2G(4, 0.5)$ を母集団として採用する. 表1は, それぞれの母集団がもつ基準化されたモーメントを情報として得られた最大エントロピー分布のパラメタ (ラグランジュ乗数) の値である. 適合度は3次元立体図と等確率線図を用いて判定する (全体図比較) とともに, 後述の他の2側面からも検討を加え, この判定の妥当性を検証する.

図2は仮定した母集団 $2G(1, 0.2)$ と, 最大エントロピー分布 $2M(3, 3, 1, 1)$, $2M(4, 4, 1, 1)$ の内 $2M(3, 3, 1, 1)$ の場合の3次元立体図****である. 3者はほぼ一致しており,

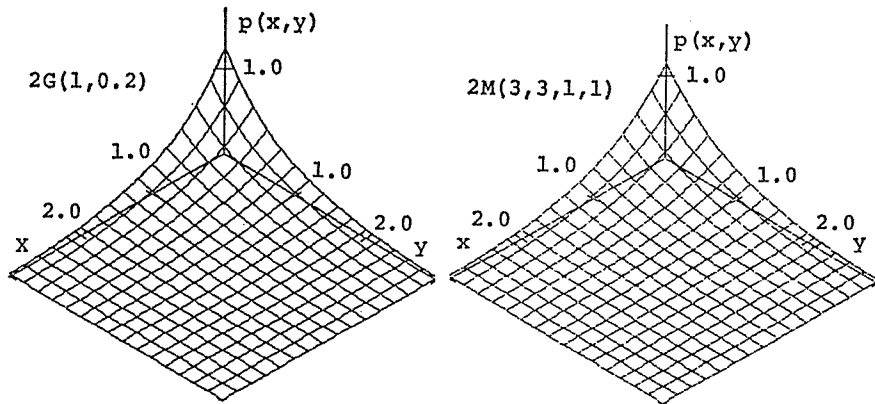
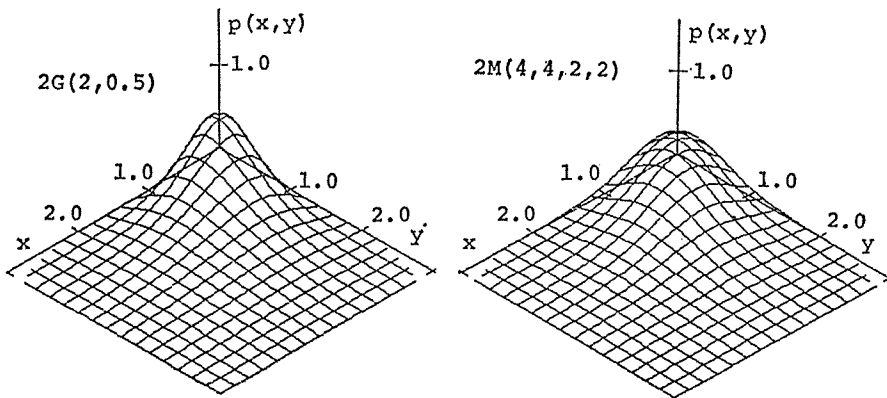
Table 1 Parameters of 2MED to $2G(1, 0.2)$, $2G(2, 0.5)$, and $2G(4, 0.5)$.

$2G(1, 0.2)$	$2M(3, 3, 1, 1)$	$2M(4, 4, 1, 1)$
α	-0.14418	-0.13749
$\beta_1 (= \gamma_1)$	1.1441	1.1308
$\beta_2 (= \gamma_2)$	-0.0029132	0.0069515
$\beta_3 (= \gamma_3)$	0.0026707	0.00050528
$\beta_4 (= \gamma_4)$	—	0.00013316
δ_{11}	-0.15045	-0.15042

$2G(2, 0.5)$	$2M(2, 2, 1, 1)$	$2M(4, 4, 1, 1)$	$2M(4, 4, 2, 2)$
α	0.57084	1.6852	1.2734
$\beta_1 (= \gamma_1)$	0.14688	-2.0921	-1.5278
$\beta_2 (= \gamma_2)$	0.77881	2.8953	3.2772
$\beta_3 (= \gamma_3)$	—	-0.70465	-0.89304
$\beta_4 (= \gamma_4)$	—	0.068628	0.077565
δ_{11}	-1.1866	-1.0480	-2.9656
$\delta_{12} (= \delta_{21})$	—	—	0.40641
δ_{22}	—	—	-0.067129

