

---

**High-Dimensional Problems in Finance and Quantitative Research**  
**Problèmes de grande dimension en finance et en recherche quantitative**  
(Org: Taehan Bae and/et Andrei Volodin (Regina))

---

---

**SYED EJAZ AHMED**, Brock university  
*Implicit Bias in Big Data Analytics*

Nowadays a large amount of data is available, and the need for novel statistical strategies to analyze such data sets is pressing. In this talk, my emphasis is on the development of statistical and computational strategies for a sparse regression model in the presence of mixed signals using machine and statistical learnings. The existing methods have often ignored contributions from weak signals and are subject unignorably bias. The amount of such information in a single predictor might be modest but helps in prediction accuracy in a meaningful way. The search for such signals, sometimes called networks or pathways, is for instance an important topic for those working on personalized medicine. We propose and implement a new “post selection shrinkage estimation/prediction strategy” that takes into account the joint impact of both strong and weak signals to improve the prediction accuracy, and opens pathways for further research in such scenarios.

---

**NAEIMA ASHLEIK**, University of Saskatchewan  
*Learning for Contingency Tables and Survival Data Using Imprecise Probabilities*

Imprecise probability theory is a generalization of the classical probability theory. A comprehensive collection of the foundations of imprecise probabilities theory is provided by Walley (1991), where the name of “imprecise probability” was proposed. The upper and lower posterior expectations of log-odds ratio are estimated and the degree of imprecision is calculated in this work. Survival data with right-censored observations are considered and represented in a sequence of  $2 \times 2$  contingency tables, one at each observed death time. A re-parametrization of odds ratio is assumed based on the feature that non-central hypergeometric distribution. Two choices (normal and beta) of imprecise priors are given to the parameters. The findings show that small values of the degree of imprecision appear when the sample size is large and the number of censored observations is small. In contrast, the large values of the degree of imprecision are observed when sample size is small and the number of censored observations is large. In short, the degree of imprecision of the parameter of interest is reduced by having more information, more data, and less censored observations as the results of this work displayed, which is intuitively what one would expect.

---

**YANGHO CHOI**, Hanyang University  
*Macroscopic Modeling of Data Breach Risk with Spatial and Temporal Autocorrelation*

Data breach risk caused by leak of private information has attracted considerable attention recently and insurers face rising necessity for predictive models to manage its risk. However, the job of modeling the risk is mainly discussed on a microscopic level in perspective of information technologies, partially due to statistically insufficient data, and it has been impeded by its unique characteristic of high correlation. This study, however, models data breach losses on a macroscopic scale in perspective of statistics, and we perform empirical analysis on their collective structure of mutual dependence in the dimension of space and time, based on the samples of data breaches in the United States during the recent decade. We discover that for data breach risk, an individual establishment or firm can be a candidate for a risk exposure unit, and we present an evidence of a medium-to-low spatial and temporal autocorrelation in the data breach events. We also find that time series of data breach events might have a covariate characterized by a market capital such as the S&P 500 index or the total private non-profit credit capitalization in the United States.

---

**XUEMIAO HAO**, University of Manitoba  
*Sharp Tail Estimate for Aggregate Critical Illness Claims in a Large Population*

Health insurance has become an essential component in our society. In most advanced countries, health insurance is jointly provided by the government and private parties through universal insurance and voluntary insurance, respectively. A fundamental question is which treatments should be covered by universal basic insurance and which by private voluntary insurance. In health economics literature, discussions focus on maximizing a population's total welfare, which is defined as the expected utility. In our research, we study the effect of variation of aggregate health claims for the population. In particular, we estimate quantities related with the tail of the aggregate critical illness claims for the population. The idea is motivated by a newly proposed credit risk model for large portfolio.

---

**SHAKHAWAT HOSSAIN**, University of Winnipeg.ca

*Estimation strategy of multilevel model for ordinal longitudinal data*

This paper considers the shrinkage estimation of multilevel models that are appropriate for ordinal longitudinal data. These models can accommodate multiple random effects and, additionally, allow for a general form of model covariates that are related to the overall level of the responses and changes to the response over time. The likelihood inference for multilevel models is computationally burdensome due to intractable integrals. A maximum marginal likelihood (MML) method with Fisher's scoring procedure is therefore followed to estimate the random and fixed effects parameters. In real life data, researchers may have collected many covariates for the response. Some of these covariates may satisfy certain constraints which can be used to produce a restricted estimate from the unrestricted likelihood function. The unrestricted and restricted MMLs can then be combined optimally to form the pretest and shrinkage estimators. Asymptotic properties of these estimators including biases and risks will be discussed. A simulation study is conducted to assess the performance of the estimators with respect to the unrestricted MML estimator. Finally the relevance of the proposed estimators will be illustrated with a real data set.

