Unified Empirical Likelihood Confidence Regions for Branching Processes with Immigration

Anand N. Vidyashankar
Department of Statistics George Mason University

Some parts are joint with Pin Ren

April 10, 2015

Contents

Statement of the problem

- Let $\{X_n : n \geq 1\}$ denote a branching process with immigration (BPI) starting with one ancestor; that is,

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,i} + I_{n+1}, \;\; n \geq 0, \tag{1}$$

  where $\{I_n : n \geq 1\}$ is a collection of i.i.d. integer valued random variables and independent of $\{\xi_{n,j}, n \geq 1, j \geq 1\}$.
- Let $m = E(\xi_{1,1}),\;\; \sigma^2 = Var(\xi_{1,1}), \lambda = E(I_1),\;\; \tau^2 = Var(I_1)$.
- Given $\mathcal{X}_n \equiv \{X_0, X_1, \cdots X_n\}$, obtain a confidence region for the parameters $\boldsymbol{\theta} = (m, \lambda, \sigma^2, \tau^2)$.

Consistent Estimability

- Let $\mathcal{F}_n = \sigma \langle X_0, X_1, \cdots X_n \rangle$.
- A generic parameter η of the BPI is said to be consistently estimable if there exists a $\mathcal{F}_n$ measurable function $\hat{\eta}_n$ such that $\hat{\eta}_n$ converges to η either in probability or with probability one.
- Using $\mathcal{X}_n$,
  - If $m > 1$ only (m, σ^2) is consistently estimable.
  - If $m = 1$, only (m, σ^2, λ) is consistently estimable.
  - If $m < 1$ then $\boldsymbol{\theta}$ is consistently estimable.
- Observation: (m, σ^2) is always estimable using $\mathcal{X}_n$.
- In today's presentation, we only focus on inference for (m, σ^2).

Historical Background: Law of Large Numbers/ Consistency

- Set $Y_n = \sum_{i=0}^{(n)} X_i$ and

$$\hat{m}_n = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=0}^{n-1} X_i} = \frac{Y_n - 1}{Y_{n-1}}. \qquad (2)$$

- Several authors have contributed to the asymptotic behavior: Harris, Lauritzen, Nagaev, Jagers, Heyde, Senata, Basawa, Klimko, Nelson, Wei, Winnicki, Venkatraman culminating in:
- Under the finite second moment hypothesis, for all $0 < m < \infty$,

$$\hat{m}_n \to m \ \text{ w.p.1 as } n \to \infty. \qquad (3)$$

- When there is no immigration, $\hat{m}_n$ is the non-parametric maximum likelihood estimator of m.

Historical Background:Central Limit Theorem/Asymptotic Normality

- Assume that $E(X_1^4) < \infty$. Then, as $n \to \infty$

$$\sqrt{Y_{n-1}}\,(\hat{m}_n - m) \xrightarrow{d} G, \tag{4}$$

where

$$G \sim \begin{cases} N(0,\sigma^2) & \text{if} \quad m \neq 1 \\ \frac{Y(1)-\lambda}{\int_0^1 Y(t)dt^{1/2}} & \text{if} \quad m = 1, \end{cases} \tag{5}$$

and $Y(t) \stackrel{d}{=} \lim_{n\to\infty} \frac{X_{[nt]}}{n}$.

Historical Background: Alternative Methods for deriving the Estimators

- Noticing that $E(X_n|\mathcal{F}_{n-1}) = mX_{n-1} + \lambda$, one can express

$$X_k = mX_{k-1} + \lambda + \delta_n, \tag{6}$$

where δ_n is a martingale difference sequence. Conditional least squares method(Klimko and Nelson) consists in minimizing

$$\sum_{k=1}^{n} (X_k - mX_{k-1} - \lambda)^2 \tag{7}$$

wrt m and λ. Noticing that $Var(\delta_n|\mathcal{F}_{n-1}) \sim X_{n-1}$, notice that the variances of the residuals behave erratically. To stabilize this Wei and Winnicki (1990) and Basawa and Vidyashankar (2005) studied minimizing

$$\frac{X_k}{\sqrt{1+X_{k-1}}} = m\frac{X_{k-1}}{\sqrt{1+X_{k-1}}} + \frac{\lambda}{\sqrt{1+X_{k-1}}} + \frac{\delta_k}{\sqrt{1+X_{k-1}}}, \tag{8}$$

Historical Background: Estimation (contd.)

- The above equation upon simplification reduces to

$$\frac{X_k}{\sqrt{1+X_{k-1}}} = m\sqrt{1+X_{k-1}} + \frac{\eta}{\sqrt{1+X_{k-1}}} + \epsilon_k. \qquad (9)$$

- where $\eta = \lambda - m$ is a new parameter and obtain the weighted conditional least squares estimators/Quasi likelihood of the parameters.
- Similar objective function can be derived for the variance (Winnicki).
- Under the finite second moment hypothesis one can show then that

$$\lim_{n\to\infty} \hat{m}_{WLS} = m \text{ with probability one for all } 0 < m < \infty. \qquad (10)$$

- A similar result holds for the variance.
- Question: What happens to the asymptotic limit distribution?

- It turns out once again that the the limit distribution for $m \neq 1$ is Gaussian with mean 0 and appropriate variance while it is once again the functional of the Feller diffusion when $m = 1$.
- Important Statistical questions arise:
  1. Can we remove this distributional discontinuity at $m = 1$?
  2. Are their some optimality properties that would suggest using one estimator over the other?

Alternative Perspective.

- Set $\boldsymbol{\Delta}_1 = (m, \eta)$

$$L_{k,1}(\boldsymbol{\Delta}_1) = \left(\frac{X_k}{\sqrt{1 + X_{k-1}}} - m\sqrt{1 + X_{k-1}} + \frac{\eta}{\sqrt{1 + X_{k-1}}} \right)^2. \tag{11}$$

- The weighted conditional least squares approach consists in solving the equation

$$\sum_{k=1}^{n} \frac{1}{n} \nabla L_k(\boldsymbol{\Delta}_1) = 0. \tag{12}$$

- The above sum on the LHS can be thought of as combining k "estimating equations" by assigning them equal weights.

Alternative Perspective (contd.)

- Alternatively, consider replacing $\frac{1}{n}$ by w_k and solving the equation
$$\sum_{k=1}^{n} w_k \nabla L_k(\boldsymbol{\Delta}_1) = 0, \tag{13}$$
where $w_k > 0$ and $\sum_{k=1}^{n} w_k = 1$.
- Is there an optimal way to choose w_k's?
- Set $\boldsymbol{w} = (w_1, w_2, \cdots w_n)$ and consider the following strategy; set
$$R_n(\boldsymbol{w}, \boldsymbol{\Delta}_1) = \prod_{i=1}^{n} \frac{w_i}{1/n}, \quad \text{where} \tag{14}$$
the weights satisfy (13) and maximize wrt w_k, m, η subject to the constraints.

Empirical Likelihood Estimators: Non-Critical cases

- The number of parameters is of the order n which is same as the equations.
- By using a Lagrangian, the problem can be converted to an optimization problem involving 3 parameters which can be solved numerically.
- The resulting estimators are referred to as empirical likelihood estimators.

Theorem

Assume $E(X_1^2) < \infty$. Then, $\hat{m}_{EL} \to m$ with probability one. Additionally, for all $m \neq 1$ as $n \to \infty$

$$-2 \log R_n(\hat{\boldsymbol{w}}, \hat{\boldsymbol{\Delta}}) \xrightarrow{d} \chi_1^2. \tag{15}$$

- Question: What happens if $m = 1$? The main result is the following:

Empirical Likelihood Estimators: Critical Case

Theorem

If $m = 1$ and $2\lambda > \sigma^2$, then assuming $E(X_1^3) < \infty$

$$-2\log R_n(\hat{\boldsymbol{w}}, \hat{\boldsymbol{\Delta}}_1) \xrightarrow{d} G^2 \tag{16}$$

where G^2 is the square of the functional of the Feller diffusion.

- Question (1) from previous slides remain.

Unification

- To handle this, let $\hat{\sigma}^2$ denote an estimate of the variance and set
$$N_c = inf\{k \geq k_0 : Y_k > c\hat{\sigma}^2\} \tag{17}$$

Theorem

Assume $E(X_1^4) < \infty$ and $2\lambda > \sigma^2$. Then

$$-2\log R_{N_c}(\hat{\boldsymbol{w}}, \hat{\boldsymbol{\Delta}}) \xrightarrow{d} \chi^2_{(1)} \text{ as } c \to \infty. \tag{18}$$

Motivating Example: Cascades Data

- Cascading failure: a failure spreads in a system of dependent components.
- Electric power transmission systems
  - Cascading failure is the main way that blackouts become more widespread.
  - A small initial disturbance can spread to a large blackout by cascading.
- The number of transmission line failure:
  - One measure of blackout size internal to the power system.
  - Useful diagnostics in monitoring the progress and extent of blackouts.

Cascades Data (contd.)

- Aim:
  - To provide precise summarizing statistics of the distribution of the number of transmission line failures as one measure of the distribution of blackout size in order to quantify the risk of cascading failures.
- Why Branching Processes model:
  - Branching processes have been used extensively to model cascading processes and their applications to the risk of cascading failure.
    - I. Dobson, B.A.Carreras, D.E.Newman, V.E.Lynch.
  - One natural way to study cascading failure is to model the failure propagation probabilistically.
  - Can capture the features of cascading blackouts.

Simulations

We conducted 2000 simulation studies of 25 generations for Supercritical BPI, Critical BPI and Subcritical BPI to examine the performances of the empirical likelihood inference for BPI.

- Supercritical BPI
  - True offspring distribution: Poi(1.6)
- Critical BPI
  - True offspring distribution: Poi(1)
  - True immigration distribution: Poi(1.6)
- Subcritical BPI
  - True offspring distribution: Poi(0.6)
  - True immigration distribution: Poi(2)

Simulation Results – Supercritical BPI

- Point Estimation:
  We compare the Joint MELE of (m, σ^2) (EL) to CLS, CWLS, MLE, Ratio estimator and Heyde's estimator.
- Interval Estimation:
  - Endpoints obtained by normality approximation: CLS, CWLS, Asy.EL.
  - Endpoints obtained by multivariate Newton iteration with backtracking: EL.

Supercritical BPI–Point Estimation for (m, σ^2)

Supercritical BPI

Offspring Mean Estimators

Estimator	Mean	Std Error	Min	Max
CLS	1.5996574	0.0053870	1.5292175	1.6362661
CWLS	1.5996250	0.0049793	1.5254409	1.6244939
MLE	1.5995014	0.0057880	1.5164204	1.6241147
Ratio	1.5995453	0.0074557	1.4970760	1.6502547
EL	1.5995564	0.0061377	1.5153504	1.6319085

Offspring Variance Estimators

Estimator	Mean	Std Error	Min	Max
CLS	0.7086591	0.7351740	-0.0121813	7.0753418
CWLS	0.4459672	22.2700942	-639.1483771	67.0293782
Heyde	1.6498455	0.5332430	0.4792931	4.9339919
EL	1.4994264	0.2233468	0.2581471	1.8796093

Table: Point Estimation for Supercritical BPI (m, σ^2)

Supercritical BPI–95% Confidence Region for (m, σ^2)

Supercritical BPI

95% Confidence Interval for Offspring Mean

Estimator	LB	UB	Avg.L	CR
CLS	1.3056195	1.8936848	0.5880654	0.9962990
CWLS	1.0680544	2.1311089	1.0630545	0.8971132
Asy.EL	1.1214110	2.0777018	0.9562908	0.9992598
EL	1.1539394	2.0541810	0.9002447	0.9918579

95% Confidence Interval for Offspring Variance

Estimator	LB	UB	Avg.L	CR
CLS	0.3157985	1.1015197	0.7857211	0.2161362
CWLS	0.1987356	0.6931989	0.4944633	0.6787565
Asy.EL	0.6681868	2.3306660	0.9407846	0.9407846
EL	0.9589265	2.5137731	0.9348631	0.9348631

Table: 95% Confidence Region for Supercritical BPI (m, σ^2)

Simulation Results – Subcritical BPI

- Point Estimation:
  - BP with Observed Immigration: EL.obs
  - BP with Unobserved Immigration: EL.
- Interval Estimation:
  - Endpoints obtained by normality approximation: CLS, CWLS, Asy.EL.
  - Endpoints obtained by multivariate Newton iteration with backtracking: EL.

