Problem set 3, ORF550

Bachir El Khadir

November 9, 2016

1 Problem 3.13a

We want to show that:

$$Ent(Z) = \inf_{t>0} E[Z \log Z - Z \log t - Z + t]$$

Which is the same as:

$$-E[Z] \log E[Z] = \inf_{t>0} -E[Z] \log t - E[Z] + t$$

Or

$$\inf_{t>0} \frac{t}{E[Z]} - \log \frac{t}{E[Z]} = 1$$

Or

$$\inf_{u>0} u - \log u = 1$$

Which is true, because $f : u \to u - \log u$ is convex ($f''(u) = \frac{1}{u^2}$), and its first derivative ($f'(u) = 1 - \frac{1}{u}$) is 0 at 1.

2 Problem 3.20

a.

$$\begin{aligned}
Ent_\nu X &= \inf_{t>0} E_\nu[X \log X - X \log t - X + t] \\
&= \inf_{t>0} E_\mu[(X \log X - X \log t - X + t)\frac{d\nu}{d\mu}] \\
&\le ||\frac{d\nu}{d\mu}||_\infty \inf_{t>0} E_\mu[X \log X - X \log t - X + t] \quad (X \log X - X \log t - X + t = X(t/X - \log(t/X) - 1) \ge 0) \\
&\le ||\frac{d\nu}{d\mu}||_\infty Ent_\mu X
\end{aligned}$$

b.

$$\begin{aligned}
\nu(\Gamma(\log f, f)) &= \mu(\frac{\Gamma(\log f, f)}{d\nu/d\mu}) \\
&\ge \frac{1}{\delta}\mu(\Gamma(\log f, f)) \\
&\ge \frac{1}{c\delta} Ent_\mu(f) \\
&\ge \frac{c\varepsilon}{\delta} Ent_\nu(f)
\end{aligned}$$

c.

$\nu(dx) = \frac{1}{Z'}e^{-V(x)+x^2}\mu(dx)$, where $\mu \sim N(0, \sqrt{2})$

$$\frac{d\nu}{d\mu} \in [\frac{1}{Z'}e^{-b}, \frac{1}{Z'}e^{-a}]$$

So:

$$Ent_\nu f^2 \leq \frac{1}{c}e^{b-a}\nu(\Gamma(\log f^2, f^2)) = \frac{1}{c}e^{b-a}\nu((2f'/f).(2ff')) = \frac{1}{4c}e^{b-a}\nu(|f'|^2)$$

d.

$$\begin{aligned}
Var_\nu(f) &= \inf_{c\in\mathbb{R}} E_\nu[(f-c)^2] \\
&= \inf_{c\in\mathbb{R}} E_\mu[(f-c)^2\frac{d\nu}{d\mu}] \\
&\leq \inf_{c\in\mathbb{R}} E_\mu[(f-c)^2]||\frac{d\nu}{d\mu}||_\infty \\
&\leq Var_\mu f||\frac{d\nu}{d\mu}||_\infty \\
&\leq c\delta\mu(\Gamma(f,f)) \\
&\leq c\delta\nu(\frac{\Gamma(f,f)}{d\nu/d\mu}) \\
&\leq \frac{c\delta}{\varepsilon}\nu(\Gamma(f,f))
\end{aligned}$$

3 Problem 4.2

a. Suppose $med(f)$ attained at x_0.

$A = \{f \leq med(f)\}$, $x_0 \in A$

$\mu(A) = \frac{1}{2}$ by definition

Let $x \in A^t$, then for all $\varepsilon > 0$, there exist $y \in A$ such that $d(x,y) \leq t+\varepsilon$.

Since f is Lip, this implies that $|f(x)-f(y)| \leq t+\varepsilon$. So that $f(x)-med(f) \leq f(x)-f(y)+f(y)-med(f) \leq t+\varepsilon$. In particular, letting $\varepsilon \to 0$ gives that $f(x) - med(f) \leq t$.

We have just proved that $1-Ce^{-t^2/2\sigma^2} \leq \mu(A^t) \leq \mathbb{P}(f(x)-med(f) \leq t)$. So $Ce^{-t^2/2\sigma^2} \geq \mathbb{P}(f(x)-med(f) \geq t)$
.

b. Let A be a set of measure $\geq \frac{1}{2}$, and consider $f(x) = d(x,A)$. Then:

- f is Lipschiz
- $med(f) = 0$

In addition to that, $A^\varepsilon = \{x, d(x,A) \geq \varepsilon\} = \{f(x) \geq \varepsilon\}$.

