# ExpByOpt for Convex Bandits

June 6, 2025

## 1 Setup and Notation

Reuse the notation from Lattimore [2024].

**Assumption 1.** The following hold:

- The losses are in $\mathcal{F}_b$.
- There is no noise so that $Y_t = f_t(X_t)$.

**Assumption 2.** $\mathcal{C}$ is finite subset of $K$ such that:

- $\log(\mathcal{C}) \leq \tilde{\mathcal{O}}(d)$.
- For all $f \in \mathcal{F}_b$ there exists $x \in \mathcal{C}$ such that $f(x) \leq \inf_{x' \in K} f(x') + \frac{1}{n}$.

## 2 Exponential Weights with Importance Sampling

*Here, it's better to think of $\hat{s}_t$ as the losses observed at time t and not the estimates of the losses.* Let $(\hat{s}_t)_{t=1}^n : \mathcal{C} \to \mathbb{R}$ be a sequence of functions and

$$q_t(x) = \frac{\exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(x))}{\sum_{y \in \mathcal{C}} \exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(y))}\,, \quad x \in \mathcal{C}\,. \tag{1}$$

The following theorem gives a bound on the regret of the exponential weights algorithm when the whole loss vector $\hat{s}_t$ (or better say $f_t$) is observed.

**Theorem 3** (Lattimore [2024], Theorem 8.11)**.** *For any $y \in \mathcal{C}$ we have*

$$\sum_{t=1}^n \langle q_t, \hat{s}_t \rangle - \hat{s}_t(y) \leq \frac{\log(|\mathcal{C}|)}{\eta} + \frac{1}{\eta} \sum_{t=1}^n \mathcal{S}_t(\eta \hat{s}_t)\,,$$

*where $\mathcal{S}_t(u) = D_{R^\star}(R'(q_t) - u, R'(q_t))$.*

The following theorem applies to the setting when $\langle f_t, X_t \rangle$ is observed and not $f_t$. The idea is to use importance sampling and estimates $\hat{s}_t$ to compute the losses.

```
args: learning rate η > 0
let C ∈ K be finite
for t = 1 to n:
    compute q_t(x) = exp(−η Σ_{u=1}^{t−1} ŝ_u(x)) / Σ_{y∈C} exp(−η Σ_{u=1}^{t−1} ŝ_u(x)) for all x ∈ C
    find distribution p_t as a function of q_t
    sample X_t ∼ p_t, and observe Y_t = f_t(X_t)
    compute ŝ_t(x) ∀x ∈ C using p_t, q_t, X_t, Y_t
```

**Algorithm 1:** Exponential Weights with Importance Sampling

**Theorem 4** (Lattimore [2024], Theorem 8.14)**.** *Let* $x_\star = \arg\min_{x\in\mathcal{C}} \sum_{t=1}^n f_t(x)$ *and* $p_\star \in \Delta(\mathcal{C})$ *be a Dirac on* $x_\star$*. The expected regret of Algorithm 1 is bounded by*

$$\mathbb{E}\left[\mathfrak{R}_n(x_\star)\right] = \frac{\log(|\mathcal{C}|)}{\eta} + \sum_{t=1}^n \mathbb{E}\left[\langle p_t - p_\star, f_t\rangle + \langle p_\star - q_t, \hat{s}_t\rangle + \frac{1}{\eta}\mathcal{S}_t(\eta\hat{s}_t)\right].$$

