GEVモデルとGPモデルを用いた日最大降水量の推定とその比較

総合数理学部　現象数理学科　池田研究室　４年

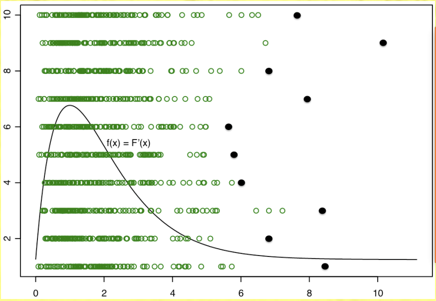
1. 序論・背景

災害や事故という現象は我々の生活の至る所に潜んでおり, 関連する観測データに異常に大きな値が現れることが多い. 従って一定期間の内にどのくらいの頻度でどれくらいの大きな値が起こるかを予測することができれば, それらの災害に備えることができるであろう.

2. 目標

神戸(1983-2013)の日合計降水量を扱い, それが従うモデルである極値統計学のGEVモデルとGPモデルを用いて解析する, そして, 再現レベル200年で最大降水量の95%信頼区間を推定し, モデルの精度を比較する事を目的とする.

3. GEVモデル



GEVモデルに当てはめる**ブロック最大データ**(黒点)

- 独立で同一分布に従う確率変数列 $X_1, X_2, \ldots, X_n$
- 分布関数 $F(x) = P(X_i \le x), i = 1, 2, \ldots, n$
- X_n を大きさの順に並べた順序統計量 $X(1:n) \le X(2:n) \le \cdots \le X(n:n)$
- 極致統計量(ブロック最大データ) $Z_n = X_{(n:n)} = \max\{X_1, X_2, \ldots, X_n\} = \max\limits_{1 \le i \le n} X_i$

$(a)_+ = \max\{a, 0\}$

- GEV(μ, σ, ξ)からの確率標本を $Z_1, Z_2, \ldots, Z_n$
- $z_1, z_2, \ldots, z_n$ をその実現値とする.
- μとσはzの基準化定数.
- 分布関数

$$G(z) = G_\xi\left(\frac{z-\mu}{\sigma}\right) = \exp\left\{-\left[1+\xi\left(\frac{z-u}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\}$$

- 密度関数

$$g(z) = \frac{1}{\sigma}\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}-1}\exp\left\{-\left[1+\xi\left(\frac{z-u}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\}$$

GEVモデルにおける最尤推定量の性質

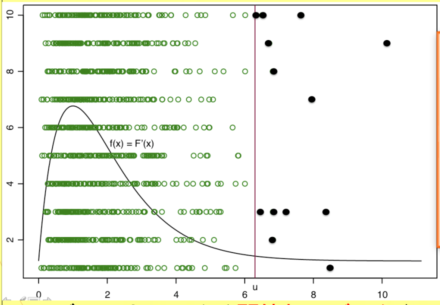
ξ≥-0.5	漸近有効で漸近正規性を持つ
-1<ξ<-0.5	一般に存在するが漸近有効かは未解決
ξ≤-1	一般に存在しない

(自然現象では-0.5<ξ<0.5が多い)

[漸近有効性]
標本数nが十分大きいとき,推定量の偏りがなく,分散が最小

[漸近正規性]
標本数nが十分大きいとき,推定量がある正規分布に従う

4. GPモデル



GPモデルに当てはめる**閾値超過データ**(黒点)

- 分布Fを持つ確率変数をXとする
- $u \to \omega_F$ (分布Fの上限)とする
- 閾値超過データ $X - u$
- $X - u \mid X > u$ の分布について考える

- GP(σ, ξ)からの確率標本を $Y_1, Y_2, \ldots, Y_n$
- $y_1, y_2, \ldots, y_n$ をその実現値とする.
- 分布関数

$$H(y) = H_\xi\left(\frac{y}{\sigma}\right) = \begin{cases} 1-\left(1+\xi\dfrac{y}{\sigma}\right)_+^{-\frac{1}{\xi}} & (\xi \ne 0) \\ 1-\exp\left(-\dfrac{y}{\sigma}\right) & (\xi = 0) \end{cases}$$

- 密度関数

$$h(y) = \frac{1}{\sigma}h_\xi\left(\frac{y}{\sigma}\right) = \begin{cases} \dfrac{1}{\sigma}\left(1+\xi\dfrac{y}{\sigma}\right)_+^{-\frac{1}{\xi}-1} & (\xi \ne 0) \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{y}{\sigma}\right) & (\xi = 0) \end{cases}$$