The result follow from the concentration inequality we assumed.

c.

$$\begin{aligned}
E_\mu[(f-med(f))_+] &= \int_0^\infty P_\mu[f-med(f) \geq t]dt \\
&\leq \int_0^\infty Ce^{-\frac{t^2}{2\sigma^2}}dt \\
&= C\sqrt{\frac{\pi}{2}}\sigma
\end{aligned}$$

$$E_\mu[f] - med(f) \leq E_\mu[(f-med(f))_+] \leq \sqrt{\frac{\pi}{2}}C\sigma$$

We conclude by considering $-f$, which is also lipschiz.

The result follow by noting that $t \to P(X \geq t)$ is non-decreasing.

d. For $t_0 = 2\sigma\sqrt{\log 2C}$, $Ce^{-t^2/2\sigma^2} = \frac{1}{2}$, and $P(f \geq E_\mu f + t_0) \leq \frac{1}{2}$.
But $P(f \geq med\ f) = \frac{1}{2}$, so $P(f \geq E_\mu f + t_0) \leq P(f \geq med\ f)$.
As a result:

$$med\ f \leq E_\mu f + t_0$$

Consider $-f$ to conclude.
$P(f - med(f) \geq t) \leq P(f - Ef \geq t - t_0) \leq Ce^{-(t-t_0)^2/2\sigma^2} \leq 2Ce^{-t^2/8\sigma^2}$
e. Let $f = d(x, B)$, $\mu_A = \mu(.|A)$, $\mu_B = \mu(.|B)$
Notice that $f(x) = 0$ on B, and $f(x) \geq d(A, B)$ on A. So:
$W_1(\mu_A, \mu_B) \geq \int f(x) d\mu_A \geq d(A, B)\mu_A(A) = d(A, B)$
But $W_1(\mu_A, \mu_B) \leq W_1(\mu_A, \mu) + W_1(\mu_B, \mu)$ and

- $W_1(\mu_A, \mu)^2 \leq 2\sigma^2 D(\mu_A||\mu) \leq 2\sigma^2 \log \frac{1}{\mu(A)}$
- $W_1(\mu_B, \mu)^2 \leq 2\sigma^2 \log \frac{1}{\mu(B)}$

Which yields the result.
f. In this case, $d(A, B) = \varepsilon$ so $\varepsilon \leq \sqrt{2\sigma^2}(\sqrt{\log 1/\mu(A)} + \sqrt{\log 1/\mu(B)}) \leq \sqrt{2\sigma^2}(\sqrt{\log 2} + \sqrt{\log 1/\mu(B)})$ so $\mu(A^\varepsilon) = 1 - \mu(B) \geq 1 - 2e^{\frac{\varepsilon^2}{8\sigma^2}}$

4 Problem 4.5

a. Choose an ε optimal coupling $M_1 \in \mathcal{C}(\rho_1, \rho_2)$ for $\inf_{M \in \mathcal{C}(\rho_1,\rho_2)} P_{(X,Y)\sim M}(X \neq Y) = ||\rho_1 - \rho_2||$ Choose an ε optimal coupling $M_2 \in \mathcal{C}(\rho_2, \rho_3)$ for $\inf_{M \in \mathcal{C}(\rho_2,\rho_2)} P_{(Y,Z)\sim M}(Z \neq Y) = ||\rho_2 - \rho_3||$
define $M \in \mathcal{C}(\rho_1, \rho_2, \rho_3)$ by:

- $M(X) = \rho_1$
- $M(Y|X) = M_1(Y|Z)$
- $M(Z|X, Y) = M_2(Z|Y)$

It is clear that M is ε -optimal.
We assume that we can take ε to 0.
b. We proceed by induction, and using part a. at each step. Assume we have the claim up to $k < n$, then we construct Z_{k+1} by applying a. as with:

$$\rho_1 = Q_{k+1}(X_k, .), \rho_2 = Q_{k+1}(Y_k, .), \rho_3 = \nu(Y_{k+1} \in .|Y_1, \dots Y_k)$$

To show the result, notice that: $X_k = \tilde{X}_k, \tilde{X}_k = Y_k \implies X_k = Y_k$, so:

$$\begin{aligned} M[X_k \neq Y_k|Z_1, \dots Z_{k-1}] &\leq M[\tilde{X}_k \neq X_k|Z_1, \dots Z_{k-1}] + M[\tilde{X}_k \neq Y_k|Z_1, \dots Z_{k-1}] && \text{(Union Bound)} \\ &= ||Q_k(Y_{k-1}, .) - \nu(Y_k \in .|Y_1, \dots Y_k)||_{TV} + ||Q_k(X_{k-1}, .) - Q_k(Y_{k-1}, .)||_{TV} \\ &\leq \sqrt{\frac{1}{2} D(\nu(Y_k \in .|Y_1, \dots Y_{k-1})||Q_k(Y_{k-1}, .))} + (1 - \alpha)1_{X_{k-1} \neq Y_{k-1}} && \text{(Bobkov-Gotze)} \end{aligned}$$

c.