$2G(4, 0.5)$	$2M(2, 2, 1, 1)$	$2M(4, 4, 1, 1)$	$2M(4, 4, 2, 2)$
α	2.1283	5.3627	4.5707
$\beta_1 (= \gamma_1)$	-1.8801	-8.4525	-7.2592
$\beta_2 (= \gamma_2)$	2.3276	9.7708	10.475
$\beta_3 (= \gamma_3)$	—	-3.1215	-3.6144
$\beta_4 (= \gamma_4)$	—	0.39584	0.42859
δ_{11}	-2.6116	-2.3235	-6.3267
$\delta_{12} (= \delta_{21})$	—	—	1.0551
δ_{22}	—	—	-0.22084

**** 3, 4章の3次元立体図には等確率線図を併記して適合性を判断しているが, 紙面の都合で省略した.

Fig. 2 Three-dimensional expression of $2G(1, 0.2)$ and $2M(3, 3, 1, 1)$.Fig. 3 Three-dimensional expression of $2G(2, 0.5)$ and $2M(4, 4, 2, 2)$.

このような指数分布に対しては $2M(3, 3, 1, 1)$ で十分よい再現が得られている。

図3は仮定した母集団 $2G(2, 0.5)$ と、最大エントロピー分布 $2M(2, 2, 1, 1)$, $2M(4, 4, 1, 1)$, $2M(4, 4, 2, 2)$ の内 $2M(4, 4, 2, 2)$ の場合の3次元立体図である。 $2G(2, 0.5)$ のモードの位置はかなり原点の近くにある。 $2M(2, 2, 1, 1)$ の適合度は $2G(2, 0.5)$ よりフラットになり、あまりよくない。 $2M(4, 4, 1, 1)$ は母集団の形状を概略的に表現しているが、まだ十分な適合度は得られていない。 $2M(4, 4, 2, 2)$ になると x 軸及び y 軸周辺で適合度の改善がみられるが、 $2G(1, 0.2)$ の場合に比べれば適合度は劣る。

図4は仮定した母集団 $2G(4, 0.5)$ と、最大エントロピー分布 $2M(2, 2, 1, 1)$, $2M(4, 4, 1, 1)$, $2M(4, 4, 2, 2)$ の内 $2M(4, 4, 2, 2)$ の場合の3次元立体図である。 $2G(4, 0.5)$ は $2G(2, 0.5)$ に比べてモード付近での密度関数の値が大きく、その位置も原点から離れている。 $2M(2, 2, 1, 1)$ は $2G(4, 0.5)$ よりフラットになり、概略的にしか母集団をとらえていないが、 $2G(2, 0.5)$ の場合の $2M(2, 2, 1, 1)$ に比べればよい適合度を示している。

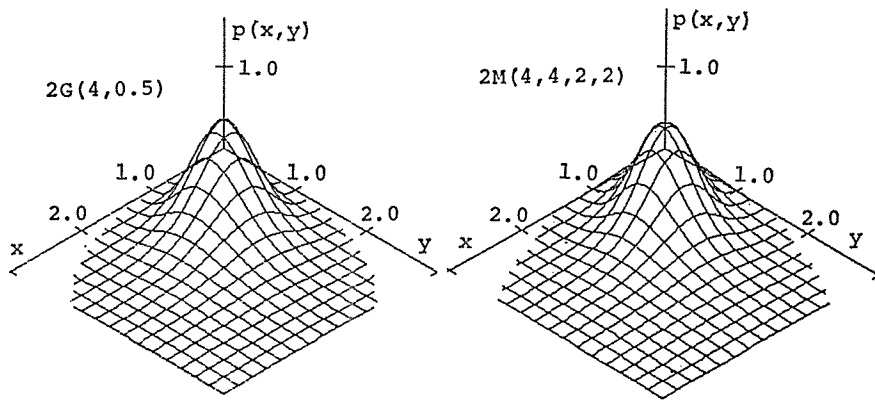


Fig. 4 Three-dimensional expression of $2G(4, 0.5)$ and $2M(4, 4, 2, 2)$.

