Final Exam

December, 21st 2024

- Time Limit: 8:30 am - 10:30 am.
- **No** books, course notes nor electronic devices are allowed.
- The problems are on the other side of the paper.
- Upon completion, the examination paper has to be submitted together with your answer book.
- Except in cases where it is explicitly specified, all the results should be justified with rigorous mathematical proofs.

Course check (50%)

1. **(10%)** Given n independent copies $X_1, \dots, X_n$ of a continuous random vector $X : \Omega \to \mathcal{X}$, that we regroup in a data set $\mathcal{D} = \{X_1, \dots, X_n\}$. Given any other continuous random vector $\theta : \Omega \to \Theta$ (dependent or not on $\mathcal{D}$), provide the expression of the density of θ conditionally on $\mathcal{D}$ as it is given by the Bayes formula. Identify in this formula the likelihood (no proof needed).

2. **(10%)** Let $X \sim \mathcal{N}(\mu, \Sigma)$ where $\mu \in \mathbb{R}^{p+q}$ and $\Sigma \in \mathbb{R}^{(p+q)\times(p+q)}$. Suppose X is partitioned into two components $X_1 : \Omega \to \mathbb{R}^p$ and $X_2 : \Omega \to \mathbb{R}^q$ such that for all $\omega \in \Omega$, $X(\omega) = (X_1(\omega), X_2(\omega))$. Express the density of X_1 conditioned on X_2 as a function of the elements of the following block decomposition:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with $\mu_1 \in \mathbb{R}^p$, $\mu_2 \in \mathbb{R}^q$, $\Sigma_{11} \in \mathbb{R}^{p\times p}$, $\Sigma_{12} \in \mathbb{R}^{p\times q}$, $\Sigma_{21} \in \mathbb{R}^{q\times p}$ and $\Sigma_{22} \in \mathbb{R}^{q\times q}$. One will rely on the following formula of the inversion of a block matrix $M = \left(\begin{smallmatrix} E & F \\ G & H \end{smallmatrix}\right)$:

$$M^{-1} = \begin{pmatrix} I & 0 \\ -H^{-1}G & I \end{pmatrix} \cdot \begin{pmatrix} (M/H)^{-1} & 0 \\ 0 & H^{-1} \end{pmatrix} \cdot \begin{pmatrix} I & -FH^{-1} \\ 0 & I \end{pmatrix}, \tag{1}$$

where $M/H \equiv E - FH^{-1}G$.

Correction: *Let us express the density thanks to* (1) *(with* $M = \Sigma$*:*

$$\begin{aligned} p_{X_1,X_2}(x_1, x_2) &\propto \exp\left(-\frac{1}{2}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \cdot \begin{pmatrix} I_p & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_q \end{pmatrix} \cdot \begin{pmatrix} (\Sigma/\Sigma_{22}) & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \cdot \begin{pmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_q \end{pmatrix} \cdot \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right) \\ &\propto \exp\left(-\frac{1}{2}\left(x_1 - \mu_{1|2}\right)^T \Sigma_{1|2}^{-1}\left(x_1 - \mu_{1|2}\right)\right) \cdot \exp\left(\frac{1}{2}(x_2 - \mu_2)^T\Sigma_{22}^{-1}(x_2 - \mu_2)\right). \end{aligned}$$

with:

- $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$
- $\Sigma_{1|2} = \Sigma/\Sigma_{22} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Then one can deduce from the identity $p_{X_1,X_2}(x_1, x_2) = p_{X_1,X_2}(x_1|x_2)p_{X_2}(x_2)$ *that:*

$$p_{X_1,X_2}(x_1|x_2) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{|\Sigma_{1|2}|}} \exp\left(-\frac{1}{2}\left(x_1 - \mu_{1|2}\right)^T \Sigma_{1|2}^{-1}\left(x_1 - \mu_{1|2}\right)\right).$$

□

3. **(10%)** Provide the expression of the mean squared error (MSE) for a regression task and of the Bayes Rule associated. Prove the validity of the expression you provided for the Bayes rule.

4. **(5%)** Give the setting and describe the method of rejection sampling (no proof needed).

