

Chapter 1

Introduction

A prototypical example of a self-normalized process is Student's t -statistic based on a sample of normal i.i.d. observations X_1, \dots, X_n , dating back to 1908 when William Gosset ("Student") considered the problem of statistical inference on the mean μ when the standard deviation σ of the underlying distribution is unknown. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ be the sample mean and $s_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ be the sample variance. Gosset (1908) derived the distribution of the t -statistic $T_n = \sqrt{n}(\bar{X}_n - \mu)/s_n$ for normal X_i ; this is the t -distribution with $n-1$ degrees of freedom. The t -distribution converges to the standard normal distribution, and in fact T_n has a limiting standard normal distribution as $n \rightarrow \infty$ even when the X_i are non-normal. When nonparametric methods were subsequently introduced, the t -test was compared with the nonparametric tests (e.g., the sign test and rank tests), in particular for "fat-tailed" distributions with infinite second or even first absolute moments. It has been found that the t -test of $\mu = \mu_0$ is robust against non-normality in terms of the Type I error probability but not the Type II error probability. Without loss of generality, consider the case $\mu_0 = 0$ so that

$$T_n = \frac{\sqrt{n}\bar{X}_n}{s_n} = \frac{S_n}{V_n} \left\{ \frac{n-1}{n - (S_n/V_n)^2} \right\}^{1/2}, \quad (1.1)$$

where $S_n = \sum_{i=1}^n X_i$, $V_n^2 = \sum_{i=1}^n X_i^2$. Efron (1969) and Logan et al. (1973) have derived limiting distributions of self-normalized sums S_n/V_n . In view of (1.1), if T_n or S_n/V_n has a limiting distribution, then so does the other, and it is well known that they coincide; see, e.g., Proposition 1 of Griffin (2002).

Active development of the probability theory of self-normalized processes began in the 1990s with the seminal work of Griffin and Kuelbs (1989, 1991) on laws of the iterated logarithm for self-normalized sums of i.i.d. variables belonging to the domain of attraction of a normal or stable law. Subsequently, Bentkus and Götze (1996) derived a Berry–Esseen bound for Student's t -statistic, and Giné et al. (1997) proved that the t -statistic has a limiting standard normal distribution if and only if X_i is in the domain of attraction of a normal law. Moreover, Csörgő et al. (2003a)

proved a self-normalized version of the weak invariance principle under the same necessary and sufficient condition. Shao (1997) proved large deviation results for S_n/V_n without moment conditions and moderate deviation results when X_i is the domain of attraction of a normal or stable law. Subsequently Shao (1999) obtained Cramér-type large deviation results when $E|X_1|^3 < \infty$. Jing et al. (2004) derived saddlepoint approximations for Student's t -statistic with no moment assumptions. Bercu et al. (2002) obtained large and moderate deviation results for self-normalized empirical processes. Self-normalized sums of independent but non-identically distributed X_i have been considered by Bentkus et al. (1996), Wang and Jing (1999), Jing et al. (2003) and Csörgő et al. (2003a).

Part I of the book presents in Chaps. 3–7 the basic ideas and results in the probability theory of self-normalized sums of independent random variables described above. It also extends in Chap. 8 the theory to self-normalized U -statistics based on independent random variables. Part II considers self-normalized processes in the case of dependent variables. Like Part I that begins by introducing some basic probability theory for sums of independent random variables in Chap. 2, Part II begins by giving in Chap. 9 an overview of martingale inequalities and related results which will be used in the subsequent chapters. Chapter 10 provides a general framework for self-normalization, which links the approach of de la Peña et al. (2000, 2004) for general self-normalized processes to that of Shao (1997) for large deviations of self-normalized sums of i.i.d. random variables. This general framework is also applicable to dependent random vectors that involve matrix normalization, as in Hotelling's T^2 -statistic which generalizes Student's t -statistic to the multivariate case. In particular, it is noted in Chap. 10 that a basic ingredient in Shao's (1997) self-normalized large deviations theory is $e^{\psi(\theta, \rho)} := E \exp\{\theta X_1 - \rho \theta^2 X_1^2\}$, which is always finite for $\rho > 0$. This can be readily extended to the multivariate case by replacing θX_1 with $\theta' X_1$, where θ and X_1 are d -dimensional vectors. Under the assumptions $EX_1 = 0$ and $E\|X_1\|^2 < \infty$, Taylor's theorem yields

$$\psi(\theta, \rho) = \log(E \exp\{\theta' X_1 - \rho(\theta' X_1)^2\}) = \left\{ \left(\frac{1}{2} - \rho + o(1) \right) \theta' E(X_1 X_1') \theta \right\}$$

as $\theta \rightarrow 0$. Let $\gamma > 0$, $C_n = (1 + \gamma) \sum_{i=1}^n X_i X_i'$, $A_n = \sum_{i=1}^n X_i$. It then follows that ρ and ε can be chosen sufficiently small so that

$$\begin{aligned} & \left\{ \exp(\theta' A_n - \theta' C_n \theta / 2), \mathcal{F}_n, n \geq 1 \right\} \\ & \text{is a supermartingale with mean } \leq 1, \text{ for } \|\theta\| < \varepsilon. \end{aligned} \quad (1.2)$$

