

## 歪度を持つ水文統計量の統計解析とその確率模型

Statistical analyses and probabilistic models of skew hydrologic statistics

中村 元\*・虫明 功臣\*

Hajime NAKAMURA and Katumi MUSIAKE

水文統計量の分布形を指数関数族として一般化することが、特に歪度を持った水文量の統計解析と確率模型の整合性に重要な意味を与えることを、最近社会科学などの応用分野においても広く用いられるようになった一般線形モデル (Generalized Linear Models) を導入することによって示した。

## 1. 水文統計量と指数関数族

指数関数族とは次の形式に変形できる確率密度関数の集まりをいう。

$$f_Y(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\} \quad (1)$$

ここで  $\phi$  はある既知数。指数関数族において次の関係が成り立つ。

$$E(Y) = b'(\theta) \quad (2)$$

$$\text{var}(Y) = b''(\theta) a(\phi) \quad (3)$$

ここで  $b''(\theta)$  は分散関数と呼ばれ平均値のみの関数である。一方  $a(\phi)$  は平均値に依存しない関数である。

正規分布、ポワソン分布、二項分布、ガンマ分布は指数関数族の例である。過去の水文統計学の文献を遡って見るとある水文統計量の分布形が指数関数族で近似されている例がきわめて多い。たとえば日降水量、無降雨日数は指数分布、月降水量はガンマ分布、年降水量は正規分布、一雨降水頻度はポワソン分布で近似されている例などがある。水文統計量の分布形を指数関数族としてとらえることが、単に数学的な一般化のみならず、特に歪度を持った水文統計解析と確率模型との整合性において重要な意味を持つことを以下に示す。

## 2. 歪度を持つ水文統計解析とその確率模型

水文統計量はその頻度分布に正の歪度を持つものが多いことが良く知られている。Eagleson(1978)、江藤、室田(1984)は入力歪度を持つガンマ分布で与え確率論等の演繹体系を用い、最終的に工学的に意味のある変量の分布形を導出しているがそれらは仮定が現実と反した時の模型への対応がそのつど成されない事において仮定依存的であるといえる。一方従来からの統計解析モデル(古典的回帰モデル、自己回帰モデル等)は実測が

\*東京大学生産技術研究所 第5部

仮定に反した時のモデルへの対応(データ変換等)が可能であるがゆえ現象描写性に弱い。元来、前者のPriori Distributionのパラメーター依存型確率モデルと後者の統計解析モデルは2つの異なった大系として理解されることが多かったが、最近社会科学等の応用分野で広く用いられるようになった一般線形モデル(GLIM)を導入することによって水文統計量の統計解析と確率模型との間に整合性を持たせ、一連のプロセスとして理解することが可能である。図1に模式的にそれらの関係を示した。

## 3. 古典的線形モデルと一般線形モデル(GLIM)

ここでは概論にとどめる。一般線形モデルの詳細についてはMcCullagh(1983)を参照されたい。一般線形モデルにおいては古典的線形モデルにおけるスケール問題が排除される。すなわち古典的線形回帰モデルで仮定される誤差項が正規分布に従い、その分散が一定で、誤差間が独立であると言う条件は、誤差間の独立性の仮定を除いて削除される。McCullaghは次の古典的線形モデルの問題例を示している。“離散データにおいて誤差がポワソン分布で近似されることがよくあるがそれに古典的線形回帰モデルを用いるには上の3条件を満足するように変換する必要がある。しかしながら誤差の正規性を満足するには  $Y^{2/3}$  という変換が最適であり、分散を一定にするには  $Y^{1/2}$ 、系統的効果への加法性を満足するには  $\log Y$  という変換を必要とする。したがって古典的線形回帰モデルの条件を満足させる単一のスケールは存在しない。”