5. **(10%)** Given a measurable mapping $h : \mathbb{R} \to \mathbb{R}$ and a density $f : \mathbb{R} \to \mathbb{R}_+$, express the density $g : \mathbb{R} \to \mathbb{R}_+$ that minimizes the variance of $\frac{h(X)f(X)}{g(X)}$ for $X \sim g$ and prove the result. Explain how it can be used for importance sampling.

6. **(5%)** Explain how to sample a given random variable X by sampling first a random variable $Y \sim \text{Unif}[0, 1]$ and applying a certain transformation $\psi : [0, 1] \to \mathbb{R}$. Express the mapping ψ and prove your result.

Problem 1 (30%): Kalman filters.

Distribution of points: 2.5% + 2.5% + 2.5% + 5% + 5% + 5% + 2.5% + 5%.

A Kalman filter aim at predicting the value of unknown variable x when the measurement y is noisy. The general idea is to take into account both the physical equation that predicts the dynamic of the system but could present some drift and the measurement that actualize the state position but could present some noise. For simplicity, let us assume that we want to track a moving object whose position $(x_i)_{i\in\mathbb{N}} \in (\mathbb{R}^p)^{\mathbb{N}}$ follows the following equation:

$$x_0 \sim \mathcal{N}(0, Q_0) \qquad \text{and} \qquad \forall i \geq 1: \quad x_i = Fx_{i-1} + v_i \qquad \text{where} \quad v_i \sim \mathcal{N}(0, Q). \tag{2}$$

We will assume that the matrices F (that accounts for the physical behavior of the system) and the covariance of noises Q_0, Q are all known. Besides the noise terms $x_0, v_1, \ldots, v_i, \ldots$ are assumed to be independent.

Q1: Give a naive estimation of x_i depending only on F. Explain why this estimation is prone to drift and accumulates errors over time.

Correction: *One could choose to define the estimator with the iterative expression:*

$$\hat{x}_0 = 0$$
$$\forall i \geq 1: \ \hat{x}_{i+1} = F\hat{x}_i$$

That would imply that for all $i > 0$: $\hat{x}_i = 0$ which can be very far from what would obtain if for instance one would set $x_0 \neq 0$, then the drift after i iterations would be equal to $F^i x_0$ which can be very big depending on F. □

We are able to solve the drift issue thanks to measurements $(y_i)_{i\in\mathbb{N}} \in (\mathbb{R}^q)^{\mathbb{N}}$ that depend on the position through the following equation:

$$y_i = Hx_i + w_i \qquad \text{where} \quad w_i \sim \mathcal{N}(0, R). \tag{3}$$

The goal is then to find an estimator $\hat{x}_i$ depending on $y_1, \ldots, y_i$ that will approximate the real position x_i. To simplify the notation, one will denote[1] $\mathcal{Y}_i = \{y_0, \ldots, y_i\}$. let us chose as estimator $\hat{x}_i$ for x_i the vector $x \in \mathbb{R}^p$ that minimizes the loss:

$$J(x) = \mathbb{E}\left[(x_i - x)^2 \mid \mathcal{Y}_i\right]. \tag{4}$$

Q2: Show that $\hat{x}_i = \mathbb{E}[x_i|\mathcal{Y}_i]$.

Correction: *Given $x \in \mathbb{R}^p$ let us bound:*

$$\begin{aligned} J(x) = \mathbb{E}\left[(x_i - x)^2 \mid \mathcal{Y}_i\right] &= \mathbb{E}\left[(x_i - \mathbb{E}[x_i|\mathcal{Y}_i] + \mathbb{E}[x_i|\mathcal{Y}_i] - x)^2 \mid \mathcal{Y}_i\right] \\ &= \mathbb{E}\left[(x_i - \mathbb{E}[x_i|\mathcal{Y}_i])^2\right] + 2\mathbb{E}\left[(x_i - \mathbb{E}[x_i|\mathcal{Y}_i])\left(\mathbb{E}[x_i|\mathcal{Y}_i] - x\right) \mid \mathcal{Y}_i\right] + \mathbb{E}\left[(\mathbb{E}[x_i|\mathcal{Y}_i] - x)^2 \mid \mathcal{Y}_i\right] \\ &= \mathbb{E}\left[(x_i - \mathbb{E}[x_i|\mathcal{Y}_i])^2\right] + \mathbb{E}\left[(\mathbb{E}[x_i|\mathcal{Y}_i] - x)^2 \mid \mathcal{Y}_i\right] \\ &\geq \mathbb{E}\left[(x_i - \mathbb{E}[x_i|\mathcal{Y}_i])^2\right] = J(\mathbb{E}[x_i|\mathcal{Y}_i]). \end{aligned}$$

