Intro DS - Group Assignment 2

Albin Åberg Dahlberg, Emil Wormbs,
Folke Hilding, Oscar Boman, Sofia Alfsson

December 27, 2023

1 Prove Collorary 3.7

Similarly to the proof of Theorem 3.6, we do not assume anything about the sign of X and denote $S_n = \sum_{i=1}^{n} X_i$ and let $s, t > 0$, where $t = n\epsilon$ for some $\epsilon > 0$

$$\begin{aligned}
&P(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq -\epsilon) \\
&= P(S_n - \mathbb{E}\left[S_n\right] \leq -t) \\
&= P(-(S_n - \mathbb{E}[S_n]) \geq t) \\
&= P(e^{-s(S_n - \mathbb{E}[S_n])} \geq e^{st}),
\end{aligned}$$

which with Markov's inequality yields

$$\begin{aligned}
P(e^{-s(S_n - \mathbb{E}[S_n])} \geq e^{st}) &\leq e^{-st}\mathbb{E}[e^{-s(S_n - \mathbb{E}[S_n])}] \\
&= e^{-st}\prod_{i=1}^{n}\mathbb{E}[e^{-s(X_i - \mathbb{E}[X_i])}].
\end{aligned}$$

Following the proof of Theorem 3.6, using Hoeffdings lemma for $\lambda = s$ for each term in the product we get

$$\begin{aligned}
= e^{-st}\prod_{i=1}^{n}\mathbb{E}[e^{-s(X_i - \mathbb{E}[X_i])}] &\leq e^{-st}e^{(-s)^2(b-a)^2 n/8} \\
&= e^{-st}e^{s^2(b-a)^2 n/8}.
\end{aligned}$$

This is the exact same setting as in the proof of Theorem 3.6, as such it follows that

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

Furthermore, using Boole's inequality for the positive and negative deviations, we get

$$\begin{aligned}
&P(\left|\overline{X}_n - \mathbb{E}[\overline{X}_n]\right| \geq \epsilon) \\
&= P(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq -\epsilon) \cup P(\overline{X}_n - \mathbb{E}[\overline{X}_n] \geq \epsilon) \\
&\leq P(\overline{X}_n - \mathbb{E}[\overline{X}_n] \leq -\epsilon) + P(\overline{X}_n - \mathbb{E}[\overline{X}_n] \geq \epsilon) \\
&\leq e^{-\frac{2n\epsilon^2}{(b-a)^2}} + e^{-\frac{2n\epsilon^2}{(b-a)^2}} = 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}.
\end{aligned}$$

$\square$

2 Prove Lemma 3.15, properties 1-4.

2.1 Property 1

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

Answer: By definition 3.11, X is said to be sub-Gaussian with parameter λ if:

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

Inserting αX yields:

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

$(\alpha s)^2\lambda^2 = \alpha^2 s^2 \lambda^2 \longrightarrow |\alpha|^2 s^2 \lambda^2$. Therefore the parameter $\lambda_\alpha = |\alpha|\lambda$, proving the property.

2.2 Property 2

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

Answer: By definition 3.12, X is said to be sub-exponential with parameter λ if:

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

Inserting αX yields:

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

$(\alpha s)^2\lambda^2 = \alpha^2 s^2 \lambda^2 \longrightarrow |\alpha|^2 s^2 \lambda^2$. As the parameter $\lambda_\alpha = \alpha\lambda$ needs to fulfill the property of $|s| \leq \frac{1}{\lambda_\alpha}$, $\lambda_\alpha = |\alpha|\lambda$, proving the property.

2.3 Property 3

A sub-Gaussian RV X with parameter λ is sub-Exponential with parameter λ. **Answer:** We want to show that a sub-Gaussian random variable X with parameter λ is also sub-Exponential with parameter λ. The definition for a sub-Gaussian RV with parameter λ:

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

Since sub-Gaussian RV is defined for all values of $s \in \mathbb{R}$ it comes with heavy restrictions on the RV. On the other hand, the definitions for a sub-exponential RV with parameter λ is:

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

Since the sub-exponential RV is defined for a stricter interval of s it's restrictions for an RV is lighter in comparison with the sub-Gaussian RV, which means a sub-Gaussian RV $\in$ sub-exponential RV. This also means that the decay for a sub-exponential RV can be slower than for a sub-Gaussian.