$$M[X_k \neq Y_k|Z_1, \dots Z_{k-1}] \leq \sqrt{\frac{1}{2} D(\nu(Y_k \in .|Y_1, \dots Y_{k-1})||Q_k(Y_{k-1}, .))} + (1 - \alpha)1_{X_{k-1} \neq Y_{k-1}}$$

Take the expectation with respect to $Z_1, \dots Z_{k-1}$ on both sides:

$$\alpha M[X_k \neq Y_k] \leq E[\sqrt{\frac{1}{2} D(\nu(Y_k \in .|Y_1, \dots Y_{k-1})||Q_k(Y_{k-1}, .))}]$$

5 Problem 4.7

a. Let

$$Y_i = \begin{cases} 1 & \text{if bin } i \text{ is empty} \\ 0 & \text{otherwise} \end{cases}$$

The Y_i are iid, and their common distribution is

$$E[Y_i] = P(Y_i = 1) = P(\forall j \in [m] \text{ ball } j \text{ missed bin } i) = (1 - \frac{1}{n})^m$$

$$E[Z] = E[Y_1 + \dots Y_n] = nE[Y_1] = n(1 - \frac{1}{n})^m$$

b. $Z = f(Y_1, \dots Y_m)$, with $f(Y) = \sum_i Y_i$. It is clear that:

- The Y_i are independent
- $||D_k f||_\infty = 1$
- $\sum ||D_k f||_\infty = m$

By McDiarmid's inequality, Z is $m/4$ -subgaussian.

b. $f_m(b) \leq f_{2m}(b_1, b'_1, \dots b_m, b'_m) = f_{2m}(b_1, \dots b_m, b'_1, \dots b'_m)$ because the number of non-empty bins can only increases if we add new balls.

Also: $f_{2m}(b_1, \dots b_m, b'_1, \dots b'_m) = f_m(b') + \sum_{i=1}^m 1_{b'_i \neq b'_j \text{ for } i<j \wedge b'_i \neq b_j \text{ for } i \leq m} \leq f_m(b') + \sum_{i=1}^m 1_{b'_i \neq b'_j \text{ for } j<i \wedge b'_i \neq b_i}$ Which proves the two inequalities.

Now we have: $f_m(b) - f_m(b') \leq \sum_{i=1}^m \underbrace{1_{b'_i \neq b'_j \text{ for } j<i}}_{c_i(b')} 1_{b'_i \neq b_i}$ Notice that $\sum_i c_i(b')^2 = \sum_i c_i(b') = f_m(b') \leq n$, which proves that Z is also n -subgaussian.

6 Problem 4.8

a. Lower bound:

- $L_n \geq \sum_i \min_{j \neq i} ||X_i - X_j||$, because at some point when we are X_i, we are going to travel to some other city X_j, and that quantity is bounded below by the minimum distance from X_i.
- Let $Z = \min_{j>1} ||X_1 - X_j|||$. $Z \geq r \iff \forall j > 1,\ ||X_j - X_i|| \geq r$
- Conditioning on X_1, $||X_j - X_1|| \leq r$ happens with probability $Surface(B(X_1, r) \cap [0,1]^2) := a(X_1, r)$.
- $a(X_1, r) \leq \pi r^2 \wedge 1$
- $P(Z \geq r | X_1) = (1 - a(X_1))^n$
- $E[Z] = E_{X_1}[\int_0^1 P(Z \geq r|X_1)dr] \geq \int_0^1 (1 - \pi r^2)_+^n dr = \frac{\Gamma(n+\frac{1}{2})}{\Gamma(n+\frac{3}{2})} \sim \frac{1}{\sqrt{n}}$
- As a result, $E[L_n] \geq nE[Z] \sim \sqrt{n}$

Upper bound:

- $L_n \leq L_{n-1} + 2\min_{k<n} ||X_k - X_n||$. This is true, because $X_{\sigma(1)}, \dots X_{\sigma(n-1)}$ is an optimal tour of the first n cities, $k^* = \arg\min_{k<n} ||X_{\sigma(k)} - X_n||$, then, $X_{\sigma(1)}, \dots X_{\sigma(k^*)}, X_n, X_{\sigma(k^*)}, \dots, X_{\sigma(n-1)}$, is a valid tour of the n cities of cost $L_{n-1} + 2\min_{k<n} ||X_k - X_n||$.

