SDS 384-11 HW3

Shivam Gupta

March 2019

1 Problem 1

Note that for $t > 0$,

$$Pr[Z_n \leq \delta n] = Pr[tZ_n \leq t\delta n] = Pr[\exp(tZ_n) \leq \exp(t\delta n)]$$

Using Markov's inequality, the above is at most

$$\frac{E[\exp(tZ_n)]}{\exp(t\delta n)} = \frac{\prod_{i=1}^n E[\exp(tX_i)]}{\exp(t\delta n)}$$

But since the X_is are i.i.d., we have that the above is

$$\frac{E[\exp(tX_1)]^n}{\exp(t\delta n)} = \exp(-n(t\delta - \log(E[\exp(tX_1)])))$$

Now, note that $\log(E[\exp(tX_1)]) = \log(\alpha e^t + (1-\alpha))$. So, we want to choose t to maximize $t\delta - \log(\alpha e^t + (1-\alpha))$. Differentiating and setting to 0, we have

$$\delta - \frac{\alpha e^t}{\alpha e^t + (1-\alpha)} = 0$$

$$\delta \alpha e^t + \delta(1-\alpha) - \alpha e^t = 0$$

$$e^t = \frac{\delta(1-\alpha)}{\alpha(1-\delta)}$$

So, $t = \log(\frac{\delta(1-\alpha)}{\alpha(1-\delta)})$. So, substituting into the equation,

$$t\delta - \log(\alpha e^t + (1-\alpha)) = \delta(\log(\frac{\delta}{\alpha}) - \log(\frac{1-\delta}{1-\alpha})) - \log[\frac{\delta(1-\alpha)}{1-\delta} + (1-\alpha)]$$

$$= \delta \log \frac{\delta}{\alpha} + (1-\delta)\log\frac{1-\delta}{1-\alpha} = KL(\delta \| \alpha)$$

Thus, we have that

$$Pr[Z_n \leq \delta n] \leq \exp(-nKL(\delta\|\alpha))$$

Now note that since $1 > 1 - \delta \geq \delta$, we have

$$(1-\delta)\log(1-\delta) \leq \delta \log(1-\delta) \leq \delta \log \delta$$

Note that Hoeffding's inequality says that

$$Pr[Z_n \leq \delta n] \leq \exp(-2n(\delta - \alpha)^2)$$

So, to show that this inequality is stronger, it suffices to show that

$$KL(\delta \| \alpha) \geq 2(\delta - \alpha)^2$$

By Pinsker's inequality, this is true.

2 Problem 2

(a) Note that since $m = \lfloor \delta n \rfloor$,

$$Pr[Z_n \leq \delta n] = \sum_{i=0}^{\lfloor \delta n \rfloor} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \geq \binom{n}{m} \alpha^m (1-\alpha)^{n-m}$$

Thus, by monotonicity of log,

$$\frac{1}{n} \log Pr[Z_n \leq \delta n] \geq \frac{1}{n} \log \binom{n}{m} + \frac{m}{n} \log \alpha + \frac{n-m}{n} \log(1-\alpha)$$

$$= \frac{1}{n} \log \binom{n}{m} + \delta' \log \alpha + (1-\delta') \log(1-\alpha)$$

(b) Using the hint, note that

$$Pr[Y = m] \geq \frac{1}{n+1}$$

since $Pr[Y = k]$ is maximized at $k = m$, and Y is supported on the $n+1$ values $0, 1, \ldots, n$. Thus, we have that

$$Pr[Y = m] = \binom{n}{m} \delta'^m (1-\delta')^{n-m} \geq \frac{1}{n+1}$$

Taking log and dividing by n, since $\delta' = m/n$ we have

$$\frac{1}{n} \log \binom{n}{m} + \delta' \log \delta' + (1-\delta') \log(1-\delta') \geq -\frac{\log(n+1)}{n}$$

Rearranging, we have the claim.

(c) Plugging in (b) into part (c), we have

$$\frac{1}{n}\log Pr[Z_n \le \delta n] \ge \delta' \log \frac{\alpha}{\delta'} + (1-\delta')\log\frac{1-\alpha}{1-\delta'} - \frac{\log(n+1)}{n}$$

$$= -KL(\delta' \| \alpha) - \frac{\log(n+1)}{n}$$

Multiplying by n and taking exp, we have

$$Pr[Z_n \le \delta n] \ge \exp(-nKL(\delta'\|\alpha) - \log(n+1)) = \frac{1}{n+1}\exp(-nKL(\delta'\|\alpha))$$

3 Problem 3

(a) Let z be such that $\inf_x f(X_{1:i-1}, x, X_{i+1:n}) = f(X_{1:i-1}, z, X_{i+1:n})$ Note that by separable convexity, we have that

$$f(X_{1:i-1}, z, X_{i+1:n}) \ge f(X) + \frac{\partial f}{\partial X_i}(X)(z - X_i)$$

Since $f(X_{1:i-1}, z, X_{i+1:n}) \le f(X)$, $f(X) - f(X_{1:i-1}, z, X_{i+1:n}) \ge 0$. Thus, both sides of the equation

$$f(X) - f(X_{1:i-1}, z, X_{i+1:n}) \le \frac{\partial f}{\partial X_i}(X)(X_i - z)$$

are non-negative.

