CSE 546: Homework 0

Andy Goldschmidt

Due: 10/4/18

1 Analysis

1.1

(a) Consider $x, y \in S := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ and $\lambda \in [0, 1]$. Compute:

$$\begin{aligned} \|\lambda x + (1-\lambda)y\| &\leq \|\lambda x\| + \|(1-\lambda)y\| && (1) \\ &= |\lambda|\|x\| + |1-\lambda|\|y\| && (2) \\ &= \lambda\|x\| + (1-\lambda)\|y\| && (3) \\ &\leq \lambda + (1-\lambda) && (4) \\ &= 1 && (5) \end{aligned}$$

Line 1 uses triangle inequality, line 2 uses absolutely scalability, and line 4 uses the definition of S.

(b) We will proceed by counterexample to show $\|x\| := (\sum_{i=1}^n |x_i|^{1/2})^2$ is not a norm. Consider x=(0,1) and y=(1,0). Compute:

$$\begin{aligned} \|x+y\| &= \|(1,1)\| && (6) \\ &= (1^{1/2} + 1^{1/2})^2 && (7) \\ &= 4 && (8) \end{aligned}$$

while

$$\begin{aligned} \|x\| + \|y\| &= \|(0,1)\| + \|(1,0)\| && (9) \\ &= (0^{1/2} + 1^{1/2})^2 + (1^{1/2} + 0^{1/2})^2 && (10) \\ &= 2 && (11) \end{aligned}$$

so that $\|x+y\| > \|x\| + \|y\|$, contradicting the triangle inequality.

1.2 Show: $\|x\|_\infty \le \|x\|_2 \le \|x\|_1$.
For the first inequality (w.l.o.g. let $\|x\|_\infty = |x_1|$):

$$\begin{aligned}
\|x\|_\infty &= (x_1^2)^{1/2} && (12)\\
&\le ({x_1}^2 + \sum_{i=2}^{n} {x_i}^2)^{1/2} && (13)\\
&= \|x\|_2 && (14)
\end{aligned}$$

For the second inequality:

$$\begin{aligned}
\|x\|_2 &= (\sum_{i=1}^{n} |x_i|^2)^{1/2} && (15)\\
&\le (\sum_{i=1}^{n} |x_i|^2 + \sum_{i\neq j} |x_i||x_j|)^{1/2} && (16)\\
&= ((\sum_{i=1}^{n} |x_i|)^2)^{1/2} && (17)\\
&= \|x\|_1 && (18)
\end{aligned}$$

where we add in the positive-valued cross terms in line 16 and factor in line 17.

1.3
Let $A, B \in \mathbb{R}^{n\times n}$ have entries labeled $(a_{ij}), (b_{ij})$. Let $c \in \mathbb{R}$. Define $f(x, y) = x^T Ax + y^T Bx + c$. Compute: $\boldsymbol{\nabla}_x$ and $\boldsymbol{\nabla}_y$. We will do this by rewriting f where repeated indices mean implied sums.

$$f(x, y) = x_j a_{ji} x_i + y_j b_{ji} x_i + c \tag{19}$$

Now compute:

$$\begin{aligned}
\boldsymbol{\nabla}_x f(x, y) &= \partial_{x_k}(x_j a_{ji} x_i + y_j b_{ji} x_i + c)\hat{k} && (20)\\
&= (\partial_{x_k}(x_j) a_{ji} x_i + x_j a_{ji} \partial_{x_k}(x_i) + y_j b_{ji} \partial_{x_k}(x_i))\hat{k} && (21)\\
&= (\delta_{kj} a_{ji} x_i + x_j a_{ji} \delta_{ki} + y_j b_{ji} \delta_{ki})\hat{k} && (22)\\
&= (a_{ki} x_i + x_j a_{jk} + y_j b_{jk})\hat{k} && (23)\\
&= Ax + A^T x + B^T y && (24)
\end{aligned}$$

