CSE 512: Homework 1 Due Wednesday September 8

1. **(1 pts, 0.25 pts each)** *Linearity.* Are the following functions linear? Justify your answer.

   (a) $f(x) = \|x\|_2^2$

   (b) $f(x) = \|x\|_1$

   (c) $f(x) = \frac{1}{2}x^TQx + p^Tx + r$

   (d) $f(x) = c^Tx + b^TAx$

2. **(1 pt, 0.25 each)** Using the properties of norms, verify that the following are norms, or prove that they are not norms

   (a) *Direct sum.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \sum_k x[k]$

   Ans. This is not a norm, since it is not nonnegative everywhere.

   (b) *Sum of square roots, squared.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \left(\sum_{k=1}^d \sqrt{|x[k]|}\right)^2$

   Ans. This is not a norm, since it is cannot satisfy triangle inequality. In particular, just take $d = 2$ and

$$f\left(\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}\right) = f\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = (\sqrt{1} + \sqrt{1})^2 = 4$$

$$f\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + f\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = 1 + 1 = 2$$

   and therefore we have shown that $f(x + y) > f(x) + f(y)$ for some choice of x, y.

   (c) *(Shifted) entropy function* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = -\sum_{i=1}^d (x[i] + \frac{1}{2})\log_2(x[i] + \frac{1}{2})$ where $-1/2 \leq x[i] \leq 1/2$ for all $i = 1, ..., d$

   Ans. No, this is not a norm. For one, the function is 0 when $x[i] = /1/2$ for any i, violating the 0 at 0 property. For another, it is not positive homogeneous anywhere. Triangle inequality can also be easily broken: pick $x = 1/2$, $y = 1/2$. Then $f(x) + f(y) = 0$ but $f(x + y) = f(1) = 1/2 > 0$. Showing any example that breaks these properties works.

   (d) *Weighted 2-norm.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \sqrt{\sum_{k=1}^d \frac{|x[k]|^2}{k}}$

   Ans. Yes, this is a norm. To see that, first note that we can write

$$f(x) = \|Wx\|_2, \qquad W = \mathbf{diag}(1, 1/2, 1/3, ..., 1/d).$$

   Then we can just go about checking the norm conditions.

   - 0 at 0? Yes, $f(0) = \|W\,0\|_2 = \|0\|_2 = 0$
   - Positive homogeneity?

$$f(\alpha x) = \|W(\alpha x)\|_2 = |\alpha|\|Wx\|_2 = |\alpha|f(x)$$

     Check.

   - Triangle inequality?

$$f(x + y) = \|W(x + y)\|_2 = \|Wx + Wy\|_2 \overset{\Delta\text{-ineq of 2-norm}}{\leq} \|Wx\|_2 + \|Wy\|_2 = f(x) + f(y)$$

     Check!

   Therefore this is a norm.

3. **(1 pt, 0.25 each)** *Independent or not independent.* Variables A and B are random variables for two distributions. Decide if A and B are independent. Justify your answer.

(a) A and B are discrete random variables and have the following p.m.f.s

$$p_A(a) = \begin{cases} 0.25, & a = \text{red} \\ 0.25, & a = \text{blue} \\ 0.5, & a = \text{green} \end{cases}, \qquad p_B(b) = \begin{cases} 0.3, & b = \text{hat} \\ 0.3, & b = \text{T-shirt} \\ 0.2, & b = \text{skirt} \\ 0.2, & b = \text{shoes} \end{cases}$$

and $p_{A,B}(a,b)$ are defined by the table below

	a = red	a = blue	a = green
b = hat	0.075	0.075	0.15
b = T-shirt	0.075	0.075	0.15
b = skirt	0.05	0.05	0.1
b = shoes	0.05	0.05	0.1

Ans. A and B are **independent**. This can be shown by systematically verifying that

$$p_{A,B}(a,b) = p_A(a)p_B(b)$$

for every combination of a and b.