一般線形モデルは次の3成分から形成される。

## 1) ランダム成分

誤差項の分布形は指数関数族から選ばれる。

## 2) 系統的成分

系統的成分が  $P$  個の説明変数  $x_j$  の線形結合で表される時、

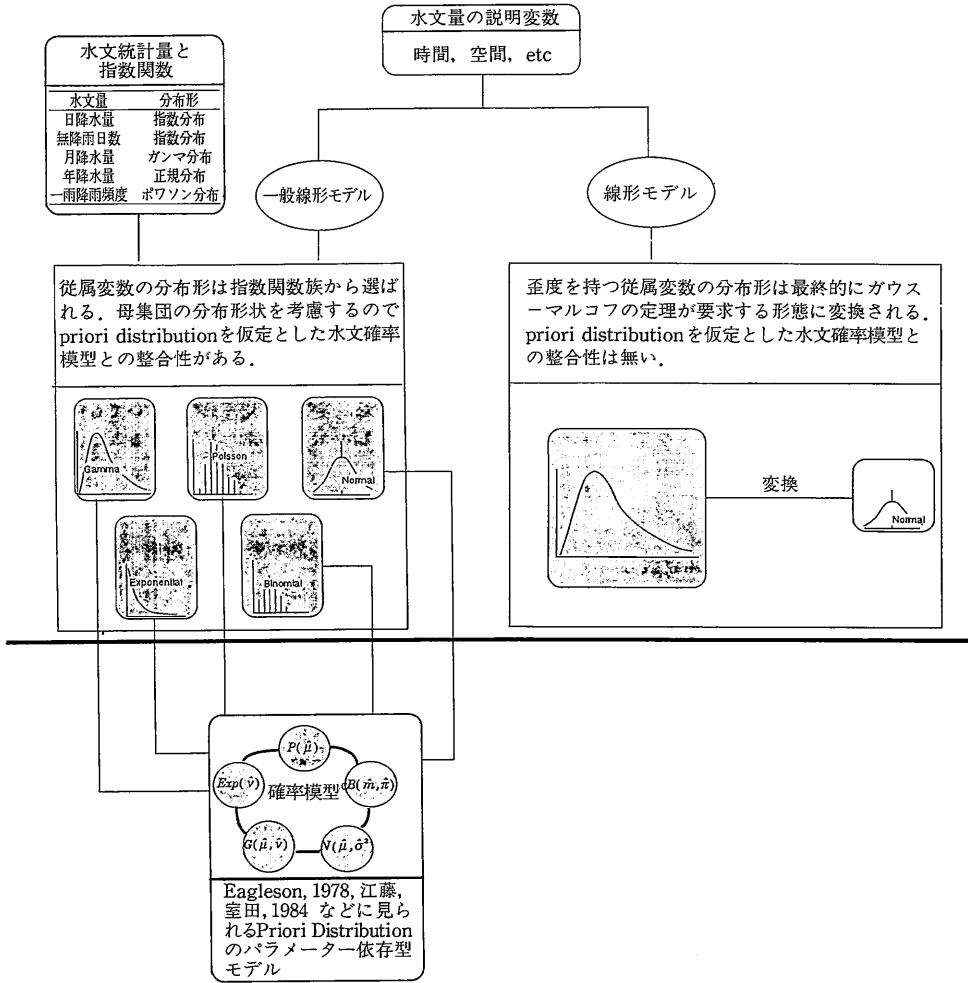


図 1 水文統計解析と確率模型

表 1 良く用いられる一変数指数関数族とその特性値 (McCullagh, 1983)

	Normal	Poisson	Binomial	Gamma
Notation	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)$	$G(\mu, \nu)$
Dispersion parameter : $\phi$	$\sigma^2$	1	1/m	$\nu^{-1}$
Cumulant function : $(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\log(1 + e^\theta)$	$-\log(-\theta)$
$c(\nu; \phi)$	$-\frac{1}{2} \left( -\frac{\nu^2}{\phi} + \log(2\pi\phi) \right)$	$-\log \nu!$	$\log \binom{m}{m\nu}$	$\nu \log(\nu y) - \log y - \log \Gamma(\nu)$
$\mu(\theta) = E(Y; \theta)$	$\theta$	$\exp(\theta)$	$e^\theta / (1 + e^\theta)$	$-1/\theta$
Canonical link : $\theta(\mu)$	identity	log	logit	reciprocal
Variance function : $V(\mu)$	1	$\mu$	$\mu(1 - \mu)$	$\mu^2$

