QUSUM: quickest quantum change-point detection

Marco Fanizza,¹ Christoph Hirche,² and John Calsamiglia¹

1 Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain
2 Center for Quantum Technologies, National University of Singapore
Zentrum Mathematik, Technical University of Munich, 85748 Garching, Germany

Online detection of a change in a sequence of events is a fundamental and practically relevant task. In this work we consider the quantum version of the problem, generalizing the well known (classical) CUSUM algorithm to detect a quantum change point. Our algorithm exploits joint measurements to amplify the trade-off between detection delay and false detections, which in the case of a quantum change point is asymptotically characterized by the quantum relative entropy. Furthermore we also show that the relative entropy is indeed asymptotically optimal under arbitrary, potentially adaptive, measurement strategies, hence providing the ultimate bound for the quickest quantum change-point detection. Finally, we discuss online change point detection in quantum channels.

The detection of sudden changes of a stochastic random variable is one of the most fundamental problems in statistical analysis. Imagine a device producing copies of an item within certain predefined parameters. If the device malfunctions it might in general be hard to immediately recognize the problem. However, it can be extremely important to identify the point from which the items have been produced by a faulty device. A point of malfunction is a particular instance of a change point, where the device changes from producing according to certain parameters to different ones. Detecting these change points has been the subject of its own field of research in classical statistical analysis [1–4]. Applications can be found in quality control [5, 6], finance [2, 7], medical diagnosis [2, 8], climate research [9, 10] and robotics [11, 12]. Recently, the concept of change point detection has been generalized to the quantum world [13–15]. Here, we have a device outputting quantum states. By default this device will output a certain state $\rho$, but from a given (random) point of time it will start producing the state $\sigma$. The goal is to identify this change point. In [13, 15] the problem has been considered as an instance of hypothesis testing, where one collects the full sequence of quantum states and then tries to determine if and where a change point occurred. Since one requires the full sequence, this is usually considered offline change point detection. For the special case of pure quantum states, solutions where given for the minimum error probability in detecting a change [14] and for exact detection [15]. For any practical application, however, one usually cannot wait for the entire sequence to be collected (it could potentially even be an infinite sequence). Therefore, it is well motivated to consider online detection or quickest change point detection: an algorithm that produces a guess on whether a change has happened after every received instance of the random variable (here, every received quantum state). This follows the same motive as recent efforts to find optimal sequential discrimination rates in quantum hypothesis testing of states [17, 18] and channels [19]. In this scenario, the natural quantities to consider are the time delay in detecting a change point versus the risk of a false alarm, i.e. falsely detecting a change when none has happened. In classical statistical analysis the most studied such algorithm is Page’s CUSUM algorithm [20]. Apart from its computational simplicity, one of its most important features is optimality under certain risk criteria, as shown in the asymptotic setting by Lorden [21]. It was later shown that the CUSUM algorithm is also optimal in a finite regime [22, 23]. In [16] online strategies for quantum change point detection have been considered in the restricted scenario of exact identification with maximum success probability given pure quantum states. As the main contribution of this work we provide a quantum version of the CUSUM algorithm, called QUSUM, and show that it is asymptotically optimal under a certain risk criteria. In particular, this algorithm uses only $k$-local (projective) measurements on the incoming quantum states, still asymptotically it cannot be outperformed by even a collective or adaptive quantum measurement strategy. The achievability proof follows from applying the analysis of the classical problem to block measurements. The crucial ingredient for the optimality proof is the use of a strong converse property in quantum hypothesis testing [24], to verify that a
certain condition that implies a converse bound in a (possibly non i.i.d.) change point problem \[10\] has to be satisfied by any candidate quantum algorithm. The paper is structured as follows: we first present the problem in its simplest form and state the two main results, which are then proven in dedicated sections. Afterwards, we comment on the implications of our results for the problem of change point detection in sequences of quantum channels. We conclude by mentioning open problems.

Setting and results:— The change point sequence is a sequence of \(d\)-dimensional states \(\{\rho^{(n)}\}, n = 1, 2, \ldots\), such that if \(n \leq \nu\), \(\rho^{(n)} = \rho\) and if \(n > \nu\), \(\rho^{(n)} = \sigma\). At each step \(n\), the algorithm receives a copy of the state \(\rho^{(n)}\). The latter is then measured together with the current state of previously received copies by a joint quantum measurement, whose outcome determines whether to continue or to stop at step \(n\) and emit an alarm that signals that the change has occurred. Let’s call \(T\) the random variable corresponding to the alarm time \(n\) at which stopping occurs, for some given strategy. Let \(E_{\infty/\nu}\) denote expectation values with respect to some measurement acting sequentially on a sequence of copies of \(\rho\) (the change point never happens, \(\nu = \infty\)) or the change point sequence for a specific finite \(\nu\). We leave the precise algorithm implicit, but it should be clear that it could be any sequential quantum measurement on the sequence \(\rho^{(n)}\), with outcomes described by random variables \(X_1, \ldots, X_\nu\), and that our figure of merit will be optimized over the choice of the measurement. We define the mean false alarm time as

\[
\bar{T}_{FA} = E_{\infty}(T). \tag{1}
\]

Furthermore, we define the worst-worst case mean delay \[10\]

\[
\bar{\tau}^* := \sup_{\nu \geq 0} \sup_{\nu > 0} E_{\nu}(T - \nu|T > \nu, X_1, \ldots, X_\nu). \tag{2}
\]

This figure of merit considers the worst delay over all possible locations of the change point and over almost all measurement outcomes before the change, measured according to the associated probability measure. A less strict figure of merit is the maximal conditional average delay of detection:

\[
\bar{\tau} := \sup_{\nu \geq 0} E_{\nu}(T - \nu|T > \nu) \leq \bar{\tau}^*. \tag{3}
\]

In the following we always assume \(\text{supp}\ \sigma \subseteq \text{supp}\ \rho\). This implies that \(D(\sigma||\rho) = \text{tr}\left[\sigma \log \sigma - \log \rho\right] < \infty\), \(D_{\max}(\sigma||\rho) = \inf\{\lambda \geq 0 : \sigma \leq 2^\lambda \rho\} < \infty\). Otherwise, there exist a projector \(\Pi\) such that \(\text{tr}[\Pi \rho] = 0\) and \(\text{tr}[\Pi \sigma] = c > 0\), therefore the change can be detected with high probability in finite time, with no dependence on \(\bar{T}_{FA}\). We prove the following main results.

