Solution for Exercise #1.

1) For $q \in (0,1)$ ie, excluding for the time being $q=0$ or $q=1$:

$p \in [0,1] \mapsto kl(p,q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}$ is twice differentiable at least on $(0,1)$,

with $\frac{\partial kl}{\partial p}(p,q) = \ln\left(\frac{p(1-q)}{q(1-p)}\right)$ and $\frac{\partial^2}{\partial p^2} kl(p,q) = \frac{1}{p(1-p)} \geqslant 4$

so that a second-order Taylor expansion ensures:

$\forall p \in [\min(p,q), \max(p,q)]$, $kl(p,q) = \cancel{kl(q,q)} + (p-q)\cancel{\frac{\partial}{\partial p} kl(q,q)} + \left(\frac{p-q}{2}\right)^2 \frac{\partial^2 kl(r,q)}{\partial p^2}$

$\geqslant 2(p-q)^2$.

We deal separately with $p \in \{0,1\}$ or $q \in \{0,1\}$.

Note: Sharper Pinsker lower bounds are possible, ie, eg, the excerpt of an article reproving the (optimal in some sense) refined Pinsker's bound by Ordentlich and Weinberger (2005)

2) By the data-processing inequality with expectations of $[0,1]$-valued random variables:

$KL(P,Q) \geqslant kl(E_P[Z], E_Q[Z]) \geqslant 2(E_P[Z] - E_Q[Z])^2$.

(by 1)

Rearranging and taking the sup over Q leads to the desired bound.

3) Note that $E(\nu) = E_\nu[X]$ where $X \sim \nu$ is indeed a $[0,1]$-valued random variable given that we consider the model $\mathcal{P}([0,1])$

Thus $KL(\nu_a, \nu) \geqslant 2(\underbrace{E(\nu_a)}_{=\mu_a < \mu^*} - \underbrace{E(\nu)}_{>\mu^*})^2 > 2(\mu^* - \mu_a)^2 = 2\Delta_a^2$

Taking the infimum over $\nu \in \mathcal{P}([0,1])$ with $E(\nu) > \mu^*$: $Kinf(\nu_a, \mu^*, \mathcal{P}([0,1])) \geqslant 2\Delta_a^2$

Thus the $\frac{\ln T}{Kinf(\nu_a, \mu^*, \mathcal{P}([0,1]))}$ bound is better than the UCB bounds of the form $\frac{c}{\Delta_a^2}$

where actually, after double-checking, it turns out that c can be made arbitrarily close to $1/2$, up to degradating the constant additive term.

Question (answer on next page):

Would you have examples of $Kinf(\nu_a, \mu^*, \mathcal{P}([0,1])) > 2\Delta_a^2$?

Yes: pick $\nu_a = \mathrm{Ber}(\mu_a)$ with $0 < \mu_a < \mu^*$

Then for all $\nu' \in \mathcal{P}([0,1])$ with $E(\nu') > \mu^*$:

$$KL(\mathrm{Ber}(\mu_a), \nu') \geq KL(\mathrm{Ber}(\mu_a), \mathrm{Ber}(E(\nu'))) \geq kl(\mu_a, \mu^*)$$

↑ data-processing inequality with expectations

↑ $kl(\mu_a, \cdot)$ is increasing on $[\mu_a, 1)$, as it is a strictly convex function with minimum achieved at μ_a

Thus, the "worst-case" ν' are given by Bernoulli distributions.

We proved: $\mathrm{Kinf}(\mathrm{Ber}(\mu_a), \mu^*) \geq kl(\mu_a, \mu^*) > 2(\mu^* - \mu_a)^2 = 2\Delta_a^2$

↑ in general (Exercise: think: in all cases but $\mu^* = \mu_a$)

The next theorem is a stronger version of Pinsker's inequality for Bernoulli distributions, that was proved[2] by Ordentlich and Weinberger [2005]. Indeed, note that the function φ defined below satisfies $\min \varphi = 2$, so that the next theorem always yields an improvement over the most classical version of Pinsker's inequality: $\mathrm{kl}(p,q) \geqslant 2(p-q)^2$.

We provide below an alternative elementary proof for Bernoulli distributions of this refined Pinsker's inequality. The extension to the case of general distributions, via the contraction-of-entropy property, is stated at the end of this section.