Subcritical BPI–Point Estimation for (m, σ^2)

Subcritical BPI

Offspring Mean Estimators

Estimator	Mean	Std Error	Min	Max
CLS	0.4068871	0.2099819	-0.3846154	0.9762500
CWLS	0.4397787	0.2206041	-0.4434389	0.9222224
MLE	0.5987485	0.0923265	0.2903226	0.9156627
Ratio	1.0341678	0.0216742	0.9833333	1.1428571
Quine	0.5615899	0.2279795	-0.3044294	1.3787879
EL.obs	0.5993934	0.1024770	0.2049674	1.3894588
EL	0.5814465	0.2284463	0.0857700	1.7436559

Table: Point Estimation for Subcritical BPI (m, σ^2)

Subcritical BPI–Point Estimation for (m, σ^2)

Subcritical BPI
Offspring Variance Estimators

Estimator	Mean	Std Error	Min	Max
CLS	0.2789816	0.4781353	-1.3801571	2.4994932
CWLS	0.6977121	5.0034361	-179.2734054	24.1664708
VarSBH	0.6251654	0.3016981	0.1534433	4.0594455
VarSV	0.5878960	0.2183728	0.1480292	1.9122667
EL.obs	0.5992628	0.1039223	0.1489673	1.4573774
EL	0.6229374	0.3781248	0.0611901	5.6458308

Table: Point Estimation for Subcritical BPI (m, σ^2)

Subcritical BPI–95% Confidence Region for (m, σ^2)

Subcritical BPI

95% Confidence Interval for Offspring Mean

Estimator	LB	UB	Avg.L	CR
CLS	0.1819753	0.6843181	0.5023428	0.4602821
CWLS	0.0953086	0.7825294	0.6872208	0.7772829
Asy.EL	0.3170494	0.9250387	0.6079893	0.9814402
EL.obs	0.2958284	0.8996697	0.6008026	0.9584261
EL	0.2949039	1.3148867	1.0199828	0.9076752

95% Confidence Interval for Offspring Variance

Estimator	LB	UB	Avg.L	CR
CLS	0.1243234	0.4336456	0.3093222	0.3533779
CWLS	0.3107357	1.0838602	0.7731245	0.8411284
Asy.EL	0.2807552	0.9792868	0.6985316	0.9747587
EL.obs	0.2693595	0.9285747	0.6620906	0.9688196
EL	0.3048490	1.4495675	1.1447185	0.8954394

Table: 95% Confidence Region for Subcritical BPI (m, σ^2)

Subcritical BPI–Point Estimation for (λ, b^2)

Subcritical BPI

Immigration Mean Estimators

Estimator	Mean	Std Error	Min	Max
CLS	3.0196735	1.1048027	0.3750000	9.3054973
CWLS	2.8301833	1.0244408	0.4180328	8.7321829
MLE	2.0745279	0.4837316	0.6400000	6.2000000
Quine	2.6508300	0.9396101	0.4216179	7.9223668
EL.obs	1.9015066	0.5750489	0.2049674	5.4748058
EL	2.3911792	0.9663029	0.0348432	5.9638341

Immigration Variance Estimators

Estimator	Mean	Std Error	Min	Max
CLS	7.0564799	4.5273718	-3.8825709	52.8482066
CWLS	5.9346225	82.0837328	-332.1321302	2972.61
VarYTD	2.7912286	1.6651569	0.4072727	18.6014294
EL.obs	1.9241260	0.5482316	0.0214932	5.4361580
EL	2.4492310	0.8442261	0.2513182	6.6314177

Table: Point Estimation for Subcritical BPI (λ, b^2)

Subcritical BPI–95% Confidence Region for (λ, b^2)

Subcritical BPI

95% Confidence Interval for Immigration Mean

Estimator	LB	UB	Avg.L	CR
CLS	2.0338164	4.0282839	1.9944675	0.5105954
CWLS	2.0209158	3.6301489	1.6092331	0.4793138
Asy.EL	1.3270262	2.3844980	1.0574718	0.8355197
EL.obs	1.3265960	2.3844124	1.0578164	0.8355197
EL	1.7870932	2.9952653	2.5876208	0.6840934

95% Confidence Interval for Immigration Variance

Estimator	LB	UB	Avg.L	CR
CLS	3.0998580	10.8124456	7.7125876	0.2522704
CWLS	1.7293116	6.0319174	4.3026059	0.5075681
Asy.EL	0.8144764	2.8970478	2.0825714	0.9112008
EL.obs	0.8928373	2.8807958	1.9879585	0.9051463
EL	1.0914466	3.8070154	2.7006976	0.9132369

Table: 95% Confidence Region for Subcritical BPI (λ, b^2)

Subcritical BPI—-95% Confidence Region for (m, σ^2)

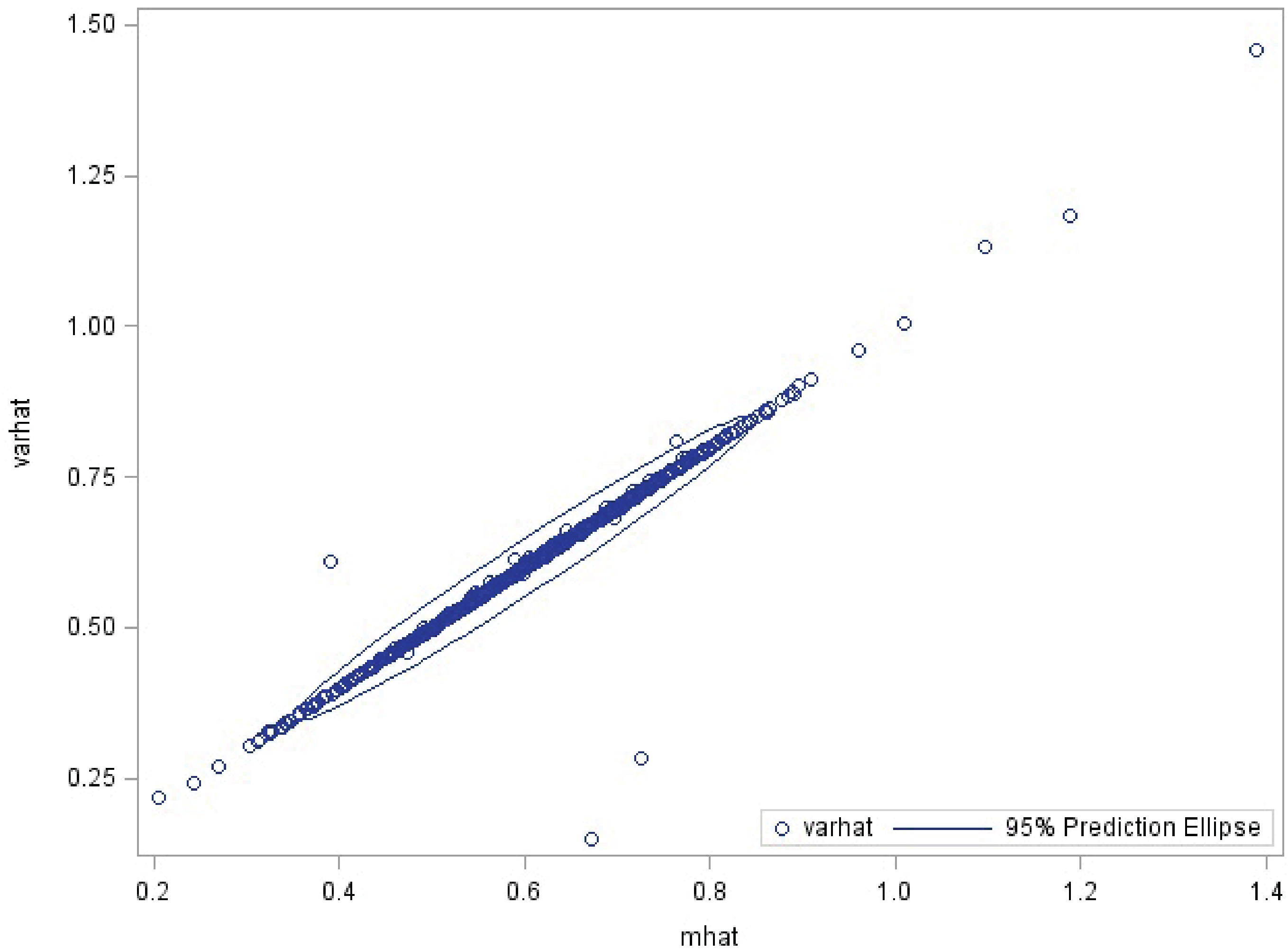


Figure: 95% Confidence Region for Subcritical BPI (m, σ^2)

Subcritical BPI—-95% Confidence Region for (λ, b^2)