**Theorem 1 (Achievability)** Given a change point problem with two finite-dimensional states \(\rho\) and \(\sigma\), \(D(\sigma||\rho) < \infty\), for any \(\epsilon > 0\), and \(\bar{T}_{FA}\) large enough there is a QUSUM algorithm such that \(\bar{\tau}^* \leq \frac{\log \bar{T}_{FA}}{D(\sigma||\rho)(1 - \epsilon)} + O(1)\)

QUSUM is, as the name suggests, a quantum version of the classical CUSUM algorithm, which in turn is based on repeated sequential probability ratio tests (SPRT), see also \[20\]. We show that the performance of QUSUM is asymptotically optimal:

**Theorem 2 (Optimality)** Any algorithm for a change point problem with two finite-dimensional states \(\rho\) and \(\sigma\), \(D(\sigma||\rho) < \infty\), with expected false alarm \(\bar{T}_{FA}\) must satisfy \(\bar{\tau}^* \geq \bar{\tau} \geq (1 - \epsilon)\frac{\log \bar{T}_{FA}}{D(\sigma||\rho)(1 + o(1))}\) for any \(\epsilon > 0\).

This implies that for large \(\bar{T}_{FA}\) optimal strategies satisfies \(\bar{\tau}^* \sim \frac{\log \bar{T}_{FA}}{D(\sigma||\rho)}\). We discuss straightforward generalizations of these results in the respective sections dedicated to their proofs.

**Achievability:**— We will prove the achievability in two steps. First we assume a simple algorithm that individually measures each incoming state with a POVM \(\{M_x\}\) applied to the \(i\)-th state, giving output probabilities

\[
p(x_i) = \text{tr}[M_x \rho], \quad q(x_i) = \text{tr}[M_x \sigma]. \tag{4}
\]

Based on these we can define the log-likelihood ratio and their partial sums,

\[
Z_i = \log \frac{q(x_i)}{p(x_i)}, \quad Z^k_j = \sum_{i=j}^k Z_i. \tag{5}
\]

It can easily be seen that

\[
E_p[Z_i] = D(q||p) \quad E_p[Z_i] = -D(p||q), \tag{6}
\]

where \(E_p\) denote expectation values with respect to the probability distributions \(p/q\) respectively. The idea is to start a one-sided SPRT on the output probability distributions at each possible change point \(j\),
meaning that we consider for each \( j \) the stopping time
\[
T_j = \min \{ k \geq j : Z_j^k \geq h \},
\]
where we define \( T_j = \infty \) if \( \{ k \geq j : Z_j^k \geq h \} = \emptyset \). Given these, we define the stopping time
\[
T^* = \min_{j \geq 1} T_j.
\]
Note that \( T^* \) can be understood as the first time when, given the current measurement record, the probability of having had a change in the past is \( e^h \)-times more likely than having no change —see Supplementary Material (SM). We can now use a result by Lorden (Theorem 2 in [21]) which states that the properties of an online change point algorithm can be deduced from the properties of a set of parallel open-ended SPRTs. In particular, in our notation, if \( P_\infty (T_1 < \infty) \leq \alpha \)
\[
\tilde{T}_{FA} = \mathbb{E}_\infty (T^*) \geq \frac{1}{\alpha}, \quad \text{and } \bar{\tau}^* \leq \mathbb{E}_0 (T_1).
\]
We will now apply this to our scenario. By the definition of our QUSUM algorithm we get, by using Wald’s identities [25], (see also Chapter 3 of [3]),
\[
P_\infty (T_1 < \infty) = \mathbb{E}_0 \left[ \frac{p}{q} I_{T_1 < \infty} \right] = \mathbb{E}_0 \left[ e^{- Z_1^T I_{T_1 < \infty}} \right]
= \mathbb{E}_0 \left[ e^{- h - x} I_{T_1 < \infty} \right] \leq e^{- h} =: \alpha,
\]
\[
\mathbb{E}_0 (T_1) = \frac{\mathbb{E}_0 Z_1^T}{\mathbb{E}_0 Z_1} = \frac{h + \mathbb{E}[x]}{D(q||p)} \to \frac{h}{D(q||p)},
\]
when \( h \to \infty \), where \( x := Z_1^T - h \) is the overshoot and the limit holds since \( Z_i < \infty \).

Putting all the results together we get
\[
\bar{\tau}^* \leq \mathbb{E}_0 (T_1) = \frac{h}{D(q||p)} + O(1) \leq \frac{\log \tilde{T}_{FA}}{D(q||p)} + O(1).
\]
Optimizing over all measurements gives us the achievable trade-off for this particular strategy,
\[
\bar{\tau}^* \leq \frac{\log \tilde{T}_{FA}}{D_M (\sigma||\rho)} + O(1),
\]
as \( h \to \infty \), in terms of the measured relative entropy
\[
D_M (\sigma||\rho) := \sup_{\{M_{\chi}\}} D(q||p) \quad [26, 27].
\]
Note that already projective measurements (PVM) achieve the measured relative entropy [28].

Now, more generally, instead of measuring each copy of \( \rho^{\otimes k} \) separately, the QUSUM algorithm is based on collecting \( k \) states and then performing a joint measurement on the gathered state which is either \( \rho^{\otimes k} \) or \( \sigma^{\otimes k} \) (assuming the change point happens at a multiple of \( k \)). The above trade-off is now easily modified to
\[
\bar{\tau}^* \leq \frac{\log (\tilde{T}_{FA} / k)}{\frac{1}{k} D_M (\sigma^{\otimes k}||\rho^{\otimes k})} + O(1).
\]
If \( \rho \) and \( \sigma \) are states of a \( d \)-dimensional Hilbert space, we have from Theorem 2 of [27] that for any \( \rho \) and \( k \) there is a PVM \( \{ M_{x_i} \} \), depending only on \( \rho \), such that if \( p^{(k)} (i) = \text{tr} [ M_{x_i} \rho^{\otimes k} ] \), \( q^{(k)} (i) = \text{tr} [ M_{x_i} \sigma^{\otimes k} ] \), \( \forall \sigma \)
\[
D (\sigma||\rho) - \frac{(d-1)\log (k+1)}{k} \leq \frac{1}{k} D (q^{(k)}||p^{(k)}) \leq D (\sigma||\rho).
\]
Choosing \( k \) such that \( D (q^{(k)}||p^{(k)}) / k \geq D (\sigma||\rho) (1 - \epsilon) \) we obtain the statement of the theorem. [29].