$2M(4, 4, 1, 1)$ は $2M(2, 2, 1, 1)$ に比べてかなり母集団に対する適合度を増し、モード付近での密度関数の値、モードの位置がほぼ母集団と一致している。 $2M(4, 4, 2, 2)$ にすると、 x, y の小さい領域では $y=x$ 軸に沿う方向に収縮され、大きい領域では拡大されて、より母集団との適合度が改善されている。

次に、 $y=0$ ($2G(1, 0.2)$ の場合のみ)、 $0.5, 1.0, 1.5, 2.0$ における x 方向の断面図を普通紙上に、 $y=0$ ($2G(1, 0.2)$ の場合のみ)、 $0.5, 1.0, 1.5, 2.0, 3.0, 4.0$ における x 方向の断面図を片対数紙上に描いて適合度を検討した (断面図比較)。さらに、図 2～4 の立体図を 0.5 区画のメッシュに切り、そのメッシュ上の体積をメッシュ面積で除した値により適合度比較を行った (メッシュ確率比較)。紙面の都合により図は省略するが、これらの比較による適合度の結果は前述の考察を支持するものであった。

3.3 $\nu_1 \neq \nu_2$ の母集団に対する適用性

ここでは、一方の変数の周辺分布が指数分布、他方は山形分布をする $2G(1, 2, 0.2)$ 、両

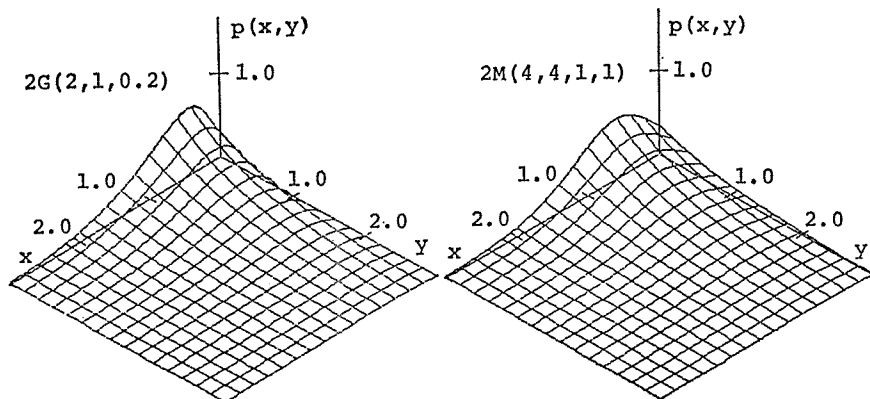


Fig. 5 Three-dimensional expression of $2G(2, 1, 0.2)$ and $2M(4, 4, 1, 1)$.

Table 2 Parameters of 2MED to 2G (2, 1, 0.2) and 2G (4, 2, 0.5).

2G(2,1,0.2)	2M(2,2,1,1)	2M(4,4,1,1)	2M(4,4,2,2)
α	0.44606	1.0496	0.98926
β_1	-0.22343	-2.7096	-2.6227
β_2	0.47515	2.8838	2.8747
β_3	—	-0.76278	-0.76496
β_4	—	0.070337	0.069914
γ_1	1.1615	1.1564	1.2405
γ_2	0.010136	0.0050997	0.0011730
γ_3	—	0.00046649	-0.00068954
γ_4	—	0.000013002	0.0000043483
δ_{11}	-0.18365	-0.16952	-0.30683
δ_{12}	—	—	0.013660
δ_{21}	—	—	0.029066
δ_{22}	—	—	-0.0016040

2G(4,2,0.5)	2M(2,2,1,1)	2M(4,4,1,1)	2M(4,4,2,2)
α	1.6458	3.9356	3.4944
β_1	-2.6333	-9.4588	-8.8829
β_2	1.9494	9.8161	9.9841
β_3	—	-3.2788	-3.4237
β_4	—	0.41419	0.41049
γ_1	0.45067	-2.0656	-1.2971
γ_2	0.59622	2.9234	2.9518
γ_3	—	-0.74772	-0.79726
γ_4	—	0.070038	0.072411
δ_{11}	-1.1016	-0.95624	-2.5731
δ_{12}	—	—	0.24403
δ_{21}	—	—	0.51503
δ_{22}	—	—	-0.066436

者とも同一でない山形分布をする 2G(4, 2, 0.5) を母集団として採用する. 表 2 は, それぞれの母集団がもつ標準化されたモーメントを情報として得られた最大エントロピー分布のパラメタの値である. 適合度の判定は 3.2 の全体図比較を用いて行う.