(b) A and B are uniform distributions, where

$$f_A(a) = \begin{cases} 1 & -1 \leq a \leq 0 \\ 0 & \text{else,} \end{cases} \qquad f_B(b) = \begin{cases} 1 & 0 \leq b \leq 1 \\ 0 & \text{else,} \end{cases}, \qquad f_{A,B}(a,b) = \begin{cases} 4/3 & |a+b| \leq 1/2 \\ 0 & \text{else,} \end{cases}$$

Ans. A and B are **not independent**. For example, taking instantiations $a = -0.1$ and $b = 0.9$,

$$f_A(a) = 1, \quad f_B(b) = 1, \quad f_{A,B}(a,b) = 0 \neq f_A(a) \cdot f_B(b).$$

(c) A follows the p.m.f.

$$p_A(a) = \begin{cases} 0.5, & a = 1 \\ 0.5, & a = -1 \end{cases}$$

and $B = A \cdot C$ where

$$p_C(c) = \begin{cases} 0.9, & c = 1 \\ 0.1, & c = -1 \end{cases}$$

Ans. A and B are **not independent**. First, use chain rule to see that

$$p_{A,B}(a,b) = p_B(b|a)p_A(a) \overset{c=b/a}{=} p_C(b/a)p_A(a).$$

Now, this quantity we can work out to be

$$p_C(b/a)p_A(a) = \begin{cases} 0.9 \cdot 0.5 & \text{if } b = 1, a = 1 \\ 0.1 \cdot 0.5 & \text{if } b = 1, a = -1 \\ 0.1 \cdot 0.5 & \text{if } b = -1, a = 1 \\ 0.9 \cdot 0.5 & \text{if } b = -1, a = -1. \end{cases}$$

Next, we use law of total probability to get

$$\begin{aligned} p_B(b) &= \sum_{a \in \{-1,1\}} p_B(b|a)p_A(a) \\ &\overset{c=b/a}{=} \sum_{a \in \{-1,1\}} p_C(b/a)p_A(a) \\ &= \begin{cases} (0.9 + 0.1) \cdot 0.5 & \text{if } b = 1 \\ (0.9 + 0.1) \cdot 0.5 & \text{if } b = -1 \end{cases} \end{aligned}$$

From this, it is clear that $p_{A,B}(a,b) \neq p_A(a)p_B(b)$.

(d) A and B are Gaussian distributions, with the following properties:

$$\mathbb{E}[A] = 0, \quad \mathbb{E}[B] = 1, \quad \mathbb{E}[A^2] = 1, \quad \mathbb{E}[(B-1)^2] = 1/2, \quad \mathbb{E}[A(B-1)] = -1.$$

Writing in terms of the usual Gaussian distribution form, if we form a random vector as $X = \begin{bmatrix} A \\ B \end{bmatrix}$, then

$$\mu = \begin{bmatrix} \mathbb{E}[A] \\ \mathbb{E}[B] \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \mathbb{E}[(A-\mathbb{E}[A])^2] & \mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] \\ \mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] & \mathbb{E}[(B-\mathbb{E}[B])^2] \end{bmatrix}$$

Ans. A and B are **not independent**. This can be seen by looking at the off-diagonal elements of Σ. In particular,

$$\mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] = \mathbb{E}[A(B-1)] = -1 \neq 0$$

and therefore Σ is not diagonal.

4. **(2 pts)** *Probability and statistics.* I have 4 children, Alexa, Siri, Googs, and Zuckie. Every morning I tell them to put on their socks.

- Alexa only listens to me on Mondays and Thursdays and puts on her socks. The rest of the days, she puts on her socks only half of the time. She either puts on both her socks or none of her socks.
- Siri always runs and gets her socks, but only puts one sock on.
- Googs tells me all this random trivia about socks, but never puts on his socks.
- Zuckie wears both his socks 4/7 of the time and sells the rest of them to CambridgeAnalytica.

Assume the children all act independently. Round all answers to at least 3 significant digits.