Since we are considering \( h \to \infty \) we can in principle choose \( k \) arbitrarily large and consider the limit
\[
\lim_{k \to \infty} \frac{\log \tilde{T}_{FA}}{k} D_M (\sigma^{\otimes k}||\rho^{\otimes k}) = D (\sigma||\rho).
\]
This implies that in the asymptotic limit of large \( h \) and \( k \) we have
\[
\bar{\tau}^* \leq \frac{\log \tilde{T}_{FA}}{D (\sigma||\rho)}.
\]
\[ \rho_x = \frac{\mathcal{M}_k(\rho)}{\text{tr}(\mathcal{M}_k(\rho))} \], where \( \text{tr}(\mathcal{M}_k(\rho)) \) corresponds to the probability of measuring \( x \) given the state \( \rho \). In our setting, at each step \( k \), we get a fresh copy of \( \rho^{(k)} \), which is either \( \rho \) or \( \sigma \). We denote a restriction of a possible sequences of the outcomes to the first \( i \) elements as \( x^i \). Let \( \rho_{x^{-1}} \) be the post-measurement state of the \( k \)-th step, then we apply the \( k \)-th quantum instrument \( \mathcal{M}_k \) (possibly depending on previous records \( x^{k-1} \)) as \( \mathcal{M}_k(\rho \otimes \rho_{x^{-1}}) \), receiving a classical output \( x_k \) and a new post-measurement state \( \rho_x = \frac{\mathcal{M}_k(\rho \otimes \rho_{x^{-1}})}{\text{tr}\mathcal{M}_k(\rho \otimes \rho_{x^{-1}})} \). We denote the post-measurement states as \( \rho_x(\infty) \) if they originate from a sequence with no change point, and \( \rho_x(\nu) \) if they come from a sequence with change point \( \nu \). We denote \( p(x_k|x^{k-1}) = \text{tr}\mathcal{M}_k(\rho \otimes \rho_{x^{-1}}) \), and \( q^{(\nu)}(x_k|x^{k-1}) = \text{tr}\mathcal{M}_k(\sigma \otimes \rho_{x^{-1}}) \).

We now define the log-likelihood ratio at step \( k \) for a candidate change point \( \nu \):

\[
Z_k^{(\nu)} = \log \frac{q^{(\nu)}(x_k|x^{k-1})}{p(x_k|x^{k-1})}, \quad \lambda_n^{(\nu)} = \sum_{k=\nu+1}^{n} Z_k^{(\nu)}. \quad (17)
\]

Note that for a fixed sequence \( x^k \) we can always write \( p(x_k) = \text{tr}\mathcal{M}_k(\rho \otimes \rho_{x^{-1}}) \) for a joint measurement \( \{\mathcal{M}_k\} \) giving a sequence of outcomes \( x^k \).

While we still get a sequence of classical measurement outcomes as a result, these can now be highly correlated and the usual techniques for i.i.d. distributions do not longer apply. In the following we will make heavy use of a result initially stated by Lai [30] and reformulated in [4], which we adapt in a form which is applicable to our case, and give the proof in the SM for completeness [31].

**Theorem 3** For a change point model with log-likelihoods \( Z_k^{(\nu)} \) and \( \epsilon > 0 \), no strategies can exceed the trade-off given by \( \tilde{\tau}^* \geq (1-\epsilon)I^{-1}\log T_{FA}(1+o(1)) \), for large \( T_{FA} \), for any \( I \) that satisfies the condition

\[
\lim_{n \to \infty} \sup_{\nu \geq 0} \text{ess sup}_{\nu} P^*_\nu = 0 \quad \text{where} \quad P^*_\nu := P^{(\nu)} \left\{ \max_{1 \leq i \leq n} \lambda^{(\nu)}_{x_i} \geq I(1+\epsilon)n \right\} \quad (18)
\]

The challenging art is to determine the smallest \( I \) such that Eq. (18) holds for any \( \epsilon > 0 \) and any strategy, considering that the underlying probability distribution can be produced by the most general kind of measurement strategy. We start by getting rid of the supremum over \( \nu \). Note that all states up to position \( \nu \) will be assumed to be \( \rho \) independent of whether we consider the case with or without change. At position \( \nu + 1 \) we will therefore try to discriminate between two states \( \sigma \otimes \rho_{x^\nu} \) and \( \rho \otimes \rho_{x^\nu} \). It is now easy to see that for any \( \nu \) and any measurement on the sequence of states, there also exists a measurement in the case of \( \nu = 0 \) that results in the same probability distribution after the change point, consisting simply preparing the state \( \rho_{x^\nu} \) and then applying the original strategy. It follows that we can without loss of generality set \( \nu = 0 \) and therefore also omit the essential supremum.

We will now bound \( P^0_0 \) based on the strong converse for quantum Stein’s Lemma [24], which states that if a sequence of binary tests \( \{M^n_0, M^n_{1}\} = I \) is such that \( \text{tr}[M_{1}^{(\nu)}(\rho \otimes n_I) \leq e^{-n(D(\sigma|\rho)+\delta)} \] for some \( \delta > 0 \), then \( \lim_{n \to \infty} \text{tr}[M_{1}^{(\nu)} \sigma^{\otimes n}] = 0 \). Let’s denote the corresponding log-likelihood sum for \( \nu = 0 \) corresponding to outcome \( x^i \) as \( \lambda_{x^i} \). Define the set \( S_i := \{x^i|\lambda_{x^i} \geq nI(1+\epsilon)\} \) and \( \lambda_{x^i} < nI(1+\epsilon) \forall j \in i \).

