Chapter 6: Maximal Inequalities and Chaining

Empirical Processes (MATH-522)

Myrto Limnios

April, 2025

This chapter focuses on providing an upperbound of $\mathbb{E}\left[\sup_{t\in T}|X_t|\right]$ and shows why it measures the *size* of a generic process $\{X_t\}_{t\in T}$, when the index set is considered to be infinite. We will present a general method, namely *chaining method*, for obtaining sharp bounds of the quantity $\mathbb{E}\left[\sup_{t\in T}|X_t|\right]$ called *maximal inequalities*. We see that if the *size* of the index set T can be analyzed w.r.t. a distance based on the process X as in the Chapter 4, then we can control the *worst* deviation of the process uniformly on T.

Notations. We will use the notations of Chapter 5 without further notice. Recall that we considered an i.i.d. samples $X_1,\dots X_n$ defined on the p.s. $(\Omega,\mathcal{A},\mathbb{P})$, valued in a measurable space $\mathcal{X}\subseteq\mathbb{R}^d$, $d\in\mathbb{N}^*$, of probability distribution P and of empirical p.d. P_n.

1 Introduction to the Chaining Method

1.1 Finite Index Set

Suppose the set T to be finite. We want to upperbound the maximum of a finite number of r.v.s. Notice that

$$\mathbb{E}[\sup_{t\in T}X_t]\leq\mathbb{E}[\sum_{t\in T}|X_t|]\leq|T|\sup_{t\in T}\mathbb{E}|X_t|\ .$$

Remark 1.1. *1. Controlling the magnitude of each of the r.v.s X-s seems unsatisfactory, and we want to take advantage of possibly some tail assumption of the r.v.s.*

2. The bound grows linearly with the size of T that is, again, unsatisfactory. It seems that we cannot get any good conclusion from this.

Suppose now that the r.v.s X have bounded p-moment, then Jensen's inequality helps us understand a refined control

$$\mathbb{E}[\sup_{t\in T}X_t]\leq\mathbb{E}[\sup_{t\in T}|X_t|^p]^{1/p}\leq|T|^{1/p}\sup_{t\in T}\mathbb{E}[|X_t|^p]^{1/p}\ .$$

Remark 1.2. *This bound is more interesting: if p is large, then the tails of the r.v.s are vanishing implying a smaller value for the expectation. In addition, the larger p and the slower growth in terms of $|T|$.*

We prove a first *maximal inequality* to see how to use more general functionals related to Cramér-Chernoff method resulting in sharper bounds. Before stating the first result, recall an important definition.

Definition 1.1. A process $(X_t)_{t\in T}$ defined on a metric space (T,d) is centered ($\nu > 0$)-sub-Gaussian if $PX_t = 0$, if, for all $\lambda > 0$, for all $t \in T$,

$$\psi_X(\lambda) = \log \mathbb{E}[e^{\lambda X_t}] \leq \frac{\lambda^2 \nu}{2} \ .$$

Lemma 1.2. *Consider a finite collection of elements $|T| < \infty$, and the process X_t be centered ν-sub-Gaussian. Suppose that we observe an i.i.d. sequence of X_t of size n of empirical measure P_n. Then it holds true that*

$$\mathbb{E}\left[\max_{t\in T} |P_n t|\right] \leq \sqrt{\frac{\nu \log(2|T|)}{n}} \ . \tag{1}$$

Remark 1.3. *1. The result is interesting as soon as we choose $n \geq \sqrt{\nu \log(2|T|)}$.*

2. We study the max because the class of functions is finite, thus the supnorm is attainable.

3. Notice that the size of the class enters into play similarly as the square-root of the entropy.

4. This is the strongest result we can have, and it is sharp: If we know the best constant v upperbounding the variance of the process, then we cannot obtain better upperbound.

5. Notice that it can be related to Massart's inequality (Massart (2000), Lemma 5.2). Notice that we consider the absolute valued here, yielding a 2 in the log.

Proof. By Jensen's inequality:

$$\begin{aligned}
\mathbb{E}\left[\max_{t\in T} |P_n t|\right] &= \frac{1}{\lambda}\mathbb{E}\left[\log \exp \lambda \max_{t\in T} |P_n t|\right] \\
&\leq \frac{1}{\lambda} \log \mathbb{E}\left[\exp \lambda \max_{t\in T} |P_n t|\right] \\
&= \frac{1}{\lambda} \log \mathbb{E}\left[\max_{t\in T} \exp \lambda |P_n t|\right] \\
&\leq \frac{1}{\lambda} \log \sum_{t\in T} \mathbb{E}\left[\exp \lambda |P_n t|\right] \ .
\end{aligned}$$

Notice that $e^{|x|} \leq e^x + e^{-x}$, so that for $\lambda \geq 0$, and using the sub-Gaussianity assumption yields

$$\mathbb{E}[\exp\{\lambda |P_n t|\}] \leq \mathbb{E}[\exp\{\lambda P_n t\}] + \mathbb{E}[\exp\{-\lambda P_n t\}] \leq 2e^{\lambda^2 \nu/2n} \ .$$

Thus

$$\mathbb{E}\left[\max_{t\in T} |P_n t_j|\right] \leq \frac{\log(2|T|)}{\lambda} + \frac{\lambda \nu}{2n} \ .$$

Minimizing w.r.t. $\lambda^* = \sqrt{2n \log 2|T|/\nu}$ yields the result.

□

We can see this result similarly to the Chernoff bound that we studied in Chapter 2, i.e., if $\log \mathbb{E}[e^{\lambda X_t}] \leq \psi(\lambda)$, then $\mathbb{P}(X_t \geq u) \leq e^{-\psi^*(u)}$, for all $u \geq 0$ and $t \in T$. We thus formulate the maximal tail inequality.

Lemma 1.3 (Maximal tail inequality)**.** *Consider the process X_t as defined in Lemma 1.2, then for all $\delta \in (0,1)$, with probability at least $1-\delta$*

$$\max_{t\in T} |P_n t| \leq \frac{\sqrt{\nu \log(2N)}}{n} + \frac{\sqrt{\nu \log(1/\delta)}}{n} \ .$$

Remark 1.4. *This result is again sharp, as soon as all the* X*-s are independent. Notice that if there is dependence, and for example all r.v.s are equal for all* t*, then essentially* $\max_{i\leq N}|P_n t_j| = |P_n t_1|$ *and thus the bound is far from being optimal.*

The next section provides a generic method for obtaining similar rate of convergence, when the index set is uncountable.