*Proof.* We have

$$\begin{aligned}
\mathbb{E}\left[\mathfrak{R}_n(x_\star)\right] &= \mathbb{E}\left[\sum_{t=1}^n \langle p_t - p_\star, f_t\rangle\right] \\
&= \mathbb{E}\left[\sum_{t=1}^n \underbrace{\langle p_t - p_\star, f_t\rangle}_{\text{actual}} - \underbrace{\langle q_t - p_\star, \hat{s}_t\rangle}_{\text{shadow}} + \langle q_t - p_\star, \hat{s}_t\rangle\right] \\
&\leq \frac{\log(|\mathcal{C}|)}{\eta} + \mathbb{E}\left[\sum_{t=1}^n \langle p_t - p_\star, f_t\rangle - \langle q_t - p_\star, \hat{s}_t\rangle + \frac{1}{\eta}\mathcal{S}_t(\eta\hat{s}_t)\right]
\end{aligned}$$

where the last inequality follows from Theorem 3. □

# 3 Exploration By Optimisation

The idea is to bound the term inside the expectation in Theorem 4 uniformly over all $t \in [n]$. For the learner, the degree of freedom is in the choice of the estimator $\hat{s}_t$ and the exploration distribution $p_t$. To this end, define

- $\mathcal{G}$ be the set of all functions $g : \mathcal{C} \to \mathbb{R}$, i.e., $\mathcal{G} = \mathbb{R}^{|\mathcal{C}|}$ (all loss vectors),
- $\mathcal{E}$ be the set of all functions $E : \mathcal{C} \times \mathbb{R} \to \mathcal{G}$ (choice of $\hat{s}$),
- $\Lambda_\eta(q, p, E, r, f) = \frac{1}{\eta}\mathbb{E}_{X\sim p}\Big[\langle p - r, f\rangle + \langle r - q, E(X, Y)\rangle + \frac{1}{\eta}\mathcal{S}_q(\eta E(X, Y))\Big]$,

where $q \in \Delta(\mathcal{C})$ is the exponential weights distribution at time $t$, $p \in \Delta(\mathcal{C})$ is the exploration distribution, $E \in \mathcal{E}$ is the estimator, $r \in \Delta(\mathcal{C})$ is the Dirac on the best action, and $f \in \mathcal{F}_b$ is the loss function. Note that the randomness is only through $X$ and $Y$, where $X \sim p$ and $Y = f(X)$.

**Remark 5.** There is a bit of waste later since we suppose that $r \in \Delta(\mathcal{C})$, but it could've been restricted to Dirac distributions.

**args**: learning rate $\eta > 0$, precision $\epsilon > 0$
let $\mathcal{C} \in K$ be finite
**for** $t = 1$ **to** $n$:
    compute $q_t(x) = \frac{\exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(x))}{\sum_{y \in \mathcal{C}} \exp(-\eta \sum_{u=1}^{t-1} \hat{s}_u(x))}$ **for** all $x \in \mathcal{C}$
    find distribution $p_t \in \Delta(\mathcal{C})$ and $E_t \in \mathcal{E}$ such that
        $\Lambda_\eta(q_t, p_t, E_t) \leq \inf_{p,E} \Lambda_\eta(q_t) + \epsilon$
    sample $X_t \sim p_t$, and observe $Y_t = f_t(X_t)$
    compute $\hat{s}_t(x) = E_t(X_t, Y_t)$

**Algorithm 2:** Exploration by Optimisation

It's worth noting that Algorithm 2 is a instantiation of Algorithm 1 where the estimator $\hat{s}_t$ and the exploration distribution $p_t$ are chosen by solving an optimisation problem.

**Theorem 6** (Lattimore [2024], Theorem 8.15)**.** *The expected regret of Algorithm 2 is bounded by*

$$\mathbb{E}\left[\mathfrak{R}_n(x_\star)\right] \leq \frac{\log(|\mathcal{C}|)}{\eta} + n\eta \sup_{q \in \Delta(\mathcal{C})} \Lambda_\eta(q) + n\eta\epsilon\,.$$

Therefore, from Theorem 6, to bound the regret of Algorithm 2 it suffices to bound

$$\Lambda^\star_{\mathcal{C}} = \sup_{\eta > 0, q \in \Delta(\mathcal{C})} \inf_{p \in \Delta(\mathcal{C}), E \in \mathcal{E}} \sup_{y \in \mathcal{C}, f \in \mathcal{F}_b} \Lambda_\eta(q, p, E, y, f)\,. \tag{2}$$

The following lemma shows that the order of the infimum and supremum can be interchanged.

