

神戸大学工学部 正 員 神 田 徹
兵庫県 正 員 福 井 克 佳

1. まえがき

時系列における一定区間内の最大値の分布に関して、ある仮定の下でその漸近近似としてグンベル分布が得られることはよく知られている。本研究は、流量時系列の模擬資料としてのショットノイズ過程に従う標本を対象として、区間最大値分布のグンベル分布への適合度を調べるとともに、漸近近似の方法によらずに極値分布を推定する方法について基礎的考察を行ったものである。

2. 区間最大値分布のグンベル分布への適合性

ショットノイズ過程は (1) 式で定義され、その変量 $X(t)$ は (2) 式のガンマ分布に従う。^{1), 2)}

$$X(t) = \sum_{m=N(-\infty)}^{N(t)} y_m \exp\{-b(t - \tau_m)\} \quad (1)$$

$$f\{X(t)\} = (1/\theta) \{1/\Gamma(\nu/b)\} \{X(t)/\theta\}^{\nu/b-1} \exp\{-X(t)/\theta\} \quad (2)$$

ここに、 $N(t)$ は事象発生率 ν のポアソン過程、 y は平均値 θ の指数分布に従い、 b は低減率である。

グンベル分布の分布関数 $F(x)$ 、平均値 μ 、分散 σ^2 、歪み係数 β は次に示す通りである。

$$F(x) = \exp[-\exp\{-c(x-d)\}] \quad (-\infty \leq x \leq +\infty, c > 0) \quad (3)$$

$$\mu \approx d + 0.5772/c \quad (4) \quad \sigma^2 \approx 1.6449/c^2 \quad (5) \quad \beta \approx 1.1395 \quad (6)$$

(1) 式の $X(t)$ を $\nu = 0.1, 0.5, 1.0, 2.0, 3.0$, $b = 0.1, 0.5, 1.0, 2.0, 3.0$, $\theta = 1.0$ の各組合せについてそれぞれ 500000 個模擬発生させ、区間 n 内の最大値を抽出する。最大値標本の大きさは、 $n = 50, 100, 500, 900$ に対してそれぞれ 10000, 5000, 1000, 550 である。最大値の分布をグンベル分布と仮定して、積率法 ((4), (5) 式) によりパラメータ c, d の値を求める。ここで、このグンベル分布への最大値標本分布の適合度を示す指標として、次のような歪み係数の相対誤差をとる。

$$\text{相対誤差(R.E.)} = \frac{\text{最大値標本の歪み係数} - 1.1395}{1.1395 \text{ (グンベル分布の歪み係数)}} \times 100 \quad (\%) \quad (7)$$

図-1 は、この相対誤差と (2) 式のガンマ分布の形状母数 ν/b 、区間長 n との関係を示したものである。 $\nu/b = 1$ 、すなわち母集団が指数分布に従う場合は、 n にかかわらず ($n \geq 50$)、相対誤差は殆ど 0 である。しかし、 $\nu/b < 1.0$ では相対誤差は正、 $\nu/b > 1.0$ では負となり、最大値分布はグンベル分布とは歪みの形状が異なった分布を示す (図-2)。また、 n が大きくなると、標本数の減少によるばらつきの影響はあるけれども全体的傾向として相対誤差が小さくなっている。上記の ν/b の範囲に対しては、 $n = 500$ 程度以上の区間長であれば、グンベル分布への適合度は良好であるといえる。