$$\eta = \sum_{j=1}^p x_j \beta_j \quad (4)$$

ここで  $\beta_j$  はパラメーター。

3) リンク関数

従属変数の平均値  $\mu_j$  と  $\eta_j$  を結び付ける関数で、微分可能な単調関数から選ばれる。

$$\eta_j = g(\mu_j) \quad (5)$$

実際の統計解析にあたっては、従属変数の分布形を指数関数族から選び、その平均値と説明変数とを結び付けるリンク関数を指定する必要がある。ランダム成分に正規分布を選び、リンク関数を  $\eta_i = \mu_i$  とすると古典的線形モデルに帰着する。

4. 一般線形モデルを用いた水文統計解析

4.1 非定常一雨降水量解析

日降水量あるいは一雨降水量の分布は一般に正の歪度を持つことが知られているためそれらの頻度解析にガンマ分布がよく使われる。したがって誤差項をガンマ分布としたGLIMを用いるのが、その水文学の分布特性を生かしたモデルに成ると思われる。季節変動を考慮した次の非定常モデルを1981年1月から1988年12月までにタイ北部で観測された降水量時系列に応用した。

$$\log(\mu(t)) = \alpha_{00} + \sum_{k=1}^m \{ \alpha_k \sin(2\pi kt/365) + \beta_k \cos(2\pi kt/365) \} \quad (6)$$

ここでリンク関数をlogとし誤差項を変動係数一定のガンマ分布とした。

すなわち、ランダム成分の密度関数は、

$$f_V(y) = G(y, \lambda) = \frac{\lambda (\lambda y)^{\nu-1} e^{-\lambda y}}{\Gamma(\nu)} \quad (7)$$

ここでνは形状母数、λは尺度母数。デビアンズ解析表を表2に、m=3としたときのパラメーター推定値を表3に示した。

調和振動成分の次数が決まるとガンマ分布の形状母数νを推定することができる。ν=σ²の2つの推定法を記す。

$$1) \hat{\nu}^{-1} = \frac{\bar{D}(6+\bar{D})}{6+2\bar{D}} \quad (8)$$

ここでD̄は

$$\bar{D} = D(y; \hat{\mu})/n$$

Dはデビアンズ関数で

表2 デビアンズ解析表を用いた調和振動成分の次数決定

Model	Deviance	First Difference	d.f.	Mean Deviance
Intercept	152.3		322	
1st harmonic	150.8	1.4	2	0.7
2nd harmonic	150.5	0.4	2	0.3
3rd harmonic	147.7	2.8	2	1.4
4th harmonic	147.5	0.2	2	0.1

表3 三次の調和振動成分まで考慮したGLIMのパラメーター推定値とその標準偏差

	coef	se(coef)	z	p
$\hat{\alpha}_{00}$	2.69861	0.11054	24.41	0.0
$\hat{\alpha}_1$	-0.38238	0.13676	-2.80	0.0052
$\hat{\beta}_1$	-0.11999	0.15887	-0.76	0.4501
$\hat{\alpha}_2$	-0.29341	0.13588	-2.16	0.0308
$\hat{\beta}_2$	0.12401	0.11013	1.13	0.2602
$\hat{\alpha}_3$	-0.09585	0.08531	-1.12	0.2612
$\hat{\beta}_3$	0.18147	0.08030	2.26	0.0238

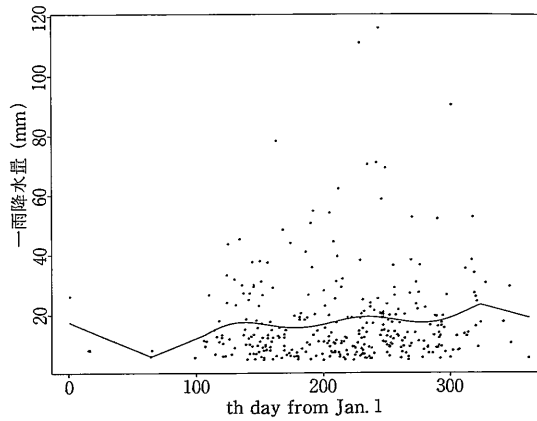


図2 観測された一雨降水量とその推定値 (m=3)

$$D(y; \hat{\mu}) = -2 \sum \omega_i \{ \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)/\hat{\mu}_i \} \quad (9)$$

