Quantum Statistical Inference - Homework 2

Guillermo Abad and Guillermo García

3 February 2025

The exercises chosen are those marked with *O2*.

Exercise 1

Section 1

Trivial Extension of Markov's inequality:

To prove Markov's inequality, we will use the suggested trivial extension: For any random variable $X \geq 0$ and any $t > 0$, if $f(x)$ is a strictly increasing non-negative function show:

$$Pr\{X \geq t\} = Pr\{f(X) \geq f(t)\} \leq \frac{E(f(X))}{f(t)} \tag{1}$$

To show this we just need to realize that:

- (I): $f(x)$ being strictly increasing, implies that $X \geq t \iff f(X) \geq f(t)$.
- (II): $f(x)$ being strictly non-negative implies that any integral of it with a distribution will be positive.

with this, let's start from the expectation value of $f(X)$, where $f_X(x)$ will be the probability density function of X:

$$\begin{aligned} \mathbb{E}(f(X)) = \int_0^\infty f(x)f_X(x)dx = \int_{x\geq t} f(x)f_X(x)dx + \int_{x<t} f(x)f_X(x)dx \overset{(II)}{\geq} \int_{x\geq t} f(x)f_X(x)dx \\ \overset{(I)}{\geq} \int_{x\geq t} f(t)f_X(x)dx = f(t)\int_{x\geq t} f_X(x)dx = f(t)Pr\{x \geq t\} \iff \boxed{Pr\{x \geq t\} \leq \frac{\mathbb{E}(f(X))}{f(t)}} \end{aligned} \tag{2}$$

Original Markov's inequality:

Now we just need to choose $f(x) = x$ (which is strictly increasing and also strictly non-negative because $X \geq 0$), and we get the originally asked Markov's inequality:

$$\boxed{Pr\{x \geq t\} \leq \frac{\mathbb{E}(X)}{t}} \tag{3}$$

Random variable that saturates Markov's inequality:

And finally, we need a random variable that saturates this inequality. For that we will choose:

$$\boxed{X = \begin{cases} t, \text{ with probability } p \\ 0, \text{ with probability } 1-p \end{cases}} \iff \begin{cases} \mathbb{E}(X) = t \cdot p + 0 \cdot (1-p) = tp \\ Pr(X \geq t) = p \quad \text{(by definition)} \end{cases} \tag{4}$$

which then gives:

$$Pr(X \geq t) = p = \frac{tp}{t} = \frac{\mathbb{E}(X)}{t} \tag{5}$$

saturating the inequality, as we wanted.

Section 3

The weak law of large numbers:

To prove the weak law of large numbers, we will use Chevysheb's inequality from Section 2 (exercises *O1*), which states that, for a random variable Y and a quantity $\epsilon > 0$:

$$Pr\{|Y - \tilde{\mu}| > \epsilon\} \leq \frac{\tilde{\sigma}^2}{\epsilon^2} \tag{6}$$

where $\tilde{\mu} = E(Y)$ and $\tilde{\sigma}^2 = E((Y - \mu)^2)$ are the mean and variance of Y.

In this case, we have a sequence of i.i.d random variables $\{X_i\}_{i=1}^n$, each of them with mean μ and variance σ^2:

$$E(X_i) = \mu, \; E((X_i - \mu)^2) = \sigma^2 \; \forall i = 1, ..., n \tag{7}$$

If we take the same mean as random variable Y:

$$Y = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{8}$$

then its mean is:

$$\tilde{\mu} = E(Y) = E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n}{n}\mu = \mu \tag{9}$$

and its variance is:

$$\tilde{\sigma}^2 = E((Y - E(Y))^2) = E((\bar{X}_n - \mu)^2) = E(\bar{X}_n^2 - 2\mu\bar{X}_n + \mu^2) = E(\bar{X}_n^2) - 2\mu^2 + \mu^2 = E(\bar{X}_n^2) - \mu^2 \tag{10}$$

To find the expected value of the square of the sample mean, we will first find the expected value of the square of each random variable X_i:

$$E((X_i - \mu)^2) = E(X_i^2 - 2\mu X_i + \mu^2) = E(X_i^2) - 2\mu^2 + \mu^2 = E(X_i^2) - \mu^2 = \sigma^2 \Longrightarrow E(X_i^2) = \sigma^2 + \mu^2 \tag{11}$$

The square of the sample mean is:

$$\bar{X}_n^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} X_i^2 + \frac{1}{n^2}\sum_{i=1, j\neq i}^{n} X_i X_j \tag{12}$$

Then, its expected value is:

$$\begin{aligned} E(\bar{X}_n^2) &= \frac{1}{n^2}\sum_{i=1}^{n} E(X_i^2) + \frac{1}{n^2}\sum_{i=1, j\neq i}^{n} E(X_i X_j) = \frac{1}{n^2}\sum_{i=1}^{n}(\sigma^2 + \mu^2) + \frac{1}{n^2}\sum_{i=1, j\neq i}^{n} E(X_i X_j) = \\ &= \frac{\sigma^2 + \mu^2}{n} + \frac{1}{n^2}\sum_{i=1, j\neq i}^{n} E(X_i X_j) \overset{(*)}{=} \frac{\sigma^2 + \mu^2}{n} + \frac{1}{n^2}\sum_{i=1, j\neq i}^{n} \mu^2 = \frac{\sigma^2 + \mu^2}{n} + \frac{n-1}{n}\mu^2 \end{aligned} \tag{13}$$

where in the $(*)$ step we have used that each random variable X_i is independent of any other random variable X_j, so the expected value of its product is the product of its expected values:

$$(*) : \; E(X_i X_j) = E(X_j \cdot E(X_i | X_j)) = E(X_j \cdot E(X_i)) = E(X_i)E(X_j) = \mu^2 \tag{14}$$

So finally, gathering all results, we obtain the variance of $Y = \bar{X}_n$:

$$\tilde{\sigma}^2 = E((Y - E(Y))^2) = E(\bar{X}_n^2) - \mu^2 = \frac{\sigma^2 + \mu^2}{n} + \frac{n-1}{n}\mu^2 - \mu^2 = \frac{\sigma^2}{n} \tag{15}$$

Applying then Chebyshev's inequality for $Y = \bar{X}_n$, $\tilde{\mu} = E(\bar{X}_n) = \mu$ and $\tilde{\sigma}^2 = E((\bar{X}_n - \tilde{\mu})^2) = \sigma^2/n$, we obtain that:

$$Pr\{|Y - \tilde{\mu}| > \epsilon\} \leq \frac{\tilde{\sigma}^2}{\epsilon^2} \Longrightarrow \boxed{Pr\{\left|\bar{X}_n - \mu\right| > \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2}} \tag{16}$$

Then, <u>for $n \to \infty$, the probability that the same mean is greater than any given quantity $\epsilon > 0$ will tend to zero.</u>

Exercise 3

We play a gambling game where we must toss a fair coin $n >> 1$ times. As we use a fair coin, the priors for each outcome are $\eta_h = \eta_t = 1/2$.

The fortune is given by B_n after n tosses. Depending on what we obtain on the n-th toss, the fortune changes as:

$$B_n = \begin{cases} \text{heads:} \quad 2B_{n-1} \\ \text{tails:} \quad B_{n-1}/3 \end{cases} \tag{17}$$

If after n tosses, we have obtained k heads and $n-k$ tails, since the product of numbers commutes, our fortune is:

$$B_n(\text{k heads}) = B_0 2^k \left(\frac{1}{3}\right)^{n-k} = 2^k \left(\frac{1}{3}\right)^{n-k} \tag{18}$$

as the initial fortune is $B_0 = 1$.