(a) **(0.5 pts)** What are the chances that either Alexa or Zuckie is wearing a sock?
Ans. This is a classic example of the **intersection rule**

$$\mathbf{Pr}(A \text{ and } B) = \mathbf{Pr}(A) + \mathbf{Pr}(B) - \mathbf{Pr}(A \cap B).$$

Specifically, since

$$\mathbf{Pr}(\text{Alexa is wearing a sock}) = \frac{2}{7} + \frac{1}{2} \cdot \frac{5}{7} = \frac{9}{14}$$

$$\begin{aligned} \mathbf{Pr}(\text{Alexa or Zuckie is wearing a sock}) &= \mathbf{Pr}(\text{Alexa is wearing a sock}) + \mathbf{Pr}(\text{Zuckie is wearing a sock}) \\ &\quad -\mathbf{Pr}(\text{both are wearing a sock}) \\ &= \frac{9}{14} + \frac{4}{7} - \frac{9}{14} \cdot \frac{4}{7} = \approx 84.7\% \end{aligned}$$

(b) **(0.5 pts)** On a random day, a girl is wearing a sock. What are the chances that it's Alexa?
Ans. Here, we use Bayes' rule:

$$\begin{aligned} & \mathbf{Pr}(\text{Alexa is wearing a sock} \mid \text{a girl is wearing a sock}) \\ = & \frac{\mathbf{Pr}(\text{a girl is wearing a sock} | \text{Alexa is wearing a sock}) \cdot \mathbf{Pr}(\text{Alexa is wearing a sock})}{\sum_{\text{child } i} \mathbf{Pr}(\text{a girl is wearing a sock} \mid \text{child } i \text{ is wearing a sock}) \cdot \mathbf{Pr}(\text{child } i \text{ is wearing a sock})} \\ = & \frac{\mathbf{Pr}(\text{Alexa is wearing a sock})}{\mathbf{Pr}(\text{Alexa is wearing a sock}) + \mathbf{Pr}(\text{Siri is wearing a sock})} \\ = & \frac{\frac{9}{14}}{\frac{9}{14} + 1} = \frac{9}{23} \approx 39.1\% \end{aligned}$$

(c) **(0.5 pts)** What is the expected number of socks being worn by each child?
Ans. We directly use the expectation definition:

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot \mathbf{Pr}(\text{no socks}) + 1 \cdot \mathbf{Pr}(1 sock) + 2 \cdot \mathbf{Pr}(2 socks).$$

For Alexa,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot \frac{1}{2} \cdot \frac{5}{7} + 1 \cdot 0 + 2 \cdot \frac{9}{14} = \frac{9}{7} \approx 1.29.$$

For Siri,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1$$

For Googs,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 = 0$$

And for Zuckie,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 3/7 + 1 \cdot 0 + 2 \cdot 4/7 = 8/7 \approx 1.14.$$

(d) **(0.5 pts)** What is the variance in the number of socks being work by each child?
Ans. We directly use the variance definition:

$$\mathbf{var}(\#\text{ socks}) = \mathbb{E}[(\#\text{ socks} - \mathbb{E}[\#\text{ socks}])^2).$$

For Siri and Googs, the mean and values are always the same, and they have 0 variance.
For Alexa,

$$\mathbf{var}(\#\text{ socks}) = (0 - \frac{9}{7})^2 \cdot \frac{5}{14} + (1 - \frac{9}{7})^2 \cdot 0 + (2 - \frac{9}{7})^2 \cdot \frac{9}{14} \approx 0.918.$$

And for Zuckie,

$$\mathbf{var}(\#\text{ socks}) = (0 - \frac{2}{7})^2 \cdot \frac{3}{7} + (1 - \frac{2}{7})^2 \cdot 0 + (2 - \frac{2}{7})^2 \cdot \frac{4}{7} \approx 1.45.$$

5. **(2 pts, 0.5 points each)** *Conditional independence vs independence.* Tom is a blue-gray cat with a bushy tail, and Jerry is a brown mouse with a rope-like tail. After many years of fighting, they both decided to settle down, and now have thriving families. Tom has 10 kids and Jerry has 40 kids. Tom's kids are all cats like him, with bushy tails. Half of Tom's kids are blue, while the other half is gray. Jerry's kids are all brown mice, with rope-like tails.

