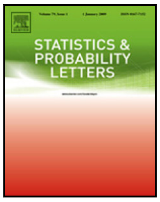
Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



Generalizations of maximal inequalities to arbitrary selection rules

Jiantao Jiao *, Yanjun Han, Tsachy Weissman

Stanford University, United States

ARTICLE INFO

Article history:
Received 29 August 2017
Received in revised form 2 January 2018
Accepted 5 January 2018
Available online xxxx

MSC:
60E15
60G70

Keywords:
Entropy
Maximal inequality
Orlicz function
Convex duality
Generalized Holder's inequality

ABSTRACT

We present a generalization of the maximal inequalities that upper bound the expectation of the maximum of n jointly distributed random variables. We control the expectation of a randomly selected random variable from n jointly distributed random variables, and present bounds that are at least as tight as the classical maximal inequalities, and much tighter when the distribution of selection index is near deterministic. A new family of information theoretic measures was introduced in the process, which may be of independent interest.

1. Introduction

Throughout this paper, we consider n random variables Z_i, $1 \leq i \leq n$ such that $\mathbb{E}[Z_i] = 0$, where n is a finite positive integer. The zero mean condition can be satisfied via the operation $Z_i' = Z_i - \mathbb{E}[Z_i]$ upon assuming that all Z_i's are integrable. The following two maximal inequalities are well known in the literature and serve as the motivational results for this work.

Lemma 1. *Let $\psi \geq 0$ be a convex function defined on the interval $[0, b)$ where $0 < b \leq \infty$. Assume that $\psi(0) = 0$. Set, for every $t \geq 0$,*

$$\psi^*(t) = \sup_{\lambda \in (0,b)} (\lambda t - \psi(\lambda)). \tag{1}$$

Suppose that $\ln \mathbb{E}[e^{\lambda Z_i}] \leq \psi(\lambda)$ for all $\lambda \in [0, b)$, $1 \leq i \leq n$. Then,

$$\mathbb{E}[\max_i Z_i] \leq \psi^{*-1}(\ln n), \tag{2}$$

where $\psi^{*-1}(y)$ is defined as

$$\psi^{*-1}(y) = \inf\{t \geq 0 : \psi^*(t) > y\}. \tag{3}$$

* Corresponding author.
E-mail addresses: jiantao@stanford.edu (J. Jiao), yjhan@stanford.edu (Y. Han), tsachy@stanford.edu (T. Weissman).

https://doi.org/10.1016/j.spl.2018.01.002

To introduce the second inequality, we say a function ψ is an Orlicz function if $\psi : [0, \infty) \mapsto [0, \infty]$ is a convex function vanishing at zero and is also not identically 0 or ∞ on $(0, \infty)$. We define the Luxemburg ψ norm of a random variable X as

$$\|X\|_\psi = \inf\left\{\sigma > 0 : \mathbb{E}\left[\psi\left(\frac{|X|}{\sigma}\right)\right] \le 1\right\}. \tag{4}$$

Lemma 2 (Pollard, 2005)**.** *Let ψ be an Orlicz function. Suppose $\|Z_i\|_\psi \le \sigma$, $1 \le i \le n$. Then,*

$$\mathbb{E}[\max_i Z_i] \le \sigma \cdot \psi^{-1}(n), \tag{5}$$

where $\psi^{-1}(y)$ is defined as $\psi^{-1}(y) = \inf\{t \ge 0 : \psi(t) > y\}$.

This paper concerns with the question of generalizing Lemmas 1 and 2 to arbitrary selection rules. Concretely, suppose $T \in \{1, 2, \dots, n\}$ is a random variable jointly distributed with $Z_1, Z_2, \dots, Z_n$. We would like to upper bound $\mathbb{E}[Z_T]$, which subsumes the maximal inequality $T = \mathrm{argmax}_i Z_i$ as a special case. Naturally, since

$$\mathbb{E}[Z_T] \le \mathbb{E}[\max_i Z_i], \tag{6}$$

we would like to obtain bounds that are at least as strong as Lemmas 1 and 2, but dependent on the joint distribution of $T, Z_1, Z_2, \dots, Z_n$. In particular, the upper bound should be zero if T is deterministic since we have already assumed that $\mathbb{E}[Z_i] = 0$ for all $1 \le i \le n$.

A generalization of Lemma 1 was achieved in Jiao et al. (2017) using the Donsker–Varadhan variational representation of the relative entropy, which is a generalization of the sub-Gaussian case in Russo and Zou (0000). Denote the entropy of a discrete random variable T as

$$H(T) = \sum_t P_T(t) \ln \frac{1}{P_T(t)}, \tag{7}$$

and the mutual information $I(X; Y)$ between X and Y as

$$I(X; Y) = \begin{cases} \int \ln \frac{dP_{XY}}{d(P_X P_Y)} dP_{XY} & \text{if } P_{XY} \ll P_X P_Y \\ \infty & \text{otherwise.} \end{cases} \tag{8}$$

The following was shown in Jiao et al. (2017).

Lemma 3. *Let $\psi \ge 0$ be a convex function defined on the interval $[0, b)$ where $0 < b \le \infty$. Assume that $\psi(0) = 0$. Set, for every $t \ge 0$,*

$$\psi^*(t) = \sup_{\lambda \in (0,b)} (\lambda t - \psi(\lambda)). \tag{9}$$

Suppose that $\ln \mathbb{E}[e^{\lambda Z_i}] \le \psi(\lambda)$ for all $\lambda \in [0, b)$, $1 \le i \le n$, and $\mathbb{E}[Z_i] = 0$, $1 \le i \le n$. Then,

$$\mathbb{E}[Z_T] \le \psi^{*-1}(I(T; \mathbf{Z})) \tag{10}$$

$$\le \psi^{*-1}(H(T)) \tag{11}$$

where $\psi^{*-1}(y)$ is defined as

$$\psi^{*-1}(y) = \inf\{t \ge 0 : \psi^*(t) > y\}. \tag{12}$$

and $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$.

Lemma 3 is clearly stronger than Lemma 1 since $I(T; \mathbf{Z}) \le H(T) \le \ln n$. It is also interesting to observe that the soft bound is maximized when T follows a uniform distribution, and it is zero when T is deterministic.

Similar attempts were made to generalize Lemma 2 in Jiao et al. (2017). However, it was not satisfactory since that even in the case of $\psi(x) = x^p$, $p \ge 1$, $x \ge 0$, the generalization bound obtained in Jiao et al. (2017) may be infinity when $1 \le p < 2$, while Lemma 2 shows that it is universally bounded by $\sigma \cdot n^{1/p}$ for every $p \ge 1$.

Our main contribution in this paper is the generalization of Lemma 2 to arbitrary selection rules. Our generalization satisfies the following properties:

1. It is at least as strong as Lemma 2: in other words, it can be shown that the worst case joint distribution of T and $\mathbf{Z}$ would not incur an upper bound larger than $\sigma \cdot \psi^{-1}(n)$, which is the upper bound in Lemma 2.
2. It admits a closed form expression for the p-norm case, i.e., the case where $\psi(x) = x^p$, $p \ge 1$, $x \ge 0$. In other words, it defines another information theoretic measure paralleling the Shannon entropy $H(T)$ in Lemma 3. Concretely, for

Please cite this article in press as: Jiao J., et al., Generalizations of maximal inequalities to arbitrary selection rules. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.01.002.