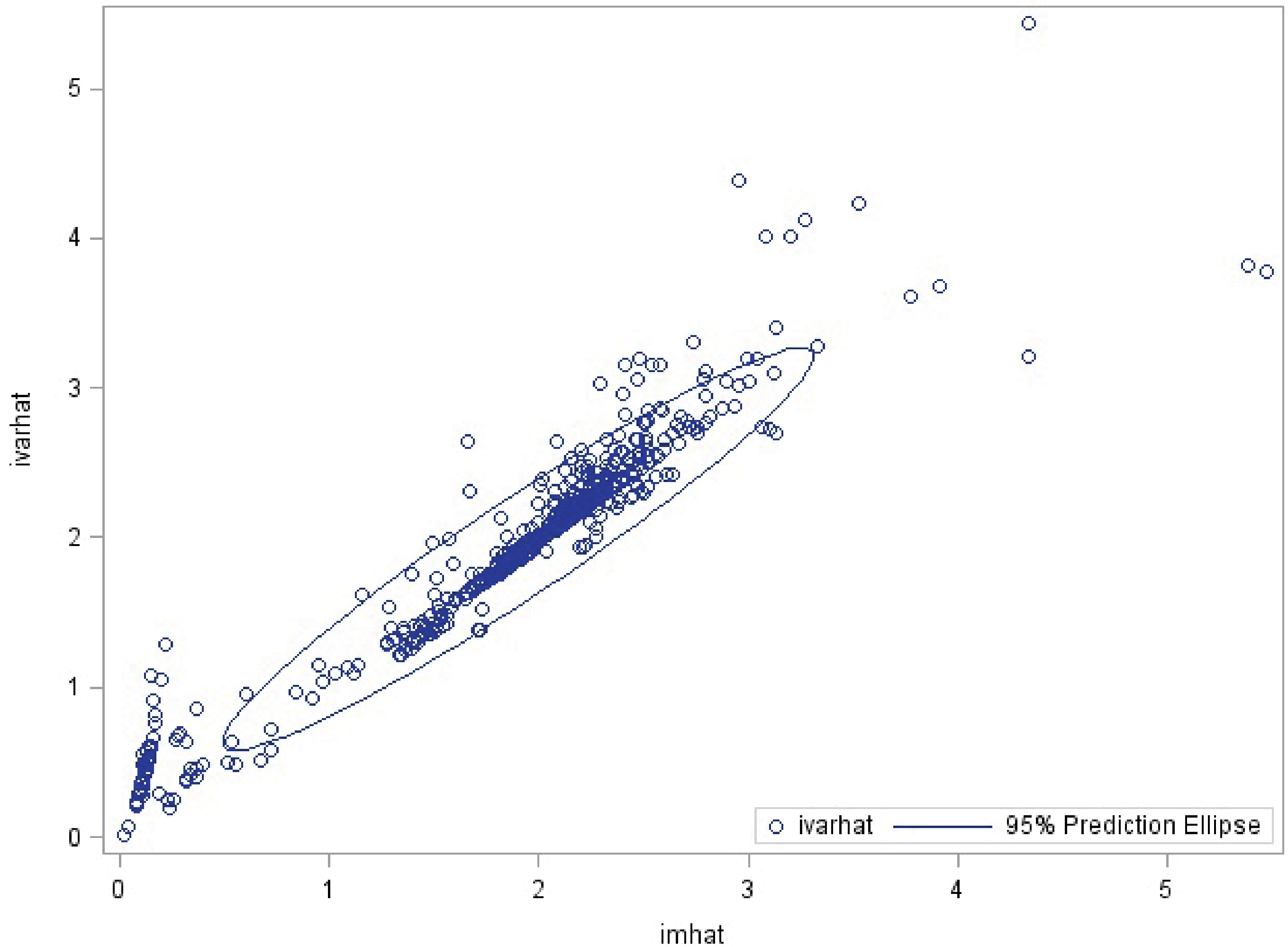


Figure: 95% Confidence Region for Subcritical BPI (λ, b^2) with Observed Immigration

Subcritical BPI—-95% Confidence Region for (λ, b^2)

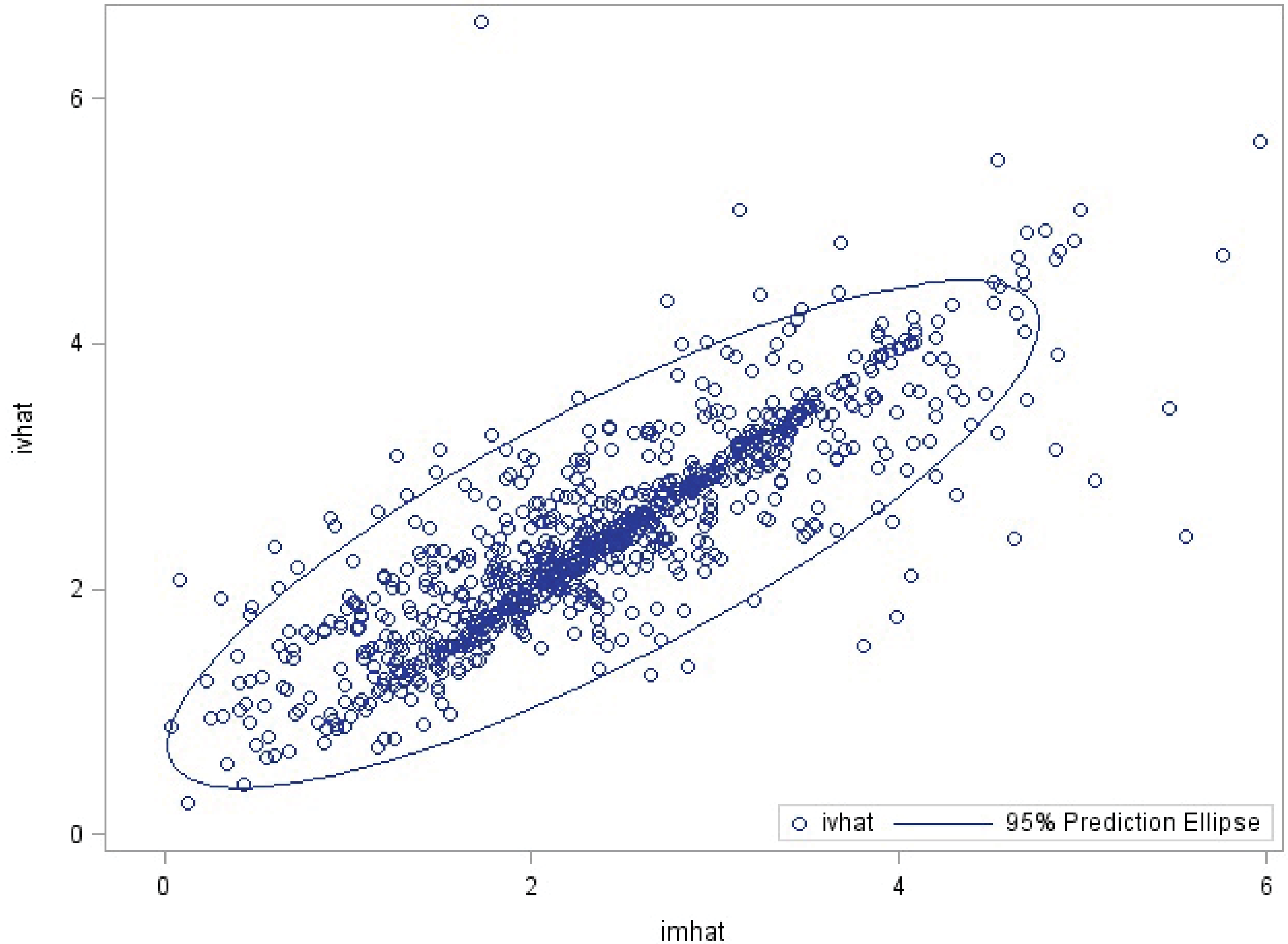


Figure: 95% Confidence Region for Subcritical BPI (λ, b^2) with Unobserved Immigration

Critical BPI–Point Estimation for (m, σ^2)

Critical BPI

Offspring Mean Estimators

Estimator	Mean	Std Error	Min	Max
CLS	0.8921324	0.1402637	0.2256241	1.2510709
CWLS	0.9232139	0.1288810	0.1938037	1.1479109
MLE	0.9735887	0.0695340	0.6203704	1.1273938
Ratio	1.0756399	0.0276707	0.9920635	1.1875000
EL.obs	0.9742001	0.0756762	0.6207906	1.5365060
EL	1.0580636	0.0826038	0.3156783	2.2288698

Offspring Variance Estimators

Estimator	Mean	Std Error	Min	Max
CLS	0.7116998	0.5244955	-0.3399267	3.7321001
CWLS	0.5246462	9.9740820	-375.2042404	62.4754245
EL.obs	0.5992628	0.1039223	0.1489673	1.4573774
EL	0.9848139	0.4900028	0.0974502	17.7483737

Table: Point Estimation for Critical BPI (m, σ^2)

Critical BPI–95% Confidence Region for (m, σ^2)

Critical BPI

Confidence Interval for Offspring Mean

Estimator	LB	UB	Avg.L	CR
CLS	0.5814326	1.2145673	0.6331347	0.8657980
CWLS	0.5373516	1.3046144	0.7672628	0.8188925
EL.obs	0.4777786	1.3500880	0.8723094	0.9921824
EL	0.6973396	1.39841412	0.7011716	0.9799465

Confidence Interval for Offspring Variance

Estimator	LB	UB	Avg.L	CR
CLS	0.2324951	0.8109535	0.7882925	0.4553746
CWLS	0.2324951	0.8109535	0.5784584	0.5140065
EL.obs	0.4425438	1.5111772	1.0686334	0.9908795
EL	0.6491466	2.2861169	1.6379064	0.8262032

Table: 95% Confidence Region for Critical BPI (m, σ^2)

Critical BPI–Point Estimation for (λ, b^2)

Critical BPI

Immigration Mean Estimators

Estimator	Mean	Std Error	Min	Max
CLS	3.3350633	1.8282804	-0.7370735	12.5887198
CWLS	2.5322309	1.1930882	0.3865971	9.3386145
MLE	1.6387712	0.4039895	0.0400000	5.3200000
EL.obs	1.5096782	0.4657211	0.0014054	3.8353091
EL	2.3928942	1.1721498	0.0273802	8.1873215

Immigration Variance Estimators

Estimator	Mean	Std Error	Min	Max
CLS	34.2399676	765.7872498	-9775.36	24178.44
CWLS	-75.6584092	2571.39	-85255.27	8720.79
VarYTD	20.5800039	17.6236148	1.3405354	157.4344275
EL.obs	1.5213745	0.4588831	0.0014015	4.1632562

Table: Point Estimation for Critical BPI (λ, b^2)

Critical BPI–95% Confidence Region for (λ, b^2)

Critical BPI

Confidence Interval for Immigration Mean

Estimator	LB	UB	Avg.L	CR
CLS	0.9434703	5.8385367	4.8950664	0.4333333
CWLS	-0.6740607	5.8065286	6.4805894	0.5264706
EL.obs	1.0186861	1.9628146	0.9441285	0.8235294
EL	0.6831852	5.0155742	5.0155742	0.8081551

Confidence Interval for Immigration Variance

Estimator	LB	UB	Avg.L	CR
CLS	24.7591358	86.3609915	61.6018557	0.0382353
CWLS	4.2515450	14.8295823	10.5780373	0.0500000
EL.obs	0.6798551	2.3575081	1.6776530	0.8598039

Table: Confidence Interval for Critical BPI (λ, b^2)

Critical BPI–95% Confidence Region for (m, σ^2)