We have the following chain of equalities,

\[
P^{(0)} \left( \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1+\epsilon) \right) = \sum_{\max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1+\epsilon)} q^{(0)}(x^n)
\]

\[
= \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} q^{(0)}(x^i) = \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} \text{tr}\mathcal{M}_x^{(\nu)} \sigma^{\otimes i}
\]

\[
= \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} \text{tr}\mathcal{M}_x^{(\nu)} \otimes \mathbb{1}^{(n-i)} \sigma^{\otimes i}. \quad (19)
\]

Defining the binary POVM \( \{\tilde{M}_i, \tilde{M}_i^\dagger = 1 - \tilde{M}_i^\dagger \} \), with \( \tilde{M}_i = \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} \mathcal{M}_x^{(\nu)} \otimes \mathbb{1}^{(n-i)} \sigma^{\otimes i} \), we get

\[
P^{(0)} \left( \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1+\epsilon) \right) \text{tr}\mathcal{M}_x^{(\nu)} \sigma^{\otimes i}.
\]

Also, since \( q^{(0)}(x^i) = e^{\lambda_{x^i}} p(x^i) \geq e^{nI(1+\epsilon)} \),

\[
\forall x^i \in S_i, \quad \text{tr}\tilde{M}_i^{\nu} p^{\otimes i} = \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} p(x^i)
\]

\[
\leq e^{-nI(1+\epsilon)} \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} q^{(0)}(x^i) = e^{-nI(1+\epsilon)} \text{tr}\tilde{M}_i^{\nu} \sigma^{\otimes i} \leq e^{-nI(1+\epsilon)}. \quad (20)
\]

By the strong converse, this means that if \( I \geq D(\sigma|\rho) \), \( \lim_{n \to \infty} P^{(0)} = \lim_{n \to \infty} \text{tr}\tilde{M}_i^{\nu} \sigma^{\otimes i} = 0 \).

The proof is extended with straightforward modifications to the case of measurement with non-discrete outcomes. This proves the optimality, \( \tilde{\tau}^* \geq \)
\[ \tilde{\tau} \geq (1 - \varepsilon) \frac{\log \tilde{T}_A}{D(\rho\|\sigma)} (1 + o(1)) \]. Even more so, optimality holds also for a collection \( S \) of possible states after the change, with \( I = \min_{\sigma \in S} D(\sigma\|\rho) \).

**Change point with channels:** One can define an analogous change point problem for quantum channels. In this case, we can leverage the achievability result for states to obtain \( \hat{\tau}^* \leq \frac{\log \tilde{T}_A}{D(\rho\|\sigma)} \), where \( D(\mathcal{M}\|\mathcal{N}) = \lim_{k \to \infty} \sup_{\rho} \frac{1}{k} D(\mathcal{M}^k(\rho)\|\mathcal{N}^k(\rho)) \) (here and in the following the input state in maximization can be any state entangled with an arbitrarily large reference system). On the other hand, we can adapt the lower bound proof using a known strong converse \([32]\), obtaining \( \hat{\tau}^* \leq \frac{\log \tilde{T}_A}{D(\rho\|\sigma)} \), where \( \tilde{D}_\infty(\mathcal{M}\|\mathcal{N}) = \lim_{k \to \infty} \sup_{\rho} \frac{1}{k} \tilde{D}_k(\mathcal{M}^k(\rho)\|\mathcal{N}^k(\rho)) \), \( \tilde{D}_k(\rho\|\sigma) = \frac{1}{\sigma} \log \text{tr} \left( \sigma^{\frac{1}{\sigma}} \rho \sigma^{\frac{1}{\sigma}} \right) \), and \( \tilde{D}_\infty(\mathcal{M}\|\mathcal{N}) = \lim_{\sigma \to 1} \tilde{D}_\infty(\mathcal{M}\|\mathcal{N}) \). The quantities in the two bounds have been conjectured to coincide \([33]\).

**Conclusions:** We showed asymptotic optimality of the QUSUM algorithm, with a tradeoff given by the relative entropy, solving the quickest change point detection problem for quantum states in the asymptotic setting. It remains unclear how to address the optimality for finite number of observations. Our results apply also to the setting when the state after the change is not known. In the asymptotic setting, it would be interesting to find achievability results in non-i.i.d. cases, especially those for which a strong converse can be found.

**ACKNOWLEDGMENTS**

This work was supported by the QuantERA grant C’MON-QSENS!, by Spanish MICINN PCI2019-11869-2 and Spanish Agencia Estatal de Investigación, project PID2019-107609GB-I00/AEI/10.13039/501100011033 and co-funded by the European Union Regional Development Fund within the ERDF Operational Program of Catalunya (project QuantumCat, ref. 001-P-001644). CH has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. H2020-MSCA-IF-2020-101025848. JC also acknowledges support from ICREA Academia award.

---

* marco.fanizza@uab.cat
† christoph.hirche@gmail.com
‡ john.calsamiglia@uab.cat

[1] M. Basseville, I. V. Nikiforov, et al., Detection of abrupt changes: theory and application, Vol. 104 (Prentice Hall Englewood Cliffs, 1993).
[2] J. Chen and A. K. Gupta, Parametric statistical change point analysis: with applications to genetics, medicine, and finance (Springer Science & Business Media, 2011).
[3] H. V. Poor and O. Hadjiliadis, Quickest detection (Cambridge University Press, 2008).
[4] A. Tartakovsky, I. Nikiforov, and M. Basseville, Sequential analysis: Hypothesis testing and change-point detection (Chapman and Hall/CRC, 2014).
[5] D. M. Hawkins, P. Qiu, and C. W. Kang, The changepoint model for statistical process control, Journal of quality technology 35, 355 (2003).
[6] T. L. Lai, Sequential changepoint detection in quality control and dynamical systems, Journal of the Royal Statistical Society: Series B (Methodological) 57, 613 (1995).
[7] J. Chen and A. K. Gupta, Testing and locating variance changepoints with application to stock prices, Journal of the American Statistical association 92, 739 (1997).
[8] D. Rosenfield, E. Zhou, F. H. Wilhelm, A. Conrad, W. T. Roth, and A. E. Meuret, Change point analysis for longitudinal physiological data: detection of cardio-respiratory changes preceding panic attacks, Biological psychology 84, 112 (2010).
[9] C. Gallagher, R. Lund, and M. Robbins, Change-point detection in climate time series with long-term trends, Journal of Climate 26, 4994 (2013).
[10] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, A review and comparison of changepoint detection techniques for climate data, Journal of applied meteorology and climatology 46, 900 (2007).
[11] S. Niekum, S. Osentoski, C. G. Atkeson, and A. G. Barto, Online bayesian changepoint detection for articulated motion models, in 2015 IEEE International Conference on Robotics and Automation (ICRA) (IEEE, 2015) pp. 1468–1475.
[12] A. Ranganathan, Pliss: labeling places using online changepoint detection, Autonomous Robots 32, 351 (2012).
[13] D. Akimoto and M. Hayashi, Discrimination of the change point in a quantum setting, Phys. Rev. A 83, 052328 (2011).
[14] G. Sentís, E. Bagan, J. Calsamiglia, G. Chiribella, and R. Munoz-Tapia, Quantum change point, Physical Review Letters 117, 150502 (2016).
[15] G. Sentís, J. Calsamiglia, and R. Munoz-Tapia, Exact identification of a quantum change point, Physical Review Letters 119, 140506 (2017).
[16] G. Sentís, E. Martínez-Vargas, and R. Muñoz-Tapia, Online strategies for exactly identifying a quantum change point, Physical Review A 98, 052305 (2018).