(a) I pick up a baby animal at random. What is the probability that ... (fill in the table)

fur \ tail	furry	rope-like
blue		
gray		
brown		

Ans.

fur \ tail	furry	rope-like
blue	10%	0
gray	10%	0
brown	0	80%

(b) Are the features "fur color" and "tail texture" independent or dependent, without knowing the type of animal? (Show mathematically.)

Ans. Yes, they are correlated. To take as an example,

$$\begin{aligned}
\mathbf{Pr}(\text{blue, furry}) &= 5/50 = 1/10 \\
\mathbf{Pr}(\text{blue}) &= 5/50 = 1/10 \\
\mathbf{Pr}(\text{furry}) &= 10/50 = 1/5 \\
\mathbf{Pr}(\text{blue}) \cdot \mathbf{Pr}(\text{furry}) &= 1/50 \neq 1/10
\end{aligned}$$

(c) Now Tom comes over and says, "I'm very proud of my baby girl, of whom you are holding." What is the probability that (fill in the table)

fur \ tail	furry	rope-like
blue		
gray		
brown		

Ans.

fur \ tail	furry	rope-like
blue	50%	0
gray	50%	0
brown	0	0

(d) Are the features "fur color" and "tail texture" independent or dependent, now that I know the animal is Tom's cherished baby daughter? (Show mathematically.)

Ans. No, now the features are uncorrelated. Specifically,

$$\begin{aligned}
\underbrace{\mathbf{Pr}(\text{blue})}_{1/2} \cdot \underbrace{\mathbf{Pr}(\text{furry})}_{1} &= \underbrace{\mathbf{Pr}(\text{blue, fuzzy})}_{1/2} \\
\underbrace{\mathbf{Pr}(\text{gray})}_{1/2} \cdot \underbrace{\mathbf{Pr}(\text{furry})}_{1} &= \underbrace{\mathbf{Pr}(\text{gray, furry})}_{1/2} \\
\underbrace{\mathbf{Pr}(\text{blue})}_{1/2} \cdot \underbrace{\mathbf{Pr}(\text{rope-like})}_{0} &= \underbrace{\mathbf{Pr}(\text{blue, furry})}_{0} \\
\underbrace{\mathbf{Pr}(\text{gray})}_{1/2} \cdot \underbrace{\mathbf{Pr}(\text{rope-like})}_{0} &= \underbrace{\mathbf{Pr}(\text{gray, furry})}_{0}
\end{aligned}$$

6. **(3 pts) Exponential distribution.** Wait time is often modeled as an exponential distribution, e.g.

$$\mathbf{Pr}(\text{I wait less than } x \text{ hours at the DMV}) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

and this cumulative density function is parametrized by some constant $\lambda > 0$. A random variable X distributed according to this CDF is denoted as $X \sim \exp[\lambda]$.

(a) **(0.25 pts)** In terms of λ, give the probability distribution function for the exponential distribution.

Ans. The PDF can be computed as just the derivative of the CDF, which comes to

$$p_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

(b) **(0.25 pts)** Show that if $X \sim \exp(\lambda)$, then the mean of X is $1/\lambda$ and the variance is $1/\lambda^2$.
(You may use a symbolic integration tool such as Wolfram Alpha. If you do wish to do the integral by hand, my hint is to review integration by parts.)

Ans. To compute the mean,

$$\mathbb{E}[X] = \int_0^\infty x p_\lambda(x) dx = \int_0^\infty \lambda x e^{-\lambda x} dx$$

Now the rest is an exercise in integration. Using Wolfram Alpha,

$$\int_0^\infty \lambda x e^{-\lambda x} dx = \frac{e^{-\lambda x}}{\lambda}(\lambda x - 1)|_0^\infty = 0 - \left(-\frac{1}{\lambda}\right) = 1/\lambda$$