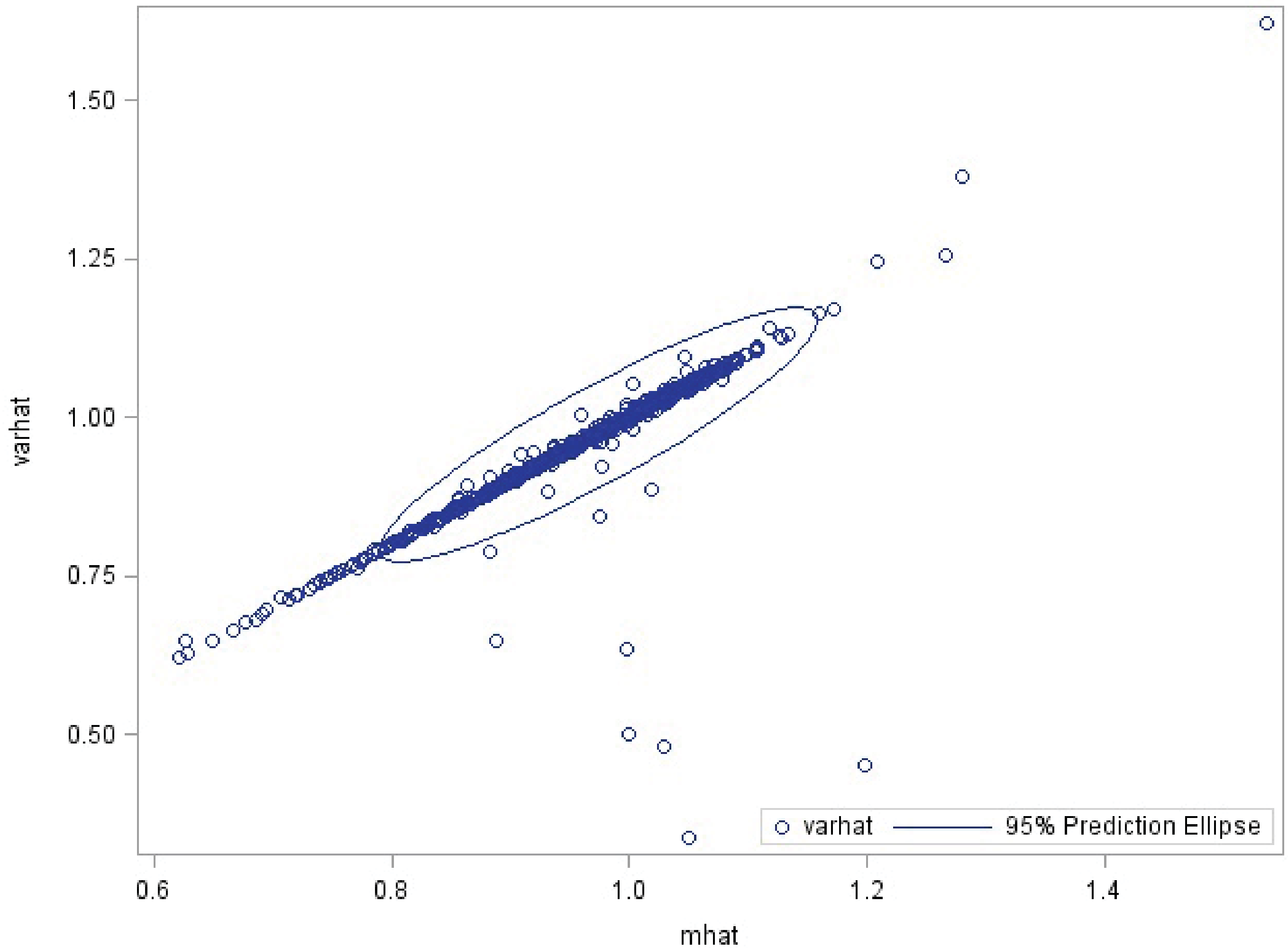


Figure: 95% Confidence Region for Critical BPI (m, σ^2)

Data Analysis

- The power system cascades are observed until are there K nontrivial cascades.
- Each nontrivial cascade has a positive number of failures in stage zero ($Z_0 > 0$).
- $Z_j^{(k)}$: the number of outages in stage j of cascade k.
- All statistics are conditioned on $Z_0 > 0$.
- A Poisson distribution with mean λ is the parametric model is assumed to describe the offspring distribution.

Data Analysis(contd.)

Traditional Data Analysis:

- An arbitrary distribution of nonzero initial failure: $P(Z_0 = z_0)$ for $z_0 = 1, 2, \cdots$.
- Y: the total number of failures.
- A mixture of Borel-Tanner distribution

$$P(Y = r) = \sum_{z_0=1}^{r} p(Z_0 = z_0) z_0 m(rm)^{r-z_0-1} \frac{e^{-rm}}{(r - z_0)!}$$

when $0 \leq m < 1$.

- $\hat{m} = \dfrac{\sum_{k=1}^{K} \left(Z_1^{(k)} + Z_2^{(k)} + \cdots \right)}{\sum_{k=1}^{K} \left(Z_0^{(k)} + Z_1^{(k)} + \cdots \right)}$

Data Analysis (contd.)

Summing over all the 226 cascades the number of outages in each stage. Of the 396 outages, 296 are in stage 0 of a cascade, 45 are in stage of 1 of a cascade and so on.

Cascades Data

Number of Outage in Each Stage

Stage Num.	0	1	2	3	4	5	6	7
Outage Num.	296	45	18	14	10	3	1	1

Number of Outage in Each Stage

Stage Num.	8	9	10	11	12	13	14	15
Outage Num.	1	1	1	1	2	1	1	0

Data Analysis (contd.)

Empirical Likelihood for Cascades Data

Empirical Likelihood for Cascades Data

ELmean	ELvar	ELLBmean	ELUBmean	ELLBvar	ELUBvar
0.25395	1.43483	.000000693	0.62712	0.40787	2.28816

Conclusions

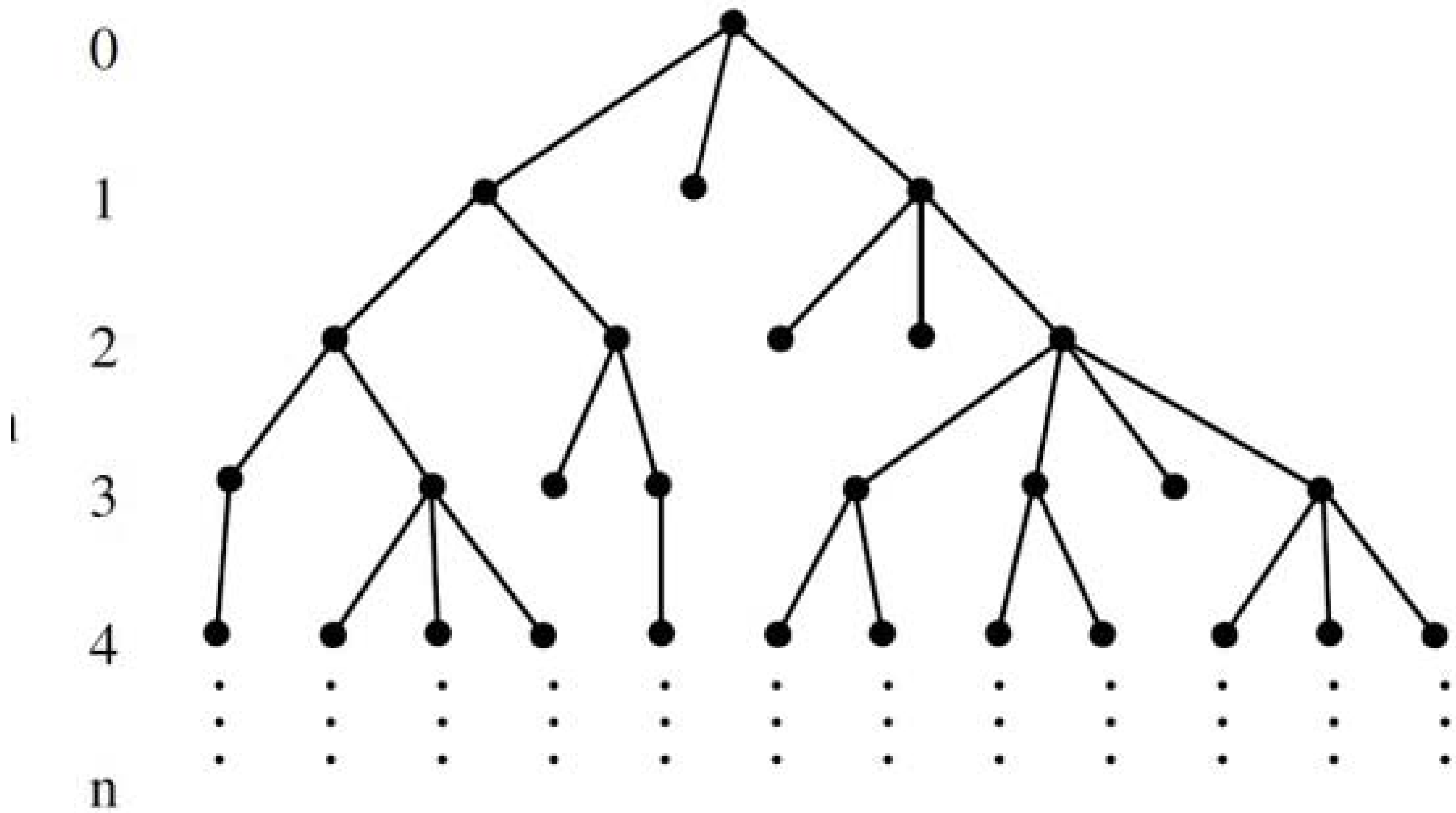
- A general background information of branching processes and branching processes with immigration.
- Described a new empirical likelihood based methodology for constructing the simultaneous confidence region for the mean and variance parameters of the branching processes.
- Provided a theoretical justification of the proposed methods and an algorithm for computations.
- Presented simulation and data analysis results.

Concluding Remarks

- We obtained a unified EL Theory for BPI which removes the distributional discontinuity at $m = 1$.
- Extensions to joint (m, σ^2) has also been carried out.
- One can replace the Empirical likelihood ratio by any divergence (for instance HD) and study the limiting properties.
- The methods can be extended to ancestoral inference for BP (Hanlon and Vidyashankar, 2012).
- Currently we are investigating extensions to other processes.

Muchas Gracias!!!

Branching Processes (contd.)

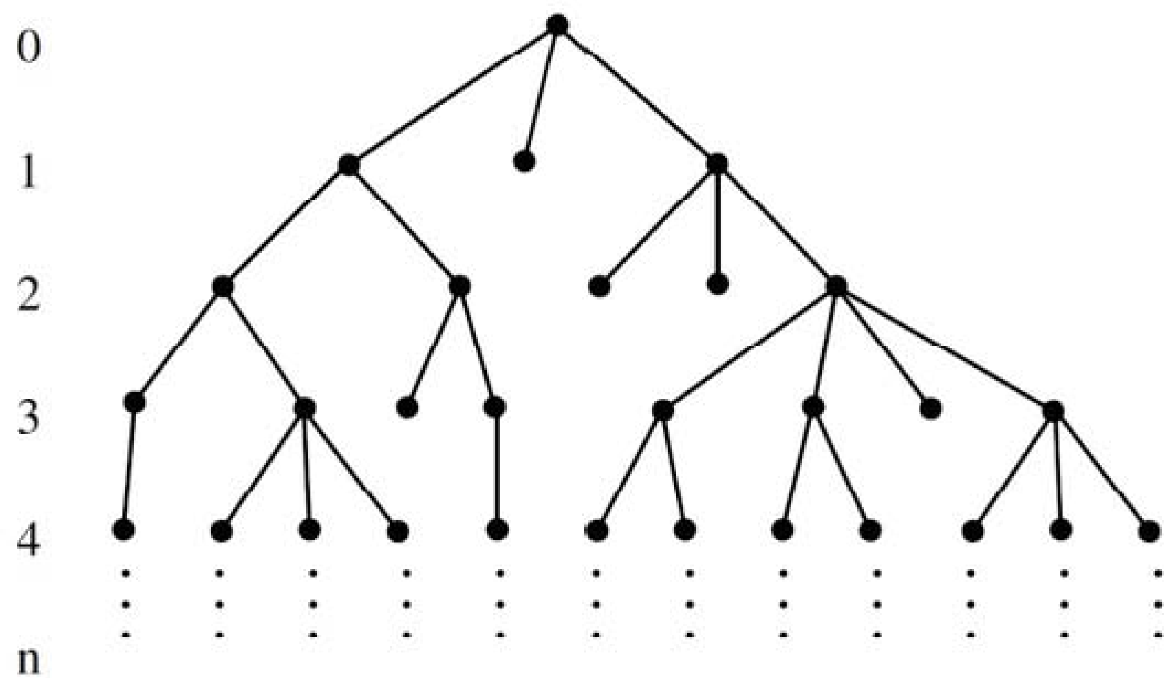


Branching Processes (contd.)

$$Z_{n+1} = \sum_{j=1}^{Z_n} \xi_{n,j}, \quad n = 1, 2, ..,$$

- $\{Z_n; n = 0, 1, 2, \cdots\}$: is a Markov chain on the nonnegative integers.
- $\xi_{n,j}$: the number of offspring produced by the j^{th} parent in the n^{th} generation, and they are $i.i.d.$ random variables with generating function $f(s)$.
- $P(\xi_{n,j} = k) = p_k, k \geq 0$: the probability distribution of $\xi_{n,j}$ for all n and j.
- $\{p_k : k \geq 0\}$: the offspring distribution.

