Bias-Corrected Maximum Likelihood Estimation in Actuarial Science

Paul H. Johnson, Jr., PhD
Yongxue Qi
Yvonne C. Chueh, PhD, ASA

Joint work between the University of Illinois at Urbana-Champaign and Central Washington University

Introduction

- The lognormal, gamma, and Weibull distributions are typically used to model financial returns
    - Lognormal: Black & Scholes (1973), Antoniou et al. (2011)
    - Gamma: Queirós (2005), Stein et al. (2005)
    - Weibull: Sazuka (2007), Gerlach & Chen (2011)
- Maximum likelihood estimation is a common method for estimating unknown distributional parameters
- Maximum likelihood estimators (MLEs) have many well known, desirable properties that depend on having a large sample size (in particular, unbiasedness)
- MLEs based on small samples can be substantially biased

Introduction (Continued)

- Cox and Snell (1968) developed an $O(n^{-1})$ formula for MLE small sample bias for independent observations
- Cox and Snell (1968) formula was re-expressed by Cordeiro and Klein (1994) for non-independent observations (henceforth referred to as CSCK bias method)
- CSCK method provides bias-corrected MLEs (BMLEs)
- Only recently have computers allowed for feasible calculation of BMLEs [Giles and Feng (2009a,b)]

Purpose

- Develop a Mathematica 8.0 module that calculates the CSCK MLE bias for a probability distribution
- Determine analytic formulas for the CSCK MLE bias for parameters of distributions commonly employed in actuarial science to model financial returns: lognormal, gamma, and Weibull
- Conduct simulation analyses for the three distributions to test whether BMLEs result in more accurate loss reserves than MLEs for an illustrative equity-linked insurance contract in small samples

Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, ..., \theta_p)'$
- Define joint cumulants based on the total loglikehood function $l(\theta)$ with n observations for i, j, l = 1, 2, ..., p:

$$\kappa_{ij} = E[\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}]$$

$$\kappa_{ijl} = E[\frac{\partial^3 l}{\partial\theta_i \partial\theta_j \partial\theta_l}]$$

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial\kappa_{ij}}{\partial\theta_l}$
- Total Fisher information Matrix of order p for θ is $K = \{-\kappa_{ij}\}$; inverse is $K^{-1} = \{-\kappa^{ij}\}$

Method

- Consider a distribution with *p* parameters: $\theta = (\theta_1, \theta_2, ..., \theta_p)'$
- Define joint cumulants based on the total loglikehood function $l(\theta)$ with *n* observations for *i*, *j*, *l* = 1, 2, ..., *p*:

$$\kappa_{ij} = E[\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}]$$

$$\kappa_{ijl} = E[\frac{\partial^3 l}{\partial\theta_i \partial\theta_j \partial\theta_l}]$$

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order *p* for θ is $K = \{-\kappa_{ij}\}$; inverse is $K^{-1} = \{-\kappa^{ij}\}$

Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, ..., \theta_p)'$
- Define joint cumulants based on the total loglikehood function $l(\theta)$ with n observations for i, j, l = 1, 2, ..., p:

$$\kappa_{ij} = E[\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}]$$

$$\kappa_{ijl} = E[\frac{\partial^3 l}{\partial\theta_i \partial\theta_j \partial\theta_l}]$$

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order p for θ is $K = \{-\kappa_{ij}\}$; inverse is $K^{-1} = \{-\kappa^{ij}\}$

Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, ..., \theta_p)'$
- Define joint cumulants based on the total loglikehood function $l(\theta)$ with n observations for $i, j, l = 1, 2, ..., p$:

$$\kappa_{ij} = E[\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}]$$
$$\kappa_{ijl} = E[\frac{\partial^3 l}{\partial\theta_i\partial\theta_j\partial\theta_l}]$$

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial\kappa_{ij}}{\partial\theta_l}$
- Total Fisher information Matrix of order p for θ is $K = \{-\kappa_{ij}\}$; inverse is $K^{-1} = \{-\kappa^{ij}\}$

Method (Continued)

- CSCK showed if all cumulants are assumed to be $O(n)$:

