Introduction to Data Science

Group Assignment 2

Group 5:
Lina Alsughair
Erik Pettersson
Elias Sanchez
Sung Pok Yau

1 Proof of Corollary 3.7

Corollary 3.7 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let $X_1, ..., X_n \overset{IID}{\sim} F$ $\mathbb{R}$-valued RVs such that $\mathbb{P}(X_i \in [a, b]) = 1$, then for any $\epsilon > 0$ we get for $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,*

$$\mathbb{P}(\bar{X}_n - \mathbb{E}[\bar{X}_n] \leq -\epsilon) \leq e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

furthermore

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \epsilon) \leq 2e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

Proof. Since $X_1, X_2, ..., X_n$ are IID, they have symmetric positive and negative values. In Theorem 3.6 proof, there were no assumptions about the sign of $(\bar{X}_n - \mathbb{E}[\bar{X}_n])$ values. In Corollary 3.7, we use the absolute value for $(\bar{X}_n - \mathbb{E}[\bar{X}_n])$, so there are no negative values to cancel their equivalent positive values, and this will double the right hand side of the inequality, as we will see in the proof.

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \epsilon) \tag{1}$$

*Assume $s, t > 0$ be positive numbers to be chosen, and $S_n = \Sigma_{i=1}^{\infty}(\bar{X}_n - \mathbb{E}[\bar{X}_n])$, then using **Theorem 3.1** we get:*

$$\mathbb{P}(|S_n| \geq \epsilon) \leq \frac{\mathbb{E}(e^{s|Sn|})}{e^{st}} \tag{2}$$

*Since $X_1, X_2, .., X_n$ are independent variables, and using **Theorem 2.36** then*

$$\mathbb{P}(|S_n| \geq \epsilon) \leq \frac{\Pi_{i=1}^{n}\mathbb{E}(e^{|S_i|)^s})}{e^{st}} \tag{3}$$

Since $X_1, X_2, .., X_n$ are identically distributed variables, they take symmetric positive and negative values. So <u>without</u> the absolute value, (S^n) can be written like this:

$$e^S = \frac{1}{2}e^s + \frac{1}{2}e^{-s} \tag{4}$$

Using ***Taylor Series*** $e^{S_i} = 1 + s + \frac{s^2}{2!} + \frac{s^3}{3!} + .. = \Sigma_{i=0}^{\infty} \frac{s^i}{i!}$ *in (4)*

$$e^s = \frac{1}{2}(1 + \underline{S} + \frac{S^2}{2!} + \underline{\frac{S^3}{3!}} + ..) + \frac{1}{2}(1 - \underline{S} + \frac{S^2}{2!} - \underline{\frac{S^3}{3!}} + ..) \tag{5}$$

The underlined terms (negative and positive odd i) cancel each other, and we get:

$$e^s = \Sigma_{1=0}^{\infty} \frac{S^{2i}}{(2i)!} \tag{6}$$

The above is the case for ***Theorem 3.6****. For our case* ***Corollary 3.7****, we use the absolute value for all random variables, so there will be no negative values, so equation (4) with absolute values becomes twice as the value in equations (5,6) above, as follows:*

$$e^s = \frac{1}{2}(1 + S + \frac{S^2}{2!} + \frac{S^3}{3!} + ..) + \frac{1}{2}(1 + S + \frac{S^2}{2!} + \frac{S^3}{3!}) = \Sigma_{i=0}^{\infty} \frac{S^i}{i!} \tag{7}$$

$$\frac{\Pi_{i=1}^{n} \mathbb{E}(e^{|S_i|^s})}{e^{st}} \leq \frac{e^{s^2(b-a)^2 n/4}}{s^{st}} \tag{8}$$