Thus,

$$(Z - Z_i)^2 = (f(X) - f(X_{1:i-1}, z, X_{i+1:n}))^2 \le (\frac{\partial f}{\partial X_i}(X)(X_i - z))^2$$

But since f has domain $[0,1]^n$ we have that $(X_i - z)^2 \le 1$. Thus,

$$(Z - Z_i)^2 \le (\frac{\partial f}{\partial X_i}(X))^2$$

So, $\sum_i E[(Z - Z_i)^2] \le \sum_i E[(\frac{\partial f}{\partial X_i}(X))^2] = E[\|\nabla f(X)\|^2]$

(b) Note that the operator norm is convex since by triangle inequality

$$\lambda\|A\|_{op} + (1-\lambda)\|B\|_{op} \ge \|\lambda A + (1-\lambda)B\|_{op}$$

for all $\lambda \in [0,1]$. Since convex implies convex along each line, the operator norm is separately convex.

Note that by triangle inequality,

$$\|A - B\|_{op} + \|B\|_{op} \ge \|A\|_{op}$$

for any two matrices A, B. Thus,

$$\|A\|_{op} - \|B\|_{op} \leq \|A - B\|_{op} \leq \|A - B\|_2$$

So, the operator norm is 1-Lipschitz. But since norm of gradient is bounded by the Lipschitz constant, we have $\|\nabla \|A\|_{op}\|^2 \leq 1$. Thus, by part (a), $var(Z) \leq 1$.

4 Problem 4

(a) Let us order the variables so that $X_1 \leq X_2 \leq \cdots \leq X_n$ and $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. We will use induction to prove that for any permutation $i_1, i_2, \ldots i_n$ of $1, 2, \ldots n$, we have that

$$X_1Y_1 + X_2Y_2 + \cdots + X_nY_n \geq X_1Y_{i_1} + X_2Y_{i_2} + \cdots + X_nY_{i_n}$$

Base case: $n = 1$. In this case, $i_1 = 1$. Thus, trivially, $X_1Y_1 \geq X_1Y_{i_1}$.

Inductive hypothesis: For all $k < n$, for any permutation $i_1, \ldots, i_k$ of $1, \ldots, k$, we have that $X_1Y_1 + \cdots + X_kY_k \geq X_1Y_{i_1} + \cdots + X_kY_{i_k}$.

Inductive case: There are two cases:

(i) $i_n = n$. In this case, we have that $i_1, \ldots, i_{n-1}$ is a permutation of $1, 2, \ldots, n-1$. Thus, by inductive hypothesis,

$$X_1Y_1 + \cdots + X_{n-1}Y_{n-1} \geq X_1Y_{i_1} + \cdots + X_{n-1}Y_{i_{n-1}}$$

and also, trivially, $X_nY_n \geq X_nY_{i_n}$. Thus,

$$X_1Y_1 + \cdots + X_nY_n \geq X_1Y_{i_1} + \cdots + X_nY_{i_n}$$

(ii) $i_n \neq n$. In this case, let $j < n$ be such that $i_j = n$, and let $l = i_n < n$. Note that $X_nY_n + X_jY_l - X_nY_l - X_jY_n = (X_n - X_j)(Y_n - Y_l) \geq 0$ since $X_n \geq X_j$, $Y_n \geq Y_l$, since $j < n$, $l < n$. Thus,

$$X_nY_n + X_jY_l \geq X_nY_l + X_jY_n$$

Now, let $i'_1, \ldots i'_{n-1}$ be such that for all $s \neq j$, $i'_s = i_s$ and $i'_j = l$. Then, $i'_1, \ldots, i'_{n-1}$ is a permutation of $1, \ldots n-1$, and so, by inductive hypothesis,

$$X_1Y_1 + \cdots + X_{n-1}Y_{n-1} \geq X_1Y_{i'_1} + \cdots + X_{n-1}Y_{i'_{n-1}}$$

$$= X_1Y_{i_1} + \cdots + X_{j-1}Y_{i_{j-1}} + X_jY_l + X_{j+1}Y_{i_{j+1}} + \cdots + X_{n-1}Y_{i_{n-1}}$$