Similarly:

$$\begin{aligned}
\boldsymbol{\nabla}_y f(x, y) &= \partial_{y_k}(x_j a_{ji} x_i + y_j b_{ji} x_i + c)\hat{k} && (25)\\
&= (\partial_{y_k}(y_j) b_{ji}(x_i))\hat{k} && (26)\\
&= (\delta_{kj} b_{ji}(x_i))\hat{k} && (27)\\
&= (b_{ki}(x_i))\hat{k} && (28)\\
&= Bx && (29)
\end{aligned}$$

1.4
Every matrix in this problem shares the same set of eigenvectors. Suppose A and B have eigenvalues a_i and b_i, respectively. The desired eigenvalues are:

$$(a)\ C = A + B : a_i + b_i$$
$$(b)\ D = A - B : a_i - b_i$$
$$(c)\ E = AB : a_i * b_i$$
$$(d)\ F = A^{-1}B : b_i/a_i$$

The quickest way to see this is to express the two operators A and B as diagonal matrices in their shared eigenbasis. After rewriting, the matrix operations remaining are straightforward and immediately yield the eigenvalues of these newly computed matrices. (For part (d), the inverse matrix is unique so the form of A^{-1} can be checked).

1.5
(a) Given $x, y \in \mathbb{R}^n$, compute:

$$x^T(yy^T)x = (x^Ty)(y^Tx) = (y^Tx)^T(y^Tx) = (y^Tx)^2 \geq 0 \tag{30}$$

We have been somewhat loose here in justifying the exchange of parenthesis that lets the matrix yy^T split to act separately on both x and x^T in their spaces. We could show this by playing with indices, but it's possibly quicker to see in the following notation: $x^T(yy^T)x \longrightarrow \langle x| (|y\rangle \langle y|) |x\rangle = \langle x|y\rangle \langle y|x\rangle$. What remains in the last result is simply a constant squared.

(b) Let $X \in \mathbb{R}^n$ be a random vector. $\forall y \in \mathbb{R}^n$, compute:

$$y^T\Sigma y = y^T\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]y \tag{31}$$
$$= \mathbb{E}[y^T(X - \mathbb{E}[X])(X - \mathbb{E}[X])^Ty] \geq 0 \tag{32}$$

In line 32, we use the linearity of expectation and note that from part (a) the quantity in the expectation is positive. The expectation of a positive quantity is positive so we have shown that Σ is postive semi-definite. (c) Suppose A is a symmetric matrix so that $A = U\text{diag}(\alpha)U^T$ with $U^TU = I$. We want to show that A is positive semi-definite $\Leftrightarrow \min\limits_i \alpha_i \geq 0$.
($\Leftarrow$) Suppose the latter, and consider the basis where A is diagonal. Note that an inner product is independent of basis so we can prove our positive semi-defininte result in any basis. $\forall x$ in this basis we have:

$$x^TAx = \sum_{i=1}^{n} x_i\alpha_i x_i = \sum_{i=1}^{n} {x_i}^2\alpha_i \geq 0 \tag{33}$$

where we use the assumption to assert the final inequality.
($\Rightarrow$) Now suppose A is positive semi-definite and that $\min\limits_i \alpha_i < 0$. W.l.o.g. let this minimal index be 1. Now $\exists x$ such that

$$x^TAx = x_1\alpha_1 x_1 = {x_1}^2\alpha_1 < 0 \tag{34}$$

This contradicts the assumption that A was positive semi-defininte.

