Point estimation in multivariate power series offspring distiributions

D. Atanasov[1], V. Stoimenova[2]

1. New Bulgarian University
2. Sofia University, Faculty of Mathematics and Informatics.

Content

1. Description of the process and definitions.
2. Family tree reconstruction
3. Estimation of the parameters.
4. Robust approach.
5. EM approach.
6. Conclusions.

The process

Type of the particles $T = \{1, 2\}$.

The process

Type of the particles $T = \{1, 2\}$.

$Z_i(t), i \in T$ is the number of type i particles in generation t.

The process

Type of the particles $T = \{1, 2\}$.

$Z_i(t), i \in T$ is the number of type i particles in generation t.

Let $\xi_k^i(t, l)$ is the number of offspring of type i in generation $t + 1$, produced from l-th particle of type k in t generation.

The process

Type of the particles $T = \{1, 2\}$.

$Z_i(t), i \in T$ is the number of type i particles in generation t.

Let $\xi_k^i(t, l)$ is the number of offspring of type i in generation $t+1$, produced from l-th particle of type k in t generation.

$$\xi_k(t, l) = \left(\xi_k^1(t, l), \xi_k^2(t, l)\right)$$

The process

Type of the particles $T = \{1, 2\}$.

$Z_i(t), i \in T$ is the number of type i particles in generation t.

Let $\xi_k^i(t, l)$ is the number of offspring of type i in generation $t + 1$, produced from l-th particle of type k in t generation.

$$\xi_k(t, l) = \left(\xi_k^1(t, l), \xi_k^2(t, l)\right)$$

$$p_{ij}^k = P(\xi_k(t, l) = (i, j)) = P\left(\xi_k^1(t, l) = i, \xi_k^2(t, l) = j\right)$$

$$Z(t) = (Z_1(t), Z_2(t))$$

$$Z(t) = (Z_1(t), Z_2(t))$$

$$Z_i^k(t+1) = \sum_{j=1}^{2} \sum_{l=1}^{Z_j(t)} \xi_k^i(t,l)$$

$Z_i^k(t)$ is the number of particles of type i in generation $t+1$ with father of type k.

$$Z(t) = (Z_1(t), Z_2(t))$$

$$Z_i^k(t+1) = \sum_{j=1}^{2} \sum_{l=1}^{Z_j(t)} \xi_k^i(t,l)$$

$Z_i^k(t)$ is the number of particles of type i in generation $t+1$ with father of type k.

$$Z_k\left(t,(i,j)\right) = \sum_{l=1}^{Z_1(t)} \sum_{h=1}^{Z_2(t)} I\left(\xi_k^1(t,l) = i, \xi_k^2(t,h) = j\right)$$

is the number of particles of type k in generation t with i offspring of type 1 and j offspring of type 2.

$$\Im_k = \{\xi_k(t, l)\}$$

$$\Im(n) = \{Z(0), \cdots, Z(n)\}$$

$$\Im_k = \{\xi_k(t, l)\}$$

$$\Im(n) = \{Z(0), \cdots, Z(n)\}$$

$$\tilde{\Im}(n) = \{Z_k(t, (i, j)); k \in T; (i, j) \in \Im_k\}$$

$$\Im_k = \{\xi_k(t, l)\}$$

$$\Im(n) = \{Z(0), \cdots, Z(n)\}$$

$$\tilde{\Im}(n) = \{Z_k(t, (i, j)); k \in T; (i, j) \in \Im_k\}$$

$$\tilde{\tilde{\Im}}(n) = \left\{\xi_j^k(t, l); k, j \in T; t = 1, \cdots, n-1; l = 1, \cdots, Z_j(t)\right\}$$

is the information, based on the observation of the entire family tree.

MPSD

$$p_{ij}^{k} = \frac{a_k(i,j)\theta_{1k}^{i}\theta_{2k}^{j}}{A_k(\theta_{1k}^{i}, \theta_{2k}^{i})}; (i,j) \in \Im_k$$

$$A_k(\theta_{1k}^{i}, \theta_{2k}^{i}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_k(i,j)\theta_{1k}^{i}\theta_{2k}^{j}$$

Doble Poisson

$$A_k(\theta_{1k}^i, \theta_{2k}^i) = \exp(\theta_{1k} + \theta_{2k})$$

$$p_{ij}^k = \frac{\theta_{1k}^i e^{-\theta_{1k}} \theta_{2k}^j e^{-\theta_{2k}}}{i!j!}$$

