# Pset 1

Noah Toyonaga

September 28, 2022

I worked with Lucy Liu on this PSET.

## 1

From the problem, we assume $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the PDF of g. We want to calculate $P(g > t)$

$$\begin{aligned}
P(g > t) &= \int_t^\infty \phi(x)\, dx \\
&= \int_t^\infty \frac{-1}{x} (-x\phi(x)) \\
&= \frac{-1}{x}\phi(x)\,|_t^\infty - \int_t^\infty \frac{1}{x^2}\phi(x) \\
&= \left(0 + \frac{\phi(t)}{t}\right) - \left(\frac{-\phi(x)}{x^3}|_t^\infty - \int_t^\infty \frac{1}{x^4}\phi(x)\right) \\
&= \frac{\phi(t)}{t} - \frac{\phi(t)}{t^3}
\end{aligned} \tag{1}$$

The first term gives an upper bound for $P(g > t)$ while the inclusion of the first correction gives a lower bound as given in the problem.

We note that this calculation is based on an expansion at $t = \infty$ and thus the series diverges as $t \to 0$. However these bounds hold for $t > 1$.

## 2

### 2.1

We consider the probability $P^* = P(X_1 X_2 > n^\epsilon Y_1 Y_2)$.

Assume that *both* $X_1 < n^{\epsilon/2} Y_1$ and $X_2 < n^{\epsilon/2} Y_2$. In this case we find, by multiplying the previous expressions, $X_1 X_2 < n^\epsilon Y_1 Y_2$. Thus, in order for the converse to be true (i.e. for $X_1 X_2 > n^\epsilon Y_1 Y_2$) one of the assumptions must be false.

This implies that the probability $P^*$ must be less than the sum of the probability of either individual event $X_1 > Y_1$ or $X_2 > Y_2$.

Finally, we have:

$$\begin{aligned}P\left(X_1X_2 > n^{\epsilon}Y_1Y_2\right) &< P(X_1 > n^{\epsilon/2}Y_1) + P(X_2 > n^{\epsilon/2}Y_2)\\&\leq n^{D_1} + n^{D_2}\\&\leq n^{D^*}\end{aligned} \tag{2}$$

Where the last line holds for $n \to \infty$. Thus we have shown $X_1 \prec Y_1$ and $X_2 \prec Y_2$ imply $X_1X_2 \prec Y_1Y_2$.

## 2.2

Let X be a random variable such that $X \prec \sqrt{n}$. We have:

$$\begin{aligned}\sqrt{n+X} - \sqrt{n} &= \left(\sqrt{n+X} - \sqrt{n}\right)\frac{\sqrt{n+X}+\sqrt{n}}{\sqrt{n+X}+\sqrt{n}}\\&= \frac{X}{\sqrt{n+X}+\sqrt{n}}\\&\leq \frac{X}{\sqrt{n}}\end{aligned} \tag{3}$$

Using the definition of stochastic dominance for $X$ and rearanging we find the desired result:

$$\begin{aligned}P\left(X > \sqrt{n}n^{\epsilon}\right) < n^{-D} &\Rightarrow P\left(\frac{X}{\sqrt{n}} > n^{\epsilon}\right) < n^{-D}\\&\Rightarrow P\left(\sqrt{n+X} - \sqrt{n} > n^{\epsilon}\right) < n^{-D}\\&\Rightarrow P\sqrt{n+X} - \sqrt{n} = \mathcal{O}_{\prec}(1)\\&\Rightarrow P\sqrt{n+X} = \sqrt{n} + \mathcal{O}_{\prec}(1)\end{aligned} \tag{4}$$

## 2.3

We note that the random variable $X_i^2$ can be written as the sum of the expectation value and a noise term: $X_i^2 = E\left(X_i^2\right) + \left(X_i^2 - 1\right)$. The expression for noise follows because $E\left(X_i^2\right) = 1$ by construction.