図 5 は仮定した母集団 2G(2, 1, 0.2) と, 最大エントロピー分布 2M(2, 2, 1, 1), 2M(4,

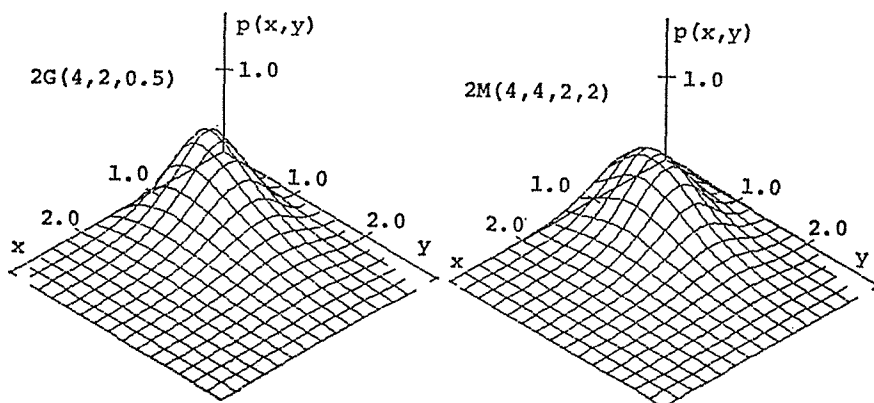


Fig. 6 Three-dimensional expression of 2G (4, 2, 0.5) and 2M (4, 4, 2, 2).

4, 1, 1), $2M(4, 4, 2, 2)$ の内 $2M(4, 4, 1, 1)$ の場合の 3 次元立体図である。 $2M(2, 2, 1, 1)$ の適合度は $2G(2, 1, 0.2)$ よりフラットになり、あまりよくない。 $2M(4, 4, 1, 1)$ にすると y 軸上付近では母集団と十分に一致していない部分があるが、全体的にはかなり適合度が改善されている。 $2M(4, 4, 2, 2)$ は $2M(4, 4, 1, 1)$ とほぼ一致し、この場合相乗モーメントの次数の増加の効果はわずかであることがわかる。

図 6 は仮定した母集団 $2G(4, 2, 0.5)$ と、最大エントロピー分布 $2M(2, 2, 1, 1)$, $2M(4, 4, 1, 1)$, $2M(4, 4, 2, 2)$ の内 $2M(4, 4, 2, 2)$ の場合の 3 次元立体図である。 $2G(4, 2, 0.5)$ の x 方向の分散は y 方向の分散より小さくなっている。 この場合も $2M(2, 2, 1, 1)$ は $2G(4, 2, 0.5)$ よりフラットになり、あまり適合度はよくない。 $2M(4, 4, 1, 1)$ にするとかなり適合度が改善され、 $2M(4, 4, 2, 2)$ にするとさらに x 軸付近での改善がみられる。

4. シミュレーションデータによるモーメントの安定性と最大エントロピー分布

4.1 モーメントの安定性

ここではモーメントのもつ誤差を検討するために、 $2G(1, 0.2)$, $2G(2, 0.5)$, $2G(4, 0.5)$ を母集団として個数 $n(n=10, 20, \dots, 150)$ のデータを 50 個発生させ、 m_{10} , m_{20} , m_{30} , m_{40} , m_{50} , m_{60} , m_{11} , m_{12} , m_{21} , m_{22} の平均と標準偏差を求めた。 図 7 は、1 例として母集団を $2G(2, 0.5)$ とした場合の m_{40} と m_{60} の安定性を示したものである。 全体的にみて、データ数が 60 個ないし 70 個付近から平均値が母集団のモーメントに近づき、標準偏差が小さくなる様子がうかがわれる。 また、高次モーメントになるに従い標準偏差が大きくなり、前章の適合性との関連から最大エントロピー分布への情報としては 4 次モーメントまでに留めることが妥当である。

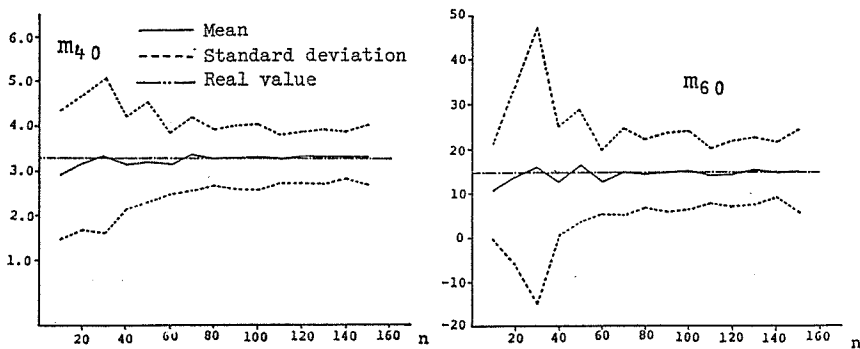


Fig. 7 Stability of moments m_{40} and m_{60} (population: $2G(4, 0.5)$).

4.2 シミュレーションデータと最大エントロピー分布

(1) データ数増加と最大エントロピー分布

ここでは母集団から発生させるデータを $n=10, 20, \dots, 150$ と増加させ、検討を要す

るデータ数に対して最大エントロピー分布を求めた.

