Research Profile

1. Empirical Likelihood.

I have been working on Empirical Likelihood for a while, in fact since my Ph.D from1990-1992 at the Australia National University with Professor Peter G. Hall as my adviser. Empirical likelihood (Owen, 1998, 1990, 2001 for overviews) was just introduced by Art Owen at the time. It is a computer intensive statistical method which can facilitate either non-parametric or semi-parametric inference. It formulates a nonparametric profile likelihood for parameters, which can be described by certain equations. The most striking aspects of the empirical likelihood are that, despite that it is formed non-parametrically without a detailed model, for many parameters of interests, the empirical likelihood ratio shares two key properties of a conventional parametric likelihood: the first order Wilks theorem and the second order Bartlett correction (BC).

I started with working on the second order property of the empirical likelihood showing that it is Bartlett correctable for quantiles (Chen and hall, 1993), parametric and nonparametric regression (Chen, 1993; Chen 1994; Chen and QIN 2000), and density estimation (1996) . The most striking and general results are obtained in Chen and Cui (2006) and (2007) for the case of general estimating equation with and without nuisance parameters. These are the most general results on Bartlett correction for the empirical likelihood so far. The representative selected papers are

[1] Chen, S.X. (1993). On the coverage accuracy of empirical likelihood confidence regions for linear regression model. Annals of Institute of Statistical Mathematics, 45, 621-637.
[2] Chen, S.X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. Annals of Statistics, 21, 1166-1181.
[3] Chen, S.X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. Journal of Multivariate Analysis, 49, 24-40.
[4] Chen, S.X. (1996c) Empirical likelihood confidence intervals for nonparametric density estimation. Biometrika, 83 , 329-341.
[5] Chen, S. X. and Qin, Yong Song (2000). Empirical Likelihood confidence interval for a local linear smoother. Biometrika , 87, 946-953.
[6] Chen, S.X. and Cui, H. J. (2006) On Bartlett Correction of Empirical Likelihood in the Presence of Nuisance Parameters. Biometrika, 93, 215-220.
Chen, S.X. and Cui, H. J. (2007) On the Second Order Properties of Empirical Likelihood for Generalized Estimation Equations. *Journal of Econometrics*, 141, 492-516.

I also worked on applications of the empirical likelihood on various inference problems, which includes goodness-of-fit tests, missing values/surrogate data, and for high dimensional data. The representative papers are

Chen, S. X., Haredle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. Journal of The Royal Statistical Society, Series B, 65, 663-678.

Chen, S. X., D. H. Y. Leung and Qin, J. (2003) Information Recovery in a Study with Surrogate Endpoints. Journal of the American Statistical Association, 98, 1052-1062.

Chen, S. X., Leung, D. Y. H. and J. Qin. (2008) Improved Semiparametric Estimation Using Surrogate Data, Journal of the Royal Statistical Society, Series B to appear.

Wang, D. and S.X. Chen (2009), Empirical Likelihood for Estimating Equation with Missing Values, The Annals of Statistics, to appear.

Chen, S. X., L. Peng and Y-L, Qin (2009) Effects of Dimensionality on empirical likelihood, *Biometrika* to appear.

2. Population Surveys

Right after my Ph.D in Canberra, I was offered a job with Australia CSIRO Marine Science Laboratories in Tasmania from 1992-1995 as a statistician working on the aerial line transect surveys to estimate the abundance of the Southern Bluefin Tuna, a highly valuable commercial fish species in danger of extinction. The nature of the work was applied but challenging with an urgent need to develop new methods to suit the survey conditions. I was responsible for designing the survey and led the aerial surveys in the summers of 1993-1995 over the Great Australian Bight. I flow with the survey crew (at the back of a four seater twin engine airplane) for 125 hours in the first year, and a total of 250 hours in the three surveys. During and off the field operation in Port Lincoln, South Australia, I formulated a nonparametric kernel estimator of animal abundance and published the findings later in a series of papers: Chen (1996, *Journal of The Royal Statistical Society, Series C*), Chen (1996, 1998, 1999, *Biometrics*) and Chen, Yip and Zhou (2002, *Biometrics*). The method and estimator are still being used by the CSIRO and have drawn interests from the International Whaling Commission for estimating the abundance of Minke whales.

Lately I have been using the insights I gained over the Great Australian Bight for US Census capture-recapture dual system surveys since 2005 funded by a previous NSF grant. We proposed a new local post-stratification approach for the census dual system estimation (Chen, Tang, Mule, 2008, under revision to JASA Application and Case Study section).

[1] Chen, S.X. (1996a) A kernel estimate for density of a biological population using line transect sampling. *Royal Statistical Society Ser. C: Applied Statistics* 45, 135-150.

[2] Chen, S.X. (1996b) Studying school size effects in line transect sampling using the kernel method. *Biometrics* 52, 1283-94.
[3] Chen, S.X. and Polacheck, T. (1996) Kernel estimates of mean school size for IWC minke whale data. *Report of International Whaling Commission*, 46 , 341-348.

