GROUP ASSIGNMENT 2

Lu Chen, Roman Denkin, William Hopper, Nicolás Martín
Introduction to Data Science
Uppsala Universitet

November 1, 2021

1 Prove Corollary 3.7.

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

$$\mathbb{P}(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq\ -\epsilon)\ \leq\ e^{-\frac{2n\epsilon^2}{(b-a)^2}} \qquad (1)$$

furthermore,

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

Proof. We begin taking into account Hoeffding's inequality Theorem:

> **Theorem.** *(Hoeffding's inequality). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a triple probability and let $X_1, \ldots, X_n \overset{IID}{\sim} F$ be $\mathbb{R}$-valued RVs such that $\mathbb{P}(X_i \in\ [a,b]) = 1$, then for any $\varepsilon\ >\ 0$ we get for $\overline{X}_n =\ \frac{1}{n}\sum_{i=1}^{n} X_i$*
>
> $$\mathbb{P}\left(\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right] \geq\ \epsilon\right)\ \leq\ e^{-\frac{2n\epsilon^2}{(b-a)^2}} \qquad (3)$$

3332Lets replace RV X with another RV $X^\star = b + a - X$, which will also be distributed on the same interval [a,b]. According to Hoeffding's inequality, it can be bounded in the same way:

$$\mathbb{P}\left(\overline{(b+a-X_n)} - \mathbb{E}\left[\overline{(b+a-X_n)}\right] \geq\ \epsilon\right)\ \leq\ e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

As b and a are constants, it can be moved out of the expectation or mean function:

$$\mathbb{P}\left(b+a-\overline{X}_n - b - a - \mathbb{E}\left[\overline{-X_n}\right] \geq\ \epsilon\right)\ \leq\ e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

Which is equivalent to:

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

If we multiply inequality in the left side of the above by -1, we will get the following equivalent inequality:

$$\mathbb{P}(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq\ -\epsilon)\ \leq\ e^{-\frac{2n\epsilon^2}{(b-a)^2}} \qquad (1)$$

And also, knowing that the right hand of the equation of (3):

$$\leq e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

is the same both in (3) as in (1), we can add up both sides of the equation between (1) and (3):

$$(\mathbb{P}(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq -\epsilon) + \mathbb{P}(\overline{X}_n - \mathbb{E}[\overline{X}_n] \geq \epsilon)) \leq (e^{-\frac{2n\epsilon^2}{(b-a)^2}} + e^{-\frac{2n\epsilon^2}{(b-a)^2}}) \qquad (4)$$

The left side of equation (4) can be re-written as having the absolute value of the probability being greater than or equal to ϵ. The right hand of the equation is the same as multiplying it by two. This gives us equation (2):

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

Which gives a proof for equation (2). $\square$

2 Prove 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. :

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

We start by the definition of a sub-Gaussian RV:

Definition 1. A $\mathbb{R}$ valued random variable X is said to be sub-Gaussian with parameter λ if

$$\mathbb{E}\left(e^{s(X-\mathbb{E}[X])}\right) \leq e^{\frac{s^2\lambda^2}{2}}, \quad \textit{for all } s. \qquad (5)$$

If we substitute αX to X in the left side and $|\alpha|\lambda$ to λ in the right side, we will get:

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

As α is a constant, we may move it out of expectation function:

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

Assuming that $\alpha \neq 0$ (in which case inequality above is satisfied as all exponents would equal 1), we can let $s = \frac{s^\star}{\alpha}$ and have our inequality rewritten in the following way:

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

which holds by the initial definition of X.

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

Definition 2. A $\mathbb{R}$ valued random variable X is said to be sub-exponential with parameter λ if

$$\mathbb{E}\left(e^{s(X-\mathbb{E}[X])}\right) \leq e^{\frac{s^2\lambda^2}{2}}, \quad \textit{for all } |s| \leq \frac{1}{\lambda}. \tag{6}$$

Proof of this part of the lemma is the same as in Part 1 with an addition that limitations on parameter s must be updated with new sub-exponential parameter $|\alpha|\lambda$:

$$|s| \leq \frac{1}{|\alpha|\lambda}.$$

2.3 A sub-Gaussian RV X with parameter λ is sub-Exponential with parameter λ

We can see from definitions (5) and (6) that sub-Exponential differ from sub-gaussian by the restriction on s. If the definition (5) holds for any S (which defines sub-gaussian RV), then it holds for $|s| \leq \frac{1}{\lambda}$ (which defines sub-exponential RV).