Family tree reconstruction

Gonzáles et al. [2]

$$P(\tilde{\tilde{\Im}}(n) \mid \Im(n), \theta_{11}, \theta_{21}) = \prod_{t=0}^{n-1} P(\tilde{\tilde{\Im}}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21})$$

$$P(\tilde{\Im}(n) \mid \Im(n), \theta_{11}, \theta_{21}) = \prod_{t=0}^{n-1} P(\tilde{\Im}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21})$$

$$P(\tilde{\Im}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21}) = \frac{\prod\limits_{k=1,2} \frac{Z_k(t)!}{\prod\limits_{(i,j)} Z_k(t,(i,j))!} \prod\limits_{(i,j)} \left[p_{ij}^k\right]^{Z_k(t,(i,j))}}{P(Z(t+1) \mid Z(t))}$$

$$P(\tilde{\Im}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21}) = \frac{\prod\limits_{k=1,2} \frac{Z_k(t)!}{\prod\limits_{(i,j)} Z_k(t,(i,j))!} \prod\limits_{(i,j)} \left[p_{ij}^k\right]^{Z_k(t,(i,j))}}{P(Z(t+1) \mid Z(t))}$$

$$P(\tilde{\Im}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21}) = \frac{\prod\limits_{k=1,2} \prod\limits_{l=1}^{Z_k(t)} p^k_{\xi_1^k(t,l), \xi_2^k(t,l)}}{P(Z(t+1) \mid Z(t))}$$

$$P(\tilde{\Im}(n) \mid Z(t), Z(t-1), \theta_{11}, \theta_{21}) =$$

$$MN\left(\sum_{s=1}^{Z_1(s)} \xi_1^1(t,s), \sum_{s=1}^{Z_2(s)} \xi_2^1(t,s), \frac{Z_1(t)\theta_{11}}{Z_1(t)\theta_{11} + Z_2(t)\theta_{21}}, \frac{Z_2(t)\theta_{21}}{Z_1(t)\theta_{11} + Z_2(t)\theta_{21}}\right) \times$$

$$MN\left(\sum_{s=1}^{Z_1(s)} \xi_1^2(t,s), \sum_{s=1}^{Z_2(s)} \xi_2^2(t,s), \frac{Z_1(t)\theta_{12}}{Z_1(t)\theta_{12} + Z_2(t)\theta_{22}}, \frac{Z_2(t)\theta_{22}}{Z_1(t)\theta_{12} + Z_2(t)\theta_{22}}\right) \times$$

$$MN\left(\sum_{s=1}^{Z_1(s)} \xi_1^1(t,s), \frac{1}{Z_1(t)}\right) \times MN\left(\sum_{s=1}^{Z_2(s)} \xi_2^1(t,s), \frac{1}{Z_2(t)}\right) \times$$

$$MN\left(\sum_{s=1}^{Z_1(s)} \xi_1^2(t,s), \frac{1}{Z_1(t)}\right) \times MN\left(\sum_{s=1}^{Z_2(s)} \xi_2^2(t,s), \frac{1}{Z_2(t)}\right)$$

Reconstruction algorithm

1. Type 1 particles

$$Z_1(t+1) = \sum_{s=1}^{Z_1(s)} \xi_1^1(t,s) + \sum_{s=1}^{Z_2(s)} \xi_2^1(t,s),$$

   according MN (binomial) (1-st row)

2. Type 2 particles

$$Z_2(t+1) = \sum_{s=1}^{Z_1(s)} \xi_1^2(t,s) + \sum_{s=1}^{Z_2(s)} \xi_2^2(t,s),$$

   according MN (binomial) (2-nd row)

Reconstruction algorithm

3. Using $\sum_{s=1}^{Z_1(s)} \xi_1^1(t,s)$, generate $\xi_1^1(t,s)$ using $MN\left(\sum_{s=1}^{Z_1(s)} \xi_1^1(t,s), \frac{1}{Z_1(t)}\right)$
4. By analogy for $\xi_2^1(t,s)$
5. By analogy for $\xi_1^2(t,s)$
6. By analogy for $\xi_2^2(t,s)$

Estimation of the parameters. ML

$$L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i, k = 1, 2) = \prod_{k=1}^{2} \prod_{t=1}^{n-1} \prod_{(i,j) \in \Im_k} p_{ij}^{k} =$$

$$= \prod_{k=1}^{2} \prod_{t=1}^{n-1} \prod_{(i,j) \in [\Im(k)]} \left[\frac{a_k(i,j)}{A_k(\theta_{1k}, \theta_{2k})} \theta_{1k}^{i} \theta_{2k}^{j} \right]^{Z_k(t,(i,j))}$$