Then we choose a value $s^* = \frac{2t}{n(b-a)^2}$ *to minimize the equation:*

$$h(s^*) = S^2 \frac{n(b-a)^2}{4} - st = -\frac{2t^2}{n(b-a)^2} \tag{9}$$

Assembling (8,9), we get:

$$\mathbb{P}(|Sn| \leq \epsilon) \leq 2e^{-\frac{2t^2}{n(b-a)^2}} \tag{10}$$

$$\mathbb{P}(|\bar{X_n} - \mathbb{E}[\bar{X_n}] \leq \epsilon) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}} \tag{11}$$

which proves ***Corollary 3.7****.* □

2 Proof of Lemma 3.14

Lemma 3.14 *The following properties hold*

1. *Let X be a sub-Gaussian RV with parameter λ , then αX is sub-Gaussian with parameter $|\alpha|\lambda$.*

2. *Let X be a sub-exponential RV with parameter λ , then αX is sub-exponential with parameter $|\alpha|\lambda$.*

3. *A sub-Gaussian RV X with parameter λ is sub-exponential with parameter λ.*

4. *A bounded RV X, i.e. $\mathbb{P}(X \in [a, b]) = 1$, then X is sub-Gaussian with parameter $(b-a)/2$. Specifically a Bernoulli RV is sub-Gaussian with parameter $1/2$.*

5. *If X is sub-Gaussian with parameter λ then $Z = X^2$ is sub-exponential with parameter 4λ.*

6. *If X, Y are independent and sub-Gaussian with parameter σ_1, σ_2, then $X + Y$ is sub-Gaussian with parameter $\sqrt{\sigma_1^2 + \sigma_2^2}$.*

Proof. According to Definition 3.10, X is sub-Gaussian with parameter λ if

$$\mathbb{E}[e^{s(X-\mathbb{E}(X)}] \leq e^{\frac{s^2\lambda^2}{2}} \text{ for all } s$$

Furthermore, X is sub-exponential with parameter λ if

$$\mathbb{E}[e^{s(X-\mathbb{E}(X)}] \leq e^{\frac{s^2\lambda^2}{2}} \text{ for all } |s| \leq \frac{1}{\lambda}$$

Property 1: Let $Y = \alpha X$.

$$\mathbb{E}[e^{s(Y-\mathbb{E}(Y)}] = \mathbb{E}[e^{s(\alpha X-\mathbb{E}(\alpha X)}] = \mathbb{E}[e^{\alpha s(X-\mathbb{E}(X)}]$$

Since X is sub-Gaussian with parameter λ, the following holds for all s and α:

$$\mathbb{E}[e^{\alpha s(X-\mathbb{E}(X)}] \leq e^{\frac{(\alpha s)^2\lambda^2}{2}} = e^{\frac{s^2(\alpha\lambda)^2}{2}}$$

Which shows that Y is sub-Gaussian with parameter $\sqrt{\alpha^2\lambda^2} = |\alpha|\lambda$.

Property 2: Let $Y = \alpha X$. If $\alpha = 0$, clearly Y is sub-exponential. So it remains to prove Property 2 for the case of $\alpha \neq 0$:

$$\mathbb{E}[e^{s(Y-\mathbb{E}(Y)}] = \mathbb{E}[e^{s(\alpha X-\mathbb{E}(\alpha X)}] = \mathbb{E}[e^{\alpha s(X-\mathbb{E}(X)}]$$

Since X is sub-exponential with parameter λ, the following holds for all s and α such that $|\alpha s| \leq \frac{1}{\lambda}$:

$$\mathbb{E}[e^{\alpha s(X-\mathbb{E}(X)}] \leq e^{\frac{(\alpha s)^2\lambda^2}{2}} = e^{\frac{s^2(\alpha\lambda)^2}{2}}$$

Furthermore,

$$|\alpha s| \leq \frac{1}{\lambda} \Rightarrow |\alpha||s| \leq \frac{1}{\lambda} \Rightarrow |s| \leq \frac{1}{|\alpha|\lambda} \text{ } (\alpha \neq 0 \text{ by assumption})$$

Which shows that Y is sub-exponential with parameter $\sqrt{\alpha^2\lambda^2} = |\alpha|\lambda$.

Property 3: Since X is sub-Gaussian, the following holds for all s:

$$\mathbb{E}[e^{s(X-\mathbb{E}(X)}] \leq e^{\frac{s^2\lambda^2}{2}}$$

In particular, it holds for all s such that $|s| \leq \frac{1}{\lambda}$. Which shows that X is sub-exponential.