ここでy<sub>i</sub>は観測値。

$$2) \hat{\nu}^{-1} = \frac{X^2}{\text{自由度}} \quad (10)$$

ここでX²はピアソン統計量で

$$X^2 = \sum \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (11)$$

V(μ̂)は表1に示したガンマ分布の分散関数である。表4に第2の手法を用いて推定した形状母数を示す。表4からGLIMを用いるとモデルによらず形状母数がほぼ一定に保たれていることがわかる。

表4 各調和振動成分ごとに推定した形状母数

1st harmonic	2nd harmonic	3rd harmonic	4th harmonic
1.402	1.388	1.425	1.422

表5 尺度母数の季節変動(形状母数=1.425)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
λ	12.28	4.39	10.32	12.11	11.42	11.97	13.43	12.95	12.63	15.50	15.81	

一雨降水量の物理特性から形状母数は時間に依存しないと仮定することができるので、形状母数と一雨降水量の推定値μ̂(t)を用いると、良く知られたガンマ分布の関係式、

$$\lambda(t) = \frac{\hat{\nu}}{\hat{\mu}(t)} \quad (12)$$

から尺度母数λの時間変化を知ることができる。表5に尺度母数の時間変化を示した。

4.2 非定常月単位降水頻度解析

降雨の継続時間が無降雨スパンに比べて充分短いと仮

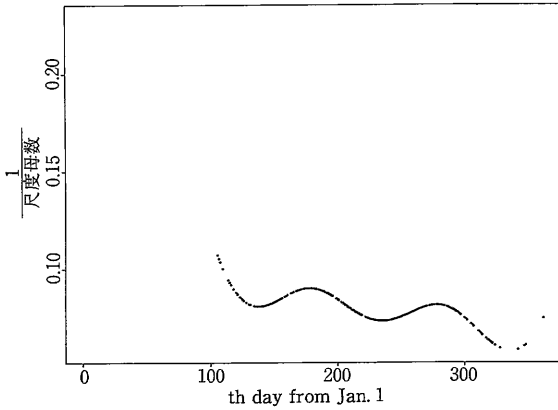


図3 尺度母数の時間変動

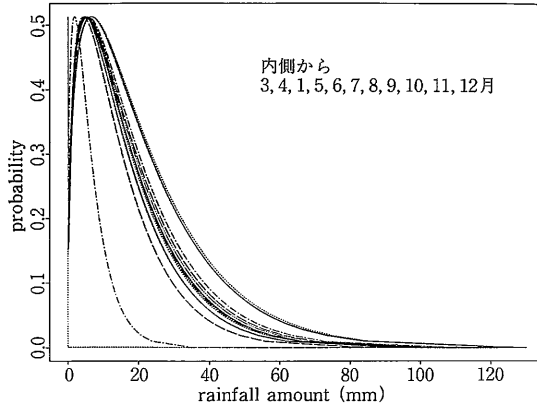


図4 一雨降水量分布の季節変化

定することにより、定常期間の降雨頻度をポワソン分布で近似することができる。ここではポワソン強度の季節変動も考慮した次の一般線形モデルを用いてタイの降雨時系列を解析した。

$$\log(\mu(t)) = \alpha_{00} + \sum_{k=1}^m \{ \alpha_k \sin(2\pi kt/12) + \beta_k \cos(2\pi kt/12) \} \quad (13)$$

ここでリンク関数を  $\log$  とした。ランダム成分の密度関数は、

$$f_Y(y) = \frac{\omega^y e^{-\omega}}{y!} \quad (14)$$

デビアンズ解析表を表6に、 $m=3$ としたときのパラメーター推定値を表7に示した。

表6 デビアンズ解析表を用いた調和振動成分の次数決定

Model	Deviance	First Difference	d.f.	Mean Deviance
Intercept	235.97		83	
1st harmonic	71.23	164.7	2	82.35
2nd harmonic	47.37	23.9	2	11.95
3rd harmonic	43.58	3.8	2	1.9
4th harmonic	42.73	0.85	2	0.425