As each toss has only two possible outcomes, the probability distribution of this game is given by a binomial distribution. The expected fortune is computed then after each possible case, from obtaining no heads to only heads:

$$\begin{aligned} \bar{B}_n = E(B_n) = \sum_{k=0}^{n} P(\text{k heads}) B_n(\text{k heads}) = \sum_{k=0}^{n} \binom{n}{k} p_h^k p_t^{n-k} B_n(\text{k heads}) = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^n B_n(\text{k heads}) = \\ = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^n 2^k \left(\frac{1}{3}\right)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{2}{2}\right)^k \left(\frac{1}{6}\right)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} 1^k \left(\frac{1}{6}\right)^{n-k} = \left(1 + \frac{1}{6}\right)^n = \left(\frac{7}{6}\right)^n \end{aligned} \tag{19}$$

The fortune rate is given by:

$$b = \lim_{n\to\infty} B_n^{1/n} \tag{20}$$

To find this quantity, we first need to use the strong law of large numbers. It says that, given an i.i.d sequence (similar to our n tosses), having each event expected value $E(X) = \mu$, then $\bar{X}_n$ converges almost surely to μ:

$$Pr\{\lim_{n\to\infty} \bar{X}_n = \mu\} = 1 \tag{21}$$

In this case, if we take the number of heads as a random variable K, the expected value to get a head in one toss is:

$$E(K) = \eta_h \cdot 1 + \eta_t \cdot 0 = \eta_h = \frac{1}{2} \tag{22}$$

as the domain of K is $\{0, 1\}$ for a single toss.

The sample mean is:

$$\bar{K}_n = \frac{1}{n} \sum_{i=1}^{n} K_i = \frac{k}{n} \tag{23}$$

where k is the number of heads after n tosses. Then, according to the strong law of large numbers, we can guarantee that:

$$\lim_{n\to\infty} \frac{k}{n} = \frac{1}{2} \tag{24}$$

which means that the number of heads will tend to be that of its prior. With this fact, we can express the fortune rate as a logarithm and, using the interchangeability of the limit going to infinity and the logarithm:

$$b = \lim_{n\to\infty} B_n^{1/n} \implies \log b = \log \lim_{n\to\infty} B_n^{1/n} = \lim_{n\to\infty} \log B_n^{1/n} = \lim_{n\to\infty} \frac{1}{n} \log B_n \tag{25}$$

Introducing our expression for B_n:

$$\begin{aligned} \log b = \lim_{n\to\infty} \frac{1}{n} \log\left(2^k (1/3)^{n-k}\right) = \lim_{n\to\infty} \frac{1}{n}(k \log(2) + (n-k) \log \frac{1}{3})) = \log 2 \lim_{n\to\infty} \frac{k}{n} + (\log \frac{1}{3})(1 - \lim_{n\to\infty} \frac{k}{n}) = \\ = \log(2)\frac{1}{2} + (\log \frac{1}{3})(1 - \frac{1}{2}) = \frac{1}{2} \log \frac{2}{3} = \log\left(\sqrt{\frac{2}{3}}\right) \implies b = \sqrt{2/3} \end{aligned} \tag{26}$$

Then:

$$\boxed{\bar{B}_n = \left(\frac{7}{6}\right)^n} \boxed{b = \sqrt{2/3}} \tag{27}$$

Thus, it is reasonable to accept to play this game as the expected fortune is higher than one for every number of tosses, so it is likely to finish with benefits over the initial fortune.

If we denote the gain factor $h = 2$ and the loss factor as $t = 1/3$, then:

$$\boxed{\bar{B}_n = \left(\frac{t+h}{2}\right)^n} \boxed{b = \sqrt{th}} \tag{28}$$

so the expected fortune is the arithmetic mean raised to the n-th power and the fortune rate is the geometric mean.

For general priors $p_t = p > 0$ and $p_h = 1 - p_t = 1 - p > 0$, gain and loss factors g and t, and an initial fortune B_0, we have that, after n tosses the fortune if k heads were obtained is:

$$B_n(\text{k heads}) = B_0 g^k t^{n-k} \tag{29}$$

The probability distribution is given again by the binomial distribution, so the expected fortune is:

$$\bar{B}_n = \sum_{k=0}^{n} \binom{n}{k} B_0 p_h^k p_t^{n-k} g^k t^{n-k} = B_0 \sum_{k=0}^{n} \binom{n}{k} (p_h g)^k (p_t t)^{n-k} = B_0 (p_h g + p_t t)^n = B_0 \big(pg + (1-p)t\big)^n \tag{30}$$

As for the case with equal priors, the strong law of large numbers can be used to derive that the rate of heads after n tosses coincides with the probability to obtain a head in a single toss:

$$E(K) = p_h \Longrightarrow \lim_{n\to\infty} \frac{k}{n} = p_h = p \tag{31}$$

so the fortune rate is computed again using logarithms:

$$\begin{aligned}\log b = \lim_{n\to\infty} \log B_n^{1/n} = \lim_{n\to\infty} \log \left(B_0 g^k t^{n-k}\right)^{1/n} = \log B_0 + \log g \lim_{n\to\infty} \frac{k}{n} + (\log t)\Big(1 - \lim_{n\to\infty} \frac{k}{n}\Big) = \\ = \log B_0 + p_h \log g + (1-p_h) \log t = \log B_0 + \log g^{p_h} + \log t^{p_t} = \log B_0 g^{p_h} t^{p_t} \Longrightarrow b = B_0 g^{p_h} t^{p_t} = B_0 g^p t^{1-p}\end{aligned} \tag{32}$$

To summarize, in the general case, the expected fortune and the fortune rate are:

$$\boxed{\bar{B}_n = B_0 \big(pg + (1-p)t\big)^n} \boxed{b = B_0 g^p t^{1-p}} \tag{33}$$

Finally, we will consider the case in which $p_h = p > 1/2$ and $h = 2$, $t = 0$. Then, if we get tails, we go bankrupt. In this case, the expected fortune after n tosses and the fortune rate are:

$$\bar{B}_n = B_0 \big(p2 + (1-p)0\big)^n = (2p)^n B_0, \; b = B_0 \cdot 2^p \cdot 0^{1-p} = 0 \tag{34}$$

for $p \neq 1$. If we had $p = 1$, we would always get heads, so our fortune would always double after each toss and we would never go bankrupt.

However, the case $p \neq 1$ has more interest as, once we get a tail, our fortune will be zero the rest of the game, so effectively the game stops. In this case, the game consists on getting always heads and keep doubling our fortune until we get a tail, when the game ends. Then, after $n-1$ tosses, if we are still playing, we will have got heads with a probability of p in each toss, and tail with probability $1-p$ in the n-th toss. Then, our probability to go bankrupt at the n-th toss is:

$$\boxed{Pr(B_n = 0) = p^{n-1}(1-p)} \tag{35}$$

We may think that, as $p > 1/2$, we can keep playing confidently as heads are more likely to appear than tails. However, we can compute the probability of loosing in any toss until the n-th toss and see if that is true:

$$P(\text{lose until n-th toss}) = \sum_{j=1}^{n} Pr(B_j = 0) = \sum_{j=1}^{n} p^{j-1}(1-p) = (1-p)\sum_{j=0}^{n-1} p^j = (1-p)\frac{1-p^n}{1-p} = 1 - p^n \tag{36}$$

Therefore, at the beginning the odds are on our side for $p > 1/2$, but for any $p \in [0, 1)$, the probability of losing is almost certain in the limit $n \to \infty$.

