

Bias-Corrected Maximum Likelihood Estimation in Actuarial Science

Paul H. Johnson, Jr., PhD

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



Outline

- 1 Introduction and Purpose
- 2 CSCK MLE Bias Calculations
- 3 Simulation Analyses
- 4 Concluding Remarks



Outline

- 1 Introduction and Purpose
- 2 CSCK MLE Bias Calculations
- 3 Simulation Analyses
- 4 Concluding Remarks



Outline

- 1 Introduction and Purpose
- 2 CSCK MLE Bias Calculations
- 3 Simulation Analyses
- 4 Concluding Remarks



Outline

- 1 Introduction and Purpose
- 2 CSCK MLE Bias Calculations
- 3 Simulation Analyses
- 4 Concluding Remarks



Introduction

- Maximum Likelihood Estimators (MLEs) have many well known, desirable properties
- Many of these properties depend on having a large sample size
- In particular, MLEs may be biased if the sample size is small ($n \leq 100$)
- This is a real issue in actuarial science, as parameters of probability distributions for financial returns are often estimated with maximum likelihood estimation using data over a small number of periods



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)
- Only recently have computers allowed for feasible calculation of bias-corrected MLEs (BMLEs). (Giles and Feng (2009a,b) for the 2-parameter gamma and generalized Pareto distributions)



Purpose

- Develop a Mathematica 8.0 module that calculates the CSCK MLE bias for any probability distribution
- Use the Mathematica module to determine analytic formulas for the CSCK MLE bias for parameters of distributions commonly employed in actuarial science to model financial returns
- Conduct simulation analyses for the Weibull distribution to illustrate the higher quality of BMLEs relative to MLEs in small samples, particularly in the valuation of an illustrative equity-linked insurance contract



Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$
- $l(\theta)$ is the total loglikelihood function, based on a sample of n observations
- $l(\theta)$ is regular with respect to all derivatives up to and including the third order
- Define joint cumulants of the total loglikelihood function derivatives for $i, j, l = 1, 2, \dots, p$:

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

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

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



Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$
- $l(\theta)$ is the total loglikelihood function, based on a sample of n observations
- $l(\theta)$ is regular with respect to all derivatives up to and including the third order
- Define joint cumulants of the total loglikelihood function derivatives for $i, j, l = 1, 2, \dots, p$:

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

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

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



Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$
- $l(\theta)$ is the total loglikelihood function, based on a sample of n observations
- $l(\theta)$ is regular with respect to all derivatives up to and including the third order
- Define joint cumulants of the total loglikelihood function derivatives for $i, j, l = 1, 2, \dots, p$:

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

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

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



Method

- Consider a distribution with p parameters: $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$
- $l(\theta)$ is the total loglikelihood function, based on a sample of n observations
- $l(\theta)$ is regular with respect to all derivatives up to and including the third order
- Define joint cumulants of the total loglikelihood function derivatives for $i, j, l = 1, 2, \dots, p$:

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

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

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



Method (Continued)

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order p for $\theta = K = \{-\kappa_{ij}\}$
- Inverse of Total Fisher information Matrix of order p for $\theta = K^{-1} = \{-\kappa^{ij}\}$
- Cox and Snell showed that for independent observations, the bias of the MLE of θ_s , $\hat{\theta}_s$, for $s = 1, 2, \dots, p$ is:

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



Method (Continued)

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order p for $\theta = K = \{-\kappa_{ij}\}$
- Inverse of Total Fisher information Matrix of order p for $\theta = K^{-1} = \{-\kappa^{ij}\}$
- Cox and Snell showed that for independent observations, the bias of the MLE of θ_s , $\hat{\theta}_s$, for $s = 1, 2, \dots, p$ is:

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



Method (Continued)

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order p for $\theta = K = \{-\kappa_{ij}\}$
- Inverse of Total Fisher information Matrix of order p for $\theta = K^{-1} = \{-\kappa^{ij}\}$
- Cox and Snell showed that for independent observations, the bias of the MLE of θ_s , $\hat{\theta}_s$, for $s = 1, 2, \dots, p$ is:

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



Method (Continued)