表7 三次の調和振動成分まで考慮したGLIMのパラメーター推定値とその標準偏差

	coef	se(coef)	z	p
$\hat{\alpha}_{00}$	0.15411	0.1818	0.85	0.3965
$\hat{\alpha}_1$	-1.6329	0.2735	-5.97	0.0
$\hat{\beta}_1$	-1.4427	0.2145	-6.72	0.0
$\hat{\alpha}_2$	-0.7098	0.2250	-3.15	0.0016
$\hat{\beta}_2$	0.3600	0.1895	1.90	0.0574
$\hat{\alpha}_3$	0.1668	0.1368	1.22	0.2228
$\hat{\beta}_3$	0.2123	0.1467	1.45	0.1480

### 4.3 日降雨生起推移確率

$n$ 年からなる日単位降雨時系列が与えられているとすると、とき年度を越えたトレンドが無視できるものと仮定することにより、年内の各日において  $n$ 個独立な反復データが観測されたとみなすことができる。ある時間（ある特定の日を示す）において降雨が観測された日を1、

表8 ポワソン強度の季節変動

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$\hat{\omega}$	0.113	0.050	0.135	1.114	4.705	5.724	5.046	5.513	4.922	2.914	1.418	0.489

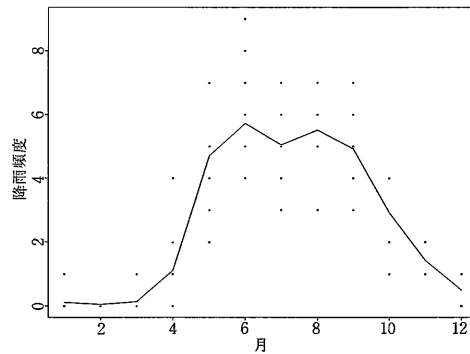


図5 月降雨頻度とその推定値

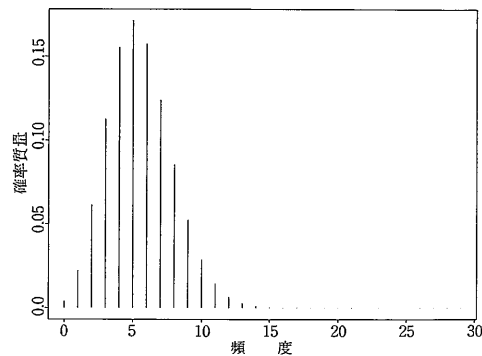


図6 推定された8月の降雨頻度分布

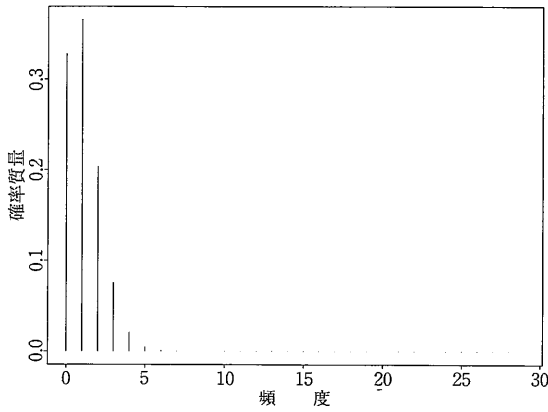


図7 推定された4月の降雨頻度分布

観測されなかった日を0とコーディングすることにより  $n \times 365$ 個のバイナリーデータが得られる。

変量を次のように定義するとき、

$\pi_0(t)$  :  $t, t-1$ 日に共に降雨が観測される確率

$\pi_1(t)$  :  $t-1$ 日に降雨が観測されなかった時に  $t$ 日に降雨が観測される確率

$n_0(t)$  :  $n$ 個の反復データの内の  $t-1$ 日に降雨が観測された回数

$n_1(t)$  :  $n$ 個の反復データの内の  $t-1$ 日に降雨が観測されなかった回数

$y_0(t)$  :  $n_0$ の内  $t$ 日に降雨が観測される回数

$y_1(t)$  :  $n_1$ の内  $t$ 日に降雨が観測される回数

我々が興味のあるのは、推移確率  $\pi_0, \pi_1$ の時間変動である。今  $Y_0$ を確率変数とするとき  $n_0$ が与えられた時  $Y_0 = y_0$ が観測される確率は二項分布にしたがうから、

$$pr(n_0; Y_0 = y_0) = \binom{n_0}{y_0} \pi_0^{y_0} (1 - \pi_0)^{n_0 - y_0} \quad (15)$$