1.6

(a) Let X and Y be random variables with PDFs f and g, respectively. Consider Z=X+Y. We will throw caution to the wind and use a continuous version of the law of total probability:

$$\begin{aligned} P(Z = z) &= P(X + Y = z) && (35) \\ &= \int \mathrm{d}x \, P(X + Y = z|X = x)P(X = x) && (36) \\ &= \int \mathrm{d}x \, P(Y = z - x)P(X = x) && (37) \\ &= \int \mathrm{d}x \, g(z - x)f(x) =: h(z) && (38) \end{aligned}$$

(b) Suppose X and Y are both independent and uniformly distributed on [0,1]. Compute:

$$\begin{aligned} h(z) &= \int_0^1 \mathrm{d}x \, f(x)g(z - x) && (39) \\ &= \int_0^1 \mathrm{d}x \, g(z - x) && (40) \\ &= \int_{z-1}^{z} \mathrm{d}u \, g(u) && (41) \\ &= \begin{cases} \int_0^z \mathrm{d}u = z, z \leq 1 \\ \int_{z-1}^1 \mathrm{d}u = 2 - z, z \geq 1 \end{cases} && (42) \\ &= 1 - |z - 1| && (43) \end{aligned}$$

(c) Compute:

$$\begin{aligned} P(X \leq 1/2|X + Y \geq 5/4) &= \frac{P(X \leq 1/2 \cap X + Y \geq 5/4)}{P(X + Y \geq 5/4)} && (44) \\ &= \frac{(1/2)(1/4)^2}{(1/2)(3/4)^2} = \frac{1}{9} && (45) \end{aligned}$$

The most effective argument comes from geometry and would be aided by pictures that are not included here.

1.7

We want the adjusted random variable: $(X - \mu)/\sigma$. With this choice we have

$$\begin{aligned} E[(X - \mu)/\sigma] &= E[(X - \mu)/\sigma] && (46) \\ &= (E[X] - \mu)/\sigma = 0 && (47) \end{aligned}$$

and

$$E[((X-\mu)/\sigma - E[(X-\mu)/\sigma])^2] \tag{48}$$
$$= E[((X-\mu)/\sigma - 0)^2] \tag{49}$$
$$= E[(X-E[X])^2]/\sigma^2 = 1 \tag{50}$$

as desired.

1.8
(a) $\forall x$, compute:

$$E[\hat{F}_n(x)] = E[\frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\}] \tag{51}$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[I\{X_i \le x\}] \tag{52}$$
$$= \frac{1}{n}\sum_{i=1}^{n} F(x) = F(x) \tag{53}$$

(b) $\forall x$, compute:

$$E[(\hat{F}_n(x) - F(x))^2] \tag{54}$$
$$= E[\hat{F}_n(x)^2] - F(x)^2 \tag{55}$$
$$= E[(\frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\})^2] - F(x)^2 \tag{56}$$
$$= \frac{1}{n^2}E[\sum_{i=1}^{n} I\{X_i \le x\} + \sum_{i\neq j} I\{X_i \le x\}I\{X_j \le x\}] - F(x)^2 \tag{57}$$
$$= \frac{1}{n}E[I\{X_1 \le x\}] + \frac{n(n-1)}{n^2}E[I\{X_1 \le x\}I\{X_2 \le x\}] - F(x)^2 \tag{58}$$
$$= \frac{1}{n}F(x) + \frac{n-1}{n}F(x)^2 - F(x)^2 \tag{59}$$
$$= \frac{1}{n}F(x)(1-F(x)) \tag{60}$$

We arrive at Eq. 57 by noting the idempotency of an indicator function. For Eq. 58, we use that each random variable is identically distributed and count the elements in each sum.

(c) Using part b., we can bound:

$$\sup_{x\in\mathbb{R}} E[(\hat{F}_n(x) - F(x))^2] \tag{61}$$

$$= \sup_{x\in\mathbb{R}} \frac{1}{n} F(x)(1 - F(x)) \tag{62}$$

$$\leq \frac{1}{n}\frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4n} \tag{63}$$

We use some calculus to argue $y(1-y)$ is maximal at $y = 1/2$.