Branching Processes (contd.)



$$Z_2 = \sum_{j=1}^{Z_1} \xi_{1,j} = \xi_{1,1} + \xi_{1,2} + \xi_{1,3} = 2 + 0 + 3 = 5.$$

$\xi_{1,j}$: number of offspring produced by the j^{th} parent in the 1^{st} generation.

Branching Processes (contd.)

- $$m = E(\xi_{n,j}) = E(Z_1 | Z_0 = 1),$$
  $$\sigma^2 = Var(\xi_{n,j}) = Var(Z_1 | Z_0 = 1) < \infty$$

- $$E(Z_n) = E\Big(E(Z_n | Z_{n-1})\Big) = mE(Z_{n-1}) = m^n$$
  and
  $$Var(Z_n) = \begin{cases} \dfrac{\sigma^2 m^{n-1}(m^n - 1)}{m-1} & \text{if } m \neq 1, \\ n\sigma^2 & \text{if } m = 1. \end{cases}$$

- $m > 1$: $Z_n \to \infty$ with positive probability; $m > 1$: $Z_n \to \infty$ with probability one if $p_0 = 0$.
- $m \leq 1$: $Z_n \to 0$ with probability one. (In the case $m = 1$, we make a further assumption that $p_1 \neq 1$.)

Branching Processes (contd.)

Trichotomy of Branching Processes

- Supercritical: $m > 1$.
- Critical: $m = 1$.
- Subcritical: $m < 1$.

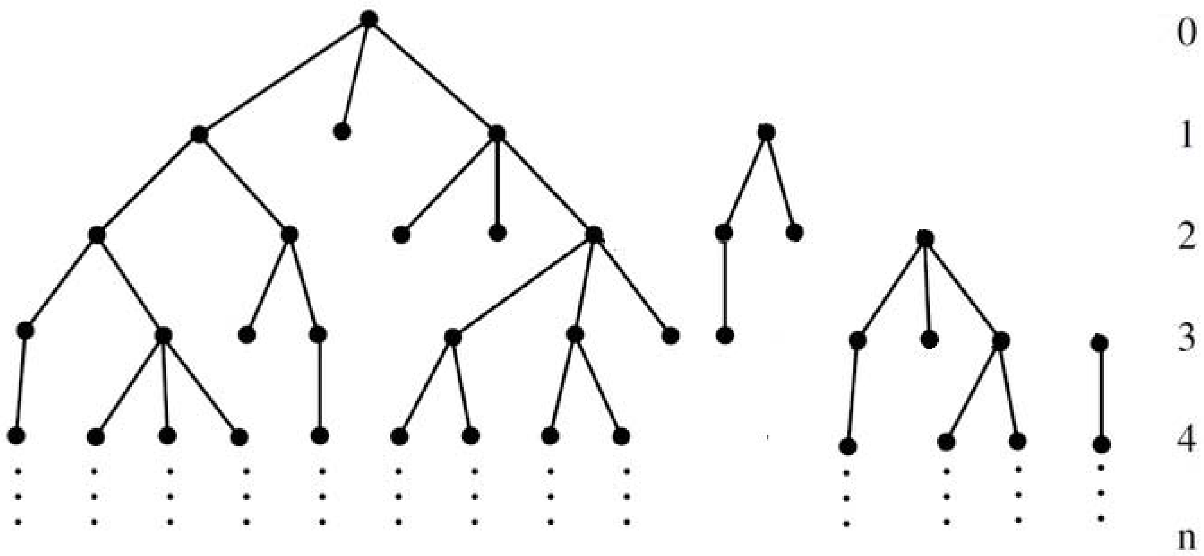
Branching Processes with Immigration

- A useful and realistic modification of the BP is the incorporation of an immigration component into the population.

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_{n,j} + I_{n+1}, \quad n = 1, 2, ..,$$

- X_n and $\xi_{n,j}$: same as defined before.
- I_{n+1}: the number of immigrants entering in the $(n+1)^{th}$ generation from an outside source, and I_1, I_2, $\cdots$ are $i.i.d$ random variables with generating function $h(s)$.
- $P(I_1 = j) = q_j, j \geq 0$: the probability distribution of I_i for all $1 \leq i \leq n$.
- $\{q_j : j \geq 0\}$: immigration distribution.
- $\{I_{n+1}\}$ is independent of $\{\xi_{n,j}\}$ and $\{X_n\}$.

Branching Processes with Immigration (contd.)



$I_1 = 1, \;\; I_2 = 1, \;\; I_3 = 1.$

$$X_2 = \sum_{j=1}^{X_1} \xi_{1,j} + I_2 = \xi_{1,1} + \xi_{1,2} + \xi_{1,3} + \xi_{1,4} + I_2 = 2+0+3+2+1 = 8.$$

$\xi_{1,j}$: number of offspring produced by the j^{th} parent in the 1^{st} generation.

I_2: number of immigration entering in the 2^{nd} generation.

Branching Processes with Immigration (contd.)

- $\lambda = E(I_1)$, $b^2 = Var(I_1)$.
- $m > 1$: the generation sizes of a BPI diverge to infinity with probability one.
- $m \leq 1$: X_n can be a transient or a recurrent Markov chain.

Literature Review

- Estimating Mean
- Estimating Variance
- Estimating Mean and Variance Jointly
- Nonexistence of consistent estimators of λ and b^2 if BPI is partially observed in the supercritical case.
    - Partially observed BPI: $\{X_n\}$ is available but $\{I_n\}$ for $n \geq 1$ is not observable.
    - The dominated contribution of offspring in the supercritical BPI case.

Literature Review–Estimating Mean

- Maximum Likelihood Estimator (MLE) (Harris(1948), Jagers(1975)).
- Ratio Estimator (Nagaev(1967)).
- Quine's Estimator (Quine(1976)).
- Conditional Least Squares Estimator (CLS) (Wei and Winnicki(1990)).
- Conditional Weighted Least Squares Estimators (Wei and Winnicki(1990)).

Estimating Mean–MLE and Ratio Estimator

Supercritical BP

- When data from all generations are all available from generation 1 through n.
  - Maximum Likelihood Estimator (Harris(1948), Jagers(1975)):

$$\tilde{m}_n = \frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} Z_{i-1}}.$$

- When the data from the last two generations are available and $Z_{n-1} \neq 0$.
  - Ratio Estimator (Nagaev(1967)):

$$\hat{m}_n = \frac{Z_n}{Z_{n-1}}.$$

Estimating Mean–Quine's Estimator

Subcritical BPI

- Basic Notation
  - $\mu = \dfrac{\lambda}{1-m}$. $c^2 = \mu\sigma^2 + b^2$. and $S_n = \sum_{i=1}^{n} X_{i-1}$.
- Preliminary Results
  - $\dfrac{1}{n} S_n \xrightarrow{\text{a.s.}} \mu$.
  - $\dfrac{1}{n} \sum_{i=1}^{n} (X_i - X_{i-1})^2 \xrightarrow{\text{a.s.}} \dfrac{2c^2}{1+m}$.
  - $\dfrac{1}{n} \sum_{i=1}^{n} X_{i-1}^2 \xrightarrow{\text{a.s.}} \dfrac{c^2}{1-m^2} + \mu^2$.
  - $Var(X_n) \rightarrow \dfrac{c^2}{1-m^2}$.
  - $\dfrac{1}{n} Var(S_n) \rightarrow \dfrac{c^2}{(1-m)^2}$.

Estimating Mean–Quine's Estimator (contd.)

Subcritical BPI

- Quines Estimators (Quine(1976)):
  - $\hat{m}_{Quine} = \dfrac{\sum_{i=1}^{n} X_{i-1}(X_i - n^{-1}S_n)}{\sum_{i=1}^{n}(X_i - n^{-1}S_n)^2}.$
  - $\hat{\lambda}_{Quine} = \left(\dfrac{S_n}{2n}\right)\left(\dfrac{\sum_{i=1}^{n}(X_i - X_{i-1})^2}{\sum_{i=1}^{n}(X_i - n^{-1}S_n)^2}\right).$

Estimating Mean–CLS

- $\mathcal{F}_{i-1} = \sigma\{X_0, X_1, \cdots, X_{i-1}\}$.
- $E(X_i|\mathcal{F}_{i-1}) = E\left((\sum_{j=1}^{X_{i-1}} \xi_{i-1,j} + I_i)|\mathcal{F}_{i-1} \right) = mX_{i-1} + \lambda$.
- $U_i = X_i - E(X_i|\mathcal{F}_{i-1}) = X_i - mX_{i-1} - \lambda$.
- Objective function: $\sum_{i=1}^{n} \left(X_i - E(X_i|\mathcal{F}_{i-1}) \right)^2$.

Estimating Mean–CLS (contd.)

- $\hat{m}_n = \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_{i-1} - n \sum_{i=1}^{n} X_{i-1} X_i}{\left(\sum_{i=1}^{n} X_{i-1}\right)^2 - n \sum_{i=1}^{n} X_{i-1}^2}.$
- $\hat{\lambda}_n = \dfrac{\sum_{i=1}^{n} X_{i-1} X_i \sum_{i=1}^{n} X_{i-1} - \sum_{i=1}^{n} X_{i-1}^2 \sum_{i=1}^{n} X_i}{\left(\sum_{i=1}^{n} X_{i-1}\right)^2 - n \sum_{i=1}^{n} X_{i-1}^2}.$

Estimating Mean–CWLS

- Heterogeneity: $Var(U_i|\mathcal{F}_{i-1}) = E(U_i^2|\mathcal{F}_{i-1}) = \sigma^2 X_{i-1} + b^2$.
- $E\left(\dfrac{U_i^2}{X_{i-1}+1}\Big|\mathcal{F}_{i-1}\right) = \dfrac{\sigma^2 X_{i-1} + b^2}{X_{i-1}+1}$
- Objective function: $\sum_{i=1}^{n}\left(\dfrac{X_i - E(X_i|\mathcal{F}_{i-1})}{\sqrt{X_{i-1}+1}}\right)^2$.

Estimating Mean–CWLS (contd.)

- $$\tilde{m}_n = \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} (1 + X_{i-1})^{-1} - n \sum_{i=1}^{n} X_i (1 + X_{i-1})^{-1}}{\sum_{i=1}^{n} (1 + X_{i-1}) \sum_{i=1}^{n} (1 + X_{i-1})^{-1} - n^2}.$$
- $$\tilde{\lambda}_n = \frac{\sum_{i=1}^{n} X_{i-1} \sum_{i=1}^{n} \frac{X_i}{1 + X_{i-1}} - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \frac{X_{i-1}}{1 + X_{i-1}}}{\sum_{i=1}^{n} (1 + X_{i-1}) \sum_{i=1}^{n} (1 + X_{i-1})^{-1} - n^2}.$$

Literature Review–Estimating Variance

- Heyde's Estimator (MLE) (Heyde(1974)).
- Conditional Least Squares Estimator (CLS) (Winnicki(1991)).
- Conditional Weighted Least Squares Estimators (CWLS) (Winnicki(1991)).

Estimating Variance–Heyde's Estimator

Supercritical BP

- Recall $Var(U_i|\mathcal{F}_{i-1}) = E(U_i^2|\mathcal{F}_{i-1}) = \sigma^2 X_{i-1} + b^2$.
- $Var\left(X_i - mX_{i-1}|\mathcal{F}_{i-1}\right) = E\left((X_i - mX_{i-1})^2|\mathcal{F}_{i-1}\right) = \sigma^2 X_{i-1}$
- Heyde's Estimators:

$$\hat{\sigma}^2_{Heyde} = \frac{1}{n}\sum_{i=1}^{n}\frac{(X_i - \hat{m}_n X_{i-1})^2}{X_{i-1}},$$

where $\hat{m}_n$ is the Ratio Estimator.