Property 4: By Hoeffdings lemma (3.5) for bounded RV:s,

$$\mathbb{E}[e^{s(X-\mathbb{E}(X)}] \leq e^{\frac{s^2(b-a)^2}{8}} \text{ for all } s$$

We can re-write the right-hand side as

$$e^{\frac{s^2(b-a)^2}{8}} = e^{\frac{s^2(\frac{b-a}{2})^2}{2}}$$

Which shows that X is sub-Gaussian with parameter $(b-a)/2$. If X is a Bernoulli RV, X is bounded by a = 0 and b = 1, so the parameter value is $(1-0)/2 = 1/2$.

Property 5: (In lecture notes...)

Property 6: Let $Z = X + Y$.

$$\mathbb{E}[e^{s(Z-\mathbb{E}(Z))}] = \mathbb{E}[e^{s(X+Y-\mathbb{E}(X+Y))}] =$$

$$= \mathbb{E}[e^{s(X+Y-\mathbb{E}(X)-\mathbb{E}(Y))}] = \mathbb{E}[e^{s(X-\mathbb{E}(X))}e^{s(Y-\mathbb{E}(Y))}] \text{ } (*)$$

We introduce two new random variables, A and B. Let A = $e^{s(X-\mathbb{E}(X))}$ and B = $e^{s(Y-\mathbb{E}(Y))}$. Then we can write (*) as

$$\mathbb{E}[e^{s(Z-\mathbb{E}(Z))}] = \mathbb{E}(AB)$$

Now, since X and Y are independent, A and B are also independent and (by Theorem 2.36),

$$\mathbb{E}(AB) = \mathbb{E}(A)\mathbb{E}(B)$$

Since X and Y are sub-Gaussian with parameters σ_1 and σ_2, and $e^x > 0$ for all x,

$$0 < \mathbb{E}(A) \leq e^{\frac{s^2\sigma_1^2}{2}} \text{ for all } s$$

and

$$0 < \mathbb{E}(B) \leq e^{\frac{s^2\sigma_2^2}{2}} \text{ for all } s$$

Combining these results, we find that the following holds for all s:

$$\mathbb{E}[e^{s(Z-\mathbb{E}(Z))}] = \mathbb{E}(AB) = \mathbb{E}(A)\mathbb{E}(B) \leq e^{\frac{s^2\sigma_1^2}{2}} e^{\frac{s^2\sigma_2^2}{2}} = e^{\frac{s^2\lambda^2}{2}}$$

where $\lambda = \sqrt{\sigma_1^2 + \sigma_2^2}$.

This shows that $Z = X + Y$ is sub-Gaussian with parameter $\sqrt{\sigma_1^2 + \sigma_2^2}$.

□

3 Exercise 9

Exercise 9 *For the Poisson distribution, we have*

$$\mathbb{E}[e^{sX}] = e^{\lambda(e^s-1)}$$

is this sub-Gaussian, sub-exponential or neither?

The Poisson distribution is not sub-gaussian, and is sub-exponential.

4 Exercise 15

Exercise 15 *What is now the statistical model for the regression problem to find the below function?*

$$r(x) = \mathbb{E}[Y|X]$$

Proof. The statistical model to the problem is given by

$$\mathcal{E} = \{f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) : \int y^2 \hat{f}_{Y|X}(y|x)\,dy < \infty, \iint y^2 f_{X,Y}(x,y)\,dx\,dy < \infty\}$$

Note that the two constraints to the set arise from the assumption we made:

$$\mathbb{E}[Y^2] < \infty, \mathbb{E}[r^2(X)] < \infty$$

□

5 Proof of Theorem 4.3

Theorem 4.3 *For any decision function $g(x)$ taking values in $0, 1$, we have*

$$R(h^*) \leq R(g)$$

Proof.

