Freedman's Inequality for Arbitrary Dependent Data

Shaocong Ma

Updated: October 27, 2021

Abstract

This note further generalizes the applicability of Freedman's inequality from the martingales to any dependent process. Based on the construction given in Harvey et al. (2019), we consider the influences of dependency, which allows the desired generalization.

1 Freedman's Inequality

The proof strategy follows Theorem 3.2, Harvey et al. (2019).

Lemma 1.1. *For every $\lambda \in (0, \frac{1}{2\alpha}]$, there exists some choices of $c = c(\lambda, \alpha)$ such that*

$$c\lambda^2 = \frac{\tilde{\lambda}^2}{2},$$

where $\tilde{\lambda} = \lambda + c\lambda^2\alpha$. More explicitly, if $\lambda = \frac{1}{2\alpha}$, then $c = 2$.

Proof. Notice that the equation $c\lambda^2 = \frac{\tilde{\lambda}^2}{2}$ is a quadratic function of c. To show that it has at least one root, it suffices to show its discriminant is not less than 0, that is

$$(2\alpha\lambda^3 - 2\lambda^2)^2 - 4\lambda^6\alpha^2 \geq 0.$$

This obviously holds when $\lambda \in (0, \frac{1}{2\alpha}]$. □

Lemma 1.2 (Supermartingale construction)**.** *Let $\{d_j\}_{j\in\mathbb{N}_+}$ be an adaptive process with respect to the filtration $\mathcal{F}$. Suppose $v_{j-1} \geq 0$ for $j \in \mathbb{N}_+$ are $\mathcal{F}_{j-1}$-measurable random variables. Assume there exists a sequence $\{C_t\}$ admitting a split $C_t = E_t - E_{t-1}$ for an increasing positive sequence $\{E_t\}$ such that*

$$\mathbb{E}[e^{\lambda d_t}|\mathcal{F}_{t-1}] \leq \exp\left(C_t\lambda + \frac{\lambda^2}{2}v_j\right) \tag{1}$$

for some $\lambda > 0$. Let $\tilde{\lambda} = (\lambda + c\lambda^2\alpha)$. Define

$$U_t = \exp\left(\sum_{j=1}^{t}(\lambda + c\lambda^2\alpha_j)d_j - \sum_{j=1}^{t}\frac{\tilde{\lambda}^2}{2}v_{j-1}\right).$$

Then

$$\left\{e^{-E_t\left(\lambda + c\lambda^2\alpha\right)}U_t\right\}$$

is a supermartingale with respect to $\mathcal{F}$, where $\alpha := \sup_j \alpha_j$.

Proof. We want to construct a supermartingale with respect to the filtration $\mathcal{F}$. Here started from the construction $\{U_t\}$ provided in Harvey et al. (2019):

$$\begin{aligned}\mathbb{E}[U_t \mid \mathcal{F}_{t-1}] &= \mathbb{E}\left[\exp\left(\sum_{j=1}^{t}(\lambda + c\lambda^2\alpha_j)d_j - \sum_{j=1}^{t}\frac{\tilde{\lambda}^2}{2}v_{j-1}\right)\Big|\mathcal{F}_{t-1}\right] \\ &= \exp\left(\sum_{j=1}^{t-1}(\lambda + c\lambda^2\alpha_j)d_j - \sum_{j=1}^{t}\frac{\tilde{\lambda}^2}{2}v_{j-1}\right)\mathbb{E}\left[\exp\left((\lambda + c\lambda^2\alpha_t)d_t\right)\big|\mathcal{F}_{t-1}\right] \\ &= U_{t-1}\exp\left(-\frac{\tilde{\lambda}^2}{2}v_{t-1}\right)\mathbb{E}\left[\exp\left((\lambda + c\lambda^2\alpha_t)d_t\right)\big|\mathcal{F}_{t-1}\right],\end{aligned}$$