Estimating Variance – CLS

- Recall $U_i = X_i - E(X_i|\mathcal{F}_{i-1}) = X_i - mX_{i-1} - \lambda$.
- $\hat{U}_i = X_i - \hat{m}_n X_{i-1} - \hat{\lambda}_n$.
- Recall $E(U_i^2|\mathcal{F}_{i-1}) = \sigma^2 X_{i-1} + b^2$.
- $V_i = U_i^2 - E(U_i^2|\mathcal{F}_{i-1}) = U_i^2 - \sigma^2 X_{i-1} - b^2$.
- Objective function: $\sum_{i=1}^{n} \left(U_i^2 - E(U_i^2|\mathcal{F}_{i-1})\right)^2$.

Estimating Variance – CLS (contd.)

- $$\hat{\sigma}_n^2 = \frac{n \sum_{i=1}^{n} X_i U_i^2 - \sum_{i=1}^{n} X_{i-1} U_i^2}{n \sum_{i=1}^{n} X_{i-1}^2 - \left(\sum_{i=1}^{n} X_{i-1} \right)^2}.$$
- $$\hat{b}_n^2 = \frac{\sum_{i=1}^{n} U_i^2 \sum_{i=1}^{n} X_{i-1}^2 - \sum_{i=1}^{n} X_{i-1} U_i^2 \sum_{i=1}^{n} X_{i-1}}{n \sum_{i=1}^{n} X_{i-1}^2 - \left(\sum_{i=1}^{n} X_{i-1} \right)^2}.$$

Estimating Variance – CWLS

- Objective function $\sum_{i=1}^{n} \left(\frac{U_i^2 - E(U_i^2|\mathcal{F}_{i-1})}{1 + X_{i-1}} \right)^2$
- $T_i = 1 + X_{i-1}$.
- $\tilde{\sigma}_n^2 = \dfrac{\sum_{i=1}^{n} T_i^{-2} \sum_{i=1}^{n} U_i^2 T_i^{-1} - \sum_{i=1}^{n} T_i^{-1} \sum_{i=1}^{n} U_i^2 T_i^{-2}}{n \sum_{i=1}^{n} T_i^{-2} - \left(\sum_{i=1}^{n} T_i^{-1} \right)^2}$.
- $\tilde{b}_n^2 = \dfrac{\sum_{i=1}^{n} X_{i-1} T_i^{-1} \sum_{i=1}^{n} U_i^2 T_i^{-2} - \sum_{i=1}^{n} X_{i-1} T_i^{-2} \sum_{i=1}^{n} U_i^2 T_i^{-1}}{n \sum_{i=1}^{n} T_i^{-2} - \left(\sum_{i=1}^{n} T_i^{-1} \right)^2}$.

Literature Review – Estimating Mean and Variance Jointly

Quasi-Likelihood Estimator (Basawa and Vidyashankar(2003)):

- The approximate quasi-likelihood score function

$$\tilde{S}_n(\theta) = \sum_{i=1}^{n} D_i(\theta)\tilde{V}_i^{-1}(\theta)g_i(\theta)$$

- $g_i(\theta) = \left(X_i - mX_{i-1} - \lambda, \quad U_i^2 - \sigma^2 X_{i-1} - b^2\right),$
- $D_i(\theta) = E\left[\frac{\partial g_i(\theta)}{\partial \theta}\middle|\mathcal{F}_{i-1}\right]$
- $\tilde{V}_i(\theta) = Var(g_i(\theta)|\mathcal{F}_{i-1}) = \begin{pmatrix} \sigma^2 X_{i-1} + b^2 & 0 \\ 0 & 2(\sigma^2 X_{i-1} + b^2) \end{pmatrix}$

Objective

- The goal is to make inference for (m, σ^2) using empirical likelihood for BPI without making assumptions concerning the value of m. This will be referred to as joint inference for branching processes with immigration.

Empirical Likelihood (EL)

- EL is developed by Owen(1988,1990) for independent and identically distributed ($i.i.d$) data, and extended by Qin and Lawless(1994) to functionally independent estimating equations.
- EL is a nonparametric inference approach based on the data-determined likelihood ratio function.
- EL incorporates the advantages of the likelihood methods and the robustness of nonparametric approach.
- The confidence regions yields by EL respect the range of the target parameters for $i.i.d.$ data.

EL for $i.i.d.$ Data (Owen(1988,1990))

Let $X_1, X_2, \ldots, X_n$ be $i.i.d$ random variables with CDF $F(\cdot)$.

- The empirical CDF (ECDF) $F_n(x) = \dfrac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$ is a non-parametric MLE of $F(\cdot)$.
- Nonparametric likelihood: $L(F) = \prod (F(X_i) - F(X_i -))$.

EL for $i.i.d.$ Data (Owen(1988,1990)) (contd.)

For hypothesis test: $H_0 : T(F_0) = \theta_0$

- Nonparametric Likelihood Ratio: $R(F) = \frac{L(F)}{L(F_n)}$.
- Profile Empirical Likelihood Ratio(ELR) Function:

$$\mathcal{R}(\theta) = \sup\{R(F) | T(F) = \theta\}.$$

- Empirical Likelihood Confidence Region: $\{\theta | \mathcal{R}(\theta) \geq c\}$

EL For The Mean of *i.i.d.* Data (Owen(1988,1990))

Example 1: Let $E(X_1) = \mu$. Consider the problem of hypothesis tests of $H_0 : \mu = \mu_0$ and constructing the confidence interval for μ.

- w_i*: weights assigned on each observation.*
- $L(F) = \prod_{i=1}^{n} w_i$.
- $L(F_n) = \prod_{i=1}^{n} \frac{1}{n}$.
- $R(F) = \prod_{i=1}^{n} n w_i$.
- $$T(F) = \sum_{i=1}^{n} w_i X_i = \mu.$$

EL For The Mean of *i.i.d.* Data (Owen(1988,1990)) (contd.)

- Profile ELR function for the mean

$$\mathcal{R}_n(\mu) = \max\left\{\prod_{i=1}^{n} nw_i \middle| w_i \geq 0,\ \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i X_i = \mu\right\}.$$

- EL confidence region for the mean

$$\{\mu | \mathcal{R}_n(\mu) \geq c\} = \left\{\sum_{i=1}^{n} w_i X_i \middle| w_i \geq 0,\ \sum_{i=1}^{n} w_i = 1, \prod_{i=1}^{n} nw_i \geq c\right\}.$$

- Under H_0, $-2log\mathcal{R}_n(\mu_0) \xrightarrow{\text{d}} \chi_1^2$.

EL for Functionally Independent Estimating Equations (Qin and Lawless(1994))

- Let $X_1, X_2, \ldots, X_n$ are $i.i.d.$ random variables with unknown distribution function $F(\cdot)$.
- Let θ be a r-dimensional parameter.
- Let $g_1(\boldsymbol{X}, \theta), g_2(\boldsymbol{X}, \theta) \ldots, g_r(\boldsymbol{X}, \theta)$ be the estimating functions.
- $E(\boldsymbol{g}(\boldsymbol{X}, \theta)) = \boldsymbol{0}$ where $\boldsymbol{g}(\boldsymbol{X}, \theta) = \left(g_1(\boldsymbol{X}, \theta), \ldots, g_r(\boldsymbol{X}, \theta)\right)'$.

EL for Functionally Independent Estimating Equations (Qin and Lawless(1994)) (contd.)

Example 2: Let $Var(X_1) = \sigma^2$ *and* $\theta = (\mu, \sigma^2)$*. Consider the problem of hypothesis tests* $H_0 : \theta = \theta_0$ *and constructing the confidence region for* θ*.*

- *Estimating equations:* $E(\mathbf{g}(X, \theta)) = (0, 0)'$*, where* $\mathbf{g}(X, \theta) = \left(X - \mu,\ (X - \mu)^2 - \sigma^2\right)'$.
- *Profile ELR function for the mean*

$$\mathcal{R}_n(\theta) = \max\left\{ \prod_{i=1}^{n} nw_i \middle| w_i \geq 0,\ \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i \mathbf{g}(X_i, \theta) = 0 \right\}.$$

- *Under* H_0, $-2log\mathcal{R}_n(\theta_0) \xrightarrow{d} \chi^2_r$.

Empirical Likelihood for Branching Process with Immigration (ELBPI)

- Our work is inspired by the joint inference for (m, σ^2) from Basawa and Vidyashankar(2003).
- EL inferential framework is well established for $i.i.d.$ data and functionally independent estimating equations.
- The classic empirical likelihood method can not be applied directly to the dependent data.
- Due to the trichotomy of BPI, and the advantage of EL method that respects the boundaries of the target parameters, we will generalize the EL method to BPI data and construct the confidence region for (m, σ^2).

ELBPI (contd.)

- Recall a branching process with immigration

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_{n,j} + I_{n+1}, \quad n = 1, 2, \ldots$$

- $E(\xi_{n,j}) = m$, $Var(\xi_{n,j}) = \sigma^2$, $E(I_1) = \lambda$, $Var(I_1) = b^2$.
- $\mathcal{F}_i = \sigma \langle X_0, X_1, \ldots, X_{i-1} \rangle$ for $1 \le i \le n$.
- $E(X_i | \mathcal{F}_{i-1}) = mX_{i-1} + \lambda$, and $Var(X_{i-1} | \mathcal{F}_{i-1}) = \sigma^2 X_{i-1} + b^2$.
- $U_i = X_i - mX_{i-1} - \lambda$ and $V_i = U_i^2 - \sigma^2 X_{i-1} - b^2$ are two martingale differences for each $1 \le i \le n$.
- $\sum_{i=1}^n U_i$ and $\sum_{i=1}^n V_i$ are two martingales.

ELBPI–Empirical Likelihood Ratio

- Estimating Equations:
$$\sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) = \sum_{i=1}^{n} (U_i, V_i)' = \left(\sum_{i=1}^{n} U_i, \sum_{i=1}^{n} V_i \right)' .$$
- Let $\theta = (m, \sigma^2), \theta_0 = (m_0, \sigma_0^2)$.
- The ELR for testing $H_0 : \theta = \theta_0$:

$$\mathcal{R}_n(\theta) = \max \left\{ \prod_{i=1}^{n} n w_i \middle| w_i \geq 0, \ \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0 \right\} .$$

ELBPI–Empirical Likelihood Ratio (contd.)

- The objective function:

$$\max \prod_{i=1}^{n} nw_i$$

subject to:

$$(i) w_i \geq 0, \ (ii) \sum_{i=1}^{n} w_i = 1, \text{ and } (iii) \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0.$$

- The above problem can be reduced to the following objective function:

$$\max \sum_{i=1}^{n} log(nw_i).$$

ELBPI–Preliminary Results

- $G = \sum_{i=1}^{n} log(nw_i) - \nu_1\Big(\sum_{i=1}^{n} w_i - 1\Big) - n\boldsymbol{\nu}_2' \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta)$, where ν_1 and $\boldsymbol{\nu}_2$ are Lagrangian multipliers.
- $0 = \sum_{i=1}^{n} \frac{\partial G}{\partial w_i} \Rightarrow w_i = \frac{1}{1 + \boldsymbol{\nu}_2' \boldsymbol{g}(X_i, \theta)}$.
- $\boldsymbol{\nu}_2$ satisfies

$$\frac{1}{N_n} \sum_{i=1}^{n} \frac{\boldsymbol{g}(X_i;\theta)}{1 + \boldsymbol{\nu}_2' \ \boldsymbol{g}(X_i, \theta)} = 0, \text{ and } N_n = \sum_{i=1}^{n} X_{i-1}.$$

- $-2 \ log\mathcal{R}_n(\theta) = -2 \sum_{i=1}^{n} log(nw_i) = A_n - B_n + R_n$.