The same result can be arrived at by using integration by parts.
To compute the variance, recall that $\mathbf{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Then

$$\mathbb{E}[X^2] = \int_0^\infty x^2 p_\lambda(x) dx = \int_0^\infty \lambda x^2 e^{-\lambda x} dx = \exp(-\lambda x)\left(-\frac{2}{\lambda^2} - \frac{2x}{\lambda} - x^2\right)|_0^\infty = \frac{2}{\lambda^2}$$

$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(c) **(0.5 pts)** Now suppose I run a huge server farm, and I am monitoring the server's ability to respond to web requests. I have m observations of delay times, $x_1, ..., x_m$, which I assume are i.i.d., distributed according to $\exp[\lambda]$ for some λ. Given these m observations, what is the maximum likelihood estimate $\hat{\lambda}$ of λ?

Ans. First, we compute the likelihood of observations $x_1, ..., x_m$ given that they are i.i.d., distributed as $\exp[\lambda]$:

$$\mathbf{Pr}(x_1, ..., x_m) = \prod_{i=1}^m \mathbf{Pr}(x_i) = \prod_{i=1}^m (\lambda e^{-\lambda x_i}) = \lambda^m \exp\left(-\lambda \sum_{i=1}^m x_i\right).$$

I would like to find λ which maximizes this quantity. However, this expression looks pretty complicated–not convex or concave.
Let's use a trick that we are now pretty familiar with: take the log.

$$\log(\mathbf{Pr}(x_1, ..., x_m)) = m \log(\lambda) - \lambda \sum_{i=1}^m x_i$$

This is a concave function of λ, so now we can find the maximum of the log probability by taking the derivative and setting it to 0:

$$\frac{\partial}{\partial \lambda} \log(\mathbf{Pr}(x_1, ..., x_m)) = \frac{m}{\lambda} - \sum_{i=1}^m x_i = 0 \Rightarrow \frac{1}{\hat{\lambda}_{\mathbf{MLE}}} = \frac{1}{m}\sum_{i=1}^m x_i \Rightarrow \hat{\lambda}_{\mathbf{MLE}} = \frac{m}{\sum_{i=1}^m x_i}.$$

(d) **(1 pt)** Given the estimate of $\hat{\lambda}$ in your previous question, is $1/\hat{\lambda}$ an unbiased estimate of the mean wait time? Is $1/\hat{\lambda}^2$ an unbiased estimate of the variance in wait time?

Ans. The term $1/\hat{\lambda}$ is indeed an unbiased estimator of the mean:

$$\mathbb{E}\left[\frac{1}{\hat{\lambda}}\right] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^m x_i\right] = \frac{1}{m}\sum_{i=1}^m \mathbb{E}[x_i] = \frac{1}{\lambda}$$

However, the term $1/\hat{\lambda}^2$ is a biased estimator of the variance:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\hat{\lambda}^2}\right] &= \mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^m x_i\right)^2\right] \\
&= \frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m \underbrace{\mathbb{E}[x_i x_j]}_{\text{i.i.d.}} \\
&= \frac{1}{m^2}\left(\sum_{i\neq j}\mathbb{E}[x_i]\mathbb{E}[x_j] + \sum_{i=j}\mathbb{E}[x_i^2]\right).
\end{aligned}$$

Noting that $\mathbb{E}[x_i] = \mathbb{E}[x_j] = \frac{1}{\lambda}$ and

$$\mathbb{E}[x_i^2] = \textbf{var}(x_i) + (\mathbb{E}[x_i])^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$$

then

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\hat{\lambda}^2}\right] &= \frac{1}{m^2}\left(\sum_{i\neq j}\frac{1}{\lambda^2} + \sum_{i=j}\frac{2}{\lambda^2}\right) \\
&= \frac{1}{m^2}\left((m^2-m)\frac{1}{\lambda^2} + m\frac{2}{\lambda^2}\right) \\
&= \frac{m+1}{m}\frac{1}{\lambda^2} \neq \frac{1}{\lambda^2}.
\end{aligned}$$

The bias, however, disappears in the limit of $m \to +\infty$.