1) $2G(1, 0.2)$ と $2M(3, 3, 1, 1)$

母集団を $2G(1, 0.2)$ とし, $n=20, 40, \dots, 140$ ごとに $2M(3, 3, 1, 1)$ を求め, 3次元立体図と等確率線図を描いた. 図8は $n=40$ の場合の3次元立体図である. 分布の形状は $n=40$ から母集団に近くなるが, その後のデータ数になっても $n=40$ の分布より適合度が劣る場合もあった.

2) $2G(4, 0.5)$ と $2M(2, 2, 1, 1), 2M(4, 4, 2, 2)$

母集団を $2G(4, 0.5)$ とし, $n=10, 20, \dots, 150$ に対する $2M(2, 2, 1, 1)$ と $n=70, 90, 100, 110, 130, 140, 150$ の $2M(4, 4, 2, 2)$ を求めた. $n=10, 20, 30, 40, 50, 60, 80, 120$ の $2M(4, 4, 2, 2)$ は求まらなかった. $2M(2, 2, 1, 1)$ の場合, 20個ごとに3次元立体図と等確率線図を描いた結果, $n=60$ 個程度から真値のモーメントを用いた $2M(2, 2, 1, 1)$ に近似していることがわかった. $2M(4, 4, 2, 2)$ の場合, 求まった7ケースとも3次元立体図と等確率線図を描いた結果, どのケースも真のモーメントを用いた $2M(4, 4, 2, 2)$ に近似していることがわかった. 図9は $n=70$ の場合の3次元立体図である.

(2) 同数の異種データと最大エントロピー分布

ここでは母集団を $2G(4, 0.5)$ とし, 1組70個のデータを5組発生させて, 各ケースごとに $2M(2, 2, 1, 1)$ と $2M(4, 4, 2, 2)$ を求めた. 図10は真値のモーメントを用いた場合と比較して最も離れた形状を示したケース3の $2M(4, 4, 2, 2)$ の3次元立体図である. ケースごとの変動はデータ数が少ない程大きくなることが予想されるが, 70個のデータ数になってもこの程度の変動が存在することを認識しておかねばならない. 又 $2M(2, 2, 1, 1)$ の場合, $2M(4, 4, 2, 2)$ より大きな変動を示すことも明らかになった.

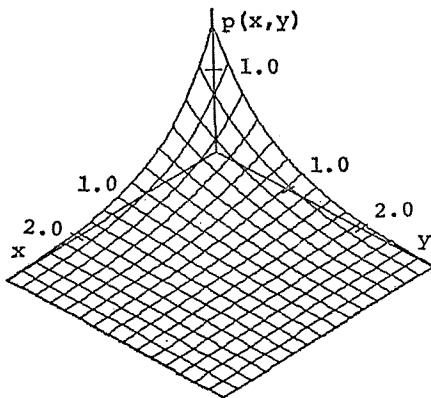


Fig. 8 Three-dimensional expression of $2M(3, 3, 1, 1)$ with $n=40$ (population: $2G(1, 0.2)$).

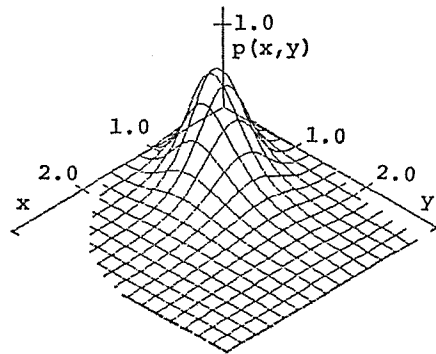


Fig. 9 Three-dimensional expression of $2M(4, 4, 2, 2)$ with $n=70$ (population: $2G(4, 0.5)$).

5. モーメントの変動による感度分析

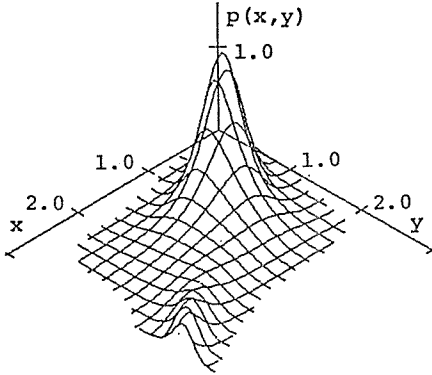


Fig. 10 Three-dimensional expression of $2M(4, 4, 2, 2)$ in the worst of five cases (population: $2G(4, 0.5)$).