Theorem 15 (A refined Pinsker's inequality by Ordentlich and Weinberger [2005])**.** *For all $p, q \in [0,1]$,*

$$\mathrm{kl}(p,q) \geqslant \frac{\ln\bigl((1-q)/q\bigr)}{1-2q}\,(p-q)^2 \stackrel{\mathrm{def}}{=} \varphi(q)\,(p-q)^2\,,$$

where the multiplicative factor $\varphi(q) = (1-2q)^{-1}\ln\bigl((1-q)/q\bigr)$ is defined for all $q \in [0,1]$ by extending it by continuity as $\varphi(1/2) = 2$ and $\varphi(0) = \varphi(1) = +\infty$.

The proof shows that $\varphi(q)$ is the optimal multiplicative factor in front of $(p-q)^2$ when the bounds needs to hold for all $p \in [0,1]$; the proof also provides a natural explanation for the value of φ.

Proof: The stated inequality is satisfied for $q \in \{0,1\}$ as $\mathrm{kl}(p,q) = +\infty$ in these cases unless $p = q$. The special case $q = 1/2$ is addressed at the end of the proof. We thus fix $q \in (0,1) \setminus \{1/2\}$ and set $f(p) = \mathrm{kl}(p,q)/(p-q)^2$ for $p \neq q$, with a continuity extension at $p = q$. We exactly show that f attains its minimum at $p = 1-q$, from which the result (and its optimality) follow by noting that

$$f(1-q) = \frac{\mathrm{kl}(1-q,q)}{(1-2q)^2} = \frac{\ln\bigl((1-q)/q\bigr)}{1-2q} = \varphi(q)\,.$$

Given the form of f, it is natural to perform a second-order Taylor expansion of $\mathrm{kl}(p,q)$ around q. We have

$$\frac{\partial}{\partial p}\,\mathrm{kl}(p,q) = \ln\left(\frac{p(1-q)}{(1-p)q}\right) \qquad \text{and} \qquad \frac{\partial^2}{\partial^2 p}\,\mathrm{kl}(p,q) = \frac{1}{p(1-p)} \stackrel{\mathrm{def}}{=} \psi(p)\,, \tag{41}$$

so that Taylor's formula with integral remainder reveals that for $p \neq q$,

$$f(p) = \frac{\mathrm{kl}(p,q)}{(p-q)^2} = \frac{1}{(p-q)^2}\int_q^p \frac{\psi(t)}{1!}(p-t)^1\,\mathrm{d}t = \int_0^1 \psi\bigl(q+u(p-q)\bigr)(1-u)\,\mathrm{d}u\,.$$

This rewriting of f shows that f is strictly convex (as ψ is so). Its global minimum is achieved at the unique point where its derivative vanishes. But by differentiating under the integral sign, we have, at $p = 1-q$,

$$f'(1-q) = \int_0^1 \psi'\bigl(q+u(1-2q)\bigr)\,u(1-u)\,\mathrm{d}u = 0\,;$$

the equality to 0 follows from the fact that the function $u \mapsto \psi'\bigl(q+u(1-2q)\bigr)u(1-u)$ is antisymmetric around $u = 1/2$ (essentially because ψ' is antisymmetric itself around $1/2$). As a consequence, the convex function f attains its global minimum at $1-q$, which concludes the proof for the case where $q \in (0,1) \setminus \{1/2\}$.

It only remains to deal with $q = 1/2$: we use the continuity of $\mathrm{kl}(p,\,\cdot\,)$ and φ to extend the obtained inequality from $q \in [0,1] \setminus \{1/2\}$ to $q = 1/2$. □

We now prove the second inequality of (13). A picture is helpful, see Figure 1.

[2] We also refer the reader to Kearns and Saul [1998, Lemma 1] and Berend and Kontorovich [2013, Theorem 3.2] for dual inequalities upper bounding the moment-generating function of the Bernoulli distributions.

Solution for Exercise #2

[written in somewhat a rush: let me know if there are typos !]

1) Fixed $K \geq 2$

→ Consider the bins $[(j-1)/K, j/K]$ for $j = 1, \dots K$

→ Master strategy

* whenever the auxiliary strategy recommends $J_t \in \{1, \dots K\}$, pick $I_t \in [0,1]$ uniformly at random in $[(J_t-1)/K, J_t/K]$
* get a reward Y_t sampled according to ν_{I_t}
* send this reward to the auxiliary strategy

→ Auxiliary strategy : UCB

* pick arms $J_t \in \{1, \dots K\}$ according to the UCB strategy
* get the associated rewards from the master strategy

The auxiliary strategy thus performs UCB on the bandit model $(\tilde{\nu}_j)_{j=1,\dots K}$ where $\tilde{\nu}_j$ is the distribution of Y_t obtained from the following two-step randomization:

- draw X uniformly at random in $[(j-1)/K, j/K]$
- draw Y at random according to ν_X (given X).