(e) **(1 pt)** Now let's consider $x_1, ..., x_m$ drawn i.i.d. from a *truncated* exponential distribution, e.g.

$$p_{\lambda,c}(x) = \begin{cases} 0 & \text{if } x > c \text{ or } x < 0 \\ \dfrac{\lambda\exp(-\lambda x)}{1-\exp(-\lambda c)} & \text{else.} \end{cases}$$

Using Hoefdding's inequality, give a range of values that account for the uncertainty in your guess. That is, as a function of x_i, m and δ, give a range of values $[\hat{\lambda}_{\min}, \hat{\lambda}_{\max}]$ such that

$$\mathbf{Pr}(\hat{\lambda}_{\min} \le \mathbb{E}[X] \le \hat{\lambda}_{\max}) \ge 1 - \delta.$$

Ans. This problem is a little bit flawed, in that the sample mean computed in your previous problem doesn't really apply to the truncated distribution.

However, hopefully you still got the intention of this problem, which is to try to apply Hoeffding's inequality to bound an estimate of the mean, given that the distribution now has a bounded support. (e.g. there exists limits l and u where if $x > u$ or $x < l$ then $\mathbf{Pr}(X = x) = 0$.)

In particular, via this truncation, we see that each x_i must be between 0 and c; otherwise, it occurs with 0 probability. Therefore, we can think of a new variable $z_i := x_i/c$, and in fact for the random variable $Z = X/c$, $\mathbb{E}[Z] = \mathbb{E}[X/c] = \frac{1}{c}\mathbb{E}[X]$.

Then, the random variable Z satisfies all the conditions needed for Hoeffding's inequality; e.g.,

$$\mathbf{Pr}\left(\frac{1}{m}\sum_{i=1}^m z_i - \mathbb{E}[Z] \ge \epsilon\right) \le e^{-2m\epsilon^2}$$

and at the same time,

$$\mathbf{Pr}\left(\frac{1}{m}\sum_{i=1}^m z_i - \mathbb{E}[Z] \le -\epsilon\right) \le e^{-2m\epsilon^2}.$$

This gives us the two-sided Hoeffding's inequality, via the union bound: [1]

$$\mathbf{Pr}\left(|\frac{1}{m}\sum_{i=1}^{m} z_i - \mathbb{E}[Z]| \leq \epsilon\right) \leq 2e^{-2m\epsilon^2}.$$

Plugging in $x_i = cz_i$ and $X = cZ$, then

$$\mathbf{Pr}\left(|\frac{1}{m}\sum_{i=1}^{m} \frac{x_i}{c} - \frac{\mathbb{E}[X]}{c}| \geq \epsilon\right) \leq 2e^{-2m\epsilon^2}.$$

Since I want an estimator for $\mathbb{E}[X]$ and not $\mathbb{E}[X]/c$, we need to rescale things:

$$\mathbf{Pr}\left(|\frac{1}{m}\sum_{i=1}^{m} x_i - \mathbb{E}[X]| \geq c\epsilon\right) \leq 2e^{-2m\epsilon^2}.$$

Ok, now we're almost there! What we basically can see now is that

$$\frac{1}{m}\sum_{i=1}^{m} x_i - c\epsilon \leq \mathbb{E}[X] \leq \frac{1}{m}\sum_{i=1}^{m} x_i + c\epsilon$$

with probability $1 - 2e^{-2m\epsilon^2}$. Taking $\delta = 2e^{-2m\epsilon^2}$, we see that

$$\epsilon = \sqrt{\frac{\log(2/\delta)}{2m}}.$$

The final answer should look like

$$\hat{\lambda}_{\min} = \frac{1}{m}\sum_{i=1}^{m} x_i - c\sqrt{\frac{\log(2/\delta)}{2m}}, \qquad \hat{\lambda}_{\max} = \frac{1}{m}\sum_{i=1}^{m} x_i + c\sqrt{\frac{\log(2/\delta)}{2m}}$$

[1] The union bound says that $\mathbf{Pr}(A \text{ or } B) \leq \mathbf{Pr}(A) + \mathbf{Pr}(B)$.