**Lemma 7.** *For all $\eta > 0$ and $q \in \Delta(\mathcal{C})$, we have*

$$\inf_{p \in \Delta(\mathcal{C}), E \in \mathcal{E}} \sup_{r \in \Delta(\mathcal{C}), f \in \mathcal{F}_b} \Lambda_\eta(q, p, E, r, f) = \sup_{r \in \Delta(\mathcal{C}), f \in \mathcal{F}_b} \inf_{p \in \Delta(\mathcal{C}), E \in \mathcal{E}} \Lambda_\eta(q, p, E, r, f)\,.$$

This lemma is wrong though... The minimax theorem only applies if the sup player should can choose from a joint distribution over $r$ and $f$.

Let $\nu \in \Delta(\Delta(\mathcal{C}) \times \mathcal{F}_b)$ be a joint distribution over $p_\star \in \Delta(\mathcal{C})$ and $f \in \mathcal{F}_b$. *$p_\star$ is the same thing as $r$...* Define the bayesian version of $\Lambda_\eta(q, p, E, \nu)$ as

$$\bar{\Lambda}_\eta(q, p, E, \nu) := \frac{1}{\eta}\mathbb{E}_{X \sim p, (p_\star, f) \sim \nu}\left[\langle p_\star - q, E(X, Y)\rangle + \langle p_\star - p, f\rangle + \frac{1}{\eta}\mathcal{S}_q(\eta E(X, Y))\right]\,.$$

It's easy to see that

$$\inf_{p \in \Delta(\mathcal{C}), E \in \mathcal{E}} \sup_{r \in \Delta(\mathcal{C}), f \in \mathcal{F}_b} \Lambda_\eta(q, p, E, r, f) \leq \inf_{p \in \Delta(\mathcal{C}), E \in \mathcal{E}} \sup_{\nu \in \Delta(\Delta(\mathcal{C}) \times \mathcal{F}_b)} \bar{\Lambda}_\eta(q, p, E, \nu)\,.$$

**Lemma 8** (Lattimore [2024], Lemma 8.21)**.** *For all $\eta > 0$ and $q \in \Delta(\mathcal{C})$, we have*

$$\inf_{p\in\Delta(\mathcal{C}),E\in\mathcal{E}} \sup_{\nu\in\Delta(\Delta(\mathcal{C})\times\mathcal{F}_b)} \bar{\Lambda}_\eta(q,p,E,\nu) = \sup_{\nu\in\Delta(\Delta(\mathcal{C})\times\mathcal{F}_b)} \inf_{p\in\Delta(\mathcal{C}),E\in\mathcal{E}} \bar{\Lambda}_\eta(q,p,E,\nu)\,.$$

In the following two subsections we will solve for the inf in the RHS of Theorem 8. In the following $q \in \Delta(\mathcal{C})$ is fixed and we will drop it from the notation.

## 3.1 Simplifying $\Lambda_\eta$

**Remark 9.** We specialize to $R$ being the unnormalised negentropy function, so that we can make the connection to the KL term in IDS.