  $b_s = E[\hat{\theta}_s - \theta_s] = \sum_{i=1}^{p} \kappa^{si} \sum_{j,l=1}^{p} [\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}]k^{jl} + O(n^{-1})$

- Bias vector = $b = E[\hat{\theta} - \theta] = K^{-1}Avec[K^{-1}] + O(n^{-2})$

  where $A = \{A^{(1)}|A^{(2)}|...|A^{(p)}\}$ and $A^{(1)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for $l = 1, 2, ..., p$

- The BMLE vector ($\tilde{\theta}$) is the difference between the MLE vector and the MLE bias vector evaluated at the MLEs:
  $\tilde{\theta} = \hat{\theta} - \hat{b}$

Method (Continued)

- CSCK showed if all cumulants are assumed to be $O(n)$:

  $b_s = E[\hat{\theta}_s - \theta_s] = \sum_{i=1}^{p} \kappa^{si} \sum_{j,l=1}^{p} [\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}]k^{jl} + O(n^{-1})$
- Bias vector = $b = E[\hat{\theta} - \theta] = K^{-1}Avec[K^{-1}] + O(n^{-2})$

  where $A = \{A^{(1)}|A^{(2)}|...|A^{(p)}\}$ and $A^{(1)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for $l = 1, 2, ..., p$
- The BMLE vector ($\tilde{\theta}$) is the difference between the MLE vector and the MLE bias vector evaluated at the MLEs:
  $\tilde{\theta} = \hat{\theta} - \hat{b}$

Method (Continued)

- CSCK showed if all cumulants are assumed to be $O(n)$:

  $b_s = E[\hat{\theta}_s - \theta_s] = \sum_{i=1}^{p} \kappa^{si} \sum_{j,l=1}^{p} [\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}]k^{jl} + O(n^{-1})$
- Bias vector = $b = E[\hat{\theta} - \theta] = K^{-1}Avec[K^{-1}] + O(n^{-2})$

  where $A = \{A^{(1)}|A^{(2)}|...|A^{(p)}\}$ and $A^{(1)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for l = 1, 2, ..., p
- The BMLE vector ($\tilde{\theta}$) is the difference between the MLE vector and the MLE bias vector evaluated at the MLEs:
  $\tilde{\theta} = \hat{\theta} - \hat{b}$

Mathematica 8.0 Module

```
b[f_, p_] := Module[{l, Gradient, Hessian, ThirdPartialDer, ExpectHessian, ExpectThirdPartialDer,
DerivativeExpectHessian, aijk, Amatrix, Kinv, vecKinv, BIAS, Expect},

Expect[x_] := Integrate[x*f, {yi, 0, ∞}, Assumptions -> {θ1 ε Reals, θ2 ε Reals, θ3 ε Reals, θ1 > 0,
θ2 > 0, θ3 > 0}];

SuperLog[On];

l = Log[Πn i=1 f];
Gradient = D[l, {p}];
Hessian = D[l, {p, 2}];
ThirdPartialDer = D[l, {p, 3}];
ExpectHessian = Map[Expect[#] &, Hessian];
ExpectThirdPartialDer = Map[Expect[#] &, ThirdPartialDer];
DerivativeExpectHessian = D[ExpectHessian, {p}];
aijk = DerivativeExpectHessian - ExpectThirdPartialDer/2;
Amatrix = Apply[Join, aijk~Join~{2}];
Kinv = Inverse[-ExpectHessian]; vecKinv = Flatten[Transpose[Kinv]];
BIAS = Simplify[Kinv.Amatrix.vecKinv];

SuperLog[Off]; BIAS]
```

CSCK MLE Bias: (θ_1 = location/shape, θ_2 = scale)

Gamma CSCK MLE Bias =

$$\left\{\frac{-2+\theta1\text{PolyGamma}[1,\theta1]-\theta1^2\text{PolyGamma}[2,\theta1]}{2n(-1+\theta1\text{PolyGamma}[1,\theta1])^2},\frac{\theta2(\text{PolyGamma}[1,\theta1]+\theta1\text{PolyGamma}[2,\theta1])}{2n(-1+\theta1\text{PolyGamma}[1,\theta1])^2}\right\}$$