2.4 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 $1/2$

Lets recall the Hoeffding lemma.

(Hoeffding's Lemma). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a triple probability and suppose that X is $\mathbb{R}$-valued RV such that $\mathbb{P}(X \in [a,b]) = 1$, then for all $\lambda \in \mathbb{R}$

$$\mathbb{E}(e^{\lambda(X-\mathbb{E}(X))}) \leq exp(\frac{\lambda^2(b-a)^2}{8})$$

If we rename λ to s and introduce parameter $\lambda^\star = \frac{b-a}{2}$, we can rewrite the lemma as:

$$\mathbb{E}\left(e^{s(X-\mathbb{E}[X])}\right) \leq e^{\frac{s^2\lambda^{\star 2}}{2}}, \quad \textit{for all } s.$$

Which is the definition of sub-gaussian RV.

For Bernoulli RV, we know that X values are bounded in $[0,1]$, therefore if we substitute a and b with this values, we will get

$$\lambda^\star = \frac{b-a}{2} = \frac{1-0}{2} = \frac{1}{2}$$

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

Proof in Lecture Notes, right under Lemma 3.14

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

Lets work with the left side of sub-gaussian RV definition for $X + V$ RV:

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

As X and Y are independent, we can freely move it into separate expectation functions:

$$\begin{aligned}\mathbb{E}\left(e^{s((X+Y)-\mathbb{E}[X+Y])}\right) &= \mathbb{E}\left(e^{s(X-\mathbb{E}[X])+s(Y-\mathbb{E}[Y])}\right)\\ &= \mathbb{E}\left(e^{s((X)-\mathbb{E}[X])}\right)\mathbb{E}\left(e^{s((Y)-\mathbb{E}[Y])}\right)\end{aligned}$$

Using the inequalities provided by sub-gaussian RV definition for X and Y separately, we can write an inequality:

$$\mathbb{E}\left(e^{s((X)-\mathbb{E}[X])}\right)\mathbb{E}\left(e^{s((Y)-\mathbb{E}[Y])}\right) \leq e^{\frac{s^2\sigma_1^2}{2}} e^{\frac{s^2\sigma_2^2}{2}} = e^{\frac{s^2(\sigma_1^2+\sigma_2^2)}{2}}$$

Which is a definition of sub-gaussian RV with parameter $\sqrt{\sigma_1^2 + \sigma_2^2}$

□

3 Solve 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?

To find if a distribution is sub-Gaussian or sub-exponential we need to simplify

$$\mathbb{E}[e^{s(X-\mathbb{E}[X]})].$$

First, for the Poisson, $\mathbb{E}[X] = \lambda$. We can then split the exponent into two terms and remove the second term since it no longer depends on X:

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

Next, we can use the given moment generating function, before again simplifying the exponent:

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

Now that we have this expression, we need to see if / when

$$e^{\lambda(e^s-s-1)} \leq e^{\frac{\sigma^2 s^2}{2}}.$$

Since both sides will be positive, we can take the natural log of both sides while maintaining the inequality, giving:

$$\lambda(e^s - s - 1) \leq \frac{\sigma^2 s^2}{2} \to e^s - s - 1 \leq \frac{\sigma^2 s^2}{2\lambda}.$$

We can proof that this inequality does not hold for all s simply by the fact that exponential function on the left grows faster than quadratic function on the right side when $s \to \infty$.

For sub-exponential case, we need to show that inequality holds for $|s| \leq \frac{1}{\sigma}$. Lets check the extreme cases. For $s = 0$, we have $-1 \leq 0$, which holds. For $|s| = \frac{1}{\sigma}$, we will have $e^{\frac{1}{\sigma}} - \frac{1}{\sigma} - 1 \leq \frac{1}{2\lambda}$ which holds as we can select σ as we see fit.
We can check that this inequality holds on an interval $s \in (0, \frac{1}{\sigma})$ by calculating the derivative of the inequality function and checking that it does not equal to 0. The above confirms that the function is sub-exponential, but not sub-gaussian.