Note that for $a \in \Delta(\mathcal{C})$ we have

$$R^\star(a) = \sum_{x\in\mathcal{C}} \exp(a(x))\,,$$

and $R'(a) = \ln(a)$, which means that

$$R^\star(R'(a)) = \sum_{x\in\mathcal{C}} a(x) = 1\,.$$

Suppose that $X \sim p$, $Y = f(X)$, and $(p_\star, f) \sim \nu$, and define

$$\Delta(\nu,p) = \mathbb{E}[\langle p_\star - p, f\rangle]\,,$$

Then we can write

$$\begin{aligned}
&\eta\Lambda_\eta(q,p,E,\nu)\\
&= \mathbb{E}\left[\langle p_\star - q, E(X,Y)\rangle + \langle p_\star - p, f\rangle + \frac{1}{\eta}\mathcal{S}_q(\eta E(X,Y))\right]\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star - q, E(X,Y)\rangle + \frac{1}{\eta}\left(R^\star\left(R'(q) - \eta E(X,Y)\right) - R^\star(R'(q)) - \nabla R^\star(R'(q))^\top(-\eta E(X,Y))\right)\right]\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star - q, E(X,Y)\rangle + \frac{1}{\eta}\left(R^\star\left(\ln(q) - \eta E(X,Y)\right) - 1 + \eta\langle q, E(X,Y)\rangle\right)\right]\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}R^\star\left(\ln(q) - \eta E(X,Y)\right)\right] - \frac{1}{\eta}\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}\sum_{z\in\mathcal{C}}\exp\left(\ln(q_z) - \eta E(X,Y)_z\right)\right] - \frac{1}{\eta}\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}\sum_{z\in\mathcal{C}} q_z\exp\left(-\eta E(X,Y)_z\right)\right] - \frac{1}{\eta}\\
&= \Delta(\nu,p) + \mathbb{E}\left[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}\langle q, \exp(-\eta E(X,Y))\rangle\right] - \frac{1}{\eta}
\end{aligned}$$

### 3.2 Unrestricted estimator class $\mathcal{E}$

The minimizer $E$ in the general case has the following close form.

**Lemma 10.** *Given $\nu \in \Delta(\Delta(\mathcal{C}), \mathcal{F}_b)$, and $p \in \Delta(\mathcal{C})$ define $G_{\nu,p} \in \mathcal{E}$ as*

$$G_{\nu,p}(x,y) = \frac{1}{\eta}\left(R'(q) - R'(\mathbb{E}[p_\star | f(x) = y])\right) ,$$

*then we have* $\inf_{E \in \mathcal{E}} \Lambda_\eta(p, E, \nu) = \Lambda_\eta(p, G_{\nu,p}, \nu)$.

*Proof.* By taking the derivative of $\Lambda_\eta(p, E, \nu)$ with respect to $E(x,y)$ we have

$$\begin{aligned}
&\nabla_{E(x,y)} \Lambda_\eta(p, E, \nu) \\
&= \tfrac{\mathbb{P}(f(x)=y, X=x)}{\eta} \nabla_{E(x,y)} \mathbb{E}\Big[\langle p_\star, E(X,Y)\rangle + \tfrac{1}{\eta}\langle q, \exp(-\eta E(X,Y))\rangle | X = x, f(x) = y\Big] \\
&= \tfrac{\mathbb{P}(f(x)=y, X=x)}{\eta} \mathbb{E}[p_\star - q \odot \exp(-\eta E(X,Y)) | X = x, f(x) = y] \\
&= \tfrac{\mathbb{P}(f(x)=y, X=x)}{\eta} \mathbb{E}[p_\star | X = x, f(x) = y] - \tfrac{1}{\eta} q \odot \exp(-\eta E(x,y)) .
\end{aligned}$$

By setting the gradient to zero, for any $z \in \mathcal{C}$ we have

$$\mathbb{E}[p_\star | X = x, f(x) = y]_z = q_z \exp(-\eta E(x,y))_z ,$$

which gives

$$\begin{aligned}
E(x,y)_z &= -\frac{1}{\eta} \ln\left(\frac{\mathbb{E}[p_\star | X = x, f(x) = y]_z}{q_z}\right) \\
&= \frac{1}{\eta}\left(\ln(q_z) - \ln\left(\mathbb{E}[p_\star | X = x, f(x) = y]_z\right)\right) ,
\end{aligned}$$

which, since $R'(q_z) = \ln(q_z)$ gives the desired result. □

### 3.3 IS estimator class–multiplicative in $y$

Restrict $E(x,y)$ to be of the form

$$E(x,y) = \frac{y}{p(x)} T(x) ,$$

where a special case could be a probability kernel as

$$\begin{aligned}
E(x,y) &= \frac{T(x|\cdot)y}{p(x)} \ (= \hat{s}_x(\cdot)) , \\
\text{and} \quad & \sum_{x \in \mathcal{C}} T(x|y) = 1 \quad \text{for all } y \in \mathcal{C} ,
\end{aligned}$$