Exercise 4

We extract $n >> 1$ balls from an urn where there are 2 red balls, 8 blue balls and 90 white balls, and the extracted ball is replaced after each extraction, so there are always $2+8+90 = 100$ balls. Then, the probability distribution of extracting a ball of color $x \in \mathcal{X} = \{r, b, w\}$ is:

$$Q(r) = \frac{2}{100} = 0.02, \quad Q(b) = \frac{8}{100} = 0.08, \quad Q(w) = \frac{90}{100} = 0.9 \tag{37}$$

Extracting each kind of ball entails a reward, $R(x)$, which is, for each color:

$$R(r) = 100, \quad R(b) = 20, \quad R(w) = 0 \tag{38}$$

Given this scheme, we want to find the probability such that the mean reward after n extractions, given by: $\bar{R}_n = \frac{1}{n}\sum_{k=1}^{n} R_k$ is greater than or equal to 5. This imposes a constraint to our problem, which can be expressed as:

$$C = \left\{\bar{R}_n = \frac{1}{n}\sum_{k=1}^{n} R(x_k) \geq 5\right\} = \left\{\frac{1}{n}\sum_{k=1}^{n} g(x_k) \geq b\right\} \text{ with } \begin{cases} g(x_k) = R(x_k) \\ b = 5 \end{cases} \tag{39}$$

where x_k represents the color obtained in the k-th extraction.

We are interested in obtaining the probability of getting a sequence of extractions that fulfills the defined constraint, so we can use Sanov's theorem to find the probability distribution that is nearest to the original probability distribution $Q(x)$ that fulfills the requirement. The corresponding set of probability distributions following the constraint is:

$$E = \left\{P : \sum_{a} P(a)g(a) \geq b\right\} = \left\{P : \sum_{x\in\mathcal{X}} P(x)R(x) \geq 5\right\} \tag{40}$$

We can solve this optimization problem with Lagrange multipliers. The Lagrangian of the problem is:

$$\mathcal{L}(P) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right) + \lambda\left(\sum_{x\in\mathcal{X}} P(x)R(x) - 5\right) + v\left(\sum_{x\in\mathcal{X}} P(x) - 1\right) \tag{41}$$

which, after minimization, yields the probability distribution:

$$P^*(x) = \frac{Q(x)e^{\lambda R(x)}}{\sum_{x\in\mathcal{X}} Q(x)e^{\lambda R(x)}} \equiv \frac{Q(x)e^{\lambda R(x)}}{\mathbb{D}} \tag{42}$$

where the parameter λ is chosen to satisfy the constraint, and with the denominator ($\mathbb{D}$) being:

$$\mathbb{D} = \sum_{x\in\mathcal{X}} Q(x)e^{\lambda R(x)} = Q(r)e^{\lambda R(r)} + Q(b)e^{\lambda R(b)} + Q(w)e^{\lambda R(w)} = 0.02e^{100\lambda} + 0.08e^{20\lambda} + 0.9 \tag{43}$$

then the constraint, with this probability distribution, becomes:

$$\sum_{x\in\mathcal{X}} P^*(x)R(x) = \sum_{x\in\mathcal{X}} \frac{Q(x)e^{\lambda R(x)}}{\mathbb{D}}R(x) \;\geq\; 5 \tag{44}$$

which can be rearranged into an inequality:

$$\begin{aligned} \frac{Q(r)e^{\lambda R(r)}}{\mathbb{D}}R(r) + \frac{Q(b)e^{\lambda R(b)}}{\mathbb{D}}R(b) + \frac{Q(w)e^{\lambda R(w)}}{\mathbb{D}}R(w) = \frac{0.02\cdot 100e^{100\lambda} + 0.08\cdot 20e^{20\lambda}}{\mathbb{D}} \quad &\geq\; 5 \\ \iff 2e^{100\lambda} + 1.6e^{20\lambda} \;\geq\; 5\mathbb{D} = 5(0.02e^{100\lambda} + 0.08e^{20\lambda} + 0.9) & \\ \iff 1.9e^{100\lambda} + 1.2e^{20\lambda} - 4.5 \;\geq\; 0 \quad \iff \quad \left(e^{20\lambda}\right)^5 + \frac{12}{19}e^{20\lambda} - \frac{45}{19} \;\geq\; 0 & \end{aligned} \tag{45}$$

And since this is non-decreasing, as its derivative is always positive:

$$f(u \equiv e^{20\lambda}) = u^5 + \frac{12}{19}u - \frac{45}{19} \quad \iff \quad f'(u) = 5u^4 + \frac{12}{19} > 0 \;\forall u \in \mathbb{R} \tag{46}$$

it crosses $f(u) = 0$ only once, for the approximate value $e^{20\lambda} \approx 1.10784$, and so our inequality becomes:

$$e^{20\lambda} \geq 1.10784 \iff 20\lambda \geq 0.10241 \iff \lambda \geq 5.1206 \cdot 10^{-3} \tag{47}$$

So if we choose $\lambda = 5.1206 \cdot 10^{-3}$, we get that the denominator is:

$$\mathbb{D} = 0.02e^{100 \cdot 5.1206 \cdot 10^{-3}} + 0.08e^{20 \cdot 5.1206 \cdot 10^{-3}} + 0.9 = 1.022 \tag{48}$$

and a probability distribution,

$$\begin{aligned} P^*(r) &= \frac{Q(r)}{D} e^{\lambda R(r)} = \frac{0.02}{1.022} e^{100 \cdot 5.1206 \cdot 10^{-3}} = 0.03265 \\ P^*(b) &= \frac{Q(b)}{D} e^{\lambda R(b)} = \frac{0.08}{1.022} e^{20 \cdot 5.1206 \cdot 10^{-3}} = 0.08672 \\ P^*(w) &= \frac{Q(w)}{D} e^{\lambda R(w)} = \frac{0.9}{1.022} = 0.88062 \end{aligned} \tag{49}$$

This probability distribution guarantees that the constraint will be fulfilled, so the asymptotic probability of getting an average reward greater than 5 is given by:

$$Q^n(E) = Pr(\bar{R}_n \geq 5) = e^{-nD(P^*||Q)} \tag{50}$$

We compute the Kullback-Leiber divergence $D(P^*||Q)$:

$$\begin{aligned} D(P^*||Q) &= \sum_{x \in \mathcal{X}} P^*(x) \log\left(\frac{P^*(x)}{Q(x)}\right) = P^*(r) \log\left(\frac{P^*(r)}{Q(r)}\right) + P^*(b) \log\left(\frac{P^*(b)}{Q(b)}\right) + P^*(w) \log\left(\frac{P^*(w)}{Q(w)}\right) = \\ &= 0.03265 \cdot \log\left(\frac{0.03265}{0.02}\right) + 0.08672 \cdot \log\left(\frac{0.08672}{0.08}\right) + 0.88062 \cdot \log\left(\frac{0.88062}{0.9}\right) \approx 0.00383 \end{aligned} \tag{51}$$

and so:

$$\boxed{Pr(\bar{R}_n \geq 5) \approx e^{-0.00383n}} \tag{52}$$

Which for $n = 10^5$ extractions:

$$Pr(\bar{R}_{n=10^5} \geq 5) = e^{-0.00383 \cdot 10^5} = e^{-383} \approx e^{-\alpha} \implies \boxed{\alpha \approx 383} \tag{53}$$

is nearly zero, meaning that <u>it is effectively impossible to have an average reward greater than 5, for 10^5 extractions.</u>