Therefore, the minimum of $J(x)$ is indeed reached at $\hat{x}_i = \mathbb{E}[x_i|\mathcal{Y}_i]$. □

This estimation is done recursively. At each time step, we will compute the conditional expectation $\hat{x}_i \equiv \mathbb{E}[x_i|\mathcal{Y}_i]$ and the conditional covariance P_i from the previous step (the use of P_i will be made clearer later).

Q3: Express the explicit value of $\hat{x}_0$ and P_0.

Correction: *This question was hard, using the next questions, you know that:*

$$x_{0|-1} \equiv \mathbb{E}[x_0] = 0 \qquad \text{and} \qquad P_{0|-1} \equiv \mathbb{E}[x_0 x_0^T] - x_{0|-1}x_{0|-1}^T = Q,$$

[1] The sequence $\mathcal{Y}_0, \ldots, \mathcal{Y}_i, \ldots$ is traditionally called a "filtration".

thus the formulas given in Question provides:

$$\hat{x}_0 = QH(HQH + R)^{-1} y_0$$
$$P_0 = Q - QH(HQH + R)^{-1} HQ$$

□

Q4: Given $i \geq 1$, assuming that $\hat{x}_{i-1}$ and P_{i-1} are known, express:

$$x_{i|i-1} \equiv \mathbb{E}[x_i|\mathcal{Y}_{i-1}] \qquad \text{and} \qquad P_{i|i-1} \equiv \mathbb{E}[x_i x_i^T|\mathcal{Y}_{i-1}] - x_{i|i-1}x_{i|i-1}^T.$$

Correction: *Let us compute from the formulas we are given:*

- $x_{i|i-1} = \mathbb{E}[Fx_{i-1} + v_i|\mathcal{Y}_{i-1}] = F\mathbb{E}[x_{i-1}|\mathcal{Y}_{i-1}] + \mathbb{E}[v_i|\mathcal{Y}_{i-1}] = F\hat{x}_{i-1}$
- $\mathbb{E}[x_i x_i^T|\mathcal{Y}_{i-1}] = \mathbb{E}[Fx_{i-1}x_{i-1}^T F + 2x_{i-1}^T F v_i + v_i^T v_i|\mathcal{Y}_{i-1}] = F\mathbb{E}[x_{i-1}x_{i-1}^T|\mathcal{Y}_{i-1}]F + \mathbb{E}[v_i^T v_i]$.

thus $x_{i|i-1} = F\hat{x}_{i-1}$ *and* $P_{i|i-1} = FP_iF + Q$. □

Q5: Let us denote $z_i = \binom{x_i}{y_i}$ Express (in block form) the mean and the covariance of z_i conditionally on $\mathcal{Y}_{i-1}$ and provide its density (one can use the notations $x_{i|i-1}$ and $P_{i|i-1}$ to stay as simple as possible).

Correction: *Let us express from previous questions:*

$$\mathbb{E}[z_i|\mathcal{Y}_{i-1}] = \begin{pmatrix} x_{i|i-1} \\ Hx_{i|i-1} \end{pmatrix}$$

To compute the covariance, one needs to express:

- $\mathbb{E}[y_i x_i^T|\mathcal{Y}_{i-1}] = \mathbb{E}[Hx_i x_i^T + w_i x_i^T|\mathcal{Y}_{i-1}] = H\mathbb{E}[x_i x_i^T|\mathcal{Y}_{i-1}]$
- $\mathbb{E}[y_i y_i^T|\mathcal{Y}_{i-1}] = \mathbb{E}[Hx_i x_i^T H + 2w_i x_i^T H + w_i x_i^T|\mathcal{Y}_{i-1}] = H\mathbb{E}[x_i x_i^T|\mathcal{Y}_{i-1}]H + R$

therefore, one can finally express:

$$\mathbb{E}[z_i z_i^T|\mathcal{Y}_{i-1}] - \mathbb{E}[z_i|\mathcal{Y}_{i-1}]\mathbb{E}[z_i|\mathcal{Y}_{i-1}]^T = \begin{pmatrix} P_{i|i-1} & P_{i|i-1}H \\ HP_{i|i-1} & HP_{i|i-1}H + R \end{pmatrix},$$