  This proves that $L_n \leq 2\sum \min_{k \leq n} ||X_k - X_n||$. Using a similar technique as in the last question by bounding $a(X_1, r)$ from below by πr^2, we get that $\min_{k \leq n} ||X_k - X_n|| \sim \frac{1}{\sqrt{i}}$.

  But $\sum_i \frac{1}{\sqrt{i}} \sim \int_1^n \frac{dt}{\sqrt{t}} \sim \sqrt{n}$, which proves the result.

b. $L_n = L_n(X_1, \dots, X_n)$. $|D_k L_n| \le 2\sqrt{2}$. Indeed:

- Let $X_{\sigma(1)}, \dots X_{\sigma(n)}$ is an optimal tour, and $i = \sigma^{-1}(k)$
- Starting with an optimal tour, changing X_k can only change the portion of the tour to and from X_k, in both cases by at most $\sqrt{2}$.
- We conclude by McDiarmid's inequality.

c. Let $x = x_1 u + x_2 v$. $x \in T \implies 0 \le x_1 \le 1, 0 \le x_2 \le 1 \implies x_1^2 \le x_1, x_2^2 \le x_2$

$$
\begin{aligned}
.||x-u||^2 + ||x-v||^2 \le ||u-v||^2 &\iff 2||x||^2 - 2\langle x, u+v\rangle \le 0 \\
&\iff \langle x_1 u + x_2 v, (x_1 - 1)u + (x_2 - 1)v\rangle \le 0 \\
&\iff x_1(x_1 - 1) + x_2(x_2 - 1) \le 0 \\
&\iff x_1^2 + x_2^2 \le x_1 + x_2
\end{aligned}
$$

And the last inequality is true.

d. We proceed by induction like the hint suggests.

- Suppose the the result true up to $n-1$, consider n points $x_1, \dots x_n$ in T
- Divide T into two right triangles S_1, S_2 until both are not empty. Without loss of generality, because the length of the path can only get shorter, we can assume $S_1 \cup S_2 \neq \emptyset$
- Apply the induction hypothesis on S_1 and S_2 to get two paths, the first one $v, y_1, \dots, y_m, O$ of length at most a, the other $O, z_1, \dots z_r$ with $m + r = n$ with length at most b.
- Consider the path $v, y_1 \dots y_m, z_1, \dots z_r, w$ that has length:

  $$||v, y_1||^2 + \dots ||y_{m-1} - y_m||^2 + ||y_m - z_1||^2 + \dots ||z_r - w||^2 \le ||v, y_1||^2 + \dots ||y_{m-1} - y_m||^2 + ||y_m - O||^2 + ||O - z_1||^2 + \dots ||z_r - w$$

  Where the first inequality comes from the fact that $||y - z||^2 \le ||y - O||^2 + ||O - z||^2$ whenever $\langle z, y\rangle \ge 0$, which is true on T.

e. By d., consider a path $v, x_{\sigma(1)}, \dots x_{\sigma(n)}, w$ of length at most 2. Consider the path $x_{\sigma(1)}, \dots x_{\sigma(n)}, x_{\sigma(1)}$, which is, for the same reason as above, has shorter length than: $x_{\sigma(1)}, \dots x_{\sigma(n)}, x_{\sigma(n-1)}, \dots x_{\sigma(1)}$, so smaller than 4.

f. We follow the hint, we start following τ until the first time we get to a point in $x \cap y$, then we follow σ until right beforewe hit $x \cup y$ again, then we follow the last portion of the path in reverse. It is clear that the length of this path has the following components:

- $l_n(y, \tau)$, if we ignore the excursions we do when we meet some element from x, because we start and end at the same point.
- $2d_i(x, \sigma)$ whenever $x_i \notin y$, because we go to and then from x_i once.

g.

- If $x \cup y \neq \emptyset$, we use f. by noting that $\min_\sigma l_n(x, \sigma) \le l_{2n}(x \cup y, \rho)$
- Otherwise, notice that $l_n(x, \sigma) \le 2\sum_i d(x_i, \sigma) = 2\sum_i 1_{x_i \in y} d(x_i, \sigma)$

h. Use Talagrand inequality with $c_i = 2d(x_i, \sigma)$, so that $||\sum c_i^2||_\infty \le 16$ by e.