Estimation of the parameters. ML

$$L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i,k=1,2) = \prod_{k=1}^{2}\prod_{t=1}^{n-1}\prod_{(i,j)\in\Im_k} p_{ij}^k =$$

$$= \prod_{k=1}^{2}\prod_{t=1}^{n-1}\prod_{(i,j)\in[\Im(k)]} \left[\frac{a_k(i,j)}{A_k(\theta_{1k},\theta_{2k})}\theta_{1k}^i\theta_{2k}^j\right]^{Z_k(t,(i,j))}$$

$$E\xi_k^1(t,l) = \frac{\theta_{1k}}{A_k(\theta_{1k},\theta_{2k})}\frac{\partial A_k(\theta_{1k},\theta_{2k})}{\partial\theta_{1k}}.$$

Estimation of the parameters. ML

$$L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i, k = 1, 2) = \prod_{k=1}^{2} \prod_{t=1}^{n-1} \prod_{(i,j)\in\Im_k} p_{ij}^{k} =$$

$$= \prod_{k=1}^{2} \prod_{t=1}^{n-1} \prod_{(i,j)\in[\Im(k)]} \left[\frac{a_k(i,j)}{A_k(\theta_{1k}, \theta_{2k})} \theta_{1k}^{i} \theta_{2k}^{j} \right]^{Z_k(t,(i,j))}$$

$$E\xi_k^1(t,l) = \frac{\theta_{1k}}{A_k(\theta_{1k}, \theta_{2k})} \frac{\partial A_k(\theta_{1k}, \theta_{2k})}{\partial \theta_{1k}}.$$

$$E\xi_k^i(t,l) = \hat{m}_{ik} = \frac{\sum_{s=1}^{n} Z_i^k(s)}{\sum_{s=1}^{n-1} Z_k(s)}$$

Robust estimation. WLTE

Vandev & Nejkov [7].

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin} \sum_{h=1}^{k} w_h f_{\nu(h)}(\theta),$$

Robust estimation. WLTE

Vandev & Nejkov [7].

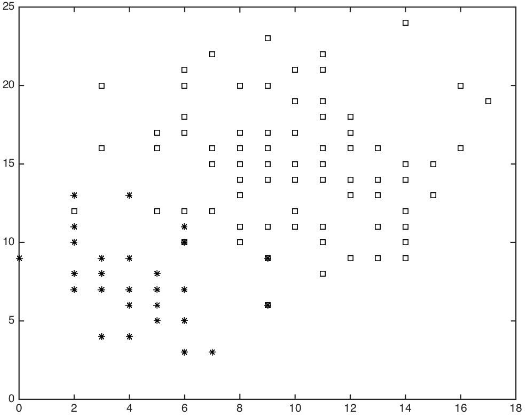
$$\hat{\theta} = \underset{\theta \in \Theta}{argmin} \sum_{h=1}^{k} w_h f_{\nu(h)}(\theta),$$

$$f_{\nu(1)}(\theta) \leq \ f_{\nu(2)}(\theta) \leq \cdots \leq f_{\nu(n)}(\theta)$$

$$f_{\nu(h)}(\theta) = -\log \phi(x_h, \theta)$$

for p.d.f $\phi(x, \theta)$ and a permutation $\nu = (\nu(1), \cdots, \nu(n))$.

An example



	θ_1	θ_2
True	10	15
Total	8.5769	13.5846
Robust	9.8800	15.3600
Outliers	4.2333	7.6667

Breakdown point

$BPoint(E) = \frac{1}{n}\max\{m : sup||E(X_m)|| < \infty\}$, where X_m is a sample, obtained from the sample X by replacing any m of the observations by arbitrary values, Hampel et al. (1986).

Breakdown point

$BPoint(E) = \frac{1}{n}\max\{m : sup||E(X_m)|| < \infty\}$, where X_m is a sample, obtained from the sample X by replacing any m of the observations by arbitrary values, Hampel et al. (1986).

Vandev & Nejkov [7].
BP of WLTE in not less than $(n-k)/n$ if
$(n+d)/2 \leq k \leq n-d$; $n \geq 3d$; $F = \{f_h(\theta)\}$ is d-full.

Breakdown point

$BPoint(E) = \frac{1}{n}\max\{m : sup||E(X_m)|| < \infty\}$, where X_m is a sample, obtained from the sample X by replacing any m of the observations by arbitrary values, Hampel et al. (1986).