[4] Chen, S.X. (1998). Measurement errors in line transect surveys. *Biometrics*, 54, 899-908.

[5] Chen, S. X. (1999a) Estimation in independent observer line transect surveys for clustered populations. *Biometrics* 55 , No. 3, 754-759.

[6] Chen, S. X. and Woolcock, J. (1999) A condition for designing bus-route type access site surveys to estimate recreational fishing effort. *Biometrics*, 55, No. 3, 799-804.

[7] Chen, S. X. (2000b) Animal abundance estimation for independent observer line transect surveys. Special Issue of *Environmental and Ecological Statistics: Statistical Ecology and Forest*

Biometry 7 , No. 3, 285-299.

[8] Chen, S. X. and Lloyd, C. J. (2000). A non-parametric approach to the analysis of two stage mark-recapture experiments. *Biometrika*, 87, 633-649.

[9] Chen, S. X. and Cowling, A. (2001). Measurement Errors in Line Transect Surveys where Detection varies with Distance and Size. *Biometrics*, 57, 732-742.

[10] Chen, S. X. , C. Y. Tang and V. T. Mule, jr.   Local post-stratification and diagnostics in Dual System Accuracy and Coverage Evaluation for US Censu, being revised for *Journal of the American Statistical Association*, *Application and Case Study Section*.

3. Nonparametric Curve Estimation

While I was with CSIRO in Tasmania, I met Bruce M. Brown who was with the Math. Department (Pitman's department) at the University of Tasmania. Using the words of Peter when I told him I had been offered a job there, ``You two will be the only two mathematical statistician on the island". Indeed, Bruce and I spend

Motivated by my works on the line transect sampling which requires estimation of functional curve with bounded support, I was interested in estimating curves with bounded support. Together with B. M. Brown, then at the University of Tasmania, we proposed smoothing using asymmetrical kernels, like Beta or Gamma kernels. The publications are the followings, several of them enjoy quite high citations from economics and finance researchers:

[1] Brown, B. M. and Chen, S. X. (1999) Beta-Bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics*, 26 , 47-59.

[2] Chen, S. X. (1999b) Beta kernel estimators for density functions. *Computational Statistics and Data Analysis*, 31 , 131-145.

[3] Chen, S. X. (2000a) Beta kernel smoothers for regression curves. *Statistica Sinica*, 10, 73-91.

[4] Chen, S. X. (2000c) Gamma kernel estimators for density functions. *Annals of Institute of Statistical Mathematics*, 52, 471-480.

[5] Chen, S. X. (2002). Local linear smoothers using asymmetric kernels. *Annals of Institute of Statistical Mathematics*, 54, 312-323.

4. Financial Applications

After I moved to Singapore (with the National University of Singapore) ,   the financial center in the South East Asia, I started conduct research on economics and finance related problems. The first piece of work he did was on specification test for regression function of time series with applications to testing the drift specification of a diffusion model.   He proposed an empirical likelihood test via simulation of a known Gaussian random field. This is the work published in Chen, Hardle and Li (2003). Recently, we have formulated a general test for diffusion processes via an empirical likelihood formulation which targets on the transitional densities of the processes. *This work on the diffusion processes is supported by a NSF grant.* Other research problems on on finance related problems inference for financial risk measures like Value at Risk and Expected Shortfall and parameter estimation for diffusion processes. The selected publication in this area are

Chen, S. X., Haredle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *Journal of The Royal Statistical Society*, Series B, 65, 663-678.

Chen, S. X. and Tang, C. (2005). Nonparametric Estimation of Value at Risk and the Standard Errors for Financial Returns. *Journal of Financial Econometrics*, *3, 227-255*.

Chen, S.X. and J. Gao (2007) An Adaptive Empirical Likelihood Test for Time Series Models. *Journal of Econometrics*, 141, 950-972.

Chen, S.X., J. Gao and Tang, C. Y. (2008) A test for model specification of diffusion processes, *The Annals of Statistics*, 36, 167-198

Chen, S.X (2008) Nonparametric Estimation of Expected Shortfall, *Journal of Financial Econometrics*, 6, 87-107.

C. Yong Tang and Chen, S. X. (2009), Parameter estimation and bias correction of diffusion processes, *Journal of Econometrics* to appear.

5. High Dimensional Data

Lately, I have started working on high dimensional data. I have been working on two projects in this area of research. One is related with the empirical likelihood. Specifically on the effect of the dimension on the asymptotic normality of the empirical likelihood ratio.
We have developed growth rates for the dimension relative to the sample size so that the EL ratio is asymptotic normally distributed. In the second project, I and my student Y-L Qin Proposed a two sample test that generalizes the Hotelling's T^2 test to large p small n situations. The selected paper are:

Chen, S. X., L. Peng and Y-L, Qin (2009) Effects of Dimensionality on empirical likelihood, *Biometrika* to appear.

Chen, S. X. and Y-L Qin, A Two Sample Test for Ultra High Dimensional Data with Applications to Gene Sets Testing. Under revision for *the Annals of Statistics*.