<u>**Check against normal distribution:**</u> This is consistent with the fact that, due to the central limit theorem and the strong law of large numbers, the sample mean of the reward will asymptotically approach the mean μ:

$$\lim_{n \to \infty} \bar{R}_n = \mu = E[X] = \sum_{x_k} Q(x_k) R(x_k) = 0.02 \cdot 100 + 0.08 \cdot 20 + 0.9 \cdot 0 = 2 + 1.6 = 3.6 \tag{54}$$

with a single shoot variance:

$$\sigma^2 = E[X^2] - E[X]^2 = \sum_{x_k} Q(x_k) R(x_k)^2 - \mu^2 = 200 + 32 - 12.96 \approx 219 \implies \sigma = \sqrt{219} \approx 14.8 \tag{55}$$

which for the average of our $n = 10^5$ extractions, gives a standard deviation (σ_n):

$$\begin{aligned} &\sigma_n^2 = \text{Var}(\overline{X}_n) = \text{Var}(\tfrac{1}{n}(X_1 + ... + X_n)) = \frac{1}{n^2} \text{Var}(X_1 + ... + X_n) = \frac{\sigma^2}{n} \iff \sigma_n = \frac{\sigma}{\sqrt{n}} = \frac{14.8}{\sqrt{10^5}} \approx 0.0468 \\ &N_\sigma = \frac{\bar{X} - \mu}{\sigma_n} = \frac{5 - 3.6}{0.0468} \approx 30 \implies Pr(\bar{R}_{10^5} \geq 5) \approx Pr(X \geq \mu + 30\sigma) = \frac{1}{2} erfc\left(\frac{30}{\sqrt{2}}\right) \boxed{\propto e^{-450}} \end{aligned} \tag{56}$$

which means that <u>getting an average of $\hat{X} = 5$ with $n = 10^5$ extractions, would require a chance of $N_\sigma \approx 30$ sigmas,</u> which is basically impossible, agreeing to some extension with the previously obtained result of $Pr(\bar{R}_{10^5} \geq 5) \propto e^{-383}$.

Exercise 5

A function $f : I \to \mathbb{R}$ is *operator convex* if, for any Hermitian operators A and B whose spectral decompositions are included in $I \subset \mathbb{R}$ and for a parameter $\lambda \in [0,1]$, it holds that:

$$\underbrace{f(\lambda A + (1-\lambda)B)}_{\equiv f(\lambda_j X_j)} \ \leq \ \underbrace{\lambda f(A) + (1-\lambda) f(B)}_{\equiv \lambda_i f(X_i)} \tag{57}$$

where we have defined $\lambda_i = \{\lambda, 1-\lambda\}$ and $X_i = \{A, B\}$, and skipped two sums over i and j using Einstein notation.

This is equivalent to say that the following matrix M, is positive semidefinite:

$$M = \lambda_i f(X_i) - f(\lambda_j X_j) \ \Big(= \lambda f(A) + (1-\lambda) f(B) - f\big(\lambda A + (1-\lambda)B\big)\Big) \ \geq \ 0 \tag{58}$$

So, we will demonstrate that the function $f(t) = t^3$ is not operator convex with a counter example, using matrices:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \tag{59}$$

and $\lambda = 1/2$, which gives a final M, to show as not being positive semidefinite:

$$M = \frac{1}{2} f(A) + \frac{1}{2} f(B) - f\left(\frac{1}{2}A + \frac{1}{2}B\right) = \underbrace{\frac{A^3 + B^3}{2}}_{=\lambda_i f(X_i)} - \underbrace{\left(\frac{A+B}{2}\right)^3}_{=f(\lambda_j X_j)} \tag{60}$$

We start by computing the value of the function for both matrices:

$$\begin{aligned} f(A) &= A^3 = A^2 A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^2 A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} A = 2AA = 2A^2 = 2 \cdot 2A = 4A = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix} \\ f(B) &= B^3 = B^2 B = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}^2 B = \begin{pmatrix} 10 & 4 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 34 & 14 \\ 14 & 6 \end{pmatrix} \end{aligned} \tag{61}$$

Then, the combination of the functions is:

$$\lambda_i f(X_i) = \frac{A^3 + B^3}{2} = \frac{1}{2} \begin{pmatrix} 38 & 18 \\ 18 & 10 \end{pmatrix} = \begin{pmatrix} 19 & 9 \\ 9 & 5 \end{pmatrix} \tag{62}$$

On the other hand, the function of the combination of matrices yields:

$$f(\lambda_j X_j) = \left(\frac{A+B}{2}\right)^3 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^3 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 13 & 8 \\ 8 & 5 \end{pmatrix} \tag{63}$$

Then, the matrix M is:

$$M = \lambda_i f(X_i) - f\big(\lambda_j X_j\big) = \begin{pmatrix} 19 & 9 \\ 9 & 5 \end{pmatrix} - \begin{pmatrix} 13 & 8 \\ 8 & 5 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 0 \end{pmatrix} = 3\mathbb{1} + 3\mathbb{Z} + 1\mathbb{X} \tag{64}$$

Which has eigenvalues $\lambda_{\pm} = 3 \pm \sqrt{3^2 + 1^2}$, and concretely has a negative eigenvalue $\lambda_- = 3 - \sqrt{10} < 0$. So the matrix $\underline{M \text{ is not positive semidefinite}}$, and:

$$\boxed{\lambda f(A) + (1-\lambda) f(B) \ \ngeq \ f(\lambda A + (1-\lambda)B)} \tag{65}$$

meaning that $\underline{f(t) = t^3 \text{ is not operator convex}}$, as we wanted to show.

Exercise 6

We need to show that if a channel fulfills the covariance property, then its Choi matrix fulfills:

$$\Lambda(U\rho U^\dagger) = V\Lambda(\rho)V^\dagger \implies (V\otimes U^*)J_\Lambda(V^\dagger\otimes U^T) = J_\Lambda \tag{66}$$

such that $UU^\dagger = U^\dagger U = \mathbb{1}_A$ and $VV^\dagger = V^\dagger V = \mathbb{1}_B$, with A as the input system and B as the output system.

To show this, we will use the vectorization of matrices as suggested, so we can express the Choi matrix J_Λ as:

$$J_\Lambda = \sum_{k=1}^{|A|} |K_k\rangle\langle K_k| = \sum_{k=1}^{|A|}(K_k\otimes\mathbb{1}_{A'})\left|\mathbb{1}\right\rangle_{AA'}\left\langle\mathbb{1}\right|_{AA'}(K_k^\dagger\otimes\mathbb{1}_{A'}) \tag{67}$$

where $\{K_k\}$ are the Kraus operators associated to the quantum channel Λ, that can be decomposed as $\Lambda(\rho) = \sum_{k=1}^{|A|} K_k\rho K_k^\dagger$, and the vector $\left|\mathbb{1}\right\rangle_{AA'}$ is the vectorization of $\mathbb{1}_A$ and represents the maximally entangled state between two systems A and A', such that $|A| = |A'|$:

$$\left|\mathbb{1}\right\rangle_{AA'} := \left|\mathbb{1}_A\right\rangle = \sum_{i=1}^{|A|}|i\rangle_A\otimes|i\rangle_{A'} = \sqrt{|A|}\left|\phi^+\right\rangle_{AA'} \tag{68}$$

Then, the ketbra corresponding to the vector $\left|\mathbb{1}\right\rangle_{AA'}$ is:

$$\left|\mathbb{1}\middle\rangle\middle\langle\mathbb{1}\right|_{AA'} = |A|\left|\phi^+\middle\rangle\middle\langle\phi^+\right|_{AA'} = |A|\sum_{i,j=1}^{|A|}|i\rangle\langle j|_A\otimes|i\rangle\langle j|_{A'} \tag{69}$$

