Correction de l'exercice 1

1) $\phi'(\eta) = \eta^{-2} \sum_i q_i \chi\big(\exp(\eta a_i)/\sum_j q_j \exp(\eta a_j)\big) \geq 0$ par étude de la fonction χ.

2) Soit $\ell_{i,t} = \ell(y_t, f_{i,t})$. Nous avons

$$\begin{aligned}
\mathbb{E}\sum_t \ell(y_t, d_t) &= \mathbb{E}\sum_t \sum_i p_{i,t}\ell_{i,t} \\
&= \mathbb{E}\sum_t \left(\frac{1}{\eta_t}\log\sum_i p_{i,t}e^{\eta_t(\sum_j p_{j,t}\ell_{j,t}-\ell_{i,t})} - \frac{1}{\eta_t}\log\sum_i p_{i,t}e^{-\eta_t\ell_{i,t}}\right) \\
&= \mathbb{E}\sum_t \left(\frac{1}{\eta_t}\log\mathbb{E}e^{\eta_t(V_t-\mathbb{E}V_t)} - \frac{1}{\eta_t}\log\frac{\sum_i e^{-\eta_t L_{i,t}}}{\sum_i e^{-\eta_t L_{i,t-1}}}\right) \qquad \mathbb{P}(V_t = -\ell_{i,t}) = p_{i,t} \\
&\leq \sum_t \frac{\eta_t}{8} - \frac{1}{\eta_n}\mathbb{E}\log\left(\frac{1}{K}\sum_j e^{-\eta_n L_{j,n}}\right) \qquad \text{Hoeffding} \\
&\quad + \sum_{t=1}^{n-1}\left(\frac{1}{\eta_{t+1}}\mathbb{E}\log\left(\frac{1}{K}\sum_j e^{-\eta_{t+1}L_{j,t}}\right) - \frac{1}{\eta_t}\mathbb{E}\log\left(\frac{1}{K}\sum_j e^{-\eta_t L_{j,t}}\right)\right) \\
&\leq \frac{\sum_{t=1}^n \eta_t}{8} + \frac{\log K}{\eta_n} + \mathbb{E}\min_j L_{j,n} \\
&\leq \frac{1}{4}\sqrt{\log K}\sum_{t=1}^n \frac{1}{\sqrt{t}} + \sqrt{\frac{n\log K}{4}} + \mathbb{E}\min_j L_{j,n} \\
&\leq \frac{1}{2}\sqrt{\log K}\sqrt{n} + \sqrt{\frac{n\log K}{4}} + \mathbb{E}\min_j L_{j,n} \\
&\leq \sqrt{n\log K} + \mathbb{E}\min_j L_{j,n}.
\end{aligned}$$

3) Par rapport à la stratégie de base à poids exponentiels: la stratégie ci-dessus n'a pas besoin de connaître l'horizon (comme celle utilisant le doubling trick). Par rapport à la procédure utilisant le doubling trick: meilleure constante + procédure évitant de réinitialiser brutalement l'algorithme aux temps de la forme 2^m, ce qui paraît inacceptable en pratique.

Correction de l'exercice 2

1) Employer la stratégie à poids exponentiels sur la classe $\mathcal{G}$ permet de garantir un regret $R_n \leq \sqrt{\frac{n\ln|\mathcal{G}|}{2}} \leq \sqrt{\frac{n\ln(K^s n^{s-1})}{2}}$.

2) Pour $s = 1$, nous retrouvons le résultat du cours, $\mathcal{G}$ étant alors l'ensemble des K experts initiaux.