E1 245 - Online Prediction and Learning, Aug-Dec 2019
Homework #5

1. *The Data Processing inequality for relative entropy*

   (a) *(Log Sum inequality)* Show that for non-negative numbers $a_1, \dots, a_n, b_1, \dots, b_n$,

   $$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{(\sum_i a_i)}{(\sum_i b_i)}.$$

   (b) Use the above to prove the following data processing inequality: If (X_1, Y_1) and (X_2, Y_2) are two pairs of random variables, with outcomes in $[n] \times [n]$, having the same conditional distribution of Y_i given X_i, i.e., $\mathbb{P}\left[Y_1 = y \mid X_1 = x\right] = \mathbb{P}\left[Y_2 = y \mid X_2 = x\right]$ $\forall (x, y) \in [n] \times [n]$, then

   $$D(P_{X_1} || P_{X_2}) \geq D(P_{Y_1} || P_{Y_2}),$$

   where P_{X_i} and P_{Y_i} are the (marginal) probability distributions of X_i and Y_i, respectively, on $[n]$.

   (c) Under what condition on the (common) conditional distribution of Y given X does the data processing inequality hold with equality?

   (d) Under what condition on the (common) conditional distribution of Y given X does the right hand side of the data processing inequality attain its least possible value?

2. *Bandit algorithms must uniformly sample arms in the beginning*
   This exercise asks you to prove that any 'intelligent' bandit algorithm must initially spend a basic amount of effort on all arms – observed widely in practice (after which the data gathered can be exploited to drive regret down to being sublinear in time).

   Consider multi-armed bandit problems with n arms and Gaussian-distributed rewards of variance $\frac{1}{2}$. Call a bandit algorithm *better-than-random* if for every bandit instance $\nu \equiv (\nu_1, \dots, \nu_n)$ (where ν_i denotes the mean reward of the ith arm), every optimal arm i^* w.r.t. ν and every time t, $\mathbb{E}_\nu\left[N_{i^*}(t)\right] \geq \frac{t}{n}$, where $N_i(t)$ is the total number of times the algorithm has played arm i until time t.

   Suppose a better-than-random bandit algorithm is run on a Gaussian bandit instance ν as above.

   (a) Show that for all arms i and all times $t \geq 1$,

   $$\mathbb{E}_\nu\left[N_i(t)\right] \geq \frac{t}{n}\left(1 - \sqrt{2t\Delta_{\nu,i}^2}\right),$$

   where $\Delta_{\nu,i} = \max_j \nu_j - \nu_i$ is the mean reward gap of arm i in the instance ν. [Hint: Argue in a manner similar to showing the asymptotic bandit regret lower bound of Lai-Robbins. Use the facts that follow.]
   Facts:

   i. $D\left(\mathcal{N}(\mu_1, \sigma^2) || \mathcal{N}(\mu_2, \sigma^2)\right) = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2}$.
   ii. $D(\text{Ber}(p)||\text{Ber}(q)) \geq \frac{1}{2q}(p-q)^2$ for $0 \leq p < q \leq 1$.
   iii. $D(\text{Ber}(p)||\text{Ber}(\cdot))$ is an increasing function in the interval $[p, 1]$.

(b) Find a time t_0 (depending on the instance ν) such that $\forall t \leq t_0$, $\mathbb{E}_\nu\left[N_i(t)\right] \geq \frac{t}{2n}$.

3. *Conjugate priors*
If the posterior distributions $\mathbb{P}\left[\theta \,\middle|\, X\right]$ are in the same probability distribution family as the prior probability distribution $\mathbb{P}\left[\theta\right]$ upon observing $X \sim \mathbb{P}_\theta$ (the sample distribution), the prior is called a conjugate prior for the likelihood (sample distribution). We have seen that a Beta prior is a conjugate prior for a Bernoulli likelihood. Show explicitly the following conjugate priors for various likelihoods[1] (sample distributions):

    (a) Beta is a conjugate prior for Binomial.

    (b) Beta is a conjugate prior for Geometric.

    (c) Gamma is a conjugate prior for Poisson.

    (d) Dirichlet is a conjugate prior for Categorical.

    (e) Normal is conjugate prior for Normal (with variance 1).

    (f) Pareto is a conjugate prior for (continuous) Uniform$(0, \theta)$, $\theta \geq 0$.

4. *Beta posterior concentration*
Analyzing Thompson sampling for Bernoulli-distributed observations and Beta-distributed priors on the Bernoulli parameters typically involves studying how a Beta prior evolves as it is updated with observations.

    Consider a simple experiment where $X_1, \ldots, X_n$ are sampled iid from a Ber(p) distribution for some fixed $p \in (0, 1)$. After this, Y_n is sampled independently from the Beta$(1 + \sum_{i=1}^n X_i, 1 + n - \sum_{i=1}^n X_i)$ distribution, which you will recall is the (random) posterior distribution of p under a Uniform$[0, 1]$ prior and observations $X_1, \ldots, X_n$. Show that for $\epsilon > 0$,
$$\mathbb{P}\left[Y_n > p + \epsilon\right] \leq c_1 e^{-c_2 n \epsilon^2}$$
for some universal constants $c_1, c_2 > 0$.

    [Hint: Here's a way to proceed: (1) Split $\mathbb{P}\left[Y_n > p + \epsilon\right] = \mathbb{P}\left[Y_n > p + \epsilon, \frac{S}{n} > p + \frac{\epsilon}{2}\right] + \mathbb{P}\left[Y_n > p + \epsilon, \frac{S}{n} \leq p + \frac{\epsilon}{2}\right]$ where $S = \sum_{i=1}^n X_i$, (2) Bound the first term by Hoeffding, (3) Bound the second term using the following relationship between the cdfs[2] of the Beta and Binomial distributions: $F_{\text{Beta}(a,b)}(y) = 1 - F_{\text{Bin}(a+b-1,y)}(a - 1)$, followed by Hoeffding.]

[1] Look up distribution definitions on Wikipedia.

[2] The cumulative distribution function (cdf) of a distribution P is $F_P(x) = \mathbb{P}_{X \sim P}[X \leq x]$.