Regarding the channel, its covariance property can be rewritten in terms of the Kraus operators as:

$$\Lambda(U\rho U^\dagger) = V\Lambda(\rho)V^\dagger \iff \sum_{k=0}^{|A|}K_kU\rho U^\dagger K_k^\dagger = \sum_{k=0}^{|A|}VK_k\rho K_k^\dagger V^\dagger \tag{70}$$

With these definitions and properties, we can finally compute the quantity $(V\otimes U^*)J_\Lambda(V^\dagger\otimes U^T)$:

$$\begin{aligned}(V\otimes U^*)J_\Lambda(V^\dagger\otimes U^T) &= (V\otimes U^*)\sum_{k=1}^{|A|}(K_k\otimes\mathbb{1}_{A'})\left|\mathbb{1}\right\rangle_{AA'}\left\langle\mathbb{1}\right|_{AA'}(K_k^\dagger\otimes\mathbb{1}_{A'})(V^\dagger\otimes U^T) = \\ &= \sum_{k=1}^{|A|}(VK_k\otimes U^*)\left(|A|\sum_{i,j=1}^{|A|}|i\rangle\langle j|\otimes|i\rangle\langle j|\right)(K_k^\dagger V^\dagger\otimes U^T) = |A|\sum_{i,j,k}(VK_k|i\rangle\langle j|K_k^\dagger V^\dagger)\otimes(U^*|i\rangle\langle j|U^T)\end{aligned} \tag{71}$$

Now, using equation 70:

$$\begin{aligned}|A|\sum_{i,j}\left(\sum_k VK_k|i\rangle\langle j|K_k^\dagger V^\dagger\right)\otimes(U^*|i\rangle\langle j|U^T) &= |A|\sum_{i,j}\left(\sum_k K_kU|i\rangle\langle j|U^\dagger K_k^\dagger\right)\otimes(U^*|i\rangle\langle j|U^T) = \\ = \sum_k(K_kU\otimes U^*)\left(|A|\sum_{i,j}|i\rangle\langle j|\otimes|i\rangle\langle j|\right)(U^\dagger K_k\otimes U^T) &= \sum_k(K_k\otimes\mathbb{1}_{A'})(U\otimes U^*)\left|\mathbb{1}\middle\rangle\middle\langle\mathbb{1}\right|(U^\dagger\otimes U^T)(K_k^\dagger\otimes\mathbb{1}_{A'})\end{aligned} \tag{72}$$

At this point, we need to use equation 2.2.111 from the Notes, which states that, for matrices A and B and a vectorized matrix $|C\rangle$, $(A\otimes B)|C\rangle = \left|ACB^T\right\rangle$. We find the bra of this vector with this property:

$$(A\otimes B)|C\rangle = \left|ACB^T\right\rangle \iff \langle C|(A^\dagger\otimes B^\dagger) = \left\langle B^*C^\dagger A^\dagger\right| \tag{73}$$

which can also be written as $\langle C|(A\otimes B) = \left\langle B^TC^\dagger A\right|$. Using these two properties and U unitarity, we can write:

$$(U\otimes U^*)\left|\mathbb{1}\middle\rangle\middle\langle\mathbb{1}\right|(U^\dagger\otimes U^T) = \left|U\mathbb{1}_AU^\dagger\middle\rangle\middle\langle U\mathbb{1}_AU^\dagger\right| = \left|\mathbb{1}_A\middle\rangle\middle\langle\mathbb{1}_A\right| = \left|\mathbb{1}\middle\rangle\middle\langle\mathbb{1}\right| \tag{74}$$

Introducing this result in our computation of $(V\otimes U^*)J_\Lambda(V^\dagger\otimes U^T)$, we arrive to our desired result:

$$(V\otimes U^*)J_\Lambda(V^\dagger\otimes U^T) = \sum_k(K_k\otimes\mathbb{1}_{A'})\left|\mathbb{1}\middle\rangle\middle\langle\mathbb{1}\right|(K_k^\dagger\otimes\mathbb{1}_{A'}) = \sum_k|K_k\rangle\langle K_k| = J_\Lambda \tag{75}$$

Exercise 7

The phase-flip error channel is:

$$\Lambda_\lambda(\rho) = (1-\lambda)\rho + \lambda\sigma_z\rho\sigma_z \text{ with } 0 \leq \lambda \leq 1 \tag{76}$$

Section 1

For

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{s}\cdot\vec{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+s_z & s_x - is_y \\ s_x + is_y & 1-s_z \end{pmatrix} \quad \text{where} \quad \begin{cases} \vec{s} = (s_x,\ s_y,\ s_z) \\ \vec{\sigma} = (\sigma_x,\ \sigma_y,\ \sigma_z) \end{cases} \tag{77}$$

the channel acts as:

$$\begin{aligned}
\Lambda_\lambda(\rho) &= (1-\lambda)\rho + \lambda\sigma_z\rho\sigma_z = (1-\lambda)\left[\frac{1}{2}(\mathbb{1} + \vec{s}\cdot\vec{\sigma})\right] + \lambda\sigma_z\left[\frac{1}{2}(\mathbb{1} + \vec{s}\cdot\vec{\sigma})\right]\sigma_z = \\
&= \frac{1}{2}\Big(\mathbb{1} + \vec{s}\cdot\Big((1-\lambda)\vec{\sigma} + \lambda(\sigma_z\vec{\sigma}\sigma_z)\Big)\Big) = \frac{1}{2}\Big(\mathbb{1} + \vec{s}\cdot\Big((1-\lambda)\vec{\sigma} + \lambda(-\sigma_x, -\sigma_y, \sigma_z)\Big)\Big) \\
&= \frac{1}{2}\left(\mathbb{1} + \vec{s}\cdot\begin{pmatrix}(1-2\lambda)\sigma_x \\ (1-2\lambda)\sigma_y \\ \sigma_z\end{pmatrix}\right) = \frac{1}{2}\left(\mathbb{1} + \begin{pmatrix}(1-2\lambda)s_x \\ (1-2\lambda)s_y \\ s_z\end{pmatrix}\cdot\vec{\sigma}\right) = \frac{1}{2}\left(\mathbb{1} + \vec{s'}\cdot\vec{\sigma}\right) = \rho'
\end{aligned} \tag{78}$$

meaning that the net effect of the channel is changing $\vec{s}$, by:

$$\boxed{\vec{s'} = \Big((1-2\lambda)s_x,\ (1-2\lambda)s_y,\ s_z\Big)} \tag{79}$$

which is a convex combination of the identity and a rotation of π over the Z axis in the Bloch sphere, meaning that our state "flips" phases, losing information in the x-y plane, squeezing ρ around the Z-axis, while maintaining all the information of its "height" in the Z axis. Graphically:

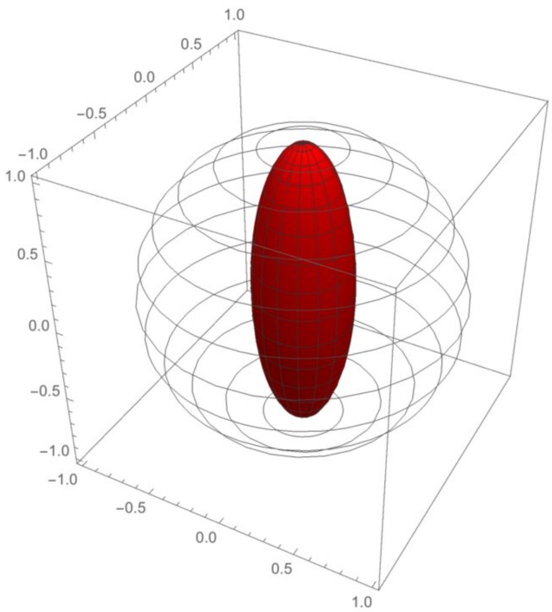

Figure 1: Visualization of the phase flip channel acting on the Bloch representation of a state ρ, by Jagadish, Vinayak & Petruccione, Francesco. (2018). An Invitation to Quantum Channels. DOI: 10.12743/quanta.v7i1.77.