2.4 Property 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* $1/2$

Answer: A bounded RV confined by Hoeffding's Lemma

$$\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \leq e^{\frac{s^2(b-a)^2}{8}}.$$

Using the definition of a sub-Gaussian RV,

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

we can put these definitions equal to each other, where λ is the parameter of the sub-Gaussian,

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

Solving for λ yields

$$\frac{\lambda^2}{2} = \frac{(b-a)^2}{8}$$
$$\lambda^2 = \frac{(b-a)^2}{4}$$
$$\lambda = \frac{b-a}{2}.$$

Thus, a bounded RV is sub-Gaussian. Furthermore, inserting the Bernoulli bounds $[0,1]$ yields the sub-Gaussian parameter

$$\lambda = 1/2$$

$\square$

3 Solve Exercise 3.16

For the Poisson distribution, we have

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

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

Answer: We can expand on the expression for the Poisson distribution. Since $\mathbb{E}[X]$ is a fixed number, i.e. non-random, we can multiply the expression for the

Poisson as

$$\begin{aligned}
&\mathbb{E}\left[e^{sX}\right] = e^{\lambda(e^s-1)} \\
\iff &\mathbb{E}\left[e^{sX}\right]e^{-s\mathbb{E}[X]} = e^{\lambda(e^s-1)}e^{-s\mathbb{E}[X]} \\
\iff &\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] = e^{\lambda(e^s-1)}e^{-s\mathbb{E}[X]} \\
&= e^{\lambda(e^s-1)-s\mathbb{E}[X]}.
\end{aligned}$$

Since $\mathbb{E}[X] = \lambda$ for a poisson distribution;

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

To determine if the distribution is sub-Gaussian, we need to determine if it fulfills the following condition, with a $\mathbb{R}$-valued random variable X and parameter σ.

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

To verify the inequality of the exponents from $e^{\lambda(e^s-1-s)}$ and $e^{\frac{s^2\sigma^2}{2}}$, we use L'Hopital's rule.

$$\frac{\lim\limits_{s\to\infty}\frac{s^2\sigma^2}{2}}{\lim\limits_{s\to\infty}\lambda(e^s-1-s)} = \frac{\lim\limits_{s\to\infty}\frac{d^2}{ds^2}\frac{s^2\sigma^2}{2}}{\lim\limits_{s\to\infty}\frac{d^2}{ds^2}\lambda(e^s-1-s)} = \lim_{s\to\infty}\frac{\sigma^2}{\lambda e^s} = 0 \tag{2}$$

for all $\lambda, \omega > 0$. Thus, $\lambda(e^s - 1 - s) > \frac{s^2\sigma^2}{2}$ violating the inequality. However, the sub-exponential bound is weaker, so we need to test that as well.

In order for the Poisson distribution to be sub-exponential, it needs to fulfill the following condition, with a $\mathbb{R}$-valued random variable X and parameter σ:

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

This yields an upper bound of;

$$e^{\lambda(e^s-1-s)} \leq e^{\frac{s^2\sigma^2}{2}} \leq e^{\frac{(\frac{1}{\sigma})^2\sigma^2}{2}} = e^{1/2} \tag{3}$$

To check if the bound holds, we need to investigate if there exists a solution to $\lambda(e^s - 1 - s) \leq 1/2$.

$$\lambda(e^s-1-s) \leq 1/2 \iff e^s - 1 - s \leq \frac{1}{2\lambda} \tag{4}$$

Expand e^s as the value of the Taylor series $e^s = 1 + s + \frac{s^2}{2!} + ...$;

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

$$\lim_{\lambda,s\to 0} e^s - 1 - s \leq \frac{1}{2\lambda} \tag{6}$$

as $\lambda, s \to 0$ we can see that the left hand side approaches 0, while the right hand side approaches ∞. Thus there exist a solution where $\lambda(e^s - 1 - s) \leq 1/2$, where thereby the poisson-distribution is sub-exponential.

4 Solve Exercise 4.7

A reasonable statistical model for the pattern recognition problem is the family of discrete distributions

$$\mathbb{F} = \left\{ F_{Y,X}(y,x) = F_{Y|X}(y,x) F_X(x), F_{Y|X} \text{ is discrete} \right\}.$$

5 Prove Theorem 4.9 with all details, basically referring to all the properties of the indicator function used, the monotonicity of measures etc

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

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

Proof: With the definition of risk and iterated expectation, i.e. the tower property, $R(g)$ can be expressed as as

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

Now we look at the inner part of the expectation. The inner expectation can be interpreted as the probability of an incorrect classification given an X. By the definition $0 - 1$ loss function, this expectation can be expressed with an indicator function, which in turn can be expressed in terms of its complement