[17] E. M. Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz-Tapia, and J. Calsamiglia, Quantum sequential hypothesis testing, Physical Review Letters 126, 180502 (2021).

[18] Y. Li, V. Y. Tan, and M. Tomamichel, Optimal adaptive strategies for sequential quantum hypothesis testing, Communications in Mathematical Physics 392, 993 (2022).

[19] Y. Li, C. Hirche, and M. Tomamichel, Sequential quantum channel discrimination, in 2022 IEEE International Symposium on Information Theory (ISIT) (2022).

[20] E. S. Page, Continuous inspection schemes, Biometrika 41, 100 (1954).

[21] G. Lorden et al., Procedures for reacting to a change in distribution, The Annals of Mathematical Statistics 42, 1897 (1971).

[22] G. V. Moustakides et al., Optimal stopping times for detecting changes in distributions, The Annals of Statistics 14, 1379 (1986).

[23] Y. Ritov, Decision theoretic optimality of the cusum procedure, The Annals of Statistics 14, 1464 (1990).

[24] T. Ogawa and H. Nagaoka, Strong converse and stein’s lemma in quantum hypothesis testing, IEEE Transactions on Information Theory 46, 2428 (2000).

[25] A. Wald, Sequential analysis (Courier Corporation, 2004).

[26] D. Petz, A variational expression for the relative entropy, Communications in Mathematical Physics 114, 345 (1988).

[27] M. Hayashi, Asymptotics of quantum relative entropy from a representation theoretical viewpoint, Journal of Physics A: Mathematical and General 34, 3413 (2001).

[28] M. Berta, O. Fawzi, and M. Tomamichel, On variational expressions for quantum relative entropies, Letters in Mathematical Physics 107, 2239 (2017).

[29] Notice also that an upper bound on the sufficient size of the blocks to get the achievability in the theorem is $k = O(d/(\epsilon D(\sigma\|\rho)))$ (hiding logarithmic factors), with an improved dependence on the dimension with respect to optimal tomography, which requires $\Theta(d^2)$ copies \cite{34,35}.

[30] T. L. Lai, Information bounds and quick detection of parameter changes in stochastic systems, IEEE Transactions on Information Theory 44, 2017 (1998).

[31] Ref. \cite{30} shows that the proof applies also when the probability in presence of a change point, conditioned on past events, can depend on the change point location. The statement of the theorem in \cite{30} has an extra condition about the convergence of the log-likelihood which is not needed for our purposes and expresses the tradeoff only as a limit for large $\log T_{FA}$ and $\epsilon \to 0$, while we state it with finite $\epsilon$.

[32] H. Fawzi and O. Fawzi, Defining quantum divergences via convex optimization, Quantum 5, 387 (2021).

[33] K. Fang, G. Gour, and X. Wang, Towards the ultimate limits of quantum channel discrimination (2021).

[34] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Transactions on Information Theory 63, 5628 (2017).

[35] R. O’Donnell and J. Wright, Efficient quantum tomography, in Proceedings of the forty-eighth annual ACM symposium on Theory of Computing (2016) pp. 899-912.

[36] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, Amortized channel divergence for asymptotic quantum channel discrimination, Letters in Mathematical Physics 110, 2277 (2020).

[37] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability for quantum channels, Physical Review Research 1, 033169 (2019).

[38] K. Fang, O. Fawzi, R. Renner, and D. Sutter, Chain rule for the quantum relative entropy, Physical Review Letters 124, 100501 (2020).
This figure illustrates how the CUSUM algorithm works in a (classical) setting where samples $x^k = \{x_1, x_2, \ldots, x_k\}$ are obtained from a Bernoulli trial (coin toss) with a bias $p = 1/5$ that at time $k = \nu = 10^4$ changes to bias $q = 1/4$. Dark green and light stochastic curves show the random walk exhibited by the log-likelihood ratio $Z^k_1$ for different measurement sequences $x^k_1$. CUSUM algorithm keeps track of $Z^k_1$, starting at $Z^0_1 = 0$ and updating its value at every time step by $Z^k_1 = Z^{k-1}_1 + Z(x_k)$ depending on the (random) outcome $x_k$. The algorithm stops and signals a change-point detection as soon as $\max_{j < k} Z^k_j \geq h$, i.e. as soon as the log-likelihood at the current time exhibits a net increase $Z^k_1 - Z^j_1 = Z^k_j$ larger or equal than $h$ with respect to some point $j$ in the past. This threshold is the only free parameter of the algorithm and regulates the trade-off between the mean detection delay and the false alarm rate. Two scenarios are show-cased: (1) a large threshold value ($h = 22$) reduces the chances of false alarms at the expense of long detection times; (2) a low threshold value ($h' = 6$) can detect the change point with a small delay, but has a high risk of producing false alarms. The orange thick line shows the average trend (over many trajectories), which is given by $E(Z^k_1) = -kD(p||q)$ before the change point ($k \leq \nu$), and at the change point the slope changes abruptly to $D(q||p)$. The detection delay $\tau$, i.e. the time required arrive to the threshold $h$ after the change has happened is given by $\tau \sim h/D(q||p)$. Note that at a particular time $k$, the likelihood corresponding to a change point at $j$ is $P^{(\nu=j)}(x^k_1) = \prod_{i=1}^{\nu} p(x_i) \prod_{i=\nu+1}^{k} q(x_i)$ and that corresponding to no change at all is
\( P(\infty)(x_k^1) = \prod_{i=1}^k p(x_i) \). It is then immediate to check that the log-likelihood ratio \( \log \frac{P(\nu=j)(x_k^1)}{P(\infty)(x_k^1)} \) is given by \( Z_j^k = \sum_{i=j}^k Z(x_i) \). Therefore, the \( j \) that attains the maximum \( \max_{j<k} Z_j^k \) is precisely the point in the past where the change has most likely happened, and the stopping condition is equivalent to fixing a minimum likelihood ratio for such change point to be accepted.