Lognormal CSCK MLE Bias =

$$\left\{0,-\frac{3\theta2}{4n}\right\}$$

Weibull CSCK MLE Bias =

$$\left\{\frac{18\theta1(\pi^2-2\text{Zeta}[3])}{n\pi^4},\right.$$
$$\left.-\frac{\theta2(\pi^4(-1+2\theta1)-6\pi^2(1+\text{EulerGamma}^2+5\theta1-2\text{EulerGamma}(1+2\theta1))-72(-1+\text{EulerGamma})\theta1\text{Zeta}[3])}{2n\pi^4\theta1^2}\right\}$$

Illustrative Insurance Contract

- Consider a 20-period equity-linked insurance contract on a policyholder (ph) age 45 (Dickson et al 2009, Chapter 12):
- Each premium of 3,000 (B.O.P.) is portioned into allocated premiums paid into a ph fund (2,850 for $t = 0$; 2,970 otherwise) and unallocated premiums paid into an insurer fund
- Initial expenses are 10% of the first premium and 0.5% of all subsequent premiums (B.O.P.)
- A management charge of 0.80% of the ph fund is transferred to the insurer fund (E.O.P.)
- Periodic rate of return on insurer fund is 6%
- Rate of return for $[t - 1, t)$ on ph fund is R_t

Illustrative Insurance Contract (Continued)

- There are two decrements: death and lapse
  - The ph's periodic death rate is 0.01
  - For t = 1, 2, ..., 5, periodic lapse rate (between 0.01 and 0.05) is negatively correlated with R_t; for t = 6, ..., 20, periodic lapse rate is 0.01
- There are three benefits (E.O.P.):
  - Death benefit = 110% of the value of ph fund at the end of the period of death: 100% from ph fund and 10% from insurer fund
  - Lapse benefit = value of ph fund at end of the period of lapse
  - Guaranteed Minimum Maturity Benefit = the greater of the ph fund value and the total premiums paid (without interest)
- Profit$_t$ = Unallocated Premium$_{t-1}$ - Expenses$_{t-1}$ + Interest$_t$@6% + Management Charge$_t$ - Expected Death Benefit$_t$ - I[t = 20](1 - 0.01)max[60,000 - ph fund at time 20, 0]

CTE Loss Reserve

- Let L denote the expected present value of the future loss at issue, such that at a risk discount rate of 14%:

$$L = -\sum_{t=1}^{20} \frac{{}_{t-1}p_{45}^{(\tau)}}{1.14^t} Profit_t$$

- 100α% conditional tail expectation (CTE) loss reserve at policy issue (Klugman et al. 2008):

$${}_0V^{\alpha} = E[L|L > Q_{\alpha}]$$

- Calculation requires identifying a distribution for $(1 + R_t)$, and simulating many realizations of L, from a small amount of periodic financial returns data

Loss Reserve Simulations: True Reserve

- Assume three "true" distributions of $(1 + R_t)$:
  - Lognormal distribution with $\theta_1 = 0.05$ and $\theta_2 = 0.30$
  - Gamma distribution with $\theta_1 = 10.0$ and $\theta_2 = 0.11$
  - Weibull distribution with $\theta_1 = 4.50$ and $\theta_2 = 1.20$
- For each of the above true distributions:
  - Simulate 1,000 data sets $\{1 + R_t\}_{t=1}^{20}$
  - For each of these data sets, simulate 5,000 values of L
  - Average of losses above the empirical 95-th quantile of the distribution of L provides a 95% CTE loss reserve value
  - Average 1,000 reserve values to get the true 95% CTE loss reserve

Loss Reserve Simulations: MLE and BMLE Reserves

- For each of the three true distributions:
  - Simulate 1,000 data sets of $\{1+R_t\}_{t=1}^{20}$ to give us “historical” rate of return data
  - Obtain 1,000 MLEs and the corresponding 1,000 BMLEs
  - Simulate 2,000 data sets $\{1+R_t\}_{t=1}^{20}$ based on above MLEs and BMLEs
  - For each MLE and BMLE data set, simulate 5,000 values of L to obtain 1,000 MLE and 1,000 BMLE 95% CTE loss reserves
  - Note the percentage of 1,000 runs that the BMLE loss reserve difference is smaller than the MLE loss reserve difference, and the average BMLE and MLE loss reserve differences