Vandev & Nejkov [7].
BP of WLTE in not less than $(n-k)/n$ if
$(n+d)/2 \leq k \leq n-d$; $n \geq 3d$; $F = \{f_h(\theta)\}$ is d-full.

F is d-full if $\sup\{F^k \in F : |F^k| = k\}$ is subcompact.

Breakdown point

$BPoint(E) = \frac{1}{n} \max\{m : sup||E(X_m)|| < \infty\}$, where X_m is a sample, obtained from the sample X by replacing any m of the observations by arbitrary values, Hampel et al. (1986).

Vandev & Nejkov [7].
BP of WLTE in not less than $(n-k)/n$ if
$(n+d)/2 \leq k \leq n-d$; $n \geq 3d$; $F = \{f_h(\theta)\}$ is d-full.

F is d-full if $\sup\{F^k \in F : |F^k| = k\}$ is subcompact.

$g(\theta)$ is subcompact if $L_g(C) = \{\theta : g(\theta) \leq C\}$ are compacts for any C.

Robustness, BPSD

$$N_h = (X_h, Y_h);\ h = 1, \cdots, n;\ w_h = 1$$

$$(\theta_1 \hat{} \theta_2) = \underset{(\theta_1,\theta_2)\in\Theta}{argmin} \sum_{h=1}^{k} - \log \left\{ \frac{A(\theta_1, \theta_2)}{a(X_{\nu(h)}, Y_{\nu(h)})\theta_1^{X_{\nu(h)}} \theta_2^{Y_{\nu(h)}}} \right\}$$

$$-log \frac{A(\theta_1, \theta_2)}{a(X_{\nu(1)}, Y_{\nu(1)})\theta_1^{X_{\nu(1)}} \theta_2^{Y_{\nu(1)}}} \leq \cdots \leq -log \frac{A(\theta_1, \theta_2)}{a(X_{\nu(n)}, Y_{\nu(n)})\theta_1^{X_{\nu(n)}} \theta_2^{Y_{\nu(n)}}}$$

- $\Im$ is the set of all posible values of the random vector (X, Y);
- $\Im^1$ the set of all possible values of the random variable X, while $\Im^2$ is the set of all possible values of the random variable Y;
- Υ is the convergence region of the series $A(\theta_1, \theta_1)$, and $\partial\Upsilon$ is its boundary;
- $MinVX \in \Im^1$ is the minimal value of the random variable X, and $MinVY \in \Im^2$ - the minimal value of the random variable Y;
- $MaxVX \in \Im^1$ is the maximal value of the random variable X (which may be finite or infinity), and $MaxVY \in \Im^2$ - the maximal value of the random variable Y;
- If $(MinVX, MinVY) \in \Im$, then $N_{MinVX,MinVY}$ is the observed number vectors in the sample with minimal values of the coordinates;
- If $MaxVX < \infty$, $MaxVY < \infty$ and $(MaxVX, MaxVY) \in \Im$, then $N_{MaxVX,MaxVY}$ is the observed number vectors in the sample with maximal values of the coordinates;

Lemma

[1]
*Let us consider a sample of n independent bivariate power series distributed observations. For the breakdown point of the $LTE(k)$ estimator (**??**) the following statements hold:*

1. *If* $|\Im| = \infty$, $(\theta_1, \theta_2) \in \Upsilon$, $\lim_{\theta_1,\theta_2 \to \partial\Upsilon} \frac{A(\theta_1,\theta_2)}{\theta_1^i \theta_2^j} = \infty \; \forall (i,j) \in \Im$ *except for* $(i,j) = (MinVX, MinVY)$, *then the* $WLT(K)$ *estimator exists and its breakdownpoint is not less than* $[n-k]/n$ *if* $n \geq 3(N_{MinVX,MinVY} + 1)$, $[n + N_{MinVX,MinVY} + 1]/2 \leq k \leq n - N_{MinVX,MinVY} - 1$.
2. *If* $|\Im| < \infty$; $\theta_1 \in (0,\infty)$, $\theta_2 \in (0,\infty)$, *then the* $WLT(K)$ *estimator exists and its breakdown point is not less than* $[n-k]/n$, *if* $n \geq 3(max\{N_{MinVX,MinVY}, N_{MaxVX,MaxVY}\} + 1)$, *and* $[n + max\{N_{MinVX,MinVY}, N_{MaxVX,MaxVY}\} + 1]/2 \leq K \leq n - max\{N_{MinVX,MinVY}, N_{MaxVX,MaxVY}\} - 1$.