We can treat the noise term as a independent random variable $Y_i \equiv X_i^2 - 1$ with zero expectation value $EY_i = 0$. We will thus use Propsition 2 from the lecture notes with the particular case $a_i = 1$.

$$\sum a_iY_i \prec \sqrt{\sum a_i} = \sqrt{n} \tag{5}$$

Plugging this into the expression for $\sqrt{\sum X_i}$ we have:

$$
\begin{aligned}
\sqrt{\sum X_i^2} &= \sqrt{\sum E\left(X_i^2\right) + Y_i} \\
&= \sqrt{n + \mathcal{O}_{\prec}\left(\sqrt{n}\right)} \\
&= \sqrt{n} + \mathcal{O}_{\prec}\left(1\right)
\end{aligned} \tag{6}
$$

Where the last line follows from the result of the second part of this problem.

# 3

## 3.1

We first consider the upper bound:

$$
\begin{aligned}
P\left(M_n > k\sqrt{ln(y)}\right) &= P\left(\bigcup\{x_j > k\sqrt{\ln u}\}\right) \\
&\leq \sum P\left(x_j > k\sqrt{\ln u}\right) \quad \text{union bound} \\
&\leq n \cdot \exp\left(-\frac{1}{2}k^2 \log n\right) \\
&= n^{-k^2/2+1} \\
&= n^{-\epsilon}
\end{aligned} \tag{7}
$$

Where the last line follows setting $k = 2+\epsilon$. This shows that $P\left(M_n > k\sqrt{ln(y)}\right) \to 0$ as $n \to \infty$.

## 3.2

We first note that $P\left(M_n > \sqrt{(2-\epsilon)\cdot ln(n)}\right) = 1 - P\left(M_n \leq \sqrt{(2-\epsilon)\cdot ln(n)}\right)$ so it suffices to show that the second expression tends to 0 as $x \to \infty$.

For the maximum element $M_n$ to be less than some value $\xi$ it must be the case that *all* elements in the sequence to be less than $\xi$. Therefore, the probability of $M_n < \xi$ should be equal to the product of probability of $g_i < \xi$ for all elements $g_i$. I.e.

$$
\begin{aligned}
P\left(M_n \leq \sqrt{(2-\epsilon)ln(n)}\right) &= \prod_i P\left(g_i \leq \sqrt{(2-\epsilon)ln(n)}\right) \\
&= \prod_i 1 - P\left(g_i > \sqrt{(2-\epsilon)ln(n)}\right) \\
&= \left(\underbrace{1 - P\left(g_i > \sqrt{(2-\epsilon)ln(n)}\right)}_{\star}\right)^n
\end{aligned} \tag{8}
$$

We now note that the last expression can be rewritten as $e^{n \ln \star}$.

This expression can be simplified by observing that $\ln 1 + x \leq x$ for all $x > 0$, $\ln\left(1 - P\left(g_i > t\right)\right) < -P\left(g_i > t\right)$.

We now have:

$$\begin{aligned}
\Bigg(\underbrace{1 - P\left(g_i > \sqrt{(2-\epsilon)ln(n)}\right)}_{\star}\Bigg)^n &= \exp\left(n \cdot \ln\left(1 - P\left(g_i > t\right)\right)\right) \\
&\leq \exp\left(-n \cdot P\left(g_i > t\right)\right) \\
&\leq \exp\left(-n \cdot \left(\frac{1}{t} - \frac{1}{t^3}\right)\phi(t)\right)
\end{aligned} \tag{9}$$

Where in the last line we plugged in our result from problem (1). In our case $t = \sqrt{(2-\epsilon)ln(n)}$ . Thus we have $\phi(t) = e^{t^2/2} = n^{(-1+\epsilon/2)}$. Finally we have:

$$\begin{aligned}
P\left(M_n \leq \sqrt{(2-\epsilon)ln(n)}\right) &\leq \exp\left(-n^{\epsilon/2}\right) \\
&= 0 \quad \text{as } n \to \infty
\end{aligned} \tag{10}$$

# 4

In this problem I index the different observations $a$ by a raised variable, i.e. the $i$th observation is $a^{(i)}$, and the elements of vectors/matrices using subscripts $a_j$.

## 4.1

We consider the case $\xi = e^{(0)} = \{1, 0, 0, \dots\}$.