If we denote p *and* q*, respectively the dimension of* X *and* Y *the density of* z *conditionally on* $\mathcal{Y}_{i-1}$ *writes:*

$$p(z|\mathcal{Y}_{i-1}) = \frac{1}{(2\pi)^{\frac{p+q}{2}}\sqrt{\left|\begin{smallmatrix} P_{i|i-1} & P_{i|i-1}H \\ HP_{i|i-1} & HP_{i|i-1}H+R \end{smallmatrix}\right|}} \exp\left(-\frac{1}{2}\left(z - \begin{pmatrix} x_{i|i-1} \\ Hx_{i|i-1} \end{pmatrix}\right)^T \begin{pmatrix} P_{i|i-1} & P_{i|i-1}H \\ HP_{i|i-1} & HP_{i|i-1}H+R \end{pmatrix}^{-1} \left(z - \begin{pmatrix} x_{i|i-1} \\ Hx_{i|i-1} \end{pmatrix}\right)\right)$$

□

Q6: Deduce the value of $\hat{x}_i = \mathbb{E}[x_i|\mathcal{Y}_i] = \mathbb{E}[x_i|y_i, \mathcal{Y}_{i-1}]$ and P_i, the covariance of x_i knowing $\mathcal{Y}_i$ (question 2 of the course check could be helpful here).

Correction: *One can deduce from the course check that:*

$$\hat{x}_i = \mathbb{E}[x_i|\mathcal{Y}_i] = x_{i|i-1} + P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}(y_i - Hx_{i|i-1})$$
$$P_i = P_{i|i-1} - P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}HP_{i|i-1}$$

□

Q7: Explain the different steps of the Kalman filter to estimate the position x_i at each time step.

Correction: *One start with the values* $\hat{x}_0 = 0$ *and* $P_i = Q_0$ *and at each time step* i*, given* $\hat{x}_{i-1}$ *and* P_{i-1}*, one can estimate consequently:*

(a) $x_{i|i-1} = F\hat{x}_{i-1} \quad$ *and* $\quad P_{i|i-1} = FP_iF + Q$

(b) $\hat{x}_i = x_{i|i-1} + P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}(y_i - Hx_{i|i-1})$

$P_i = P_{i|i-1} - P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}HP_{i|i-1}$

□

Q8: Let us assume that the law governing the position x_i is now:

$$\forall k \geq 1 : \quad x_i = Fx_{i-1} + Gu_i + v_i \qquad \text{where} \;\; v_i \sim \mathcal{N}(0, Q),$$

for a certain known sequence of inputs $u_1, \ldots, u_i, \ldots$. Deduce the Kalman filter procedure in this new setting.

Correction: *In this setting, the good choice for the first step is* $\hat{x}_0 = Gu_0$ *and* $P_i = Q_0$*. Then, assuming that* $\hat{x}_{i-1}$ *and* P_{i-1} *are known, one can evaluate:*

$$x_{i|i-1} = F\hat{x}_{i-1} + Gu_{i-1}$$

and removing directly the projections with zero mean independent gaussian vectors, one gets:

$$\begin{aligned}
\mathbb{E}[x_i x_i^T | \mathcal{Y}_{i-1}] &= \mathbb{E}\left[Fx_{i-1}x_{i-1}^T F + Gu_{i-1}u_{i-1}^T G + Fx_{i-1}u_{i-1}^T G + Gu_{i-1}x_{i-1}^T F + v_i v_i^T \;|\mathcal{Y}_{i-1}\right] \\
&= F\mathbb{E}\left[x_{i-1}x_{i-1}^T \;|\mathcal{Y}_{i-1}\right] F + Gu_{i-1}u_{i-1}^T G + F\hat{x}_{i-1}u_{i-1}^T G + Gu_{i-1}\hat{x}_{i-1}^T F + Q \\
x_{i|i-1}x_{i|i-1}^T &= F\hat{x}_{i-1}\hat{x}_{i-1}^T F + Gu_{i-1}u_{i-1}^T G + F\hat{x}_{i-1}u_{i-1}^T G + Gu_{i-1}x_{i-1}^T F,
\end{aligned}$$