H_ξ は標準一般パレート分布 $GP(1,\xi)$ の分布関数

$$H_\xi(y) = 1-(1+\xi y)_+^{-\frac{1}{\xi}}$$

GEVモデルにおける最尤推定量の性質

$\xi > -0.5$ で漸近有効性を持つ

5. GEVモデル・GPモデルの性質

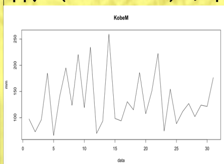
パラメータの値	近似できる分布	分布の性質
$\xi < 0$	Weibull分布	取りうる値の上限が有界
$\xi = 0$	Gumbel分布	裾が指数減衰
$\xi > 0$	Fréchet分布	裾が厚い

母集団分布Fの吸引領域	近似できる分布	分布の性質
Weibull分布 $(G_\xi, \xi < 0)$	ベータ分布	取りうる値の上限が有界
Gumbel分布 (G_0)	指数分布	裾が指数減衰
Frechet分布 $(G_\xi, \xi > 0)$	パレート分布	裾が厚い

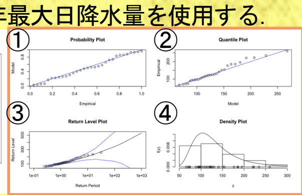
6. 実測値データによるフィッティング

・GEVモデル

神戸(1983-2013)の年最大日降水量を使用する.



実測値データ



分布への当てはめ診断図

パラメーターの最尤推定値(標準誤差)

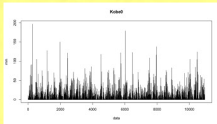
$\hat{\mu} = 108.1(7.636)$
$\hat{\sigma} = 35.82(6.02)$
$\hat{\xi} = 0.1429(0.1865)$

・①と②は分布への当てはまり具合を表している.
・③はデータが分布の95%信頼区間に収まっている事を表す.
・④はデータのヒストグラムと分布の類似性を表している.
診断の結果は良好であるとみなせる.

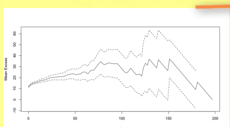
$\xi > 0$ より神戸の日最大降水量はFréchet分布に従い, 再現期間200年の再現レベル z_{200} の95%信頼区間は[193.2261, 1248.500]

・GPモデル

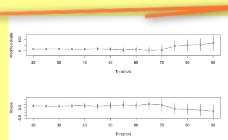
閾値を定め, それを超過したデータを使用する.



実測値データ



標本平均超過関数

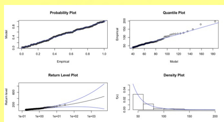


各uでのパラメータの推定値

このニつの図から30<u<70とわかる. さらにデータ数を考慮した上で本研究ではu=40として解析を行う. この方法は恣意的であるが下の分布当てはめ診断図で問題がなければ, 選択した閾値uは妥当であったものとみなす.

パラメーターの最尤推定値(標準誤差)

$\hat{\sigma} = 19.15(2.161)$
$\hat{\xi} = 0.1317(0.1865)$



分布への当てはめ診断図

$\xi > 0$ より神戸の日最大降水量はパレート分布に従い, 再現期間200年の再現レベル z_{200} の95%信頼区間は[138.10, 506.425]であった.

再現レベル200年の推定値とはその値を超える現象が発生するまでの平均期間が200年であるということを意味する.

結果

・GEVモデルとGPモデルのどちらも $\xi > 0$ となり, データは**裾の厚い分布**に従うという**同様の結果**が得られた.
・形状パラメータと再現レベル200年の推定値に注目すると, 本研究ではGPモデルの方がGEVモデルに比べて**標準誤差が小さく**推定の**精度が良い**ことがわかった.
・本研究では, 実測値データは神戸の1983年から2013年までのものを使用し, GEVモデルのブロック最大データのブロックの大きさを1年としたため**データ数は30個**となった. それに対してGPモデルでは, 閾値u=40を超える全てのデータを用いため, **データ数は184**となった. 故にニつのモデルの精度の差は**標本数による**ものであった考えられる.

7. 結論

・GEVモデルとGPモデルを用いて最大降水量を再現レベル200年で推定することができた.
・本研究ではGEVモデルとGPモデルであれば, 扱うデータ数がより多いGPモデルを用いた方が精度が良いという結果を得ることができた.