ELBPI–Representation Formulas

$$
\begin{aligned}
A_n &= N_n \left(\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right)' S_n^{-1} \left(\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right) \\
&= \left(\sqrt{N_n} \left[\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right] S_n^{-\frac{1}{2}} \right)' \left(\sqrt{N_n} \left[\frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \right] S_n^{-\frac{1}{2}} \right)
\end{aligned}
$$

$$B_n = N_n \left(\beta_n^*\right)' S_n^{-1} \beta_n^*,$$

$$R_n = \frac{2}{3} \sum_{i=1}^{n} r_i^*, \quad r^* \text{is the reminder term in the Taylor expansion.}$$

$$S_n = \frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \boldsymbol{g}(X_i, \theta)',$$

$$\beta_n^* = \frac{1}{N_n} \sum_{i=1}^{n} \boldsymbol{g}(X_i, \theta) \left(\frac{r_i^2}{1 + r_i} \right)', \text{ and } r_i = \boldsymbol{\nu}_2' \boldsymbol{g}(X_i, \theta).$$

ELBPI–Notations

- $\bar{U}_n = \frac{1}{n}\sum_{i=1}^{n} U_i, \ \bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i.$
- $\bar{U}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} U_i^2, \ \bar{V}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} V_i^2.$
- $S_n = \frac{1}{N_n}\begin{pmatrix} \sum_{i=1}^{n} U_i^2 & \sum_{i=1}^{n} U_i V_i \\ \sum_{i=1}^{n} U_i V_i & \sum_{i=1}^{n} V_i^2 \end{pmatrix}.$
- $det(S_n) = \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n^{(2)}\right)\left(\bar{V}_n^{(2)}\right) - \left(\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i\right)^2.$
- $A_n = N_n \left(\frac{n}{N_n}\right)^2 \left(\bar{U}_n, \bar{V}_n\right) S_n^{-1} \left(\bar{U}_n, \bar{V}_n\right)'.$

ELBPI–Main Results for Non-Critical BPI

Theorem

Assuming that $E(X_1^4) < \infty$ and $E(I_1^4) < \infty$. Under H_0, the following hold:

1. *If $m \neq 1$, $\hat{m}_{n,EL} \xrightarrow{a.s.} m$, and $\hat{\sigma}^2_{n,EL} \xrightarrow{a.s.} \sigma^2$.*
2. $$\sqrt{N_n}(\hat{m}_{n,EL} - m, \hat{\sigma}^2_{n,EL} - \sigma^2) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}\right),$$
   where $\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$.
3. *Consider the testing problem, $H_0 : (m, \sigma^2) = (m_0, \sigma_0^2)$,*
   $$-2 \log \mathcal{R}_n(\theta_0) \xrightarrow{d} \chi^2_2.$$

ELBPI–Main Results for Critical BPI

Theorem

Assuming that $m = 1$, $E(X_1^4) < \infty$ and $E(I_1^4) < \infty$, then the following hold:

1. *If $m = 1$, $\hat{m}_{n,EL} \xrightarrow{p} m$, $\hat{\sigma}^2_{n,EL} \xrightarrow{p} \sigma^2$.*
2. *Consider the testing problem $H_0 : (m, \sigma^2) = (m_0, \sigma_0^2)$.*
   1. *When $2\lambda > \sigma^2$, then under H_0,*

$$-2 \log \mathcal{R}_n(\theta_0) \xrightarrow{d} D_1^2 + D_2^2.$$

D_1^2 and D_2^2 are two independent random variables which are distributed as functionals of Feller-diffusion with $D_1^2 = \dfrac{(Y(1) - \lambda)^2}{\sigma^2 \int_0^1 Y(t)dt}$ and $D_2^2 = \dfrac{(\int_0^1 Y(t)dB(t))^2}{\int_0^1 Y^2(t)dt}$, where $B(t)$ is a Brownian motion.

   2. *When $2\lambda \leq \sigma^2$, $-2 \log \mathcal{R}_n(\theta_0) \xrightarrow{p} 0$.*

ELBPI-Joint Maximum Empirical Likelihood Eestimator (MELE)

- $(\hat{m}_{n,EL}, \hat{\sigma}^2_{n,EL}) = \left(\dfrac{\sum_{i=1}^{n} w_i X_i - \hat{\lambda}_n}{\sum_{i=1}^{n} w_i X_{i-1}}, \dfrac{\sum_{i=1}^{n} w_i \hat{U}_i^2 - \hat{b}_n^2}{\sum_{i=1}^{n} w_i X_{i-1}} \right).$
- $\hat{U}_i = X_i - \hat{m}_{n,EL} X_{i-1} - \hat{\lambda}_n.$
- $\hat{\lambda}_n$ and $\hat{b}_n^2$ can be any estimator in the literature studies.

ELBPI–Proof Sketches for Non-Critical BPI

$$-2log\mathcal{R}_n = A_n - B_n + R_n$$

If $m \neq 1$:

- $A_n \xrightarrow{\text{d}} \chi^2_2$,
  - $\hat{m}_{n,EL} \xrightarrow{\text{a.s.}} m$, and $\hat{\sigma}^2_{n,EL} \xrightarrow{\text{a.s.}} \sigma^2$.
  - $\sqrt{N_n}(\hat{m}_{n,EL} - m, \hat{\sigma}^2_{n,EL} - \sigma^2) \xrightarrow{\text{d}} N\left(\mathbf{0}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}\right)$.
- $B_n \xrightarrow{\text{p}} 0$,
- $R_n \xrightarrow{\text{p}} 0$.

ELBPI–Proof Sketches for Non-Critical BPI (contd.)

- $(\hat{m}_{n,EL}, \hat{\sigma}^2_{n,EL}) \xrightarrow{\text{a.s.}} (m, \sigma^2)$.

$$\lim_{n\to\infty} \hat{m}_{n,EL} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} w_i X_i - \hat{\lambda}_n}{\sum_{i=1}^{n} w_i X_{i-1}} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - \hat{\lambda}_n}{\sum_{i=1}^{n} X_{i-1}} = m \text{ a.s.}$$

$$\begin{aligned}
\lim_{n\to\infty} \hat{\sigma}^2_{n,EL} &= \lim_{n\to\infty} \frac{\sum_{i=1}^{n} w_i \hat{U}_i^2 - \hat{b}_n^2}{\sum_{i=1}^{n} w_i X_{i-1}} \\
&= \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \hat{U}_i^2 - \hat{b}_n^2}{\sum_{i=1}^{n} X_{i-1}} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{U}_i^2 - \hat{b}_n^2}{X_{i-1}} \\
&= \sigma^2 \text{ a.s.}
\end{aligned}$$

ELBPI–Main Technics for The Proof of Non-Critical BPI

- Supercritical BPI:
  - Toeplitz Lemma.
  - $\sum_{i=1}^{n} E\left[\frac{U_i}{\sqrt{N_n}}\frac{V_i}{\sqrt{N_n}}\middle|\mathcal{F}_{i-1}\right] \le \frac{c}{N_n^{\delta}}$ for some for some $0 < \delta < \frac{1}{2}$.
  - $\sqrt{n}\bar{U}_n \xrightarrow{\text{d}} N(0,\sigma^2)$, $\sqrt{n}\bar{V}_n \xrightarrow{\text{d}} N(0,2\sigma^4)$.
  - Martingale Central Limit Theorem.
- Subcritical BPI:
  - $\{X_n\}$ is stationary and ergodic.
  - Ergodic Theorem: Under finite moment conditions,
    $\frac{1}{n}\sum_{i=1}^{n} X_{i-1}^{\alpha} \to EX^{\alpha}$ a.s.
    $\frac{1}{n}\sum_{i=1}^{n} X_{i-1}^{\alpha}U_i^2 \to E\left[X^{\alpha}(\sigma^2 X + b^2)\right]$ a.s. for $\alpha = 0, 1, 2,$
  - $\sqrt{n}\bar{U}_n \xrightarrow{\text{d}} N(0,\sigma^2)$, $\sqrt{n}\bar{V}_n \xrightarrow{\text{d}} N(0,2\sigma^4)$.
  - Martingale Central Limit Theorem.

ELBPI-Proof Sketches for Critical BPI

- If $m = 1$ and $2\lambda > \sigma^2$:
  - $A_n \xrightarrow{\text{d}} D_1^2 + D_2^2$.
  - $B_n \xrightarrow{\text{p}} 0$,
  - $R_n \xrightarrow{\text{p}} 0$.
- If $m = 1$ and $2\lambda \leq \sigma^2$:
  - $A_n \xrightarrow{\text{p}} 0$, $B_n \xrightarrow{\text{p}} 0$, $R_n \xrightarrow{\text{p}} 0$.

ELBPI-Some Important Convergence Rates for The Proof of Critical BPI

- $$\frac{1}{N_n}\sum_{i=1}^{n} U_i V_i = \begin{cases} o_p(\sqrt{n}), & \text{if } \tau > 1; \\ o_p(n^{\beta+\frac{1}{2}}), & \text{if } \beta > 0 \text{ and } \tau = 1; \\ o_p(n^{\frac{3}{2}-\tau}), & \text{if } \tau < 1; \end{cases}$$

- $$\frac{\sum_{i=1}^{n} U_i^3}{N_n} = \begin{cases} O_p(1) + o_p(n^{\frac{1}{2}}) = o_p(\sqrt{n}), & \text{if } \tau > 1; \\ O_p(1) + o_p(n^{\beta+\frac{1}{2}}) = o_p(n^{\beta+\frac{1}{2}}), & \text{if } \beta > 0 \text{ and } \tau = 1; \\ O_p(1) + o_p(n^{\frac{3}{2}-\tau}) = o_p(n^{\frac{3}{2}-\tau}), & \text{if } \tau < 1; \end{cases}$$

- $$\frac{\sum_{i=1}^{n} U_i V_i^2}{N_n} + \frac{\sum_{i=1}^{n} U_i^3}{N_n} = \begin{cases} O_p(n), & \text{if } \tau > 1; \\ O_p(n), & \text{if } 0 < \beta \leq \frac{1}{2} \text{ and } \tau = 1 \\ o_p(n^{\beta+\frac{1}{2}}), & \text{if } \beta > \frac{1}{2} \text{ and } \tau = 1; \\ O_p(n), & \text{if } \frac{1}{2} \leq \tau < 1 \\ o_p(n^{\frac{3}{2}}), & \text{if } \ 0 < \tau < \frac{1}{2}; \end{cases}$$

ELBPI-Main Technics for The Proof of Critical BPI

- $D^+[0,\infty)$: the space of nonnegative functions which are right continuous and having left limits defined on $[0,\infty)$.
- Let $Y_n(t) = \frac{X_{[nt]}}{n}$, then $Y_n(t) \xrightarrow{\text{d}} Y(t)$.
- Y is a diffusion with generator $Af(x) = \frac{1}{2}\sigma^2 x f''(x) + \lambda f'(x)$, $f \in C_c^\infty[0,\infty)$, and $Y(0) = 0$.
- When $m = 1$ and $EX^k \leq \infty$,
  - $\frac{1}{n^{k+1}} \sum_{i=1}^{n} X_i^k \xrightarrow{\text{d}} \int_0^1 Y^k(t)dt$ for $k \geq 1$.
  - The joint convergence

    $(T_1, T_2, \ldots, T_n) \xrightarrow{\text{d}} \left(Y(1), \int_0^1 Y(t)dt, \ldots, \int_0^1 Y^k(t)dt \right)$,

    where $T_1 = \frac{X_n}{n}$, $T_2 = \frac{1}{n^2} \sum_{i=1}^{n} X_i$, $T_n = \frac{1}{n^{k+1}} \sum_{i=1}^{n} X_i^k$.

