Commun. Math. Res.
doi: 10.4208/cmr.2020-0041

Vol. **37**, No. 1, pp. 1-85
February 2021

REVIEW ARTICLE

Concentration Inequalities for Statistical Inference

Huiming Zhang[1,3,4] and Song Xi Chen[1,2,3,*]

[1] *School of Mathematical Sciences, Peking University,*
Beijing 100871, P.R. China.
[2] *Guanghua School of Management, Peking University,*
Beijing 100871, P.R. China.
[3] *Center for Statistical Sciences, Peking University,*
Beijing 100871, P.R. China.
[4] *Department of Mathematics, Faculty of Science and Technology,*
University of Macau, P.R. China.

Received 4 November 2020; Accepted 12 December 2020

Abstract. This paper gives a review of concentration inequalities which are widely employed in non-asymptotical analyses of mathematical statistics in a wide range of settings, from distribution-free to distribution-dependent, from sub-Gaussian to sub-exponential, sub-Gamma, and sub-Weibull random variables, and from the mean to the maximum concentration. This review provides results in these settings with some fresh new results. Given the increasing popularity of high-dimensional data and inference, results in the context of high-dimensional linear and Poisson regressions are also provided. We aim to illustrate the concentration inequalities with known constants and to improve existing bounds with sharper constants.

AMS subject classifications: 60F10, 60G50, 62E17

Key words: Constants-specified inequalities, sub-Weibull random variables, heavy-tailed distributions, high-dimensional estimation and testing, finite-sample theory, random matrices.

*Corresponding author. *Email addresses:* zhanghuiming@pku.edu.cn (H. Zhang), csx@gsm.pku.edu.cn (S. X. Chen)

1 Introduction

In probability theory and statistical inference, researchers often need to bound the probability of a difference between a random quantity from its target, usually the error bound of estimation. Concentration inequalities (CIs) are tools for attaining such bounds, and play important roles in deriving theoretical results for various inferential situations in statistics and probability. The recent developments in high-dimensional (HD) statistical inference, and statistical and machine learning have generated renewed interests in the CIs, as reflected in [29,47,84,86]. As the CIs are diverse in their forms and the underlying distributional requirements, and are scattered around in references, there is an increasing need for a review which collects existing results together with some new results (sharper and constants-specified CIs) from the authors for researchers and graduate students working in statistics and probability. This motivates the writing of this review.

CIs enable us to obtain non-asymptotic results for estimating, constructing confidence intervals, and doing hypothesis testing with a high-probability guarantee. For example, the first-order optimized condition for HD linear regressions should be held with a high probability to guarantee the well-behavior of the estimator. The concentration inequality for error distributions is to ensure the concentration from first-order optimized conditions to the estimator. Our review focuses on four types of CIs:

$$P(Z_n > \mathrm{E}Z_n + t), \quad P(Z_n < \mathrm{E}Z_n - t), \quad P(|Z_n - \mathrm{E}Z_n| > t), \quad \mathrm{E}(\max_{i=1,\dots,n} |X_i|),$$

where $Z_n := f(X_1,\cdots,X_n)$ and $X_1,\cdots,X_n$ are random variables. We present two types of CIs: distribution-free and distribution-dependent. Distribution free CIs are free of distribution assumptions, while the distribution-dependent CIs are based on exponential moment conditions reflecting the tail property for the particular class of distributions. Concentration phenomenons for a sum of sub-Weibull random variables will lead to a mixture of two tails: sub-Gaussian for small deviations and sub-Weibull for large deviations from the mean, and it is closely related to Strong Law of Large Numbers, Central Limit Theorem, and Law of the Iterative Logarithm. We provide applications of the CIs to empirical processes and high-dimensional data settings. The latter includes the linear and Poisson regression with a diverging number of covariates. We organize the materials in the forms of lemmas, corollaries, propositions, and theorems. Lemmas and corollaries are on existing results usually without proof except for a few fundamental ones. Propositions are also for existing results but with sharper or more precise constants and sometimes come with proofs. Theorems are for new results. This review contains 26 lemmas, 21 corollaries, 14 propositions, and 4 theorems.