where the second equality comes from the measurability of $\{d_j\}_{j\le t}$ and $\{v_{j-1}\}_{j\le t}$, and the third equality holds by inserting the definition of U_{t-1}. Plugging the condition (1) (the upper bound of $\mathbb{E}\left[\exp\left(\lambda d_t\right)\big|\mathcal{F}_{t-1}\right]$), we can further obtain:

$$\begin{aligned}\mathbb{E}[U_t \mid \mathcal{F}_{t-1}] &= U_{t-1}\exp\left(-\frac{\tilde{\lambda}^2}{2}v_{t-1}\right)\mathbb{E}\left[\exp\left((\lambda + c\lambda^2\alpha_t)d_t\right)\big|\mathcal{F}_{t-1}\right] \\ &\le U_{t-1}\exp\left[C_t\left(\lambda + c\lambda^2\alpha\right) + \left(\frac{(\lambda + c\lambda^2\alpha)^2}{2} - \frac{\tilde{\lambda}^2}{2}\right)\cdot v_{t-1}\right] \\ &= U_{t-1}\exp\left[(E_t - E_{t-1})\left(\lambda + c\lambda^2\alpha\right)\right]\exp\left[\left(\frac{(\lambda + c\lambda^2\alpha)^2}{2} - \frac{\tilde{\lambda}^2}{2}\right)\cdot v_{t-1}\right]\end{aligned}$$

Divided by $\exp\left[E_t\left(\lambda + c\lambda^2\alpha\right)\right]$ on both sides, the inequality becomes

$$\begin{aligned}\mathbb{E}[e^{-E_t\left(\lambda + c\lambda^2\alpha\right)}U_t \mid \mathcal{F}_{t-1}] &\le e^{-E_{t-1}\left(\lambda + c\lambda^2\alpha\right)}U_{t-1}\cdot\exp\left[\left(\frac{(\lambda + c\lambda^2\alpha)^2}{2} - \frac{\tilde{\lambda}^2}{2}\right)\cdot v_{t-1}\right] \\ &\le e^{-E_{t-1}\left(\lambda + c\lambda^2\alpha\right)}U_{t-1},\end{aligned}$$

where to make the equality hold, we set $\tilde{\lambda} = (\lambda + c\lambda^2\alpha)$. It concludes that $\{e^{-E_t\left(\lambda + c\lambda^2\alpha\right)}U_t\}$ is a supermartingale with respect to $\mathcal{F}$. □

Theorem 1.3 (Freedman)**.** *Let $\{d_j\}_{j\in\mathbb{N}_+}$ be an adaptive process with respect to the filtration $\mathcal{F}$. Suppose $v_{j-1} \ge 0$ for $j \in \mathbb{N}_+$ are $\mathcal{F}_{j-1}$-measurable random variables. Assume there exists a sequence $\{C_t\}$ admitting a split $C_t = E_t - E_{t-1}$ for an increasing positive sequence $\{E_t\}$ such that*

$$\mathbb{E}[e^{\lambda d_t}|\mathcal{F}_{t-1}] \le \exp\left(C_t\lambda + \frac{\lambda^2}{2}v_j\right) \tag{2}$$

for all $\lambda > 0$. Then for any $\alpha_j \ge 0$ and $\beta > 0$,

$$\mathbb{P}\left[\bigcup_{t=1}^{T}\left\{\sum_{j=1}^{t}d_j \ge x \text{ and } \sum_{j=1}^{t}v_{j-1} \le \sum_{j=1}^{t}\alpha_j d_j + \beta\right\}\right] \le \exp(E_T\left(\lambda + c\lambda^2\alpha\right))\cdot\exp(-\lambda x + c\lambda^2\beta),$$

holds for all $\lambda \in (0, \frac{1}{2\alpha}]$, where $\alpha := \sup_j \alpha_j$.