Note that (1.2) implies that $\{\int_{\|\theta\| < \varepsilon} e^{\theta' A_n - \theta' C_n \theta / 2} f(\theta) d\theta, \mathcal{F}_n, n \geq 1\}$ is also a supermartingale, for any probability density f on the ball $\{\theta : \|\theta\| < \varepsilon\}$.

In Chap. 11 and its multivariate extension given in Chap. 14, we show that the supermartingale property (1.2), its weaker version $E\{\exp(\theta' A_n - \theta' C_n \theta / 2)\} \leq 1$ for $\|\theta\| < \varepsilon$, and other variants given in Chap. 10 provide a general set of conditions from which we can derive exponential bounds and moment inequalities for self-normalized processes in dependent settings. A key tool is the *pseudo-maximization*

method which involves Laplace's method for evaluating integrals of the form $\int_{\|\theta\| < \varepsilon} e^{\theta'A_n - \theta'C_n\theta/2} f(\theta) d\theta$. If the random function $\exp\{\theta'A_n - \theta'C_n\theta/2\}$ in (1.2) could be maximized over θ inside the expectation $E\{\exp(\theta'A_n - \theta'C_n\theta/2)\}$, taking the maximizing value $\theta = C_n^{-1}A_n$ would yield the expectation of the self-normalized variable $\exp\{A_n C_n^{-1} A_n / 2\}$. Although this argument is not valid, integrating $\exp\{\theta'A_n - \theta'C_n\theta/2\}$ with respect to $f(\theta) d\theta$ and applying Laplace's method to evaluate the integral basically achieves the same effect as in the heuristic argument. This method is used to derive exponential and L_p -bounds for self-normalized processes in Chap. 12. The exponential bounds are used to derive laws of the iterated logarithm for self-normalized processes in Chap. 13.

Student's t -statistic $\sqrt{n}(\bar{X}_n - \mu)/s_n$ has also undergone far-reaching generalizations in the statistics literature during the past century. Its generalization is the *Studentized statistic* $(\hat{\theta}_n - \theta)/\hat{s}_n$, where θ is a functional $g(F)$ of the underlying distribution function F , $\hat{\theta}_n$ is usually chosen to be the corresponding functional $g(\hat{F}_n)$ of the empirical distribution, and \hat{s}_n is a consistent estimator of the standard error of $\hat{\theta}_n$. Its multivariate generalization, which replaces $1/\hat{s}_n$ by $\hat{\Sigma}_n^{-1/2}$, where $\hat{\Sigma}_n$ is a consistent estimator of the covariance matrix of the vector $\hat{\theta}_n$ or its variant, is ubiquitous in statistical applications. Part III of the book, which is on statistical applications of self-normalized processes, begins with an overview in Chap. 15 of the distribution theory of the t -statistic and its multivariate extensions, for samples first from normal distributions and then from general distributions that may have infinite second moments. Chapter 15 also considers the asymptotic theory of general Studentized statistics in time series and control systems and relates this theory to that of self-normalized martingales. An alternative to inference based on asymptotic distributions of Studentized statistics is to make use of bootstrapping. Chapter 16 describes the role of self-normalization in deriving approximate pivots for the construction of bootstrap confidence intervals, whose accuracy and correctness are analyzed by Edgeworth and Cornish–Fisher expansions. Chapter 17 introduces generalized likelihood ratio statistics as another class of self-normalized statistics. It also relates the pseudo-maximization approach and the method of mixtures in Part II to the close connections between likelihood and Bayesian inference. Whereas the framework of Part I covers the classical setting of independent observations sampled from a population, that of Part II is applicable to time series models and stochastic dynamic systems, and examples are given in Chaps. 15, 17 and 18. Moreover, the probability theory in Parts I and II is related not only to samples of fixed size, but also to sequentially generated samples that are associated with asymptotically optimal stopping rules. Part III concludes with Chap. 18 which considers self-normalized processes in sequential analysis and the associated boundary crossing problems.

Self-Normalized Processes

Limit Theory and Statistical Applications

Peña, V.H.; Lai, T.L.; Shao, Q.-M.

2009, XIV, 275 p., Hardcover

ISBN: 978-3-540-85635-1