And as one would suspect, from looking at Eq. (76) (where the λ part changes our state, while the $(1-\lambda)$ part leaves it unchanged), the result of Eq. (79) shows that the bigger the lambda is, the bigger the "squeeze" will be. And also since in Eq. (76) no x or y appears, and there are only changes with z's, we would expect such symmetry to also appear in the final net effect, which is clearly what we get in Eq. (79), and what Fig. 1 shows.

Section 2

We have already seen that $\sigma_z \rho_{\vec{s}} \sigma_z = \rho_{(-s_x,-s_y,s_z)}$, the generalization of this is:

$$\sigma_i \rho_{\vec{s}} \sigma_i = \rho_{\vec{s}\odot\left[\sum_j (-1)^{\delta_{ij}+1}\vec{e}_j\right]} = \begin{cases} i=x: & \rho_{\vec{s}\odot(+1,-1,-1)} = \rho_{(s_x,-s_y,-s_z)} \\ i=y: & \rho_{\vec{s}\odot(-1,+1,-1)} = \rho_{(-s_x,s_y,-s_z)} \\ i=z: & \rho_{\vec{s}\odot(-1,-1,+1)} = \rho_{(-s_x,-s_y,s_z)} \end{cases} \equiv \rho_{\vec{s}\odot[\vec{\pm}^i]} \tag{80}$$

where $\odot$ represents an elementwise product between the vectors, that give another vector: $\vec{a}\odot\vec{b} = \sum_j a_j b_j \vec{e}_j$.

Then we also have seen that our channel performs another elementwise product:

$$\Lambda_\lambda(\rho_{\vec{s}}) = \rho_{\vec{s}\odot(1-2\lambda,1-2\lambda,1)} \quad \equiv \quad \rho_{\vec{s}\odot[\lambda^{\vec{x}y}]} \tag{81}$$

So putting both this facts together and going back to the question, we get:

$$V_i \Lambda_\lambda(\sigma_i \rho \sigma_i) V_i^\dagger = V_i \Lambda_\lambda\left(\rho_{\vec{s}\odot[\vec{\pm}^i]}\right) V_i^\dagger = V_i \left(\rho_{\vec{s}\odot[\vec{\pm}^i]\odot[\lambda^{\vec{x}y}]}\right) V_i^\dagger \tag{82}$$

and since the product of numbers is associative and commutes, so is its elementwise product. Making it obvious that performing the same rotation $\sigma_i \rho' \sigma_i$, will cancel the part of the elementwise product coming from the first one, since $[\vec{\pm}^i]\odot[\vec{\pm}^i] = \left[\sum_j (+1)^{\delta_{ij}+1}\vec{e}_j\right] = (1,1,1)$, and $\vec{s}\odot(1,1,1) = \vec{s}$. Giving:

$$\sigma_i \left(\rho_{\vec{s}\odot[\vec{\pm}^i]\odot[\lambda^{\vec{x}y}]}\right)\sigma_i = \rho_{\vec{s}\odot\cancel{[\vec{\pm}^i]\odot[\vec{\pm}^i]}\odot[\lambda^{\vec{x}y}]} = \rho_{\vec{s}\odot[\lambda^{\vec{x}y}]} = \Lambda_\lambda(\rho_{\vec{s}}) \tag{83}$$

Which means that the channel is teleportation-covariant, and our 4 "correcting" unitaries are: $\boxed{V_i = \sigma_i}$

Also, even though we did not include index 0 in the computation to make it easier, it's obvious that it will also fulfill $V_0 = \sigma_0$ since, $\sigma_0 = \mathbb{1}$, and if you don't do anything before, you don't need to do anything later.

Finally, we can understand this graphically, since $\sigma_i \rho \sigma_i$ is a rotation of concretely π around the i axis, whenever you do one this rotations, the "squeeze" will always stay on the z-axis. You can capsize it, with $i = x, y$, or rotate it in its own axis with $i = z$, but the deformation will always remain in the same axis. Meaning, you can rotate, squeeze and undo the first rotation, and get the same result as with a single squeeze:

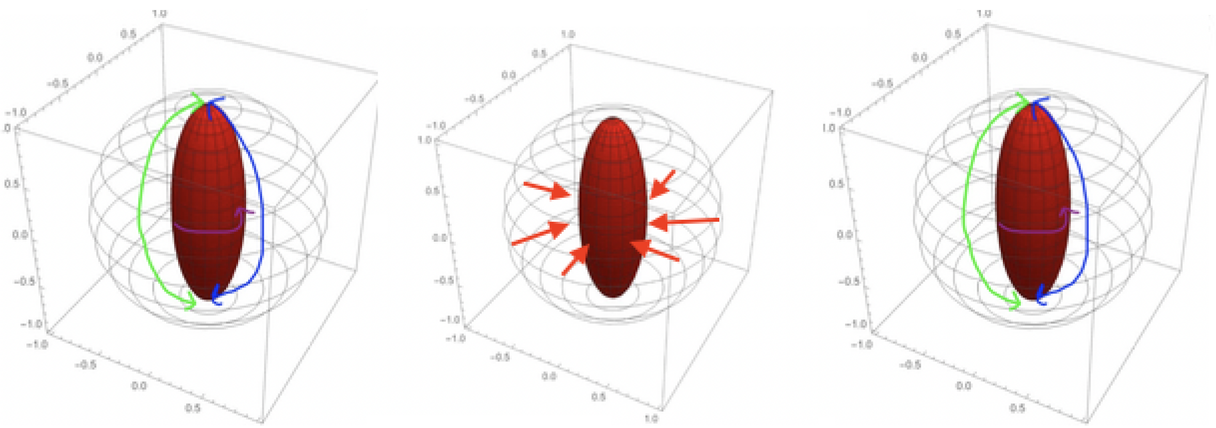

Figure 2: Using the same figure as in Fig. 1, but this time illustrating what each $\sigma_i \rho \sigma_i$ does before and after the "squeeze". Where blue is $i = x$, green $i = y$ and purple $i = z$, all of them leaving the "squeezes" in the same axis.

If instead, we did rotations different than π, the "squeeze" after the first rotation would not be in the same z-axis direction, so you would not be able to undo the rotation and get the same z-axis "squeeze" than in the case without rotations.