Then, adding the previous equation to this equation, we have

$$X_1Y_1+\dots X_nY_n \geq X_1Y_{i_1}+\dots+X_{j-1}Y_{i_{j-1}}+X_{j+1}Y_{i_{j+1}}+\dots+X_{n-1}Y_{n-1}+X_nY_l+X_jY_n$$

But since $l = i_n$ and $i_j = n$, we have that $X_nY_l = X_nY_{i_n}$ and $X_jY_n = X_jY_{i_j}$. Thus, the RHS of the above equation is

$$X_1Y_{i_1} + \dots + X_nY_{i_n}$$

and we have proved the claim.

Now, we will use this to prove that order statistics are 1-Lipschitz. Let $f(X)$ be the k^{th} smallest element of the vector X. Let $X_1 \leq X_2 \leq \dots \leq X_n$ and $Y_1 \leq Y_2 \leq \dots \leq Y_n$ so that vectors X and Y are some permutation of $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ respectively. Note that $(f(X)-f(Y))^2 = (X_k - Y_k)^2 \leq \sum_{j=1}^n (X_j - Y_j)^2$ and that $\|X-Y\|^2 = \sum_{j=1}^n (X_j - Y_{i_j})^2$ for some permutation $i_1, \dots, i_n$ of $1, 2, \dots, n$. So, we need to prove that $\sum_{j=1}^n (X_j - Y_j)^2 \leq \sum_{j=1}^n (X_j - Y_{i_j})^2$ for any permutation $i_1, \dots i_n$ of $1, \dots, n$. But note that since the X_j^2 and Y_j^2 terms cancel out, this is equivalent to proving that $-2\sum_{j=1}^n X_jY_j \leq -2\sum_{j=1}^n X_jY_{i_j}$ which follows from our first claim ($\sum_{j=1}^n X_jY_j \geq \sum_{j=1}^n X_jY_{i_j}$)

(b) (i) Note that by Jensen's, and since $\exp(tY) \leq \sum_i \exp(tX_i)$ for $t > 0$,

$$\exp(tE[Y]) = \exp(E[tY]) \leq E[\exp(tY)] \leq E[\sum_i \exp(tX_i)] = \sum_i E[\exp(tX_i)]$$

But $E[\exp(tX_i)]$ is the MGF of $N(0,1)$. Thus, it is equal to $e^{t^2/2}$. Setting $t = \sqrt{2\log n}$, we have that

$$\exp(tE[Y]) \leq \sum_i e^{t^2/2} = \sum_i e^{\log n} = \sum_i n = n^2$$

Thus,

$$E[Y] \leq \frac{\log n^2}{t} = \frac{2\log n}{\sqrt{2\log n}} = \sqrt{2\log n}$$

(ii) A. Note that for $\delta > 0$, $Y \geq \delta I_{Y\geq\delta} + YI_{Y<\delta}$. Now, if $Y \geq 0$, then, $YI_{Y\leq\delta} \geq 0$ and if $Y < 0$, then, $YI_{Y<\delta} = Y$. Thus, $YI_{Y\leq\delta} \geq \min(Y, 0)$. So, we have that

$$Y \geq \delta I_{Y\geq\delta} + \min(Y, 0)$$

Taking expectation on both sides,

$$E[Y] \geq \delta Pr[Y \geq \delta] + E[\min(Y, 0)]$$

B. Since $Y = \max_i X_i \geq X_1$, $\min(Y, 0) \geq \min(X_1, 0)$. Thus, $E[\min(Y, 0)] \geq E[\min(X_1, 0)]$

C. Note that

$$Pr[Y \geq \delta] = 1 - Pr[Y < \delta] = 1 - Pr[X_1 < \delta \wedge X_2 < \delta \wedge \cdots \wedge X_n < \delta]$$

By independence, the above is equal to

$$1 - Pr[X_1 < \delta]^n = 1 - (1 - Pr[X_1 \geq \delta])^n$$

D. Note that by AM-GM,

$$(a+b)^2 = a^2 + b^2 + 2ab \leq a^2 + b^2 + 2\frac{a^2+b^2}{2} = 2a^2 + 2b^2$$

Thus, we have that

$$x^2 = ((x-\delta)+\delta)^2 \leq 2(x-\delta)^2 + 2\delta^2$$

Thus,

$$Pr[X_1 \geq \delta] = \frac{1}{\sqrt{2\pi}} \int_{\delta}^{\infty} e^{-x^2/2} dx \geq \frac{1}{\sqrt{2\pi}} \int_{\delta}^{\infty} e^{-(x-\delta)^2/2} e^{-\delta^2/2} dx$$