In particular,

$$\tilde{\mu}_j = E(\tilde{\nu}_j) = K \int_{(j-1)/K}^{j/K} f(t)\, dt \qquad \text{where } f(t) = E(\nu_t)$$

Performance of the (auxiliary) strategy as indicated by the distribution-free bound on UCB we exhibited in our earlier exercise:

$$T \max_{j=1\dots K} \tilde{\mu}_j - E\left[\sum_{t=1}^{T} Y_t\right] \leq \sqrt{KT(8 \ln T + 2)}$$

To get the performance of the (master) strategy, we only need to control the

 Lecture 3 — February 22, 2017 Lecture 3 — February 22, 2017

approximation error $\max_{x \in [0,1]} f(x) - \max_j \tilde{\mu}_j$

But $\forall x \in [(j-1)/K, j/K]$, $\quad |\tilde{\mu}_j - f(x)| \leqslant K \int_{\frac{j-1}{K}}^{j/K} |f(t) - f(x)| \, dt$

$$\leqslant L \times K \int_{(j-1)/K}^{j/K} |t - x| \, dt$$

worst-case (largest) value is when $x = j/K$ or $x = \frac{j-1}{K}$

$$\leqslant L \times K \int_0^{1/K} t \, dt = \frac{L}{2K}$$

in particular,

$$\left| \max_j \tilde{\mu}_j - \max_{x \in [0,1]} f(x) \right| \leqslant T \frac{L}{2K}$$

The (total) regret is therefore:

$$T \max_{x \in [0,1]} f(x) - \mathbb{E}\left[\sum_{t=1}^{T} Y_t \right] \leqslant \frac{LT}{2K} + \sqrt{KT(8 \ln T + 2)}$$

2) <u>How should we pick K?</u>

→ If T is known, we can set K s.t. T/K is of the same order of magnitude as $\sqrt{KT}$ (the bound needs to hold $\forall L$, so we cannot have K depend on L): e.g., $K = \lceil T^{1/3} \rceil \leqslant 1 + T^{1/3}$, in which case the regret bound is $\leqslant \left(\frac{L}{2} + \sqrt{8 \ln T + 2} \right) \left(T^{2/3} + \sqrt{T} \right)$.

→ Otherwise, we resort to a (dirty) "doubling trick," by restarting the strategy of question (1) after times $t = 2^{r+1}$, with $r = 0, 1, 2, \ldots$, for 2^r rounds and with $K = \lceil (2^r)^{1/3} \rceil$

The total regret is equal to the sum of the regrets over these regimes:

$$\bar{R}_T \leqslant 2 + \sum_{r=1}^{r_T} \left(\frac{L}{2} + \sqrt{8 \ln 2^r + 2} \right) \left((2^r)^{2/3} + \underbrace{\sqrt{2^r}}_{\leqslant (2^r)^{2/3}} \right)$$

regime r_T might be incomplete but the bound also holds for $t \leqslant T$

with r_T s.t. $2^{r_T} + 1 \leqslant T \leqslant 2^{r_T + 1}$

$$\bar{R}_T \leqslant 2 + \left(\frac{L}{2} + \sqrt{8 \ln T + 2} \right) \times \underbrace{\left(\sum_{r=0}^{r_T - 1} (2^{2/3})^r \right)}_{\leqslant (2^{r_T})^{2/3} / 2^{2/3} - 1 \;\leqslant\; T^{2/3} / 2^{2/3} - 1} \times 2 \times 2^{2/3}$$

That is, $\overline{R}_T \leq 2 + \left(\frac{L}{2} + \sqrt{8 \ln T + 2}\right) \times \underbrace{\frac{2 \times 2^{2/3}}{2^{2/3} - 1}}_{\leq 6} \times T^{2/3}$

Final clean bound: $\overline{R}_T \leq \left(3L + 6\sqrt{8 \ln T + 2}\right) T^{2/3} + 2$

Note it can be shown that the $T^{2/3}$ order of magnitude is optimal; the $\sqrt{\ln T}$ term can be dropped by resorting to more efficient auxiliary strategies than UCB.