thus, as before:

$$P_{i|i-1} = \mathbb{E}[x_i x_i^T | \mathcal{Y}_{i-1}] - x_{i|i-1}x_{i|i-1}^T = FP_{i-1}F + Q.$$

The second step is the same:

$$\begin{aligned}
\hat{x}_i &= x_{i|i-1} + P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}(y_i - Hx_{i|i-1}) \\
P_i &= P_{i|i-1} - P_{i|i-1}H(HP_{i|i-1}H + R)^{-1}HP_{i|i-1}.
\end{aligned}$$

□

Problem 2 (20%): Classification.

Distribution of points: 7.5% + 5% + 5% + 2.5%.

Q1: Given k-class classification setting with an observation $X : \Omega \to \mathbb{R}^p$ and a label $Y : \Omega \to [K]$ provide the Bayes rule $g^* : \mathbb{R}^p \to [k]$ that minimizes the risk associated to the mis-classification loss and justify.

Correction: *Let us denote the misclassification loss* $l(z, y) = \mathbb{1}_{z \neq y}$*, and introduce:*

$$g^*(x) = \arg\max_{a \in \mathcal{Y}} P(Y = a \mid X = x).$$

Given a decision function $f : \mathcal{X} \to \mathcal{Y}$*, one can bound from below:*

$$\begin{aligned}
R(f) &= \mathbb{E}[l(f(X), Y)] = \mathbb{E}[\mathbb{1}_{f(X) \neq Y}] = \mathbb{E}[\mathbb{E}[\mathbb{1}_{f(X) \neq Y} \mid X]] = \mathbb{E}\left[\sum_{k=1}^{K} \mathbb{E}[\mathbb{1}_{f(X)=a_k}\mathbb{1}_{Y \neq a_k} \mid X]\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{K} \mathbb{E}[\mathbb{1}_{f(X)=a_k}]\mathbb{E}[\mathbb{1}_{Y \neq a_k} \mid X]\right] = \mathbb{E}\left[\sum_{k=1}^{K} \mathbb{P}(f(X) = a_k)\left(1 - \mathbb{P}(Y = a_k \mid X)\right)\right] \\
&= \sum_{k=1}^{K} \mathbb{P}(f(X) = a_k) - \mathbb{E}\left[\sum_{k=1}^{K} \mathbb{1}_{f(X)=a_k}\mathbb{P}(Y = a_k \mid X)\right] \\
&\geq \mathbb{E}[1 - \mathbb{P}(Y = g^*(X) \mid X)] = \mathbb{P}(Y \neq g^*(X)) = \mathbb{E}[\mathbb{1}_{Y \neq g^*(X)}] = R(g^*)
\end{aligned}$$

since:

- $\sum_{k=1}^{K} \mathbb{P}(f(X) = a_k) = 1$
- $\forall k \in [K]$: $\mathbb{P}(Y = g^*(X) \mid X) \geq \mathbb{P}(Y = a_k \mid X)$.

The fact that $R(f) \geq R(g^*)$ for any decision function f exactly means that g^* is the Bayes rule. □

Let us then consider the following model for $k = 3$, $p = 2$ and given $q \in [0, 1]$:

$$\mathbb{P}(Y = 1) = \frac{q}{2}; \qquad \mathbb{P}(Y = 2) = 1 - q \qquad \text{and} \qquad \mathbb{P}(Y = 3) = \frac{q}{2},$$

and the distribution of the continuous random vector $X : \Omega \to \mathbb{R}^2$ is defined followingly:

$$X|Y = 1 \sim \mathcal{N}((0, 0), I_2); \qquad X|Y = 2 \sim \text{Unif}([0, 1] \times [0, 1]) \qquad \text{and} \qquad X|Y = 3 \sim \mathcal{N}((1, 1), I_2),$$

Q2: Provide the Bayes rule for this model (and for the mis-classification loss).