Proof. Define the stopping time $\mathscr{T} := \min\{t : \sum_{j=1}^{t} d_j \geq x \text{ and } \sum_{j=1}^{t} v_{j-1} \leq \sum_{j=1}^{t} \alpha_j d_j + \beta\}$.

$$
\begin{aligned}
&\mathbb{P}\left[\bigcup_{t=1}^{T}\left\{\sum_{j=1}^{t} d_j \geq x \text{ and } \sum_{j=1}^{t} v_{j-1} \leq \sum_{j=1}^{t} \alpha_j d_j + \beta\right\}\right] \\
&= \mathbb{P}\left[\sum_{j=1}^{\mathscr{T}\wedge T} d_j \geq x \text{ and } \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1} \leq \sum_{j=1}^{\mathscr{T}\wedge T} \alpha_j d_j + \beta\right] \\
&= \mathbb{P}\left[\lambda \sum_{j=1}^{\mathscr{T}\wedge T} d_j \geq \lambda x \text{ and } c\lambda^2 \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1} \leq c\lambda^2 \sum_{j=1}^{\mathscr{T}\wedge T} \alpha_j d_j + c\lambda^2\beta\right] \\
&\leq \mathbb{P}\left[\sum_{j=1}^{\mathscr{T}\wedge T} (\lambda + c\lambda^2\alpha_j) d_j - c\lambda^2 \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1} \geq \lambda x - c\lambda^2\beta\right] \\
&\leq \mathbb{E}\left[\exp\left(\sum_{j=1}^{\mathscr{T}\wedge T} (\lambda + c\lambda^2\alpha_j) d_j - c\lambda^2 \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1}\right)\right] \cdot \exp(-\lambda x + c\lambda^2\beta),
\end{aligned}
$$

where the last inequality applies the Chernoff's inequality. By Lemma 1.1, there exists c such that $c\lambda^2 = \frac{\tilde{\lambda}^2}{2}$. Then the following bound holds:

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(\sum_{j=1}^{\mathscr{T}\wedge T} (\lambda + c\lambda^2\alpha_j) d_j - c\lambda^2 \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1}\right)\right] &= \mathbb{E}\left[\exp\left(\sum_{j=1}^{\mathscr{T}\wedge T} (\lambda + c\lambda^2\alpha_j) d_j - \frac{\tilde{\lambda}^2}{2} \sum_{j=1}^{\mathscr{T}\wedge T} v_{j-1}\right)\right] \\
&= \mathbb{E}\left[e^{E_{\mathscr{T}\wedge T}\cdot\left(\lambda + c\lambda^2\alpha\right)} \cdot e^{-E_{\mathscr{T}\wedge T}\left(\lambda + c\lambda^2\alpha\right)} U_{\mathscr{T}\wedge T}\right] \\
&\leq e^{E_T\cdot\left(\lambda + c\lambda^2\alpha\right)} \mathbb{E}\left[e^{-\mathscr{T}\wedge T\cdot C\left(\lambda + c\lambda^2\alpha\right)} U_{\mathscr{T}\wedge T}\right] \\
&\leq e^{E_T\cdot\left(\lambda + c\lambda^2\alpha\right)}.
\end{aligned}
$$

Then we obtain the final bound:

$$
\begin{aligned}
&\mathbb{P}\left[\bigcup_{t=1}^{T}\left\{\sum_{j=1}^{t} d_j \geq x \text{ and } \sum_{j=1}^{t} v_{j-1} \leq \sum_{j=1}^{t} \alpha_j d_j + \beta\right\}\right] \\
&\leq \exp(E_T \cdot \left(\lambda + c\lambda^2\alpha\right)) \cdot \exp(-\lambda x + c\lambda^2\beta)
\end{aligned}
$$

□

References

Harvey, N. J., Liaw, C., Plan, Y., and Randhawa, S. (2019). Tight analyses for non-smooth stochastic gradient descent. In Conference on Learning Theory, pages 1579–1613. PMLR.