- Cumulant derivative: $\kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l}$
- Total Fisher information Matrix of order p for $\theta = K = \{-\kappa_{ij}\}$
- Inverse of Total Fisher information Matrix of order p for $\theta = K^{-1} = \{-\kappa^{ij}\}$
- Cox and Snell showed that for independent observations, the bias of the MLE of θ_s , $\hat{\theta}_s$, for $s = 1, 2, \dots, p$ is:

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



Method (Continued)

- Cordeiro and Klein later showed that the assumption of independence in the formula for b_s can be relaxed as long as 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 = $b = E[\hat{\theta} - \theta] = K^{-1} \text{Avec}[K^{-1}] + O(n^{-2})$
 where $A = \{A^{(1)} | A^{(2)} | \dots | A^{(p)}\}$ and $A^{(l)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for $l = 1, 2, \dots, p$
- 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)

- Cordeiro and Klein later showed that the assumption of independence in the formula for b_s can be relaxed as long as 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 = $b = E[\hat{\theta} - \theta] = K^{-1} \text{Avec}[K^{-1}] + O(n^{-2})$
 where $A = \{A^{(1)} | A^{(2)} | \dots | A^{(p)}\}$ and $A^{(l)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for $l = 1, 2, \dots, p$

- 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)

- Cordeiro and Klein later showed that the assumption of independence in the formula for b_s can be relaxed as long as 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 = $b = E[\hat{\theta} - \theta] = K^{-1} \text{Avec}[K^{-1}] + O(n^{-2})$
 where $A = \{A^{(1)}|A^{(2)}|\dots|A^{(p)}\}$ and $A^{(l)} = \{\kappa_{ij}^{(l)} - 0.5\kappa_{ijl}\}$ for $l = 1, 2, \dots, p$
- 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

- We developed a module in Mathematica 8.0 that computes the MLE bias for a probability distribution
- Our module uses the Symbolic Maximum Likelihood Estimation package (SMLE.m) provided in Rose and Smith (2000), that allows for an analytic representation of the total loglikelihood function



Mathematica 8.0 Module (Continued)

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

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

```
SuperLog[On];
```

```
l = Log[∏i=1nf];
```

```
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]
```



Distribution pdfs (Parameters are θ_1, θ_2)

- Using the module, we determined analytic expressions for the MLE bias for three probability distributions: 2- parameter gamma (considered by David E. Giles and co-authors), lognormal, Weibull
- Gamma ($\theta_1 = \alpha, \theta_2 = \theta$):

$$f(y) = \frac{1}{\theta_2^{\theta_1} \Gamma(\theta_1)} y^{\theta_1-1} \exp[-y/\theta_2] \text{ for } y, \theta_1, \theta_2 > 0$$

- Lognormal ($\theta_1 = \mu, \theta_2 = \sigma$):

$$f(y) = \frac{1}{\theta_2 y \sqrt{2\pi}} \exp\left[-\frac{(\ln(y)-\theta_1)^2}{2\theta_2^2}\right], y, \theta_2 > 0$$

- Weibull ($\theta_1 = \tau, \theta_2 = \theta$):

$$f(y) = (1/y)\theta_1(y/\theta_2)^{\theta_1} \exp[-(y/\theta_2)^{\theta_1}] \text{ for } y, \theta_1, \theta_2 > 0$$



CSCK MLE Bias: (θ_1, θ_2)

Gamma CSCK MLE Bias =

$$\left\{ \frac{-2 + \theta_1 \text{PolyGamma}[1, \theta_1] - \theta_1^2 \text{PolyGamma}[2, \theta_1]}{2n(-1 + \theta_1 \text{PolyGamma}[1, \theta_1])^2}, \frac{\theta_2 (\text{PolyGamma}[1, \theta_1] + \theta_1 \text{PolyGamma}[2, \theta_1])}{2n(-1 + \theta_1 \text{PolyGamma}[1, \theta_1])^2} \right\}$$

Lognormal CSCK MLE Bias =

$$\left\{ 0, -\frac{3\theta_2}{4n} \right\}$$

Weibull CSCK MLE Bias =

$$\left\{ \frac{18\theta_1(\pi^2 - 2\text{Zeta}[3])}{n\pi^4}, \frac{\theta_2(\pi^4(-1 + 2\theta_1) - 6\pi^2(1 + \text{EulerGamma})^2 + 5\theta_1 - 2\text{EulerGamma}(1 + 2\theta_1)) - 72(-1 + \text{EulerGamma})\theta_1\text{Zeta}[3]}{2n\pi^4\theta_1^2} \right\}$$