**QUSUM for general change point**

If the change point does not happen at a multiple of \( k \), the first inequality of Eq. (9) still holds, while we can replace the second one with

\[
\bar{\tau}^* = k \sup_{\nu} \sup_{\nu \geq 0} \mathbb{E}_\nu (T^* - \nu | T^* > \nu, X_1, ..., X_\nu) \leq k \sup_{\nu} \sup_{\nu \geq 0} \mathbb{E}_0 (T_{\nu/k} + 2 - \nu | T^* > \nu, X_1, ..., X_\nu) = k \mathbb{E}_0 (T_2) = k(1 + \mathbb{E}_0 (T_1)),
\]

(S1)

where the inequality comes from the definition of \( T^* \) (Eq. (8) and the equalities from the fact that \( T_{\nu/k} + 2 \) does not depend on the outcomes before \( t = \nu + 1 \). Therefore, we get again Eq. (15).

We also discuss how we can generalize this achievability result in the case where we have that the state after the change point is unknown and belonging to a family of states \( S \). In this case, we consider the random variable \( S_i = \sup_{\sigma \in S} Z_i^k(\sigma) \), where \( Z_i^k(\sigma) \) is the one from Eq. (5), and we generalize the stopping time accordingly, stopping as soon as \( S_i \geq h \). In particular we consider the parallel stopping times \( T_j^{(\sigma)} \), \( j \geq 1 \), \( \sigma \in S \), which are defined as in the main text but depend on the different \( \sigma \), and define \( T_j = \min_{\sigma \in S} T_j^{(\sigma)} \)

Since the POVM \( M_{x_i}^{(k)} \) for which Eq. (14) holds does not depend on \( \sigma \), we get asymptotically

\[
\bar{\tau}^* \leq h \frac{1}{\min_{\sigma \in S} D(\sigma || \rho)}.
\]

(S2)

On the other hand, using that \( I_{T_1 < \infty} = \cup_{\sigma \in S} I_{T_1^{(\sigma)} < \infty} \) and the union bound, Eq. (10) becomes

\[
P_\infty (T_1 < \infty) \leq \sum_{\sigma \in S} P_{\infty} (T_{1}^{(\sigma)} < \infty) = \sum_{\sigma \in S} \mathbb{E}_0 \left[ \frac{p_{\sigma}^{(\sigma)}}{q_{\sigma}^{(\sigma)}} I_{T_{1}^{(\sigma)} < \infty} \right] = \sum_{\sigma \in S} \mathbb{E}_0 \left[ e^{-Z_{T_{1}^{(\sigma)}}(\sigma)} I_{T_{1}^{(\sigma)} < \infty} \right] = \sum_{\sigma \in S} \mathbb{E}_0 \left[ e^{-h - x(\sigma)} I_{T_{1}^{(\sigma)} < \infty} \right] \leq |S| e^{-h} =: \alpha,
\]

(S3)

meaning that we can take \( h = \log \bar{T}_{FA} + \log |S| \), obtaining the desired statement as \( \bar{T}_{FA} \to \infty \).
Proof of Theorem 3

We restate the theorem and prove it.

**Theorem 4** For a change point model with log-likelihoods \( Z^{(\nu)} \) and \( \epsilon > 0 \), no strategies can exceed the trade-off given by 
\[
\tau^* \geq (1 - \epsilon) T \left(1 + o(1)\right),
\]
for large \( T \), for any \( I \) that satisfies the condition
\[
\lim_{n \to \infty} \sup_{\nu \geq 0} \text{ess sup}^* P^*_\nu = 0 \quad \text{(S4)}
\]

**Proof.** First notice that for any \( \nu \geq 0 \)
\[
\tau^* \geq \tau \geq E_\nu(T - \nu|T > \nu),
\]
and by Markov inequality
\[
E_\nu(T - \nu|T > \nu) \geq mP_\nu(T - \nu \geq m|T > \nu) \quad \text{(S6)}
\]

We start from observing that, for the event \( E_\nu := \nu < T < \nu + m \) we have
\[
P_\infty(E_\nu) = E_\nu[I_{E_\nu} e^{-\lambda_T^{(\nu)}}],
\]
where
\[
\lambda^{(\nu)}_T = \log \frac{\text{tr}[M_{x_1, \ldots, x_T}(\rho \otimes^{\nu-1} \sigma \otimes^{T-\nu})]}{\text{tr}[M_{x_1, \ldots, x_T} \rho \otimes^{T}]}, \quad \text{if } T > \nu,
\]
\[
\lambda_{\nu}^{(\nu)} = 0, \quad \text{if } T \leq \nu,
\]
and the change of measure is justified since \( D_{max}(\sigma||\rho) < \infty \). Then we have
\[
E_\nu[I_{E_\nu} e^{-\lambda_T^{(\nu)}}] \geq E_\nu[I_{E_\nu, \lambda^{(\nu)}_T < c} e^{-\lambda_T^{(\nu)}}] \geq e^{-c} P_\nu(E_\nu, \lambda^{(\nu)}_T < c) \geq e^{-c} P_\nu(E_\nu, \max_{\nu < n < \nu + m} \lambda^{(\nu)}_n < c) \geq e^{-c} P_\nu(E_\nu) - e^{-c} P_\nu \left( \max_{\nu < n < \nu + m} \lambda^{(\nu)}_n \geq c \right),
\]
\[
\text{(S9)}
\]