---

**ALEXANDER MELNIKOV**, University of Alberta

*On Option Pricing Methods in Modern Mathematical finance*

Option pricing is one of the main research areas of modern Mathematical Finance. Hence, new valuable developments in this area remain well-motivated and highly desirable. The aim of the talk is to present some comprehensive issues that can be interesting also for a wider audience besides those experts who are directly working in Mathematical Finance. Moreover, the developments in option pricing can be considered as a reasonable source of new problems for further studies and research. In the talk a dual theory of option pricing will be developed by means of market completions as an alternative of the well-known option price characterization via martingale measures. Beside that we present another approach in option pricing which is based on comparison theorems for solutions of stochastic differential equations. It will be shown also how to use in option pricing the so-called partial/imperfect hedging technique that is concentrated around a statistical notion of "loss functions" and a financial notion of "risk measures". Finally, we will pay an attention to extensions of probability distributions of stock returns using orthogonal polynomials techniques. Going in this way we get a possibility to see what can happen beyond the Black-Scholes model.

---

**SHANOJA NAIK**, RAO, Toronto

*On Wishart Process and Sovereign Credit Risk Modelling*

The Wishart Process has got attention in financial modeling due to its specific characteristics. In this talk, I will explain the Cox-Ingersoll-Ross (CIR) process and its matrix variate extension as Wishart process. Application of Wishart process in Sovereign Credit Risk model will be discussed. Additionally, the calibration of the model using the exponential matrix variate form will be explained with challenges and future directions to expand the theory of matrix variate Wishart model.

---

**THUNTIDA NGAMKHAM**, University of Calgary

*Confidence intervals for a ratio of binomial proportions*

A general problem of the interval estimation for a ratio of two proportions according to data from two independent samples is considered. Each sample may be obtained in the framework of direct or inverse binomial sampling. Asymptotic confidence

intervals are constructed in accordance with different types of sampling schemes with an application, where it is possible, of unbiased estimations of success probabilities and also their logarithms. Since methods of constructing confidence intervals in the situations when values for the both samples are obtained for identical sample schemes are already developed and well known, the main purpose of this paper is the investigation of constructing confidence intervals in two cases that correspond to different sampling schemes. In this situation it is possible to plan the sample size for the second sample according to the number of successes in the first sample. This, as it is shown by the results of statistical modeling, provides the intervals with confidence level which closer to the nominal value.

My goal is to show that the normal approximations for estimates of the ratio of proportions and their logarithms are reliable for a construction of confidence intervals. The main criterion of our judgment is the closeness of the confidence coefficient to the nominal confidence level. It is proved theoretically and shown by statistically modeled data that the scheme of inverse binomial sampling with planning of the size in the second sample is preferred.

---

**SALMA SAAD**, University of Regina

*Asymptotic Analysis of Method of Moments Estimators of Parameters  $p$  and  $m$  for the Binomial Distribution*

An estimation of parameters of the Binomial distribution by a sample of fixed size  $n$ , when both parameters  $m$  and  $p$  are unknown, has remained an important statistical problem for more than three quarters of a century. Known estimates of  $m$  usually underestimate the true value. We consider only the Method of Moments and its modifications for estimation of parameters  $m$  and  $p$  of the Binomial distribution. We also apply the delta method is for the proof of asymptotic normality of the joint distribution of the estimators of  $m$  and  $p$  by the Method of Moments. The main difficulty here is that the estimators do not have moments of all orders and hence the parameters of asymptotic normality do not have direct interpretations as characteristics of accuracy properties of these estimators. We are mostly interested in the bias and variance of the Method of Moments and its modifications estimators. To achieve these goals it is necessary to solve the following problems:

1. Derivation of estimates of parameters of Binomial distribution by the Method of Moments;
2. Derivation of the parameters of asymptotic normality by the delta-method;
3. Comparison of the derived asymptotic with the true probabilistic characteristics of the estimators by the method of statistical simulations