Correction: *The identity:*

$$\forall y \in \{1, 2, 3\} : \qquad \mathbb{P}(Y = y|x) = \frac{p(X = x, Y = y)}{p_X(x)}$$

allows us to express the Bayes rule as:

$$g^*(x) = \arg \max_{a \in \{1,2,3\}} r_a(x) \qquad \textit{with} \quad \forall x \in \mathbb{R}^2, a \in \{1, 2, 3\} : r_a(x) = p(X = x, Y = y).$$

To be more precise, let us express thanks to the formula $p(X = x, Y = y) = p(X = x|Y = y)\mathbb{P}(Y = y)$:

- $r_1(x) = \frac{q}{4\pi} e^{-\|x\|^2/2}$
- $r_2(x) = (1 - q)\mathbb{1}_{x \in [0,1]^2}$
- $r_3(x) = \frac{q}{4\pi} e^{-\|x - \mathbb{1}\|^2/2}$.

□

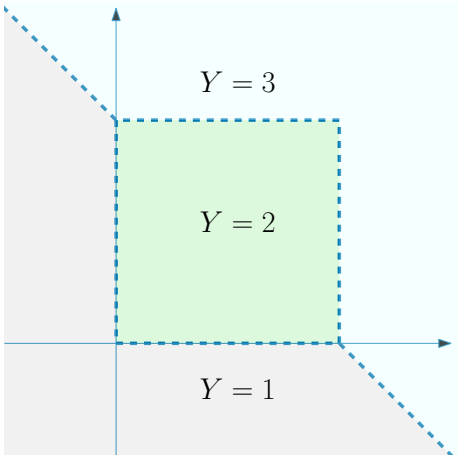
Q3: Draw the decision boundaries for $q = \frac{1}{2}$.

Correction: *from the formula given in previous question, one sees that when $q = \frac{1}{2}$:*

- *for $x \in [0, 1]^2$ $r_1(x), r_3(x) \leq \frac{1}{8\pi} \leq \frac{1}{2} = r_2(x)$*
- *for $x \notin [0, 1]^2$, $r_2(x) = 0$ and besides:*

$$r_1(x) \geq r_3(x) \qquad \Longleftrightarrow \qquad \|x\|^2 \leq \|x - \mathbb{1}\|^2$$
$$\Longleftrightarrow \qquad x_1^2 + x_2^2 \leq x_1^2 + x_2^2 - 2x_1 - 2x_2 + 2 \qquad \Longleftrightarrow \qquad x_1 + x_2 \leq 1$$

That gives us the following decision boundary:



□

Q4: For which value of q the Bayes rule will only output 2 classes.

Correction: *The Bayes rule strictly output 2 classes if and only if for all* $x \in \mathbb{R}^2$*:* $\max(r_1(x), r_3(x)) \geq r_2(x)$*. Given* $x \in [0,1]^2$ *:*

$$\max(r_1(x), r_3(x)) = \frac{q}{4\pi} \exp\left(-\frac{1}{2}\min(\|x\|^2, \|x - \mathbb{1}\|^2)\right),$$

and besides, for $x_0 = (1,0)$ *(one could also take* $x_0 = (0,1)$*):*

$$\min(\|x\|^2, \|x - \mathbb{1}\|^2) \leq 1 = \min(\|x_0\|^2, \|x_0 - \mathbb{1}\|^2).$$

Thus

$$\max(r_1(x), r_3(x)) \geq r_1(x_0) = r_3(x_0) = \frac{q}{4\pi} e^{-1/2}.$$

And since r_2 *is constant on* $[0,1]^2$ *and equal to* $1 - q$*, one can deduce that:*

$$\begin{aligned} \left(\forall x \in [0,1]^2 : \ r_2(x) \leq \max(r_1(x), r_3(x))\right) &\iff r_2(x_0) \leq r_1(x_0) \\ &\iff 1 - q \leq \frac{q}{4\pi} e^{-1/2} \iff q \geq \frac{1}{1 + \frac{1}{4\pi} e^{-1/2}}. \end{aligned}$$

□