which includes all importance sampling estimators that are **multiplicative** in y. Let $\nu \in \Delta(\Delta(\mathcal{C}) \times \mathcal{F}_b)$ and $p \in \Delta(\mathcal{C})$. We are interested in the $T$ that minimizes

$$\begin{aligned}
\eta\bar{\Lambda}(T) &:= \Delta(\nu, p) + \mathbb{E}\left[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}\langle q, \exp(-\eta E(X,Y))\rangle\right] - \frac{1}{\eta} \\
&:= \Delta(\nu, p) + \mathbb{E}\left[\mathbb{E}\left[\langle p_\star, \tfrac{T(X|\cdot)Y}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta \tfrac{T(X|\cdot)Y}{p(X)})\rangle \Big| X\right]\right] - \frac{1}{\eta} .
\end{aligned}$$

Note that we drop the sampling notation but it's important to remember that $f, p_\star, X$ and $Y$ are random. Define the shorthand $T_x = T(x|\cdot) \in \mathbb{R}^{|\mathcal{C}|}$. Then through differentiation we have

$$\begin{aligned}
\nabla_{T_x}\bar{\Lambda}(T) &= \tfrac{p(x)}{\eta}\nabla_{T_x}\mathbb{E}\Big[\langle p_\star, \tfrac{T_x Y}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta \tfrac{T_x Y}{p(X)})\rangle | X = x\Big] \\
&= \tfrac{p(x)}{\eta}\mathbb{E}\Big[\tfrac{Y}{p(X)}p_\star - \tfrac{Y}{p(X)} q \odot \exp(-\tfrac{\eta Y}{p(X)}T_X) | X = x\Big] \\
&= \tfrac{1}{\eta}\mathbb{E}\Big[Y\left(p_\star - q \odot \exp(-\tfrac{\eta Y}{p(X)}T_X)\right) | X = x\Big] \;,
\end{aligned}$$

where the expectation is over $(p_\star, f) \sim \nu$. We want to solve for $T_x$ such that the gradient is zero, so for coordinate $z \in \mathcal{C}$ we have

$$q(z)\mathbb{E}\Big[f(X)\exp(-\tfrac{\eta f(X)}{p(X)}T(X|z))|x\Big] = \mathbb{E}[f(X)p_{\star,z}|x] = \mathbb{P}(x_\star = z)\mathbb{E}[f(X)|x_\star = z, x] \tag{3}$$

If $p_\star$ is a Dirac on $z$ almost surely, then the RHS is $\mathbb{E}[f(X)|X = x]$, while if $T(x|z) \geq 0$ then the LHS is bounded by

$$q(z)\mathbb{E}\Big[f(X)\exp(-\tfrac{\eta f(X)}{p(X)}T(X|z))|X = x\Big] \leq q(z)\mathbb{E}[f(X)|X = x] \;.$$

**Remark 11.** If $T(x|\cdot) \geq 0$, the gradient can't be zero in the general case.

**Remark 12.** The LHS is the first derivative of the moment generating function of $f(X)|X = x$ evaluated at $-\frac{\eta}{p(X)}T(x|z)$.

- **Closed form for $T(x|z)$?** It can be seen that it's unlikely that $T(x|z)$ has a closed form. However, there might be a chance to use Eq. (3) when this is evaluated on $\bar{\Lambda}$:

$$\begin{aligned}
\eta\bar{\Lambda}(T) &= \Delta(\nu, p) + \mathbb{E}\Big[\langle p_\star, E(X,Y)\rangle + \frac{1}{\eta}\langle q, \exp(-\eta E(X,Y))\rangle\Big] - \frac{1}{\eta} \\
&= \Delta(\nu, p) + \mathbb{E}\Big[\langle p_\star, \tfrac{T(X|\cdot)Y}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T(X|\cdot)Y}{p(X)})\rangle\Big] - \frac{1}{\eta} \\
&= \Delta(\nu, p) + \mathbb{E}\Big[\langle f(X)p_\star, \tfrac{T(X|\cdot)}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T(X|\cdot)Y}{p(X)})\rangle\Big] - \frac{1}{\eta} \;,
\end{aligned} \tag{4}$$