次に  $\pi_0$ の不偏推定値  $\hat{\pi}_0$ は

$$\hat{\pi}_0 = \frac{Y_0}{n_0} \quad (16)$$

与えられ同様に二項分布に従う確率変数である。

ここで中心極限定理としてよく知られているように  $n_0 \rightarrow \infty$ の時  $\hat{\pi}_0$ の分布は平均値  $\pi_0$ 分散  $\frac{\pi_0(1-\pi_0)}{n_0}$ の正規分布に漸近する。したがって大サンプルではそれほど問題とはならないが小サンプルにおいては不偏推定値  $\hat{\pi}_0$ の分布形を二項分布として取り扱う必要があるためランダム成分に二項分布を選んだGLIMを用いるべきである。季節変動を考慮して次のモデルを採用した。リンク関数には *logit*を選んだ。

$$\text{logit}(\pi_0(t)) = \alpha_0 + \sum_{k=1}^m \{ \alpha_{0k} \sin(2\pi kt/365) + \beta_{0k} \cos(2\pi kt/365) \} \quad (17)$$

タイ北部において観測された降雨時系列データを用いて解析した結果得られたデビアンズ解析表を表9に、

表9 デビアンズ解析表を用いた調和振動成分の次数決定

Model	Deviance	First Difference	d. f.	Mean Deviance
Intercept	299.6		212	
1st harmonic	290.9	8.7	2	4.35
2nd harmonic	289.7	1.2	2	0.6
3rd harmonic	289.7	0	2	0
4th harmonic	286.4	3.3	2	1.65

表10 二次の調和振動成分まで考慮したGLIMのパラメーター推定値とその標準偏差

	coef	se(coef)	z	p
$\hat{\alpha}_0$	-0.3642	0.2820	-1.29	0.1965
$\hat{\alpha}_{01}$	-0.6590	0.3184	-2.07	0.0385
$\hat{\beta}_{01}$	-0.6231	0.3763	-1.66	0.0977
$\hat{\alpha}_{02}$	-0.1955	0.2459	-0.80	0.4264
$\hat{\beta}_{02}$	-0.1548	0.1928	-0.80	0.4221

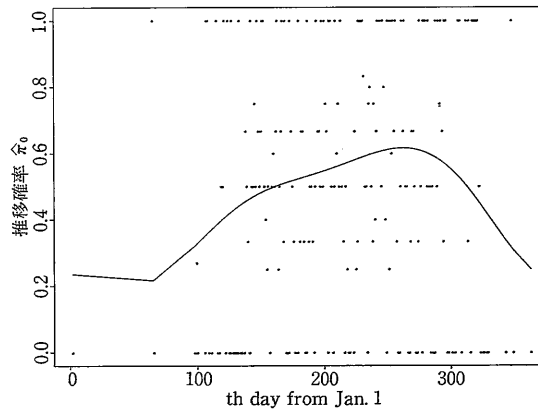


図8 推定確率  $\pi_0$  とその推定値

$m=2$ としたときのパラメーター推定値を表10に示した。

## 5. 一般線形モデルを用いた統計解析から水文確率モデルへの移行

### 5.1 Eagleson (1978) の概略

Eagleson (1978) は年内で水文量の特性がほぼ定常と見なせる期間を取り出し次の確率モデルを用いて年降水量の分布型を導出した。

#### 1) 一雨降雨事象の到着過程

定常期間  $t$ において一雨降水頻度  $K$ はポワソン分布に従うとした。すなわち  $K$ の密度関数は、

$$P_{K|t}(k) = \frac{(\omega t)^k e^{-\omega t}}{k!} \quad (18)$$

#### 2) 一雨降水量

一雨降水量の分布はガンマ分布に従うと仮定した。すなわち  $H$ を一雨降水量とする時

$$P_H(h) = G(\nu, \lambda) = \frac{\lambda (\lambda h)^{\nu-1} e^{-\lambda h}}{\Gamma(\nu)} \quad (19)$$