2 Programming

2.9 For (a) and (b), we are to produce Fig. 1. We choose $n = 40,000$ in the inequality in Eq. 63 for the bound $\sup_{x\in\mathbb{R}} \sqrt{E[(\hat{F}_n(x) - F(x))^2]} \leq 0.0025$ to hold. Fig. 1 shows that the sum of k independent, zero-mean random variables with variance 1/k converges to a standard-normal distribution as $k \longrightarrow \infty$. The figure shows this by plotting the empirical cumulative distribution functions of such sums of random variables and comparing these to the cumulative distribution function of a standard-normal-distributed random variable.

My code:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as ss

# Set-up: Iterations and random seed
n = 40000
seed = 1
np.random.seed(1)

# Define a very literal FnHat.
# n.b. could also have done this by sorting x-data and
    labelling, as was suggested.
Fn_Hat = lambda Z,X: [np.sum(Z <= xi)/len(Z) for xi in X]

# Create plot resolution (x values) and figure
    environment
x = np.linspace(-3,3,1000)
fig = plt.figure()
ax = fig.gca()
legend = []

# (a) Normally-distributed data
aZ = np.random.normal(loc=0, scale=1, size=n)
```

```
# (b) Central limit theorem data: Generate and plot
for k in [1,8,64,512]:
    # Creates a vector of n versions of Y_k. The Y_k are
        made from k summed variables B_i
    vecYk = np.sum(np.random.choice([-1,1],size=(k,n)),
        axis=0)/np.sqrt(k)
    ax.plot(x, Fn_Hat(vecYk, x))
    legend.append('k={}'.format(k))

# Plot the part (a) data last.
ax.plot(x,Fn_Hat(aZ,x))
legend.append('N(0,1)')

# Plot asthetics
ax.legend(legend, fontsize=12)
ax.grid(alpha=0.33)
plt.xlabel('Observations', fontsize=12)
plt.ylabel('Probability', fontsize=12)
plt.savefig('./plot_hw0p9.png')
```

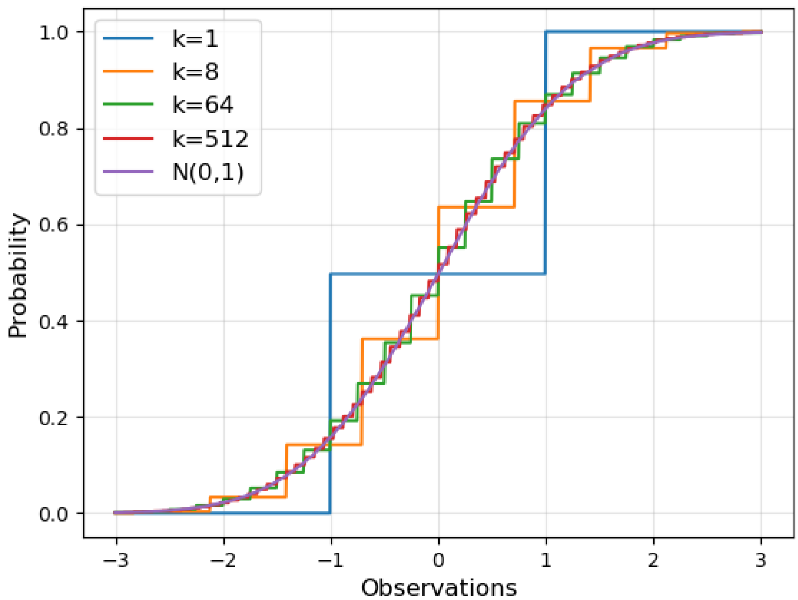

Figure 1: The figure shows $\hat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} I\{Y_j^{(k)} \leq x\}$ where $Y_j^{(k)} = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} B_i$ for n=40,000 for various values of k. For reference the plot includes $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i \leq x\}$ with standard-normal random variables $Z_i \sim N(0,1)$ because these will have exactly the empirical cumulative distribution function that the central limit theorems asserts $Y^{(k)}$ approaches in the large k limit.