4 Solve Exercise 15

Exercise. *15 What is now the statistical model? [for regression functions]*

Since we are now dealing with regression, the statistical model will be an indexed family of regression functions. Since the exercise is asking for the generic regression context, we then get the generic statistical model:

$$\mathcal{R} = \{r(x) = \int y \cdot dF(y|x),\ F(y|x) \text{ is a conditional distribution function}\}$$

5 Explain Respective Equations in the Proof Process of Theorem 4.3

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

$$R(h^\star) \le R(G).$$

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

$$R(h^*) \le R(g).$$

Proof. Note that we can write

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

we will work only with the inner part, i.e. now

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

Now

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

This immediately implies the statement of the theorem. □

Figure 1: Original proof

The following properties, theorems and lemmas are all from Lecture notes.
As for the inner part:
The first equation is according to the first property of Lemma 2.7, which is referring to page 25;
The second one is according to the second property of Lemma 2.7, which is referring to page 25;
The third one is according to the second property of Theorem 2.36, which is referring to page 45;
The fourth one is according to the fifth property of Theorem 2.36, which is referring to page 45;
As for the next part:
The first equation is to plug the conclusion of the above inner part into the expression, and then let $g(x) = h^*(x)$ and then plug it in the expression, too;
The second equation is to extract the common factor r(x) and (1-r(x)) respectively;

The third equation is according to the first property of Lemma 2.7, which is referring to page 25;
The fourth equation is to extract the common factor $(\mathbb{1}_{1=h^*(x)} - \mathbb{1}_{1=g(x)})$;
The inequation is according to Definition 4.2, which is referring to page 80.

6 Solve Exercise 20

Exercise. ***20.*** *If you use the equality:*

$$E(X) = \int_0^\infty P(X > t)dt$$

(valid for non-negative RV's) can you improve upon the constant in Lemma 5.13?

(Lemma 5.13). Let Y be a RV satisfying the estimate for fixed $c_0 \geq 1$ and for all $\epsilon > 0$

$$P(|Y| \geq \epsilon) < 2e^{-c_0\epsilon^2}$$

Then

$$E(|Y|^2) \leq \frac{5}{c_0}$$

Proof. . Let $X = |Y|^2$ given $c_0 \geq 1$ fixed, then $E(X) = \int_0^\infty P(X > t)dt$ is

$$E(|Y|^2) = \int_0^\infty P(|Y|^2 > t)dt = \int_0^\infty P(|Y| > \sqrt{t})dt$$

Let ϵ given by $\sqrt{t} > 0$, that is, $\epsilon^2 = t$
Recall Lemma 5.13, we have

$$P(|Y| \geq \epsilon) < 2e^{-c_0\epsilon^2}$$

Due to $\epsilon^2 = t$, we have

$$P(|Y| \geq \sqrt{t}) < 2e^{-c_0 t}$$

Due to $dt/d\epsilon = 2\epsilon$, we have

$$E(|Y|^2) = \int_0^\infty P(|Y| > \sqrt{t})dt < \int_0^\infty 2e^{-c_0 t}dt$$

That is,

$$E(|Y|^2) < \int_0^\infty 2e^{-c_0 t}dt = \frac{2e^{-c_0 t}}{-c_0}|_0^\infty = \frac{2}{c_0}$$

$\square$

7 Solve 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. Relative entropy risk:*

$$R(G) = \int ln(\frac{f^\star(x)}{g(x)})f^\star(x)dx$$

$$R(\alpha) = -\int ln(p_\alpha(z))p_{\alpha^\star}(z)dz$$

Proof. Lets use the basic properties of the ln function: $ln(\frac{a}{b}) = ln(a) - ln(b)$

$$\begin{aligned} R(G) &= \int ln(\frac{f^\star(x)}{g(X)})f^\star(x)dx \\ &= \int (ln(f^\star(x)) - ln(g(x))f^\star(x)dx \\ &= \int ln(f^\star(x))f^\star(x)dx - \int ln(g(x))f^\star(x)dx = I - II \end{aligned}$$

Here 'I' does not depend of the function that we measure risk of ($G(x)$) and is therefore a constant. '$-II$' is the same as the risk described in section 4.2 with respect that $p_\alpha(z)$ is a special (parametric) case of general non-parametric g(x). □