Section 3

The Choi matrix is:

$$J_\lambda = \sum_{i,j=1}^{d_{in}} \Lambda_\lambda\big(\,|i\rangle\langle j|\,\big) \otimes |i\rangle\langle j| = \begin{pmatrix} \Lambda_\lambda\big(\,|0\rangle\langle 0|\,\big) & \Lambda_\lambda\big(\,|0\rangle\langle 1|\,\big) \\ \Lambda_\lambda\big(\,|1\rangle\langle 0|\,\big) & \Lambda_\lambda\big(\,|1\rangle\langle 1|\,\big) \end{pmatrix} \tag{84}$$

And the action of the channel on each matrix element is:

$$\begin{cases} \Lambda_\lambda(|0\rangle\langle 0|) = (1-\lambda)\,|0\rangle\langle 0| + \lambda\sigma_z\,|0\rangle\langle 0|\,\sigma_z = (1-\lambda)\,|0\rangle\langle 0| + \lambda\,|0\rangle\langle 0| = |0\rangle\langle 0| \\ \Lambda_\lambda(|0\rangle\langle 1|) = (1-\lambda)\,|0\rangle\langle 1| + \lambda\sigma_z\,|0\rangle\langle 1|\,\sigma_z = (1-\lambda)\,|0\rangle\langle 1| - \lambda\,|0\rangle\langle 1| = (1-2\lambda)\,|0\rangle\langle 1| \\ \Lambda_\lambda(|1\rangle\langle 0|) = (1-\lambda)\,|1\rangle\langle 0| + \lambda\sigma_z\,|1\rangle\langle 0|\,\sigma_z = (1-\lambda)\,|1\rangle\langle 0| - \lambda\,|1\rangle\langle 0| = (1-2\lambda)\,|1\rangle\langle 0| \\ \Lambda_\lambda(|1\rangle\langle 1|) = (1-\lambda)\,|1\rangle\langle 1| + \lambda\sigma_z\,|1\rangle\langle 1|\,\sigma_z = (1-\lambda)\,|1\rangle\langle 1| + \lambda\,|1\rangle\langle 1| = |1\rangle\langle 1| \end{cases} \tag{85}$$

which going back into the Choi matrix of eq. (84), gives:

$$\begin{aligned} J_\lambda &= \begin{pmatrix} |0\rangle\langle 0| & (1-2\lambda)\,|0\rangle\langle 1| \\ (1-2\lambda)\,|1\rangle\langle 0| & |1\rangle\langle 1| \end{pmatrix} = \boxed{\begin{pmatrix} 1 & 0 & 0 & 1-2\lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-2\lambda & 0 & 0 & 1 \end{pmatrix}} = \\ &= |00\rangle\langle 00| + (1-2\lambda)\big(\,|00\rangle\langle 11| + |11\rangle\langle 00|\,\big) + |11\rangle\langle 11| \end{aligned} \tag{86}$$

Section 4

Since for both λ_1 and λ_2, the result is the same:

$$\Lambda_{\lambda_1}\big(\,|0\rangle\langle 0|\,\big) = \Lambda_{\lambda_2}\big(\,|0\rangle\langle 0|\,\big) = |0\rangle\langle 0| \tag{87}$$

You can't distinguish them, so the error should be maximum possible. As the channels are equiprobable, the error probability for distinguishing between the two output states ρ_{λ_1} and ρ_{λ_2} is:

$$P_e = \frac{1}{2}\left(1 - \frac{1}{2}\|\rho_{\lambda_1} - \rho_{\lambda_2}\|_1\right) \tag{88}$$

And as $\rho_{\lambda_i} = \Lambda_{\lambda_i}(|0\rangle\langle 0|) = |0\rangle\langle 0|$ for $i = 1, 2$, then $\|\rho_{\lambda_1} - \rho_{\lambda_2}\|_1 = 0$, so we get $\boxed{P_e = 1/2}$, meaning that we would need to completely guess which channel was used.

Section 5

Considering that the maximally entangled state is:

$$\rho^+_{AB} = \big|\phi^+\big\rangle\big\langle\phi^+\big|_{AB} = \frac{1}{2}\sum_{i,j=0}^{1} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B \tag{89}$$

applying the channel to the first subspace (A) of this state will give as a result, precisely the Choi Matrix:

$$\rho'_{AB} = \Lambda_\lambda(\rho^+_{AB})_A = \frac{1}{2}\sum_{i,j=0}^{1} \Lambda_\lambda\big(\,|i\rangle\langle j|\,\big) \otimes |i\rangle\langle j| = \frac{1}{2}J_\lambda \tag{90}$$

getting the full information of the channel in the output state. And since then:

$$\big\|\rho'_{AB,\lambda_1} - \rho'_{AB,\lambda_2}\big\|_1 = \frac{1}{2}\big\|J_{\lambda_1} - J_{\lambda_2}\big\|_1 = \frac{1}{2}\Big\|2(\lambda_2-\lambda_1)\cancelto{2}{\big(|00\rangle\langle 11| + |11\rangle\langle 00|\big)}\Big\|_1 = 2|\lambda_2 - \lambda_1| \tag{91}$$

The probability of error for distinguishing both states is given by:

$$\boxed{P_e = \frac{1}{2}\left(1 - \frac{1}{2}\big\|\rho'_{AB,\lambda_1} - \rho'_{AB,\lambda_2}\big\|_1\right) = \frac{1 - |\lambda_2 - \lambda_1|}{2}} \tag{92}$$

If $\lambda_1 = \lambda_2$, then both channels are indistinguishable, so the probability of error is $1/2$. If channel two takes the greatest possible value, $\lambda_2 = 1$, and channel 1 the smallest one, $\lambda_1 = 0$, then $P_e = 0$ and they can be perfectly distinguishable.

Section 6

For a general state ρ, we compute its probability of error. First, we start by finding the Schatten-1 norm of the difference between the outputs of the two channels:

$$\|\rho_{\lambda_1} - \rho_{\lambda_2}\|_1 = \left\| \begin{pmatrix} 0 & (\lambda_2 - \lambda_1)(s_x - is_y) \\ (\lambda_2 - \lambda_1)(s_x + is_y) & 0 \end{pmatrix} \right\|_1 = 2|\lambda_1 - \lambda_2|\sqrt{s_x^2 + s_y^2} \tag{93}$$

so the probability of error is:

$$P_e = \frac{1}{2}\left(1 - \frac{1}{2}2|\lambda_2 - \lambda_1|\sqrt{s_x^2 + s_y^2}\right) = \frac{1 - |\lambda_2 - \lambda_1|\sqrt{s_x^2 + s_y^2}}{2} \tag{94}$$

We would like now to minimize this quantity considering that Bloch vectors are confined inside the Bloch sphere:

$$s_x^2 + s_y^2 + s_z^2 \leq 1 \tag{95}$$

As we need to maximize the quantity $s_x^2 + s_y^2$ inside this region, we require that the Bloch vector lies on the surface and that its z-component is zero:

$$\vec{s}_{min} = (s_x, s_y, 0) \text{ such that } s_x^2 + s_y^2 = 1 \tag{96}$$

Therefore, the states that minimize the error are those lying on the equator of the Bloch sphere.

Then, we find that the minimal error probability, using any state is:

$$\boxed{P_e = \frac{1 - |\lambda_2 - \lambda_1|}{2}} \tag{97}$$

which coincides with the probability of error found in the previous section. Thus, using a maximally entangled state is the optimal strategy to discriminate between this teleportation-covariant channel.

This would be what one would have guessed, since as we saw, the maximally entangled state made that the final state was exactly the Choi matrix, meaning, that all the channel information was codified in the final state itself.

Section 7

When n uses of the channel are available, then the optimal probability of error is the same as that of discriminating n copies of the states $\sigma_\lambda^{\otimes n}$. In particular one can make use of the known results in quantum state discrimination, e.g. the Helstrom bound, which says that for two channels $\Lambda_{1,2}$ with priors $\eta_{1,2}$:

$$P_{err} = \frac{1}{2}\left(1 - \|\eta_1\sigma_{\Lambda_1}^{\otimes n} - \eta_2\sigma_{\Lambda_2}^{\otimes n}\|_1\right) \xrightarrow{\eta_1=\eta_2=\frac{1}{2}} \frac{1}{2}\left(1 - \frac{1}{2}\|\sigma_{\Lambda_1}^{\otimes n} - \sigma_{\Lambda_2}^{\otimes n}\|_1\right) \tag{98}$$

where each $\sigma_{\Lambda_i}^{\otimes n}$ state is all n qubits first entangled between them, where then we use the Λ_i channels in parallel in each qubit individually, to extract the most information possible (optimal = parallel channel uses with entanglement).