Two Simulation Analyses

- (1) For the Weibull distribution, we measured the “quality” of the BMLEs by comparing the percent bias and MSE relative to the distribution’s true parameters vs the same quantities for the (unadjusted) MLEs
- (2) We simulated rates of return from the Weibull distribution using BMLEs to determine the impact on the valuation of an equity-linked life insurance contract



Simulation Analyses (1)

- We determined different pairs of parameter values the corresponded to the true distribution ($\tau = 2.0, \theta = 1.2$; $\tau = 3.5, \theta = 1.2$; $\tau = 5.0, \theta = 1.2$)
- We then estimated MLEs and BMLEs using the Mathematica 8.0 module for each pair using sample sizes of $n = 25, 50, 75, 100$
- 5000 replications were used per pair per sample size
- This is a similar simulation analysis to that employed by Giles and co-authors (2009a,b)



Simulation: Weibull with $\tau = 2.0$, $\theta = 1.2$

n	%Bias: $\hat{\tau}$	%Bias: $\tilde{\tau}$	%Bias: $\hat{\theta}$	%Bias: $\tilde{\theta}$
	[%MSE: $\hat{\tau}$]	[%MSE: $\tilde{\tau}$]	[%MSE: $\hat{\theta}$]	[%MSE: $\tilde{\theta}$]
25	6.37	0.50	0.00	0.19
	[3.50]	[2.76]	[1.08]	[1.09]
50	2.73	-0.10	-0.17	-0.08
	[1.44]	[1.29]	[0.56]	[0.56]
75	1.89	0.02	-0.11	-0.05
	[0.91]	[0.84]	[0.36]	[0.36]
100	1.49	0.09	-0.06	-0.02
	[0.67]	[0.63]	[0.28]	[0.28]



Simulation: Weibull with $\tau = 3.5$, $\theta = 1.2$

n	%Bias: $\hat{\tau}$	%Bias: $\tilde{\tau}$	%Bias: $\hat{\theta}$	%Bias: $\tilde{\theta}$
	[%MSE: $\hat{\tau}$]	[%MSE: $\tilde{\tau}$]	[%MSE: $\hat{\theta}$]	[%MSE: $\tilde{\theta}$]
25	5.90	0.06	-0.23	0.01
	[3.54]	[2.85]	[0.35]	[0.35]
50	2.92	0.09	-0.23	-0.11
	[1.50]	[1.34]	[0.18]	[0.18]
75	1.89	0.01	-0.08	-0.00
	[0.88]	[0.82]	[0.12]	[0.12]
100	1.45	0.05	-0.11	-0.05
	[0.67]	[0.63]	[0.09]	[0.09]



Simulation: Weibull with $\tau = 5.0$, $\theta = 1.2$

n	%Bias: $\hat{\tau}$	%Bias: $\tilde{\tau}$	%Bias: $\hat{\theta}$	%Bias: $\tilde{\theta}$
	[%MSE: $\hat{\tau}$]	[%MSE: $\tilde{\tau}$]	[%MSE: $\hat{\theta}$]	[%MSE: $\tilde{\theta}$]
25	5.82	-0.02	-0.23	-0.03
	[3.46]	[2.79]	[0.18]	[0.18]
50	2.67	-0.16	-0.08	0.02
	[1.43]	[1.28]	[0.09]	[0.09]
75	2.06	0.18	-0.09	-0.02
	[0.92]	[0.84]	[0.06]	[0.06]
100	1.28	-0.12	-0.05	0.00
	[0.67]	[0.63]	[0.05]	[0.05]



Simulation Analyses (2): Background

- Consider a 20-period equity-linked term insurance contract on a policyholder (ph) age 50. Following the approach in Dickson et al 2009, Chapter 12:
- The probability the ph dies in any period is 0.005
- There are two funds: a ph fund and an insurer fund
- The ph pays premiums of 6000 at the beginning of each period
- Each premium is portioned into allocated premiums (paid into ph fund) and unallocated premiums (paid into insurer fund). The allocated premium for period one is 5700; the allocated premium for other periods is 5940