ガンマ分布の再現性より  $K$  回の降水量の和,

$$P(k) = \sum_{j=1}^k h_j \quad (20)$$

の分布は

$$f_{P(k)}(y) = G(\nu k, \lambda) \quad (21)$$

3) 複合過程

$$\begin{aligned} f_{P(k)}(y) &= \sum_{k=1}^{\infty} f_{P(k)}(y) P_{K|t}(k) \\ &= \sum_{k=1}^{\infty} \frac{\eta \nu (\eta \nu y)^{\nu k - 1} e^{-\eta \nu y}}{\Gamma(\nu k)} \frac{(\omega \tau)^k e^{-\omega \tau}}{k!} \end{aligned} \quad (22)$$

ここで,  $\eta = \frac{\lambda}{\nu}$

4) 年降水量の分布関数

Eagleson は最終的に年降水量の分布関数を,

$$\begin{aligned} Prob\left[\frac{P_A}{m_{P_A}} \leq z\right] \\ = e^{-\omega m_{P_A}} \left\{ 1 + \sum_{y=1}^m \frac{(\omega m_{P_A})^y}{y!} P[\nu k, \omega m_{P_A} \nu z] \right\} \end{aligned} \quad (23)$$

で与えた。

ここで  $P_A$  は年降水量,  $m_{P_A}$  は年平均降水量,  $P[.,.]$  はピアソンの不完全ガンマ関数。

観測データの一部を用いてパラメータを推定した時の観測データ全体とのフィット性を, 湿潤地域において年降水量を正規分布と仮定した場合と乾燥域で対数正規分布と仮定した場合の 2 ケースで比較したところ両者において Eagleson のモデルの優越性が認められた。

5.2 Eagleson (1978) の非定常モデル

4 章で水文データの分布形状を考慮した GLIM は種々の誤差分布のパラメータの逆推定をも可能にしめることを示し, その例としてガンマ分布の尺度母数, ポワソン強度の季節変動を求めた。ここでは Eagleson のモデルに 4 章で逆推定されたパラメータを導入し, タイ北部の年降水量分布を月単位まで非定常化する一連のプロセスを例示する。図 9-10 にシミュレートされた月降水量分布を示す。

6. スケール変化に伴う降水量の分布形状変化に関する一考察

5 章で述べた Eagleson の年降水量モデルと月降水量モデルを用いて次の 3 ケースのスケール変化に伴う降水量の分布形状変化構造を説明することができる。どのケースにおいても正規漸近速度がポワソン強度  $\omega$  と定常期間長  $m_r$  の積に比例することが示される。

(a) 月単位降水量の正規漸近性

5.2 での 2 つのシミュレーションにおいて表 8 より 8 月では  $\omega m_r = 5.513$  12 月では  $\omega m_r = 0.489$  である。一方, ガンマ分布の正規漸近性として知られているように,