For the first term, we can use Eq. (3) to get

$$\begin{aligned}
\mathbb{E}\Big[\langle f(X)p_\star, \tfrac{T(X|\cdot)}{p(X)}\rangle | X = x\Big] &= \Big\langle \mathbb{E}[f(X)p_\star | X = x] \,,\, \tfrac{T(x|\cdot)}{p(x)}\Big\rangle \\
&= \Big\langle \mathbb{E}\Big[f(x) q \odot \exp(-\tfrac{\eta f(x)}{p(x)}T(x|\cdot))\Big] \,,\, \tfrac{T(x|\cdot)}{p(x)}\Big\rangle \\
&= \Big\langle q \odot \mathbb{E}\Big[f(x)\exp(-\tfrac{\eta f(x)}{p(x)}T(x|\cdot))\Big] \,,\, \tfrac{T(x|\cdot)}{p(x)}\Big\rangle \\
&= \mathbb{E}\Big[Y\Big\langle q, \tfrac{T(x|\cdot)}{p(x)} \odot \exp(-\tfrac{\eta Y}{p(X)}T(X|\cdot))\Big\rangle | X = x\Big] \;,
\end{aligned}$$

putting this back into the expectation in Eq. (4) gives

$$
\begin{aligned}
&\mathbb{E}\left[\langle f(X)p_\star, \tfrac{T(X|\cdot)}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta \tfrac{T(X|\cdot)Y}{p(X)})\rangle\right] \\
=&\mathbb{E}\left[Y\left\langle q, \tfrac{T(X|\cdot)}{p(X)} \odot \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)})\right\rangle + \frac{1}{\eta}\langle q, \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)})\rangle\right] \\
=&\mathbb{E}\left[Y\left\langle q \odot \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)}), \tfrac{T(x|\cdot)}{p(X)}\right\rangle + \frac{1}{\eta}\langle q \odot \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)}), 1_{\mathcal{C}}\rangle\right] \\
=&\mathbb{E}\left[\left\langle q \odot \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)}), Y\tfrac{T(X|\cdot)}{p(X)} + \frac{1}{\eta}1_{\mathcal{C}}\right\rangle\right] ,
\end{aligned}
$$

which makes the whole thing

$$
\bar{\Lambda}(T) = \eta^{-1}\Delta(\nu, p) + \eta^{-2}\left(\mathbb{E}\left[\left\langle q \odot \exp(-\tfrac{\eta T(X|\cdot)Y}{p(X)}), \tfrac{\eta T(X|\cdot)Y}{p(X)} + 1_{\mathcal{C}}\right\rangle\right] - 1\right) .
$$

Recall that $q$ was fixed through the minimax step, so we still have

$$
\begin{aligned}
q_z &= \frac{\exp(-\eta\sum_{u=1}^{t-1}\hat{s}_u(z))}{\sum_{y\in\mathcal{C}}\exp(-\eta\sum_{u=1}^{t-1}\hat{s}_u(y))} \\
&= \frac{\exp(-\eta\sum_{u=1}^{t-1}E_u(X_u, Y_u)_z)}{\sum_{h\in\mathcal{C}}\exp(-\eta\sum_{u=1}^{t-1}E_u(X_u, Y_u)_h)} \\
&= \frac{\exp(-\eta\sum_{u=1}^{t-1}Y_u\frac{T_u(X_u|z)}{p_u(z)})}{\sum_{h\in\mathcal{C}}\exp(-\eta\sum_{u=1}^{t-1}Y_u\frac{T_u(X_u|h)}{p_u(h)})} ,
\end{aligned}
$$

This should play well with the new exponential term.