Simulation Analyses (2) Background (Continued)

- Initial expenses are 10% of the first premium and 0.5% of all subsequent premiums, incurred at the beginning of the period
- At the end of each period, a management charge of 0.80% of the ph fund is transferred to the insurer fund
- The end-of-period death benefit is 110% of the value of the ph fund at the end of the period of death: 100% from ph fund and 10% from insurer fund
- If ph survives 20 periods, the ph receives the greater of the ph fund value and the total premiums paid (GMMB)



Simulation Analyses (2) Background (Continued)

- $\text{Profit}_t = \text{Unallocated Premium}_{t-1} - \text{Expenses}_{t-1} + \text{Interest}_t @ 6\% + \text{Management Charge}_t - \text{Expected Death Benefit}_t$
- The insurer fund earns interest of 6% per period
- The ph fund earns interest at R_t during the t -th period
- For Profit_{20} , there is an additional charge of $(1 - 0.005)\max[120,000 - \text{ph fund at time } 20, 0]$
- Management Charge, Expected Death Benefit, and GMMB depend on $\{R_t\}_{t=1}^{20}$



Simulation Analyses (2) Reserve

- Let L denote the expected present value of the future loss at issue, such that at a risk discount rate of 15%:

$$L = - \sum_{t=1}^{20} \frac{(1-0.005)^{t-1}}{1.15^t} Profit_t$$

- A common way of determining the insurer reserve at issue is to use the conditional tail expectation (CTE). If Q_α is the 100α -th quantile of the loss distribution L :

$$\text{CTE Reserve at time 0} = E[L|L > Q_\alpha]$$

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



Simulation Analyses (2): Strategy

- Assume the true distribution of $(1 + R_t)$ is Weibull($\tau = 5, \theta = 1.2$)
- Simulate 20 values of $(1 + R_t)$ to obtain past experience:
(1.26076, 1.30661, 0.859027, 1.29311, 1.15941, 1.40232,
1.05463, 1.15122, 1.08061, 0.831058, 0.63808, 1.18715,
0.976797, 0.842122, 1.15623, 1.02031, 0.832913, 1.30698,
1.54043, 1.54232)
- MLEs using data: $\hat{\tau} = 5.32$ and $\hat{\theta} = 1.22$
- BMLEs using data: $\tilde{\tau} = 4.95$ and $\tilde{\theta} = 1.22$
- For each of the pairs of τ, θ (true values, MLEs, BMLEs), simulate 5000 L_s , and calculate the CTE Reserve@0 for $\alpha = 0.95$



Simulation Analyses (2): Results

Parameters	% Positive Losses	Mean Loss	Median Loss	Std Dev Loss
True	6.82	-3,197	-3,201	1,949
MLE	2.20	-3,955	-3,747	1,853
BMLE	3.90	-3,742	-3,565	2,009

Parameters	95-th Quantile Loss	95-th CTE Loss Reserve
True	615	1,787
MLE	-1,468	64
BMLE	-583	1,032



Summary of Current Research

- Developed a Mathematica 8.0 module which calculates the CSCK MLE bias for probability distributions with up to three parameters
- We calculated the CSCK MLE bias for important distributions in actuarial science
- We demonstrated the quality of BMLEs relative to MLEs, particularly with regard to insurance contract valuations.



Next Steps

- Improve Mathematica module's functionality to extend to distributions (or mixtures of distributions) with more than 3 parameters
- Conduct analyses with real financial returns data
- Integrate our findings with the Actuarial Model Outcome Optimal Fit (AMOOOF) project
- Consideration of alternative methods for minimizing finite sample MLE bias, such as the “preventive” approach discussed in Firth (1993)



References

- Actuarial Model Outcome Optimat Fit (AMOOOF). 2008.
<http://www.cwu.edu/~chueh/AMOOOF.html>.
- 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
- 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



References (Continued)

- 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”
- Rose, C. and M.D. Smith. 2000. “Symbolic Maximum Likelihood Estimation with Mathematica.” *The Statistician*, 49(2): 229-240
- Wolfram Research, Inc. 2010. “Mathematica Edition: Version 8.0.” Wolfram Research, Inc: Champaign, IL