where the last inequality comes from the fact that for any two events \( A \) and \( B \), \( P(A \cap B) \geq P(A) - P(B^c) \). From this chain of inequalities and Eq. (S7) we get
\[
P_\nu(\nu < T < \nu + m) \leq e^c P_\infty(T < \nu + m) + P_\nu \left( \max_{\nu < n < \nu + m} \lambda^{(\nu)}_n \geq c \right)
\]
\[
\text{(S10)}
\]
This inequality is valid also if we substitute probabilities with conditional probabilities with respect to the event \( T > \nu \). In order to see this, we observe that \( P_\infty(T > \nu) = 1 - P_\infty(T \leq \nu) = 1 - P_\nu(T \leq \nu) = P_\nu(T > \nu) \),
since the probability of stopping at \( \nu \) depends only on measurement outcomes up to \( \nu \), whose distribution is identical for the cases where the change point is at \( \nu \) or at infinity. Moreover, \( P_\infty(T > \nu) > 0 \) if we require \( \bar{T}_{\text{FA}} \) sufficiently large. We can then divide both members of the inequality \((S10)\) by \( P_\infty(T > \nu) \) and get by Bayes’ rule

\[
P_\nu(T < \nu + m|T > \nu) \leq e^c P_\infty(T < \nu + m|T > \nu) + P_\nu \left( \max_{\nu \leq n < \nu + m} \lambda^{(\nu)}_n \geq c|T > \nu \right) \quad (S11)
\]

We next show that the RHS, and hence the conditional probability go to zero as we increase \( m \). We first concentrate on the first term of the right hand side. For this purpose, we write

\[
\bar{T}_{\text{FA}} = \mathbb{E}_\infty[T] = \sum_{j=0}^{\infty} P_\infty(T > j) = \sum_{i=0}^{m-1} \sum_{k=0}^{\infty} P_\infty(T > i + km)
\]

\[
= \sum_{i=0}^{m-1} \sum_{k=0}^{\infty} P_\infty(T > i) P_\infty(T > i + km|T > i) \quad (S12)
\]

Defining

\[
\Delta := \sup_{i \geq 0} P_\infty(T > i + m|T > i),
\]

we have

\[
P_\infty(T > i + km|T > i) = P_\infty(T > i + (k-1)m + m|T > i)
\]

\[
= P_\infty(T > i + (k-1)m + m|T > i) P(T > i + (k-1)m|T > i)
\]

\[
= \prod_{j=0}^{k-1} P_\infty(T > i + jm + m|T > i + jm) \leq \Delta^k. \quad (S14)
\]

Therefore

\[
\bar{T}_{\text{FA}} \leq \sum_{i=0}^{m-1} \sum_{k=0}^{\infty} P_\infty(T > i) \Delta^k = \frac{1}{1 - \Delta} \sum_{i=0}^{m-1} P_\infty(T > i) \leq \frac{m}{1 - \Delta}
\]

From which we obtain \( \Delta \geq 1 - m/\bar{T}_{\text{FA}} \). If the superior in the definition of \( \Delta \) is actually a maximum, then we can choose \( \nu \) such that

\[
P_\infty(T \leq \nu + m|T > \nu) = 1 - P_\infty(T > \nu + m|T > \nu) = 1 - \Delta \leq m/\bar{T}_{\text{FA}}.
\]

Otherwise, the last inequality in Eq. \((S14)\) is strict and we get \( \Delta > 1 - m/\bar{T}_{\text{FA}} \). Since \( \Delta > 1 - m/\bar{T}_{\text{FA}} \) implies that there exists \( \nu \) such that \( P_\infty(T > \nu + m|T > \nu) > 1 - m/\bar{T}_{\text{FA}} \), and for the same \( \nu \) it holds that \( P_\infty(T > \nu + m|T > \nu) \geq 1 - m/\bar{T}_{\text{FA}} \).
We now take \( m = (1 - \epsilon) I^{-1} \log \bar{T}_{FA} \), and \( c = (1 + \epsilon) m I = (1 - \epsilon^2) \log \bar{T}_{FA} \) and upper-bound the first term in RHS of inequality (S11) by

\[
e^{-c} P_{\infty}(T < \nu + m|T > \nu) \leq (\bar{T}_{FA})^{1-\epsilon^2} (1 - \epsilon) I^{-1} \log \bar{T}_{FA} \leq (1 - \epsilon) I^{-1} \log \bar{T}_{FA} \quad (\bar{T}_F)^{\epsilon} \quad (S15)
\]

Putting all together, we conclude that the LHS of (S11) is upper-bounded by a quantity that goes to zero:

\[
P_{\nu}(T < \nu + m|T > \nu) \leq \frac{(1 - \epsilon) I^{-1} \log \bar{T}_{FA}}{T_F^{\epsilon}}
\]

\[+ P_{\nu} \left( \max_{\nu \leq n < \nu + m} \lambda_n^{(\nu)} \geq (1 + \epsilon) m I|T > \nu \right) \rightarrow 0, \quad (m \rightarrow \infty) \quad (S16)
\]

The first term goes to zero since \((\log \bar{T}_{FA})/T_{FA}^{\epsilon}\) goes to zero as \( m \) goes to infinity by definition of \( m \), and the second term goes to zero by hypothesis of the theorem.

Recalling Eq. (S6), this means that there exists \( \nu \) such that

\[
\mathbb{E}_{\nu}(T - \nu|T > \nu) \geq (1 - \epsilon) \frac{\log \bar{T}_{FA}}{I}(1 + o(1)), \quad (S17)
\]

and therefore, by Eq. (S5)

\[
\bar{\tau}^* \geq \bar{\tau} \geq (1 - \epsilon) \frac{\log \bar{T}_{FA}}{I}(1 + o(1)), \quad (S18)
\]

\[
\text{Change point with channels}
\]

In this section we will extend the previous results to quantum channels. From now on we consider a process that is modeled by a sequences of quantum channel \( N^{(k)}_{A_k \rightarrow B_k} \) which can be either \( M_{A_k \rightarrow B_k} \) (if \( k > \nu \)) or \( N_{A_k \rightarrow B_k} \) (if \( k \leq \nu \)). Let us first consider a modification of QUSUM to show an achievable rate. For every instance of the process we input a state \( \rho \) into the channel, leading to an output state \( \rho_i = N^{(k)}(\rho) \). We can now simply run the previously used state version of QUSUM on these output states and generalizing Equation (13) we get

\[
\bar{\tau}^* \leq \frac{(\log \bar{T}_{FA}/k)}{D_M(M||N)}, \quad (S19)
\]

where

\[
D_M(M||N) = \sup_{\rho} D_M(M(\rho)||N(\rho)), \quad (S20)
\]
by optimizing over input states $\rho$. The same way collective measurements lead to an advantage in the state case, also entangled input states can improve the performance. Using joint states on $k$ systems combined with collective measurements on the same, we get

$$\tilde{\tau}^* \leq \frac{\log \tilde{T}_{FA}}{kD_M(M^\otimes k \Vert N^\otimes k)}.$$  \hfill (S21)