- **The other way around?** Continuing from Eq. (3) we have

$$\begin{aligned}
q(z)\mathbb{E}\Big[f(X)\exp(-\tfrac{\eta f(X)}{p(X)}T(X|z))|X=x\Big] &= \mathbb{E}[f(X)p_{\star,z}|X=x] \\
\Rightarrow \sum_{z\in\mathcal{C}} q(z)\mathbb{E}\Big[f(X)\exp(-\tfrac{\eta f(X)}{p(X)}T(X|z))|X=x\Big] &= \sum_{z\in\mathcal{C}} \mathbb{E}[f(X)p_{\star,z}|X=x] \\
\Rightarrow \Big\langle q, \mathbb{E}\Big[f(X)\exp(-\tfrac{\eta f(X)}{p(X)}T(X|\cdot))|X=x\Big]\Big\rangle &= \mathbb{E}[f(X)|X=x] \\
\Rightarrow \mathbb{E}\Big[f(x)\Big\langle q, \exp(-\tfrac{\eta f(x)}{p(x)}T(x|\cdot))\Big\rangle\Big] &= \mathbb{E}[f(x)]
\end{aligned}$$

Evaluating the expectation in Eq. (4) gives

$$\mathbb{E}\Big[\langle p_\star, \tfrac{T(X|\cdot)f(X)}{p(X)}\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T(X|\cdot)f(X)}{p(X)})\rangle\Big] = \mathbb{E}\Big[\langle p_\star, \tfrac{T(X|\cdot)f(X)}{p(X)}\rangle + \frac{1}{Y\eta}Y\langle q, \exp(-\eta\tfrac{T(X|\cdot)f(X)}{p(X)})\rangle\Big]$$

### 3.4 IS estimator–linear in $y$

Restrict $E(x,y)$ to be of the form

$$E(x,y) = \frac{y}{p(x)}T(x) + b(x)\,,$$

which includes all importance sampling estimators that are **linear** in $y$. We are interested in the $T$ that minimizes

$$\eta\bar{\Lambda}(T,b) := \Delta(\nu,p) + \mathbb{E}\Big[\mathbb{E}\left[\langle p_\star, \tfrac{T(X)Y}{p(X)} + b(X)\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T(X)Y}{p(X)} + b(X))\rangle\Big| X\right]\Big] - \frac{1}{\eta}\,.$$

Then through differentiation we have

$$\begin{aligned}
\nabla_{T_x}\eta\bar{\Lambda}(T) &= \nabla_{T_x}\mathbb{E}\Big[\langle p_\star, \tfrac{T_xY}{p(X)} + b_x\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T_xY}{p(X)} + b_x)\rangle\Big] \\
&= \mathbb{E}\Big[\tfrac{Y}{p(x)}p_\star - \tfrac{Y}{p(x)}q\odot\exp(-\tfrac{\eta Y}{p(x)}T_x + b_x)\Big]\,,
\end{aligned}$$

and

$$\begin{aligned}
\nabla_{b_x}\eta\bar{\Lambda}(T) &= \nabla_{b_x}\mathbb{E}\Big[\langle p_\star, \tfrac{T_xY}{p(X)} + b_x\rangle + \frac{1}{\eta}\langle q, \exp(-\eta\tfrac{T_xY}{p(X)} + b_x)\rangle\Big] \\
&= \mathbb{E}\Big[p_\star - \tfrac{1}{\eta}q\odot\exp(-\tfrac{\eta Y}{p(x)}T_x + b_x)\Big]\,,
\end{aligned}$$

where the expectation is over $(p_\star, f)\sim\nu$. We want to find $T_x$ such that the gradient is zero, so for coordinate $z\in\mathcal{C}$ we have

$$q(z)\mathbb{E}\Big[f(x)\exp(-\tfrac{\eta f(x)}{p(x)}T_{x,z} + b_{x,z})\Big] = \mathbb{E}[f(x)p_{\star,z}] = \mathbb{P}(x_\star = z)\mathbb{E}[f(x)|x_\star = z]\,.$$