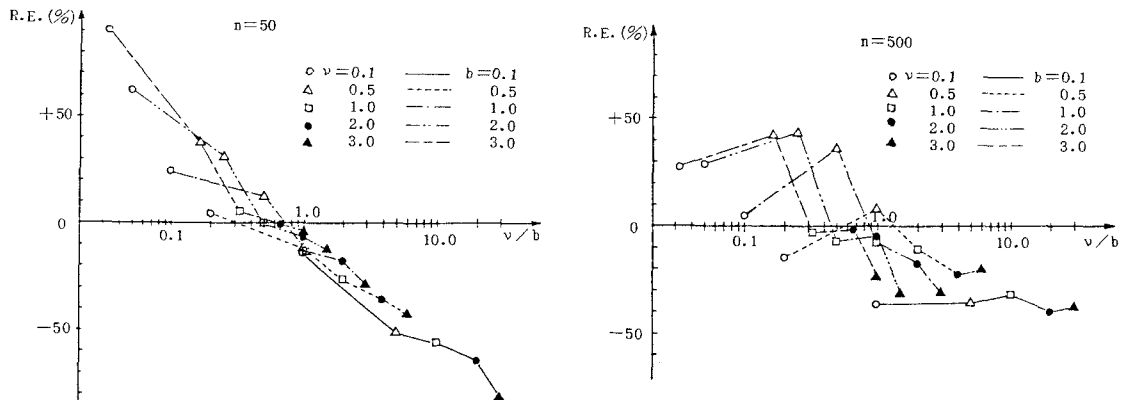


図-1 歪み係数の相対誤差

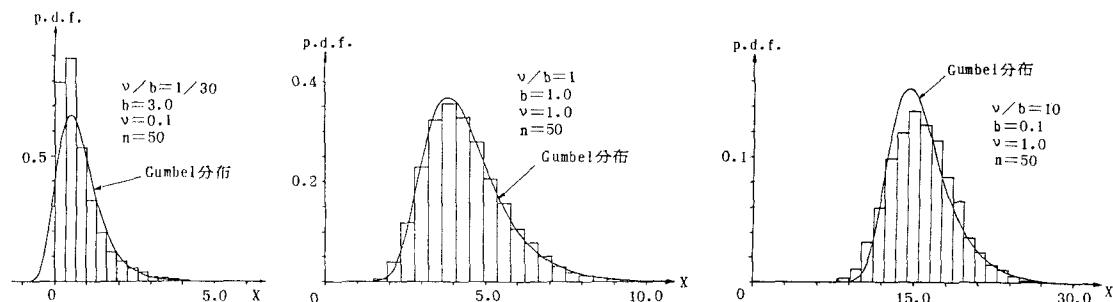


図-2 区間最大値の分布

3. 超過事象によるグンベル分布の推定

$X(t)$ においてある基準量 X_0 を超過する離散的な事象に注目する。この超過事象の発生が (8) 式に示すポアソン過程に従い、その大きさが (9) 式に示す指数分布に従う場合には、超過事象の区間内最大値の分布は (10) 式に示すようにグンベル分布となる。

$$P(k) = \lambda^k \exp(-\lambda) / k! \quad (k=0, 1, 2, \dots) \quad \text{区間内発生回数の平均値 } \mu_k = \lambda, \quad \text{分散 } \sigma_k^2 = \lambda \quad (8)$$

$$F(\eta) = 1 - \exp\{-u(\eta - v)\} \quad (\eta \geq v) \quad \text{平均値 } \mu_\eta = 1/u + v, \quad \text{分散 } \sigma_\eta^2 = 1/u^2 \quad (9)$$

$$F(x) = \{F(\eta)\}^k = \sum_{k=0}^{\infty} [1 - \exp\{-u(\eta - v)\}]^k \lambda^k \exp(-\lambda) / k! = \exp[-\exp\{-c'(x-d')\}] \quad (10)$$

ここに、 $c' = u$, $d' = v + \ln(\lambda)/u$

いま、基準量 X_0 としてガンマ分布の分布関数の値が $p(0 \leq p < 1)$ となるような値 X_p をとってこれを v とし、超過事象が指数分布になるかを確かめ、その平均値 μ_η と v から u を求める。また、一定区間の事象発生回数がポアソン分布に従うかを確かめ、その平均回数 μ_k から λ を推定する。このとき、区間 n 内の最大値の分布 (グンベル分布) が (10) 式として定められる。この方法をショットノイズ過程の時系列 ($v=0.1, b=0.1$) に適用した結果、 $X_0 = X_{0.95}$, $X_{0.99}$ とすれば指数分布に関しては良好な適合性が得られ (図-3), またポアソン分布 ($n=500$) に関しては $b=0.1$ の場合若干適合性が落ちるが他は良好である (図-4)。これより、区間 $n=500$ の最大値に対するグンベル分布は図-5の実線のように得られる。区間内最大値の分布 (ヒストグラム) とこのグンベル分布との対応は良好である。同図にはヒストグラムの積率から求めたグンベル分布も破線で併示している。 b , v の値の他の組合せについても、これら2つのグンベル分布の一致度は良い。この結果、超過事象を用いる上述の方法によって最大値分布 (グンベル分布) を良い精度で推定し得ることが確かめられた。ただし、前者のグンベル分布は後者のグンベル分布よりもピークが若干低く、分散が大きいという傾向がある。極値分布の推定においては、一般に小標本を扱うので、上述の方法が有効となりうると考えられるが、この点については今後検討を加えたい。

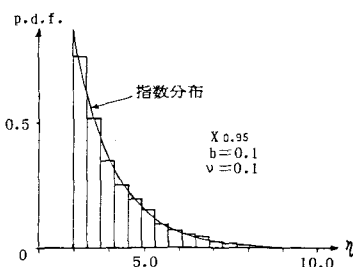


図-3 指数分布

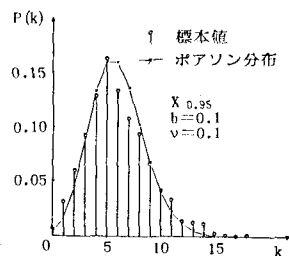


図-4 ポアソン分布

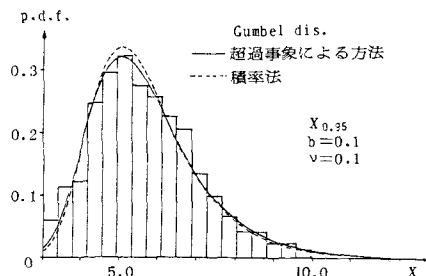


図-5 推定されたGumbel分布

参考文献

- 1) Weiss, G.: Shot noise models for the generation of synthetic streamflow data, W.R.R., Vol. 13, No. 1, 1977.
- 2) 神田徹: ショット・ノイズ過程の基礎特性と流量時系列への適用性, 第27回水理講演会論文集, 1983.