ここでは、モーメントの人為的変動が最大エントロピー分布に及ぼす影響を分析する。母集団として $2G(1, 0.2)$, $2G(4, 0.5)$ を採用し、変動させるモーメントは最高次数の周辺分布に関するモーメントと最高次数の相乗モーメントとし、それ以外はすべて真値に固定した。

このとき、採用した最大エントロピー分布が、3次元立体図から判断して真のモーメントを使った最大エントロピー分布に形状的に近似していた区間は次のとおりであった。

$2G(1, 0.2)$ に対する $2M(3, 3, 1, 1)$

$$6.0 \leq m_{30} (\text{真値} 6.0) \leq 10.0, \quad 1.01 \leq m_{11}(1.2) \leq 1.75 \quad (20)$$

$2G(1, 0.2)$ に対する $2M(4, 4, 1, 1)$

$$23.0 \leq m_{40}(24.0) \leq 24.39, \quad 0.8 \leq m_{11}(1.2) \leq 1.3 \quad (21)$$

$2G(4, 0.5)$ に対する $2M(2, 2, 1, 1)$

$$1.20 \leq m_{20}(1.25) \leq 1.35, \quad 1.05 \leq m_{11}(1.125) \leq 1.175 \quad (22)$$

$2G(4, 0.5)$ に対する $2M(4, 4, 2, 2)$

$$3.22 \leq m_{40}(3.2813) \leq 3.795, \quad 2.35 \leq m_{22}(2.3828) \leq 2.43 \quad (23)$$

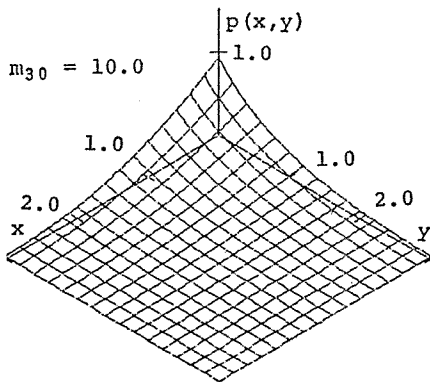


Fig. 11 Three-dimensional expression of $2M(3, 3, 1, 1)$ with the upper limit of m_{30} population: $2G(1, 0.2)$.

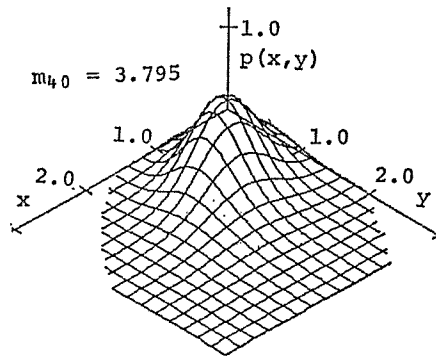


Fig. 12 Three-dimensional expression of $2M(4, 4, 2, 2)$ with the upper limit of m_{40} (population: $2G(4, 0.5)$).

図11は(20)式で示した m_{30} の上限値での $2M(3, 3, 1, 1)$, 図12は(23)式で示した m_{40} の上限値での $2M(4, 4, 2, 2)$ の3次元立体図である。

6. 実データへの適用性

6.1 2地点流量

石原・長尾¹²⁾は合流河川における治水計画に2変数正規分布を適用した。ここでは、この目的のために2変数最大エントロピー分布を求める。

使用したデータは、図13の立ヶ花地点29年分の年最大流量に対応する小市、杭瀬下の最大流量である。ここでは、両者の値を各々の平均で除した基準化した値をそれぞれ確率変数 x, y として $2M(2, 2, 1, 1)$ と $2M(4, 4, 1, 1)$ を求めた。このときのパラメタの値を表3に、 $2M(4, 4, 1, 1)$ とデータのメッシュ確率を図14に示す。データ数が少ないので適合度に関する厳密な議論はできないが、この図より $2M(4, 4, 1, 1)$ は比較的良好な適合を示していることがわかる。

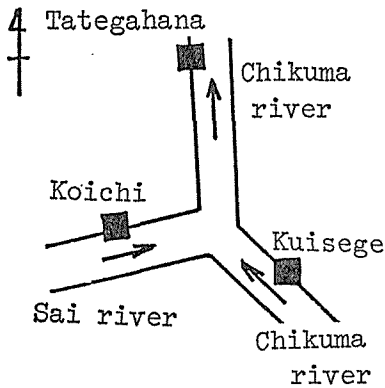


Fig. 13 Location of gauging stations.

Table 3 Parameters of 2MED to discharge data of two stations.