The matrix σ_{Λ_λ} has the following spectral decomposition:

$$\sigma_{\Lambda_\lambda} = \frac{J_{\Lambda_\lambda}}{2} = (1-\lambda)\left|\phi^+\middle\rangle\!\middle\langle\phi^+\right| + \lambda\left|\phi^-\middle\rangle\!\middle\langle\phi^-\right| \tag{99}$$

so the tensor product $\sigma_{\Lambda_\lambda}^{\otimes n}$ is:

$$\sigma_{\Lambda_\lambda}^{\otimes n} = \sum_{k=0}^{n}(1-\lambda)^k\lambda^{n-k}\sigma(k)\left|\phi^+\middle\rangle\!\middle\langle\phi^+\right|^{\otimes k} \otimes \left|\phi^-\middle\rangle\!\middle\langle\phi^-\right|^{\otimes n-k} \tag{100}$$

where $\sigma(k)$ represents all possible permutations of tensor products of k states $|\phi^+\rangle\langle\phi^+|$ and $n-k$ states $|\phi^-\rangle\langle\phi^-|$. Whit this, then:

$$\sigma_{\Lambda_1}^{\otimes n} - \sigma_{\Lambda_2}^{\otimes n} = \sum_{k=0}^{n}\left((1-\lambda_1)^k\lambda_1^{n-k}(1-\lambda_2)^k\lambda_2^{n-k}\right)\sigma(k)\left|\phi^+\middle\rangle\!\middle\langle\phi^+\right|^{\otimes k} \otimes \left|\phi^-\middle\rangle\!\middle\langle\phi^-\right|^{\otimes n-k} \tag{101}$$

so the norm of the difference is:

$$\left\|\sigma_{\Lambda_1}^{\otimes n} - \sigma_{\Lambda_2}^{\otimes n}\right\|_1 = \sum_{k=0}^{n} \binom{n}{k} \left|(1-\lambda_1)^k \lambda_1^{n-k} - (1-\lambda_2)^k \lambda_2^{n-k}\right| \tag{102}$$

where the terms $\binom{n}{k}$ comes from adding up all possible permutations of $\sigma(k)$.

Then, the probability of error is:

$$\boxed{P_{err} = \frac{1}{2}\left(1 - \frac{1}{2}\sum_{k=0}^{n}\binom{n}{k}\left|(1-\lambda_1)^k \lambda_1^{n-k} - (1-\lambda_2)^k \lambda_2^{n-k}\right|\right)} \tag{103}$$

which for $n = 1$, gives back the previous result:

$$P_{err}^{\,n=1} = \frac{1}{2}\left(1 - \frac{1}{2}\sum_{k=0}^{1}\binom{1}{k}\left|(1-\lambda_1)^k \lambda_1^{1-k} - (1-\lambda_2)^k \lambda_2^{1-k}\right|\right) = \frac{1 - \frac{1}{2}\cancel{\left(\binom{1}{0}+\binom{1}{1}\right)}|\lambda_2 - \lambda_1|}{2} = \boxed{P_{err}^{\text{ section 6}}} \tag{104}$$

and which gets smaller the more uses you do. Gaining more information with each try, as one would expect:

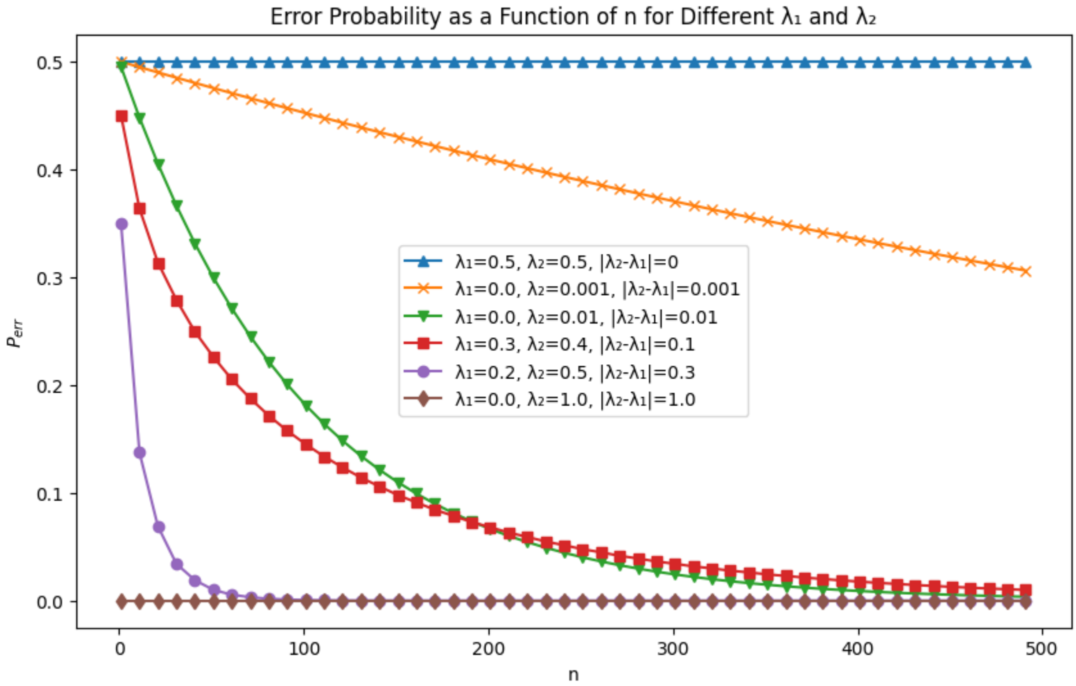

Figure 3: Probability of error (P_{err}) as a function of n, for different values of λ_1 and λ_2.

Also we see a convergence $P_{err} \xrightarrow{n\to\infty} 0$, which agrees with the asymptotic limit expected from the quantum Chernoff bound, that says that for a large enough n the probability of error (P_{err}) will decay as an exponential, given by:

$$P_{err} \approx e^{-nQ_{Ch}(\Lambda_1,\Lambda_2)} \quad \text{with} \quad Q_{Ch}(\Lambda_1,\Lambda_2) = -\lim_{n\to\infty}\frac{1}{n}\log P_{err} = -\min_{0\le s\le 1}\log\left(tr(\rho_{\Lambda_1}^{s}\rho_{\Lambda_2}^{1-s})\right) \tag{105}$$

which in this case, would be obtained minimizing in s, the following expression:

$$-\min_{0\le s\le 1}\log\left((1-\lambda_1)^s(1-\lambda_2)^{1-s} + \lambda_1^s\lambda_2^{1-s}\right)\begin{cases}\lambda_1,\lambda_2 = 0.3, 0.4: -\log\left(\min_s 0.7^s 0.6^{1-s} + 0.3^s 0.4^{1-s}\right) = 2.4\cdot 10^{-3}\\ \lambda_1,\lambda_2 = 0.2, 0.5: -\log\left(\min_s 0.8^s 0.5^{1-s} + 0.2^s 0.5^{1-s}\right) = 2.3\cdot 10^{-2}\end{cases} \tag{106}$$

where to end, we obtain that in the asymptotic limit the purple line in Fig. 3, will decay with an exponent 10 times bigger than that of the red line, in agreement with the observed result in the figure.