Correction of the exercise on convex loss functions

Convex loss functions and comparison to the best convex vector

Strategy at hand: $\eta > 0$ and

$$p_t = \int_{\mathcal{X}} p \, e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)} \, d\mu(p) \Big/ \int_{\mathcal{X}} e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)} \, d\mu(p)$$

$$= \int_{\mathcal{X}} p \, d\mu_t(p) \qquad \text{where} \qquad \frac{d\mu_t}{d\mu}(p) = \frac{e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)}}{\int_{\mathcal{X}} e^{-\eta \sum_{s=1}^{t-1} \ell_s(q)} \, d\mu(q)}$$

1) $$\ell_t(p_t) = \ell_t\left(\int p \, d\mu_t(p) \right) \underset{\text{Jensen}}{\leq} \int_{\mathcal{X}} \ell_t(p) \, d\mu_t(p)$$

$$\underset{\substack{\text{Hoeffding,} \\ \text{as for EWA}}}{\leq} -\frac{1}{\eta} \underbrace{\ln \int e^{-\eta \ell_t(p)} d\mu_t(p)}_{\ln \frac{\int e^{-\eta \sum_{s=1}^{t} \ell_s(p)} d\mu(p)}{\int e^{-\eta \sum_{s=1}^{t-1} \ell_s(p)} d\mu(p)}} + \frac{(M-m)^2}{8} \eta$$

Summing over $t = 1, \ldots T$, a telescoping sum appears:

$$\sum_{t=1}^{T} \ell_t(p_t) \leq \underbrace{-\frac{1}{\eta} \ln \frac{\int e^{-\eta \sum_{t=1}^{T} \ell_t(p)} d\mu(p)}{1}}_{} + \frac{(M-m)^2}{8} \eta T$$

can be bounded using the same techniques as for exp. concave losses, but the proof needs to be slightly adapted:

$\delta > 0$ and p^*_δ s.t. $\inf_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p) \leq \delta + \sum_{t=1}^{T} \ell_t(p^*_\delta)$

$\varepsilon > 0$ and $\Delta^*_{\delta, \varepsilon} = \{ (1-\varepsilon) p^*_\delta + \varepsilon r, \quad r \in \mathcal{X} \}$

We still have $\mu(\Delta^*_{\delta,\varepsilon}) = \varepsilon^{N-1}$

 Lecture 3 — Winter 2022 term Lecture 3 — Winter 2022 term

But for $p = (1-\varepsilon)\, p^*_\delta + \varepsilon r$ we can only resort to convexity:

$$\begin{aligned} \ell_t(p) &\leqslant (1-\varepsilon)\, \ell_t(p^*_\delta) + \varepsilon\, \ell_t(r) \\ &\leqslant \ell_t(p^*_\delta) + \varepsilon \big(\underbrace{\ell_t(r) - \ell_t(p^*_\delta)}_{\leqslant M-m} \big) \end{aligned}$$

$\leqslant M-m$ since ℓ_t takes values in $[m, M]$ by assumption

$$e^{-\eta \ell_t(p)} \geqslant e^{-\eta \ell_t(p^*_\delta)} \quad e^{-\eta \varepsilon (M-m)}$$

Putting all things together:

$$\int_{\mathcal{X}} e^{-\eta \sum_{t=1}^{T} \ell_t(p)} \, d\mu(p) \geqslant e^{-\eta \sum_{t=1}^{T} \ell_t(p^*_\delta)} \times e^{-\eta \varepsilon (M-m) T} \times \varepsilon^{N-1}$$

integral only over $\Delta^*_{\delta,\varepsilon}$

Substituting above and taking $\inf_\varepsilon$:

$$\sum_{t=1}^{T} \ell_t(p_t) \leqslant \underbrace{\sum_{t=1}^{T} \ell_t(p^*_\delta)}_{\leqslant \delta + \inf_p \sum_{t=1}^{T} \ell_t(p)} + \inf_{\varepsilon \in (0,1)} \left\{ \varepsilon (M-m) T - \frac{N-1}{\eta} \ln \varepsilon \right\} + \frac{(M-m)^2}{8} \eta T$$

We let $\delta \downarrow 0$ to conclude:

$$\sum_{t=1}^{T} \ell_t(p_t) - \inf_{p \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(p) \leqslant \frac{(M-m)^2}{8} \eta T + \inf_{\varepsilon \in (0,1)} \left(\varepsilon (M-m) T - \frac{N-1}{\eta} \ln \varepsilon \right)$$

2) Optimize first over ε:

$$g(\varepsilon) = \varepsilon (M-m) T - \frac{N-1}{\eta} \ln \varepsilon$$

g strictly convex,
$$\begin{cases} g'(\varepsilon) = (M-m) T - \dfrac{N-1}{\eta \varepsilon} \\ g''(\varepsilon) = \dfrac{N-1}{\eta \varepsilon^2} > 0 \end{cases}$$

a unique minimizer on $(0, +\infty)$ at ε s.t. $g'(\varepsilon) = 0 \iff \underbrace{\varepsilon = \frac{N-1}{\eta (M-m) T}}$

but question is whether this ε is in $(0,1)$!

 Lecture 3 — Winter 2022 term Lecture 3 — Winter 2022 term

I tried with this value of ε (which is ok for large T) but couldn't get a simple and readable $O(\sqrt{NT\ln T})$ bound.

Let's not optimize over ε and take an arbitrary choice: $\varepsilon = 1/\sqrt{T}$

The bound is $\leqslant \dfrac{(M-m)^2}{8}\eta T + (M-m)\sqrt{T} + \dfrac{N-1}{2\eta}\ln T$

Optimal value for η: η^* s.t. (as seen in class) $\dfrac{(M-m)^2}{8} T \eta^* = \dfrac{N-1}{2\eta^*}\ln T$

and for this η^*, the sum is $2 \times \sqrt{\text{the product}}$

$$= \underbrace{\frac{2}{\sqrt{16}}}_{=1/2} (M-m)\sqrt{(N-1)T\ln T}$$

Final bound: $\dfrac{1}{2}(M-m)\sqrt{(N-1)T\ln T} + (M-m)\sqrt{T}$.

$\rightsquigarrow$ If you can proceed better, please send me your solution (and you may be rewarded with bonus points at the exam).