	$2M(2, 2, 1, 1)$	$2M(4, 4, 1, 1)$
α	1.8875	9.5384
β_1	-4.1926	-33.053
β_2	2.6726	39.458
β_3		-18.184
β_4		2.9935
γ_1	1.4446	-1.9346
γ_2	0.22082	5.3094
γ_3		-2.5014
γ_4		0.37225
δ_{11}	-1.0654	-0.95042

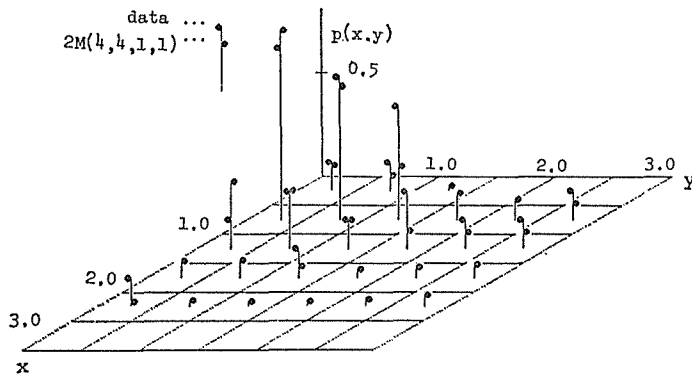


Fig. 14 Probabilities of Koichi-Kuissege data and $2M(4, 4, 1, 1)$ in each mesh.

6.2 年最大日降雨とそれに対応した日流量

ここでは、上田の28年間の年最大日降雨とそれに対応する日流量の値を各々の平均で除した基準化した値をそれぞれ確率変動 x, y として最大エントロピー分布を求めた. 表4は $2M(2, 2, 1, 1)$ のパラメタの値である. モーメントの最高次数を3次以上に増加させた最大エントロピー分布は求まらなかった. 図15は $2M(2, 2, 1, 1)$ とデータのメッシュ確率を比較したものである. $2M(2, 2, 1, 1)$ はデータの全体的傾向をとらえているが, 両者のモードの位置が異っており, データは必ずしも求った分布のように y 軸方向に指数分布をしていない等, よい適合度をもつとは言えない. 又, 6.1の場合も同様であるが, データ数が少ないためデータ1個に付与される確率が大きくなり, 分布形の裾の部分の適合度は考察しにくい.

Table 4 Parameters of 2MED to rainfall-discharge data.

	$2M(2, 2, 1, 1)$
α	3.9403
β_1	-9.1643
β_2	5.4547
γ_1	1.6623
γ_2	0.37273
δ_{11}	-1.7157

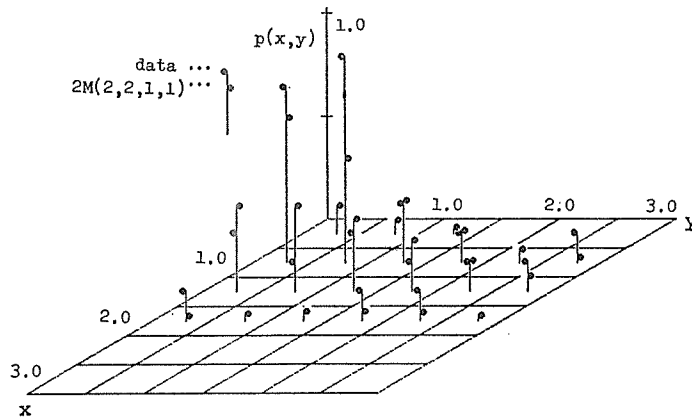


Fig. 15 Probabilities of Ueda-Kuissege data and $2M(2, 2, 1, 1)$ in each mesh.

7. あとがき

本稿では2変数最大エントロピー分布の実務レベルでの利用をめざして適用性について種々の側面から検討した. 得られた結果を要約すると次のようになる.

1) 2変数ガンマ分布を母集団と仮定して, この分布の真値のモーメントを用いて適合度を検討した結果, 仮定した母集団と適合度のよい最大エントロピー分布との関係は次のとおりであった.