Simulation Results

Table: **Illustrative Insurance Simulation Results Summary.**

Probability Distribution	**True Reserve**	**% Runs BMLE Reserve Closer to True Reserve**	**Mean MLE Reserve Diff.**	**Mean BMLE Reserve Diff.**
Lognormal ($\theta_1 = 0.05$, $\theta_2 = 0.30$)	873.95	53	-248.28	-204.29
Gamma ($\theta_1 = 10.0$, $\theta_2 = 0.11$)	1,029.63	32	-244.84	527.81
Weibull ($\theta_1 = 4.50$, $\theta_2 = 1.20$)	821.49	50	-270.53	-125.78

Summary/Next Steps

- Developed a Mathematica 8.0 module which calculates the CSCK MLE bias for important distributions in actuarial science
- For illustrative insurance contract and small sample data, demonstrated that 95% CTE loss reserves were more accurately calculated using BMLEs instead of MLEs for lognormal and Weibull, but not gamma, ph fund rate of return distributions
- In future, relax iid assumption for $(1 + R_t)$ and consider a more sophisticated rate of return model, such as in Hardy 2001
- In future, consider alternative methods for minimizing small sample MLE bias, such as a "preventive" approach [Firth (1993)]
- In future, integrate our findings with the Actuarial Model Outcome Optimal Fit (AMOOF) project

References

- Actuarial Model Outcome Optimat Fit (AMOOF). 2008. `http://www.cwu.edu/~chueh/AMOOF.html`
- Antoniou, I., Vi.V. Ivanov, Va.V. Ivanov, and P.V. Zrelov. 2011. “On the Log- normal Distribution of Stock Market Data.” *Physica A: Statistical Mechanics and its Applications*, 231(3-4): 617-638
- Black, F. and M. Scholes. 1973. “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy*, 81(3): 637-654
- Cordeiro, G.M. and R. Klein. 1994. “Bias Correction in ARMA Models.” *Statistics and Probability Letters*, 19: 169-176
- Cox, D.R. and E.J. Snell. 1968. “A General Definition of Residuals.” *Journal of the Royal Statistical Society*, B, 30: 248-275

References (Continued)

- Dickson, D.C.M., M.R. Hardy, and H.R. Waters. 2009. "Actuarial Mathematics for Life Contingent Risks," Cambridge: Cambridge University Press
- Firth, D. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika*, 80: 27-38
- Gerlach, R., and Q. Chen. 2011. "The Two-sided Weibull Distribution and Forecasting Financial Tail Risk." *OME Working Paper 01/2011*, 1-36
- Giles, David E. 2009. `http://web.uvic.ca/~dgiles/downloads/working_papers/wpapers.html`. a corresponds to "Bias of the Maximum Likelihood Estimators of the Two-Parameter Gamma Distribution Revisited," and b corresponds to "Bias-Corrected Maximum Likelihood Estimation of the Parameters of the Generalized Pareto Distribution"

References (Continued)

- Klugman, S.A., H.H. Panjer, and G.E. Willmot. 2008. "Loss Models: From Data to Decisions, Third Edition." Hoboken, NJ: Wiley
- Queirós, S.M.D. 2005. "On the Emergence of a Generalised Gamma Distribution. Application to Traded Volume in Financial Markets." *Europhysics Letters*, 71(3): 339-345
- Sazuka, N. 2007. "On the Gap between an Empirical Distribution and an Exponential Distribution of Waiting Times for Price Changes in a Financial Market." *Physica A: Statistical Mechanics and its Applications*, 376: 500-506
- Stein, H.J., P.P. Carr, and H. Apollo. 2005. "Time for a Change: The Variance Gamma Model and Option Pricing." *Social Science Research Network: http://ssrn.com/abstract=956625*, 1-11
- Wolfram Research, Inc. 2010. "Mathematica Edition: Version 8.0." Wolfram Research, Inc: Champaign, IL