$$R(g) = \mathbb{E}[L(Y, g(X))] = \mathbb{E}[\mathbb{E}[L(Y, g(X)) \mid X]]$$

Working on the inner part, by definition of E[L(y,g)] and the complement rule:

$$\mathbb{E}[L(Y, g(X)) \mid X = x] = 1 - \mathbb{E}\left[\mathbb{1}_{\{y=g(x)\}} \mid X = x\right]$$

by Properties of intersection and union of indicator function we have

$$= 1 - \mathbb{E}\left[\mathbb{1}_{\{1=g(x)\}}\mathbb{1}_{\{y=1\}} + \mathbb{1}_{\{0=g(x)\}}\mathbb{1}_{\{y=0\}} \mid X = x\right]$$

by Linearity of Expectation

$$= 1 - \mathbb{1}_{\{1=g(x)\}}\mathbb{E}\left[\mathbb{H}_{\{y=1\}} \mid X = x\right] - \mathbb{1}_{\{0=g(x)\}}\mathbb{E}\left[\mathbb{1}_{\{y=0\}} \mid X = x\right]$$

by definition of r(x)

$$= 1 - \mathbb{1}_{\{1=g(x)\}}r(x) - \mathbb{1}_{\{0=g(x)\}}(1 - r(x))$$

Now

$$\begin{aligned}
&\mathbb{E}[L(Y, g(X)) \mid X = x] - \mathbb{E}\left[L\left(Y, h^*(X)\right) \mid X = x\right] \\
&= -\mathbb{1}_{\{1=g(x)\}}r(x) - \mathbb{1}_{\{0=g(x)\}}(1 - r(x)) + \mathbb{1}_{\{1=h^*(x)\}}r(x) + \mathbb{1}_{\{0=h^*(x)\}}(1 - r(x))\big) \\
&= r(x)\left(\mathbb{1}_{\{1=h^*(x)\}} - \mathbb{1}_{\{1=g(x)\}}\right) + (1 - r(x))\left(\mathbb{1}_{\{0=h^*(x)\}} - \mathbb{1}_{\{0=g(x)\}}\right) \\
&= r(x)\left(\mathbb{1}_{\{1=h^*(x)\}} - \mathbb{1}_{\{1=g(x)\}}\right) - (1 - r(x))\left(\mathbb{1}_{\{1=h^*(x)\}} - \mathbb{1}_{\{1=g(x)\}}\right) \\
&= (2r(x) - 1)\left(\mathbb{1}_{\{1=h^*(x)\}} - \mathbb{1}_{\{1=g(x)\}}\right) \geq 0
\end{aligned}$$

□

6 Exercise 20

Exercise 20 *If you use the equality*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{P}(X > t)\, dt$$

can you improve upon the constant in Lemma 5.13?

Proof. Using the equality, we have $\mathbb{E}[|Y|^2] = \int_{-\infty}^{\infty} \mathbb{P}(|Y|^2 > t)\, dt \leq \int_{-\infty}^{\infty} \mathbb{P}(|Y|^2 \geq t)\, dt = \int_0^{\infty} \mathbb{P}(|Y| \geq \sqrt{t})\, dt$ (since $t > 0$).

By Formula (5.4), the above inequality becomes $\int_0^{\infty} \mathbb{P}(|Y| \geq \sqrt{t})\, dt < \int_0^{\infty} 2e^{-c_0 t}\, dt = \frac{-2e^{-c_0 t}}{c_0}|_0^{\infty} = \frac{2}{c_0} < \frac{5}{c_0}$. □

7 Exercise 22

Exercise 22 *Show that the relative entropy risk is the same risk as we saw in Section 4.2, it only differs by a constant.*

Proof. The relative entropy risk is given by $R(G) = \int log(\frac{f^*(x)}{g(x)}) f^*(x) dx$ by the law of large numbers this becomes $R(G) = \frac{1}{n} \sum log(\frac{f^*(x)}{g(x)}) = \frac{1}{n} \sum log(f^*(x)) - \frac{1}{n} \sum log(g(x)$

because $\frac{1}{n} \sum log(f^*(x))$ is a constant C, since we are looking for the risk with respect to g(x). Therefore we have

$$R(G) = C - \frac{1}{n} \sum log(g(x))$$

since we are minimizing R(G) the constant does not matter. Replace g(x) by p(X).

□