Numerical Methods for Empirical Likelihood – *i.i.d.* Data

- Point estimation: Newton Iteration(Owen(1988, 1990)).
- Point estimation: Newton Iteration with Backtracking (Owen(2013)).
- Interval estimation: Multivariate Newton Iteration(Hall and La Scala(1990)).

Numerical Methods for Empirical Likelihood – BPI

- Recall that

$$w_i = \frac{1}{n} \frac{1}{1 + \boldsymbol{\nu}_2' \boldsymbol{g}(X_i, \theta)}$$

- The Lagrangian multiplier $\boldsymbol{\nu}_2$ satisfy

$$\frac{1}{N_n} \sum_{i=1}^{n} \frac{\boldsymbol{g}(X_i; \theta)}{1 + \boldsymbol{\nu}_2' \, \boldsymbol{g}(X_i, \theta)} = 0, \text{ where } N_n = \sum_{i=1}^{n} X_{i-1}.$$

- Objective function

$$\mathcal{R}(\theta) = \max \prod_{i=1}^{n} \left(1 + \boldsymbol{\nu}_2' \boldsymbol{g}(X_i, \theta)\right)^{-1}$$

subject to $w_i \geq 0, \; \sum_{i=1}^{n} w_i = 1, \; \sum_{i=1}^{n} w_i \boldsymbol{g}(X_i, \theta) = 0.$

Numerical Methods for Empirical Likelihood–BPI (contd.)

- Observed Immigration.
  - Because of the independence of the offspring distribution and immigration distribution, $\hat{\lambda}_{n,EL}$ and $\hat{b}^2_{n,EL}$ can be estimated based on the observed $i.i.d.$ $\{I_n\}$ for $n \geq 1$.
- Unobserved Immigration.
  - If $\{I_n\}$ for $n \geq 1$ is not observed, $\hat{\lambda}_{n,EL}$ and $\hat{b}^2_{n,EL}$ can only be estimated based on the observed $\{X_n\}$.

Point Estimation–Newton Iteration with Backtracking for Offspring Parameters

$$\min f(\boldsymbol{\nu}_2) = -\sum \log(1 + \boldsymbol{\nu}_2' \boldsymbol{g}(X_i, \theta))$$

- Given a starting point $\boldsymbol{\nu}_2 = \mathbf{0}$ and $\epsilon = 10^{-4}$.
- Calculate the gradient $\nabla(f(\boldsymbol{\nu}_2))$ and Hessian $\nabla^2(f(\boldsymbol{\nu}_2))$.
- Repeat
  1. Newton step $\Delta(\boldsymbol{\nu}_2) = -\left(\nabla^2 f(\boldsymbol{\nu}_2)\right)^{-1} \nabla f(\boldsymbol{\nu}_2)$.
  2. Stopping Criteria: stop if $\left| \frac{\hat{m}_{n,EL}^{(k)}}{\hat{m}_{n,EL}^{(k-1)}} - 1 \right| < \epsilon$.
  3. Line search: Choose step size t by backtracking line search.
  4. Update: $\boldsymbol{\nu}_2 \leftrightarrow \boldsymbol{\nu}_2 + t\Delta(\boldsymbol{\nu}_2)$, $t \leftrightarrow \beta t$ starts with $t = 1$ while $f(\boldsymbol{\nu}_2 + t\Delta(\boldsymbol{\nu}_2)) \geq f(\boldsymbol{\nu}_2) + \alpha t \left(\nabla f(\boldsymbol{\nu}_2)\right)' \Delta(\boldsymbol{\nu}_2)$.
- $\alpha \in (0, 0.5), \beta \in (0, 1)$.
- $\hat{m}_{EL}^{(1)}, \hat{\sigma}_{EL}^{2(1)}$.

Point Estimation–Multivariate Newton Iteration with Backtracking for Immigration Parameters

- $\hat{\boldsymbol{g}}(X_i,\theta) = (X_i - \hat{m}_{EL}^{(1)}X_{i-1} - \hat{\lambda}, \hat{U_i}^2 - \hat{\sigma}_{EL}^{2(1)}X_{i-1} - \hat{b}^2)$
- Three equations to solve $\nu_{21}, \nu_{22}, \lambda$ and b^2.

$$h_1(\boldsymbol{\nu_2}, \lambda, b^2) = \sum_{i=1}^{n} \frac{\hat{U}_i}{1 + \boldsymbol{\nu}_2'\hat{\boldsymbol{g}}(X_i,\theta)} = 0$$

$$h_2(\boldsymbol{\nu_2}, \lambda, b^2) = \sum_{i=1}^{n} \frac{\hat{V}_i}{1 + \boldsymbol{\nu}_2'\hat{\boldsymbol{g}}(X_i,\theta)} = 0$$

$$h_3(\boldsymbol{\nu_2}, \lambda, b^2) = \sum_{i=1}^{n} log(1 + \boldsymbol{\nu}_2'\hat{\boldsymbol{g}}(X_i,\theta)) = \frac{1}{2}c$$

- Reparameterize $\nu_{21} = \nu sin(u), \nu_{22} = \nu cos(u)$, where $u \in [0, 2\pi)$ to solve ν, λ and b^2.
- c is the critical value from either a χ^2 distribution or a distribution of the functional of Feller-diffusion.

Point Estimation–Multivariate Newton Iteration with Backtracking for Immigration Parameters (contd.)

Branching Processes with Unobserved Immigration.

$$\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), h_3(\boldsymbol{x})) = 0$$

$$\boldsymbol{x} = (\nu, \lambda, b^2).$$

- Calculate the Jacobian matrix $\boldsymbol{J} = (J_{ik})$, where $J_{ik} = \dfrac{\partial h_i}{\partial x_k}$ for $i = 1, 2, 3$ and $k = 1, 2, 3$.
- Repeat
  1. Compute the Newton step $\Delta(\boldsymbol{x}) = -\boldsymbol{J}^{-1}\boldsymbol{h}(\boldsymbol{x})$.
  2. Stopping Criteria: $\left|\frac{\hat{m}_{n,EL}^{(k)}}{\hat{m}_{n,EL}^{(k-1)}} - 1\right| < \epsilon.$
  3. Update: $t \leftrightarrow \frac{t}{2}$ starts with $t = 1$. while $\|\boldsymbol{h}(\boldsymbol{x} + t\Delta(\boldsymbol{x}))\| \geq (1 - \frac{t}{2})\|\boldsymbol{h}(\boldsymbol{x})\|$
- $\hat{\lambda}_{EL}^{(1)}, \hat{b}_{EL}^{2(1)}$.

Interval Estimation–Multivariate Newton Iteration with Backtracking for Offspring Parameters

- $\boldsymbol{h}(\boldsymbol{x}) = (h_1(x), h_2(x), h_3(x)) = 0$ to find two sets of $\boldsymbol{x}_1 = (\nu^{(1)}, m_1, \sigma_1^2)$ and $\boldsymbol{x}_2 = (\nu^{(2)}, m_2, \sigma_2^2)$.
- $$h_1(\nu, m, \sigma^2) = \sum_{i=1}^{n} \frac{\hat{U}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i}.$$
  $$h_2(\nu, m, \sigma^2) = \sum_{i=1}^{n} \frac{\hat{V}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i}.$$
  $$h_3(\nu, m, \sigma^2) = \sum_{i=1}^{n} \log\{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i\} - \tfrac{1}{2}c.$$
- $\hat{U}_i = X_i - \hat{m}_n X_{i-1} - \hat{\lambda}_n = \hat{U}_i - \hat{\sigma}_n^2 X_{i-1} - \hat{b}_n^2$.
- $u \in [0, 2\pi)$.
- Take $(\hat{m}_{L,EL}, \hat{\sigma}^2_{L,EL})$ and $(\hat{m}_{U,EL}, \hat{\sigma}^2_{U,EL})$ to be the smallest and largest respectively of (m_1, σ_1^2) and (m_2, σ_2^2).

Interval Estimation–Multivariate Newton Iteration with Backtracking for Immigration Parameters

- $\boldsymbol{h}(\boldsymbol{x}) = (h_1(x), h_2(x), h_3(x)) = 0$ to find two sets of $\boldsymbol{x}_1 = (\nu^{(1)}, \lambda_1, b_1^2)$ and $\boldsymbol{x}_2 = (\nu^{(2)}, \lambda_2, b_2^2)$.
- $$h_1(\nu, \lambda, b^2) = \sum_{i=1}^{n} \frac{\hat{U}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i}.$$
  $$h_2(\nu, \lambda, b^2) = \sum_{i=1}^{n} \frac{\hat{V}_i}{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i}.$$
  $$h_3(\nu, \lambda, b^2) = \sum_{i=1}^{n} \log\{1 + \nu sin(u)\hat{U}_i + \nu cos(u)\hat{V}_i\} - \frac{1}{2}c.$$
- $\hat{U}_i = X_i - \hat{m}_{n,EL}X_{i-1} - \hat{\lambda}_n = \hat{U}_i - \hat{\sigma}^2_{n,EL}X_{i-1} - \hat{b}^2_n$.
- $u \in [0, 2\pi)$.
- $(\hat{\lambda}_{L,EL}, \hat{b}^2_{L,EL})$ and $(\hat{\lambda}_{U,EL}, \hat{b}^2_{U,EL})$ to be the smallest and largest respectively of (λ_1, b_1^2) and (λ_2, b_2^2).

Monte Carlo Simulations for Critical BPI

Simulate the critical value from a distribution of a functional of a Feller diffusion.

- Recall that $-2log\mathcal{R}_n(\theta) \xrightarrow{\text{d}} \frac{(Y(1)-\lambda)^2}{\sigma^2\int_0^1 Y(t)dt} + \frac{(\int_0^1 Y(t)dB(t))^2}{\int_0^1 Y^2(t)dt}$
- Recall that $\frac{\sum x_i^k}{n^{k+1}} \xrightarrow{\text{d}} \int_0^1 Y^k(t)$ and $Y(1) = \lim \frac{X_n}{n}$.
- Riemann sums:
  $\int_0^1 Y(t)dt = \lim \frac{1}{n}\sum_{i=1}^N X_{[nt_i]}(t_{i+1}-t_i)$
  $\int_0^1 Y^2(t)dt = \lim \frac{1}{n^2}\sum_{i=1}^N X_{[nt_i]}^2(t_{i+1}-t_i)$
  $\int_0^1 Y(t)dB(t) = \lim \frac{1}{n}\sum_{i=1}^N X_{[nt_i]}(B(t_{i+1})-B(t_i))$
- N: number of partition.
- t_i: random numbers generated from $[0, 1]$.
- $B(t_{i+1}) - B(t_i) \xrightarrow{\text{d}} N(0, t_{i+1}-t_i)$.

Simulations

We conducted 2000 simulation studies of 25 generations for Supercritical BPI, Critical BPI and Subcritical BPI to examine the performances of the empirical likelihood inference for BPI.

- Supercritical BPI
    - True offspring distribution: Poi(1.6)
- Critical BPI
    - True offspring distribution: Poi(1)
    - True immigration distribution: Poi(1.6)
- Subcritical BPI
    - True offspring distribution: Poi(0.6)
    - True immigration distribution: Poi(2)