$$f_{X_s} = G(k\nu, \lambda)$$

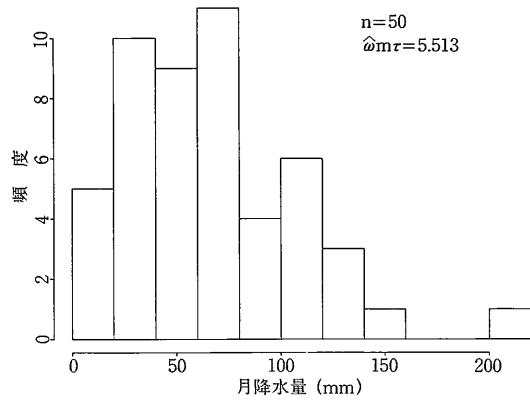


図9 シミュレートされた8月の降水量分布

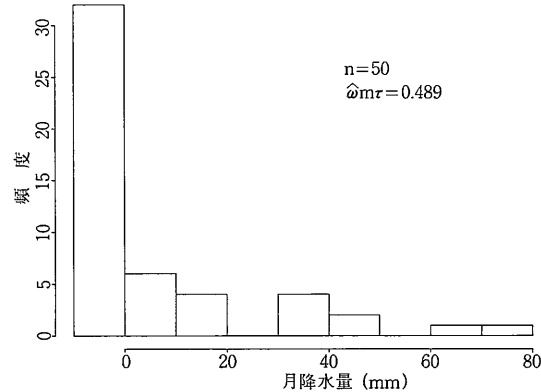


図10 シミュレートされた12月の降水量分布

とすると,

$$\frac{X_k - k\nu\lambda}{\lambda(k\nu)^{1/2}} \xrightarrow{D} N(0, 1), \text{ as } k \rightarrow \infty \quad (24)$$

したがって, 月単位降水量の正規漸近性は定常期間長が一定なので, ポワソン強度に比例すると考えられる。図 9-10 に見られるようにポワソン強度の大きな 8 月における月降水量分布は小さな 12 月の分布より正規分布に近い。一般に定常期間の長さに違いがあるときは, 正規漸近速度は  $\omega m_r$  に比例すると考えられる。

(b) 一定常区間に支配される年降水量の正規漸近性

Eagleson (1978) は湿潤域, 乾燥域の例共に年降水量に寄与する期間として単一の定常期間しか考えていないので (a) と同様に年降水量の正規漸近性は  $\omega m_r$  に支配される。Eagleson (1978) の例で湿潤域での  $\omega m_r$  が 109 と乾燥域の 17.2 と比べて 1 オーダー大きいのはポワソン強度の違いによるよりも雨期の長さの違いによるものと思われる。雨期の長い湿潤域で年降水量が正規分布で近似されるのはこの漸近特性によるものである。Linsley は乾燥域での年降水量分布を対数正規分布で近似しているが, ここでは正規分布に漸近過程にあるガンマ分布と説明される。一般に降水量分布が時間スケールの増大によって正規化されるには次のどれかの条件を満たす必要がある。

- i. ガンマ分布の形状母数が大きいこと。
  - ii. 1 定常区間内でのポワソン強度が大きいこと。
  - iii. 1 定常区間が長いこと。
- (c) 複数個の定常期間の和から構成される年降水量の正規漸近特性

一般に、月によって  $\omega m_r$  が違うので図 9-10に見られるようにある月では正規分布で近似できても他の月に正規分布とまったくかけ離れた分布を持つことが多い。これらの和の分布形を解析的に求めるのは至難の技である。しかしながらある地域で年降水量の分布が正規分布で近似できるとき、その漸近過程は非定常区間の和にあるのではなく (a), (b) に述べた定常期間内での正規漸近過程より支配的であると思われる (すなわち年降水量が正規分布になるにはそれぞれの定常スパン内で (a), (b) の漸近過程により充分正規化される必要がある)。

#### 7. これらの研究課題

本論では年降水量分布を月単位まで非正常化する一連

のプロセスを示すのみにとどまった。

- (a) 月単位降水量モデルにおいても年降水量モデルの結論が当てはまるかどうか。月降水量をガンマ分布に従うと仮定した時と月単位降水量モデルとの精度差を議論する必要がある。
- (b) 月単位降水量モデルと年降水量モデルを用いたスケール変化に伴う水文統計量の分布形状変化の解明。
- (c)  $\omega m_r$  に基づく地域分類, (1992年1月24日受理)

#### 参 考 文 献

- 1) McCullagh and Nelder, 1983: Generalized Linear Models, Chapman and Hall, NY.
- 2) Eagleson, P.: Climate, soil, and vegetation, 2. The distribution of annual precipitation derived from observed storm sequences, WRR, Vol. 14, No. 5, pp. 713-721, 1978.
- 3) 江藤剛治, 室田 明: 一雨降雨の一確率模型, 土木学会論文集, No. 345/2-1, pp. 101-109, 1984.