$2G(1, 0.2) : 2M(3, 3, 1, 1)$, $2G(2, 0.5) : 2M(4, 4, 2, 2)$, $2G(2, 0.5) : 2M(4, 4, 2, 2)$,
 $2G(2, 1, 0.2) : 2M(4, 4, 1, 1)$, $2G(4, 2, 0.5) : 2M(4, 4, 2, 2)$

2) $2G(1, 0.2)$, $2G(2, 0.5)$, $2G(4, 0.5)$ からシミュレートしたデータを用いてモーメ

ントの安定性を検討した結果、データ数が60個ないし70個から安定するが、高次モーメントになる程その値の信頼性が乏しく、水文統計に採用するモーメントの次数は4次までが適当であった。

3) シミュレーションデータを用いて求めた最大エントロピー分布と真のモーメントを用いて求めた最大エントロピー分布を比較した結果、 $2G(1, 0.2)$ の $2M(3, 3, 1, 1)$ で40個、 $2G(4, 0.5)$ の $2M(2, 2, 1, 1)$ で60個、 $2M(4, 4, 2, 2)$ で70個のデータ数から両者の分布の形状が近似していることがわかった。

4) $2G(4, 0.5)$ から70個のデータを5組発生させ、 $2M(2, 2, 1, 1)$ と $2M(4, 4, 2, 2)$ を求めた結果、同じデータ数でも生起データの相違により、4.2(2)に示した程度の変動が存在することが明らかになった。

5) 最高次数の周辺分布に関するモーメントと最高次数の相乗モーメントを変動させて、 $2G(1, 0.2)$ の $2M(3, 3, 1, 1)$ と $2M(4, 4, 1, 1)$ 、 $2G(4, 0.5)$ の $2M(2, 2, 1, 1)$ と $2M(4, 4, 2, 2)$ の感度分析をした結果、許容できるモーメントの変動域を明示することができた。

6) 2地点の29個の流量データを用いて $2M(2, 2, 1, 1)$ と $2M(4, 4, 1, 1)$ を求めた。この場合、データ数29個に対し $2M(4, 4, 1, 1)$ は比較的よい適合度を示していることがわかった。

7) 1地点の年最大日降雨とそれに対応した1地点日流量の28個のデータから最大エントロピー分布を求めた結果、 $2M(2, 2, 1, 1)$ のみパラメータ同定がなされた。これはデータの傾向を概略的にとらえていたが、適合性は十分ではなかった。

今後は $g_r(x, y)$ の与え方を種々検討することにより、さらに実用的に有効 (たとえばリターンピリオドに妥当で安定した値を与え得る) な分布へと改善していくとともに、適合度判定のための客観的指標の導入を考えたい。

尚、本研究を行うにあたり、当時信州大学大学院生佐藤健次君 (現在(株)東京建設コンサルタント)、及び当時学部学生刑部 博君 (現在(株)玉野総合コンサルタント) の協力を頂いた。記して謝意を表する。

計算にあたっては、信州大学情報処理センター HITAC M-240H を使用した。

参 考 文 献

- 1) Freund, J.E. : A Bivariate Extension of the Exponential Distribution ; American Statistical Association Journal, 56, 971~977 (1961).
- 2) 井沢竜夫 : 二変数 Γ -分布について (降水量分布第2報) ; 気象と統計, 4, (1), 9-15 (1953).
- 3) 井沢竜夫 : 二変数 Γ -分布について (続) ; 気象と統計, 4, (2), 15-19 (1953).
- 4) たとえば, 神田 徹, 藤田睦博 : 水文学—確率論的手法とその応用— ; 65-66, 技報堂出版, 1982.
- 5) 長尾正志, 角屋 睦 : 二変数ガンマ分布とその適用に関する研究(1)—二変数指数分布の基礎理論—, 京都大学防災研究所年報, 13, (B), 105~133 (1970).
- 6) 長尾正志, 角屋 睦 : 二変数ガンマ分布とその適用に関する研究(2)—二変数ガンマ分布の母数推定法—京都大学防災研究所年報, 14, (B), 43~56 (1971).

- 7) 端野道夫, 杉 雄司: 複数の2変数指数型分布の結合とその応用に関する研究; 徳島大学工学部研究報告, 29, 49~57 (1984).
- 8) 寒川典昭, 荒木正夫, 寺島 彰: 2変数 MEP 分布とその特性に関する研究; 水理講演会論文集, 28, 397-402 (1984).
- 9) 寒川典昭, 荒木正夫: 水文事象の頻度分析への MEP 導入について; 土木学会論文報告集, 335, 89~95 (1983).
- 10) 寒川典昭, 荒木正夫, 佐藤健次: 多変数最大エントロピー分布とその基礎特性に関する研究; 土木学会論文集, 375/II-6, 89~98 (1986).
- 11) Noriaki Sogawa, Masao Araki, and Toshihico Imai: Studies on Multivariate Conditional Maximum Entropy Distribution and its Characteristics; J. of Hydrosience and Hydraulic Engineering, 4, (1), 79~97 (1986).
- 12) 石原安雄, 長尾正志: 計画高水流量の二変数確率論的研究; 京都大学防災研究所年報, 6, 181~188 (1953).