Finally we observe that

$$\lim_{k \to \infty} \frac{1}{k} D_M(M^\otimes k \Vert N^\otimes k) = D^\infty(M \Vert N),$$  \hfill (S22)

where

$$D^\infty(M \Vert N) = \lim_{k \to \infty} \sup_{\rho} \frac{1}{k} D(M^\otimes k(\rho) \Vert N^\otimes k(\rho)).$$  \hfill (S23)

This can be seen as follows. The $\leq$ direction is a simple consequence of data processing. For the $\geq$ direction fix $k = ml$ and consider

$$\frac{1}{ml} D_M(M^\otimes ml \Vert N^\otimes m) \geq \sup_{\rho} \frac{1}{ml} D(M^\otimes m(\rho) \Vert N^\otimes m(\rho)).$$  \hfill (S24)

Taking first the limit $l \to \infty$ gives us the relative entropy, then taking $m \to \infty$ gives the desired regularization. We remark here that the regularized channel relative entropy is known to be the optimal achievable rate in the Stein’s Lemma for quantum channels [36–38] and that the regularization is generally necessary [38]. Overall, this leads to the asymptotic achievable tradeoff

$$\tilde{\tau}^* \leq \frac{\log \tilde{T}_{FA}}{D^\infty(M \Vert N)}.$$  \hfill (S25)

For the converse bound, we again mostly follow the state case, with the difference that the set of allowed strategies is larger. In the case of channels, any sequential strategy can be described as follows: first, we are allowed to prepare any initial state $\rho^{(0)}_{LA_1}$ (possibly infinite dimensional); assume that at step $k$ we have a access to a state $\rho_{LA_k}^{k-1}$, first apply the channel $N_{A_k \rightarrow B_k}$ to $\rho_{LA_k}^{k-1}$, and then we apply the instrument $E_{LA_k}^{x_k}$, corresponding to CP maps $E_{LB_k \rightarrow LA_{k+1}}$ (where we drop the dependence on $x^{k-1}$ for readability); the state conditioned on the new outcome $x_k$ is then denoted as $\rho_{LA_{k+1}}^{x_k}$. The probability of getting outcome $x_k$ is then recursively defined as

$$p(x_k | x^{k-1}) = \text{tr}_{LA_{k+1}} [E_{LB_k \rightarrow LA_{k+1}} \circ N_{A_k \rightarrow B_k} [\rho_{LA_k}^{k-1}]],$$  \hfill (S26)

for $k \leq \nu$ or if there is no change point, and

$$q^{(\nu)}(x_k | x^{k-1}) = \text{tr}_{LA_{k+1}} [E_{LB_k \rightarrow LA_{k+1}} \circ M_{A_k \rightarrow B_k}^{(k)} [\rho_{LA_k}^{k-1}]],$$  \hfill (S27)

for $k = \nu$. The probability of getting outcome $x_k$ is then recursively defined as

$$p(x_k | x^{k-1}) = \text{tr}_{LA_{k+1}} [E_{LB_k \rightarrow LA_{k+1}} \circ N_{A_k \rightarrow B_k} [\rho_{LA_k}^{k-1}]],$$  \hfill (S26)

for $k \leq \nu$ or if there is no change point, and

$$q^{(\nu)}(x_k | x^{k-1}) = \text{tr}_{LA_{k+1}} [E_{LB_k \rightarrow LA_{k+1}} \circ M_{A_k \rightarrow B_k}^{(k)} [\rho_{LA_k}^{k-1}]],$$  \hfill (S27)
if the change point is at \( \nu \) and \( k > \nu \).

Using the same notation as in the state case, we get

\[
P^{(0)} \left( \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1 + \epsilon) \right) = \sum_{x^n; \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1 + \epsilon)} q^{(0)}(x^n) \tag{S28}
\]

\[
= \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} q^{(0)}(x^i) = \sum_{1 \leq i \leq n} \sum_{x^i \in S_i} \tilde{q}^{(0)}(x^i) \tag{S29}
\]

where \( \tilde{q}^{(0)}(x^i) \) are the outcomes probabilities corresponding to a modified strategy, where upon obtaining an output \( x^i \in S_i \), a fixed state is send through the next channels and the output is discarded. Clearly, \( \tilde{q}^{(0)}(x^i) \) can still be defined recursively as \( q^{(0)}(x^i) \).

At fixed \( n \), the test obtained by accepting if the strategy finds \( x^i \in S_i \) at some step \( i \) is a valid binary test, which has probability of success when performed on \( n \) accesses to \( \mathcal{M} \) equal to \( P^{(0)} \left( \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1 + \epsilon) \right) \).

In the same way we define \( \tilde{p}(x^i) \), obtained applying the same strategy to accesses to \( \mathcal{N} \). By construction, we have \( \tilde{q}^{(0)}(x^i) \geq e^{nI(1+\epsilon)} \tilde{p}(x^i) \). A strong converse bound resulting from [32] applies to this strategy and the pair \( (\mathcal{M},\mathcal{N}) \), and the same argument of the state case gives that if \( I \geq \hat{D}_1^\infty(\mathcal{M}\|\mathcal{N}) \), \( \lim_{n \to \infty} P^{(0)} \left( \max_{1 \leq i \leq n} \lambda_{x^i} \geq nI(1 + \epsilon) \right) = 0 \).

However, note that we do not know whether \( \hat{D}_1^\infty(\mathcal{M}\|\mathcal{N}) = D^\infty(\mathcal{M}\|\mathcal{N}) \), but we conjecture that they are equal. Combining the above with the QUSUM optimality condition we have that, asymptotically

\[
\tilde{T}^* \geq \frac{\log \tilde{T}_{FA}}{\hat{D}_1^\infty(\mathcal{M}\|\mathcal{N})}. \tag{S30}
\]

Our achievability and optimality results give close bounds on the asymptotic performance and they do indeed match if the aforementioned conjecture holds.