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

We can decompose the indicator function. Note that y and $g(x)$ take values in $\{0,1\}$, so the two possible events are disjoint. Thus, we'll write the indicator function as the union of all possible combinations of correct classifications, that is, $y = g(x) = 1$ or $y = g(x) = 0$.

$$\begin{aligned}
&1 - \mathbb{E}\left[\mathbb{1}_{\{y=g(x)\}}|X = x\right] \\
&= 1 - \mathbb{E}\left[\mathbb{1}_{\{g(x)=1\}}\mathbb{1}_{\{y=1\}} + \mathbb{1}_{\{g(x)=0\}}\mathbb{1}_{\{y=0\}}|X = x\right] \\
&= 1 - \mathbb{E}\left[\mathbb{1}_{\{g(x)=1\}}\mathbb{1}_{\{y=1\}}|X = x\right] - \mathbb{E}\left[\mathbb{1}_{\{g(x)=0\}}\mathbb{1}_{\{y=0\}}|X = x\right].
\end{aligned}$$

Since these are conditional expectations, we can extract all factors dependent on x.

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

From the definition of the regression setting of a classification problem with two possible events, we get

$$\mathbb{E}\left[\mathbb{1}_{\{y=1\}}|X = x\right] = P(Y = 1|X = x) = \mathbb{E}\left[Y|X = x\right] = r(x).$$

Thus we can replace the expectations from the previous expression

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

As a checkpoint, we emphasize that we now see that

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

Now we want to see that Bayes classification rule h^* actually optimizes the pattern recognition problem

$$\begin{aligned} & R(h^*) \leq R(g) \\ \iff & R(g) - R(h^*) \geq 0. \end{aligned}$$

With $(\star)$ we can rewrite the left-hand expression to

$$\begin{aligned} & 1 - \mathbb{1}_{\{g(x)=1\}} r(x) - \mathbb{1}_{\{g(x)=0\}} (1 - r(x)) - \left(1 - \mathbb{1}_{\{h^*(x)=1\}} r(x) - \mathbb{1}_{\{h^*(x)=0\}} (1 - r(x))\right) \\ = & -\mathbb{1}_{\{g(x)=1\}} r(x) - \mathbb{1}_{\{g(x)=0\}} (1 - r(x)) + \mathbb{1}_{\{h^*(x)=1\}} r(x) + \mathbb{1}_{\{h^*(x)=0\}} (1 - r(x)) \\ = & r(x) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) - (1 - r(x)) \left(\mathbb{1}_{\{h^*(x)=0\}} - \mathbb{1}_{\{g(x)=0\}}\right). \end{aligned}$$

The complement of the indicator function of any decision function $f(x)$ can be written as

$$\mathbb{1}_{\{f(x)=0\}} = 1 - \mathbb{1}_{\{f(x)=1\}},$$

this gives us

$$\begin{aligned} & r(x) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) - (1 - r(x)) \left(1 - \mathbb{1}_{\{h^*(x)=1\}} - 1 + \mathbb{1}_{\{g(x)=1\}}\right) \\ = & r(x) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) - (1 - r(x)) \left(-\mathbb{1}_{\{h^*(x)=1\}} + \mathbb{1}_{\{g(x)=1\}}\right) \\ = & r(x) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) + (1 - r(x)) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \\ = & r(x) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) + (1 - r(x)) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \\ = & (2r(x) - 1) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \geq 0. \end{aligned}$$

We check this by inspecting what happens when $r(x) > \frac{1}{2}$ and $r(x) \leq \frac{1}{2}$. For $r(x) > \frac{1}{2} \to h^*(x) = 1$ we get

$$\begin{aligned} & (2r(x) - 1) > 0 \\ & \mathbb{1}_{\{h^*(x)=1\}} = 1 \\ & \mathbb{1}_{\{g(x)=1\}} \in \{0, 1\} \\ & \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \in \{0, 1\} \end{aligned}$$

which means that

$$(2r(x) - 1) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \geq 0$$

For $r(x) \leq \frac{1}{2} \to h^*(x) = 0$ yields

$$\begin{aligned} & (2r(x) - 1) \leq 0 \\ & \mathbb{1}_{\{h^*(x)=1\}} = 0 \\ & \mathbb{1}_{\{g(x)=1\}} \in \{0, 1\} \\ & \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}}\right) \in \{0, -1\} \end{aligned}$$

which means that

$$(2r(x) - 1) \left(\mathbb{1}_{\{h^*(x)=1\}} - \mathbb{1}_{\{g(x)=1\}} \right) \geq 0$$

As such, the theorem is proven. $\square$