Letting $t = x - \delta$, we have that the above integral is

$$\frac{e^{-\delta^2/2}}{\sqrt{2\pi}} \int_0^{\infty} e^{-t^2/2} dt$$

But since $N(0,1)$ is symmetric around 0, we have that

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-t^2/2} dt = 1/2$$

Thus, we have

$$Pr[X_1 \geq \delta] \geq \frac{1}{2} e^{-\delta^2/2}$$

E. Now, setting $\delta = \sqrt{2 \log n/2}$, we have that

$$Pr[X_1 \geq \delta] \geq \frac{1}{2} e^{-\log n/2} = 1/n$$

Now, by symmetry,

$$E[\min(X_1, 0)] = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-x^2/2} x dx = -\frac{1}{2\sqrt{2\pi}}$$

So,

$$E[Y] \geq \sqrt{2\log(n/2)}(1 - (\frac{n-1}{n})^n) - \frac{1}{2\sqrt{2\pi}} \geq c\sqrt{\log n}$$

for small enough $c > 0$.

5 Problem 5

Using the method of martingale differences, let $V_i = E[\phi|X_1^i] - E[\phi|X_1^{i-1}]$.

Let U_i be V_i conditioned no X_1^{i-1} so that

$$U_i = \sum_{x_{i+1}^n} P(x_{i+1}^n) \sum_{x_i, x_i'} P(x_i)P(x_i')(\phi(X_1^{i-1}, x_i, x_{i+1}^n) - \phi(X_1^{i-1}, x_i', x_{i+1}^n))$$

Now, by Jensen's inequality,

$$E[e^{\lambda V_i}|X_1^{i-1}] = E[e^{\lambda(E[\phi|X_1^i] - E[\phi|X_1^{i-1}])}|X_1^{i-1}] \leq E[(E[e^{\lambda\phi}|X_1^i])(-E[e^{\lambda\phi}|X_1^{i-1}])|X_1^{i-1}]$$

$$= \sum_{x_{i+1}^n} P(x_{i+1}^n) \sum_{x_i, x_i'} P(x_i)P(x_i')(\exp(\lambda\phi(X_1^{i-1}, x_i, x_{i+1}^n) - \lambda\phi(X_1^{i-1}, x_i', x_{i+1}^n)))$$

Now, let $F(y) = \phi(X_1^{i-1}, y, x_{i+1}^n)$ for fixed X_1^{i-1}, x_{i+1}^n and now, since F is 1-Lipschitz wrt the metric ρ_i, we have that since $e^t + e^{-t} = 2\cosh(t) \leq \cosh(s)$ for all $|t| \leq s$,

$$e^{\lambda(F(y) - F(y'))} + e^{\lambda(F(y') - F(y))} \leq \frac{1}{2}(e^{\lambda\rho_i(y,y')} - e^{-\lambda\rho_i(y,y')})$$

Thus,

$$\sum_{y,y'} P(y)P(y')e^{\lambda(F(y)-F(y'))} \leq \frac{1}{2}[\sum_{y,y'} P(y)P(y')e^{\lambda\rho_i(y,y')} + \sum_{y,y'} P(y)P(y')e^{-\lambda\rho_i(y,y')}]$$

$$= E[e^{\lambda d(\mathcal{X}_i)}] \leq \exp(\lambda^2 \Delta_{SG}^2(\mathcal{X}_i)/2)$$

where d is the symmeterized distance and Δ_{SG} is subgaussian diameter.

Now, using Markov,

$$Pr[\phi - E[\phi] > t] = P[\sum_{i=1}^n V_i > t] \leq e^{-\lambda t} E[\prod_{i=1}^n e^{\lambda V_i}] \leq e^{-\lambda t} E[\prod_{i=1}^n E[e^{\lambda V_i}|X_1^{i-1}]]$$

Using the previous step, the above is at most

$$e^{-\lambda t} E[\prod_{i=1}^n \exp(\lambda^2 \Delta_{SG}^2(\mathcal{X}_i)/2)] = \exp(\frac{1}{2}\lambda^2 \sum_{i=1}^n \Delta_{SG}^2(\mathcal{X}_i) - \lambda t)$$

Optimizing λ, we get that the above is at most

$$\exp(-\frac{t^2}{2\sum_{i=1}^n \Delta_{SG}^2(\mathcal{X}_i)})$$

Applying the same argument to the lower tail and taking union bound gives the result.