We begin by computing $y^{(i)} = \langle \xi, a^{(i)} \rangle = a_0^{(i)}$

We can plug this into the expression for $\Psi$ and take the expectation value.

$$E\left(\Psi\right) = E\left[\frac{1}{n}\sum_i \underbrace{\left(a_0^{(i)}\right)^2 a_j^{(i)} a_k^{(i)}}_{\star}\right] \tag{11}$$

The expectation value of $\star$ is given by:

$$E\left[\left(a_0^{(i)}\right)^2 a_j^{(i)} a_k^{(i)}\right] = \begin{cases} 3 & j = k = 0 \\ 1 & j = k \neq 0 \\ 0 & j \neq k \end{cases} \tag{12}$$

'

Thus we have

$$E\left(\Psi\right) = \begin{pmatrix} 3 & 0 & \cdots & \\ 0 & 1 & 0 & \cdots \\ \vdots & & \ddots & \\ \cdots & & & 1 \end{pmatrix} \tag{13}$$

For convenience in the next problem, we note that this can be written as:

$$2e^{(0)}e^{(0)\intercal} + \boldsymbol{I}$$

## 4.2

We now consider the case of arbitrary $\xi$. We would like to use the rotation trick to simplify our computation, so we assume we have a matrix (constructable from hausholder method) that rotates $\xi$ into $\vec{e}^{(0)}$. Dropping the vector notation for convenience, we have: $\xi = \|\xi\|\boldsymbol{H}^T e^{(0)}$ We also note that $\boldsymbol{H}$ in general rotates $a^{(i)}$ to a new vector $\tilde{a}^{(i)}$.

As before we first compute $y^{(i)}$:

$$\begin{aligned} y^{(i)} &= \langle \xi, a^{(i)} \rangle^2 \\ &= \langle \|\xi\|\boldsymbol{H}^\intercal e^{(i)}, \boldsymbol{H}^\intercal \tilde{a}^{(i)} \rangle^2 \\ &= \|\xi\| \langle e^{(i)}, \tilde{a}^{(i)} \rangle^2 \\ &= \|\xi\|^2 \left(\tilde{a}_0^{(i)}\right)^2 \end{aligned} \tag{14}$$

We now turn to $\Phi$:

$$\begin{aligned} E\left(\Psi\right) &= E\left[\frac{1}{n}\sum_i \|\xi\|^2 \left(a_0^{(i)}\right)^2 \boldsymbol{H}^\intercal \tilde{a}_j^{(i)} \left(\boldsymbol{H}^\intercal \tilde{a}^{(i)}\right)_k^\intercal\right] \\ &= E\left[\frac{1}{n}\sum_i \|\xi\|^2 \boldsymbol{H}^\intercal \underbrace{\left(\tilde{a}_0^{(i)}\right)^2 \tilde{a}^{(i)}\tilde{a}^{(i)}}_{\star} \boldsymbol{H}^\intercal\right] \end{aligned} \tag{15}$$

Since $a$ is a random variables, *tildea* is equivalent in distribution by rotational invariance. Thus we can treat $\tilde{a}$ as $a$ and note that we can replace $\star$ with the expectation values found in part 1. Thus we find:

$$\begin{aligned} E\left(\Phi\right) &= \|\xi\|^2 \boldsymbol{H}^\intercal \left(2e^{(0)}e^{(0)\intercal} + \boldsymbol{I}\right)\boldsymbol{H} \\ &= 2\xi\xi^\intercal + \|\xi\|^2 \boldsymbol{I} \end{aligned} \tag{16}$$

Where, in the second line we have again used the definition of our matrix $\boldsymbol{H}$ to rotate $e^{(0)}$ back.

$$E\left(\Phi\right)=\begin{pmatrix} 3\xi_0^2 & \xi_0\xi_1 & \cdots & \xi_0\xi_n \\ \xi_1\xi_0 & 3\xi_1^2 & \xi_1\xi_2 & \cdots \\ \vdots & & \ddots & \\ \xi_n\xi_1 & \dots & & 3\xi_n^2 \end{pmatrix} \tag{17}$$

### 4.3

WLOG we can consider the case $\xi=\|\xi\|e^{(0)}$. In this case we see that In this case the matrix $E\Phi$ will be close to diagonal with the largest eigenvalue in the upper left, $i=j=0$. Assuming there is some correlation among the observations $a$, then the