   If $(MaxVX, MaxVY) \notin \Im$, *then the statement in 1. hold.*
3. *If* $(\theta_1, \theta_2) \in \Phi \subset \Upsilon_1$, *where* Φ *is a compact set, then the* $WLT(k)$ *estimator exists and its breakdown point is not less than* $[n-k]/n$ *if* $n \geq 3$, $[n+1]/2 \leq k \leq n-1$.

LTE estimation

1. Estimating the vector θ, for example using the Harris estimators on an available subset of the family tree, free of outliers.
2. Generating a family tree, using the observed generation sizes and the parameter value from 1.
3. Estimating the vector θ, using the $WLT(K)$ estimator over the entire family tree, and detecting the 'errors'.

LTE estimation

1. Estimating the vector θ, for example using the Harris estimators on an available subset of the family tree, free of outliers.
2. Generating a family tree, using the observed generation sizes and the parameter value from 1.
3. Estimating the vector θ, using the $WLT(K)$ estimator over the entire family tree, and detecting the 'errors'.

This algorithm can be bootstrapped large number of times.

EM estimation. M-step

[4]

$$\log L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i, k = 1, 2) = \sum_{k=1}^{2} \sum_{t=1}^{n-1} \sum_{(i,j) \in [\Im_k]} Z_k(t, (i, j)) \times$$

$$\times \left(\log \theta_{1k}^{i} + \log \theta_{2k}^{j} - \theta_{1k} - \theta_{2k} - \log i! j! \right)$$

EM estimation. M-step

[4]

$$\log L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i,k=1,2) = \sum_{k=1}^{2} \sum_{t=1}^{n-1} \sum_{(i,j)\in[\Im_k]} Z_k(t,(i,j)) \times$$

$$\times \left(\log \theta_{1k}^{i} + \log \theta_{2k}^{j} - \theta_{1k} - \theta_{2k} - \log i!j! \right)$$

$$E_Z \log L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i,k=1,2) \approx$$

$$\approx \sum_{k=1}^{2} \sum_{(i,j)\in[\Im_k]} EZ_k(i,j) \times (i \log \theta_{1k} + j \log \theta_{2k} - \theta_{1k} - \theta_{2k})$$

.

EM estimation. E-step

$$EZ_k(i,j) = \sum_{t=1}^{n-1} Z_1(t)Z_2(t)\hat{p}_{ij}^k$$

where

$$\hat{p}_{ij}^k = \frac{\hat{\theta}_{1k}^i e^{-\hat{\theta}_{1k}} \hat{\theta}_{2k}^j e^{-\hat{\theta}_{2k}}}{i!j!}.$$

EM algorithm

1. Set a starting values for $\{\theta_{ik}, i, k = 1, 2\}$
2. Perform E-step - calculate $EZ_k(i, j)$
3. Substitute $EZ_k(i, j)$ in $E_Z \log L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i, k = 1, 2)$
4. Perform M-step - maximize $E_Z \log L(\tilde{\tilde{\Im}}(n) \mid \theta_{ik}; i, k = 1, 2)$ and obtain new $\{\theta_{ik}, i, k = 1, 2\}$
5. Go to 2. if the $E_Z \log L$ is changed.

References

Atanasov D., Staneva A., Stoimenova V., Robust estimators for the bivaiate power series offspring distributios. $3^r d$ *SMTDA Conference Proceedings*. In press.

González, M., Martín, J., Martínez, R., Mota, M. Non-parametric Bayesian estimation for multitype branching processes through simulation-based methods *Computational statistics & data analysis*, 52, 1281–1291, 2008.

Hampel, F. R., Ronchetti, E.M., Rousseeuw, P. J., Stahel W. A. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York, 1986.

McLachlan G., T. Krishnan. *The EM Algorithm and Extensions. Second Edition*. John Weley. 2008.

Staneva A., Stoimenova V. Statistical Estimation In Branching Processes With Bivariate Poisson Offspring Distribution. *Pliska Stud. Math. Bulgar.* **24**. 2015. In press.

Vandev D.L. A Note on Breakdown Point of the Least Median of Squares and Least Trimmed Estimators. *Statistics and Probability Letters* **16**, 117–119, 1993.

Vandev, D. L. and Neykov, N. M. Robust Maximum Likelihood in the Gaussian Case. *In: New Directions in Statistical Data Analysis and Robustness*, S. Morgenthaler, E. Ronchetti, and W. A. Stahel (eds.), (Basel: Birkhauser Verlag), 257–264, 1993.