# 4 Archive(useless for now)

$$
\begin{aligned}
&\nabla_{T_x}\Lambda(T)\\
&= \nabla_{T_x}\mathbb{E}\left[\langle p - p_\star, f\ \rangle + \langle p_\star - q, \tfrac{T(X|\cdot)Y}{p(X)}\rangle + \frac{1}{\eta}\mathcal{S}_q\left(\eta\tfrac{T(X|\cdot)Y}{p(X)}\right)\middle| X = x\right]\\
&= \nabla_{T_x}\mathbb{E}\left[\langle p_\star - q, \tfrac{T_xY}{p(X)}\rangle + \frac{1}{\eta}\left(R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right) - R^\star(R'(q)) - \nabla R^\star(R'(q))^\top(-\eta\tfrac{T_xY}{p(X)})\right)\middle| X = x\right]\\
&= \nabla_{T_x}\mathbb{E}\left[\langle p_\star, \tfrac{T_xY}{p(X)}\rangle - \tfrac{Y}{p(X)}q^\top T_x + \frac{1}{\eta}\left(R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right) + \tfrac{\eta Y}{p(X)}q^\top T_x\right)\middle| X = x\right]\\
&= \nabla_{T_x}\mathbb{E}\left[\langle p_\star, \tfrac{T_xY}{p(X)}\rangle + \frac{1}{\eta}R^\star\left(R'(q) - \eta\tfrac{T_xY}{p(X)}\right)\middle| X = x\right]\\
&= \mathbb{E}\left[\frac{Y}{p(x)}\Big(p_\star - \nabla R^\star\Big(R'(q) - \eta\,\tfrac{T_x\,Y}{p(x)}\Big)\Big)\,\middle|\, X = x\right]\\
&= \frac{1}{p(x)}\mathbb{E}\left[f(x)\Big(p_\star - \exp\Big(R'(q) - \eta\,\tfrac{T_x\,f(x)}{p(x)}\Big)\Big)\right]\\
&= \frac{1}{p(x)}\left(\mathbb{E}[f(x)p_\star] - \mathbb{E}\left[f(x)\exp(R'(q) - \tfrac{\eta}{p(x)}f(x)T_x)\right]\right)\\
&= \frac{1}{p(x)}\left(\mathbb{E}[f(x)p_\star] - \mathbb{E}\left[f(x)q\exp(-\tfrac{\eta}{p(x)}f(x)T_x)\right]\right)
\end{aligned}
$$

we want to solve for $T_x$ such that the gradient is zero. Therefore, for coordinate $z \in \mathcal{C}$ we have

$$
\mathbb{E}\left[f(x)q_z\exp\Big(-\tfrac{\eta}{p(x)}f(x)T_{x,z}\Big)\right] = \mathbb{E}[f(x)p_{\star,z}] = \mathbb{P}(x_\star = z)\mathbb{E}[f(x)|x_\star = z]\,,
$$

which means we need to set $T_{x,z}$ such that the following equation holds

$$
\mathbb{E}\left[f(x)\exp\Big(-\tfrac{\eta}{p(x)}f(x)T_{x,z}\Big)\right] = \tfrac{\mathbb{P}(x_\star = z)}{q_z}\mathbb{E}[f(x)|x_\star = z]
$$

while

$$
\begin{aligned}
T_{x,z} = 0 \quad &\text{then } \mathbb{E}\left[f(x)\exp\Big(-\tfrac{\eta}{p(x)}f(x)T_{x,z}\Big)\right] = \mathbb{E}[f(x)]\\
\text{and} \quad T_{x,z} \to \infty \quad &\text{then } \mathbb{E}\left[f(x)\exp\Big(-\tfrac{\eta}{p(x)}f(x)T_{x,z}\Big)\right] \to 0\,.
\end{aligned}
$$

# References


Tor Lattimore. Bandit convex optimisation. *arXiv preprint arXiv:2402.06535*, 2024.