Lecture 5

Min-relative entropy, conditional max-entropy, and hypothesis-testing relative entropy

In this lecture we will discuss a few additional generalized entropy measures, namely the min-relative entropy, the conditional max-entropy, and the hypothesis-testing relative entropy. We will discuss various properties of these quantities, and relate them to the other entropic quantities we have previously discussed.

5.1 Min-relative entropy

We will begin with the min-relative entropy.

Definition 5.1 (Quantum min-relative entropy)**.** Let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator and let $Q \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator, for $\mathcal{X}$ a complex Euclidean space. The *quantum min-relative entropy* (or min-relative entropy, for short) of ρ with respect to Q is defined as

$$\mathrm{D}_{\min}(\rho\|Q) = -\log\bigl(\mathrm{F}(\rho, Q)^2\bigr), \tag{5.1}$$

where

$$\mathrm{F}(\rho, Q) = \Bigl\| \sqrt{\rho}\sqrt{Q} \Bigr\|_1 \tag{5.2}$$

is the fidelity between ρ and Q.

Remark 5.2. In the case that $\mathrm{F}(\rho, Q) = 0$, which is equivalent to $\mathrm{im}(\rho) \perp \mathrm{im}(Q)$, one is to interpret that $\mathrm{D}_{\min}(\rho\|Q) = \infty$.

Elementary observations

Here are a couple of relevant properties of the min-relative entropy that follow directly from known properties of the fidelity function.

1. A variant of Klein's inequality holds for the min-relative entropy. That is, if $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ are density operators, then $\mathrm{D}_{\min}(\rho \| \sigma) \geq 0$, with equality if and only if $\rho = \sigma$.
2. The min-relative entropy is monotonic with respect to the action of channels. That is, for every choice of $\rho \in \mathrm{D}(\mathcal{X})$, $Q \in \mathrm{Pos}(\mathcal{X})$, and $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, it is the case that
$$\mathrm{D}_{\min}(\Phi(\rho) \| \Phi(Q)) \leq \mathrm{D}_{\min}(\rho \| Q).$$
This is true, in fact, for all positive and trace-preserving maps $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$.

One can identify additional properties of the min-relative entropy through its very direct connection to the fidelity function, which we know to have many interesting and remarkable properties.

Relationship to the quantum relative entropy

The following theorem reveals that the min-relative entropy is upper-bounded by the ordinary quantum relative entropy.

Theorem 5.3. *Let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator and let $Q \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator, for $\mathcal{X}$ a complex Euclidean space. It is the case that*

$$\mathrm{D}_{\min}(\rho \| Q) \leq \mathrm{D}(\rho \| Q). \tag{5.3}$$

Proof. The theorem is trivial in the case $\mathrm{im}(\rho) \not\subseteq \mathrm{im}(Q)$, as the right-hand side of (5.3) is infinite in this case, so the remainder of the proof is focused on the case $\mathrm{im}(\rho) \subseteq \mathrm{im}(Q)$. There is no loss of generality in assuming that Q is positive definite in this case, as the values of $\mathrm{D}_{\min}(\rho \| Q)$ and $\mathrm{D}(\rho \| Q)$ then do not change if $\mathcal{X}$ is replaced by $\mathrm{im}(Q)$.

Define a function $\phi : (-1, 1) \rightarrow \mathbb{R}$ as

$$\phi(\alpha) = -\ln \mathrm{Tr}(\rho^{1-\alpha} Q^{\alpha}). \tag{5.4}$$

We are using the natural logarithm because it will simplify the calculus that will soon be considered. It is not really important to the proof that this function is defined on the entire interval $(-1, 1)$, we only require that the function is defined on the interval $[0, 1/2]$ and is differentiable at $\alpha = 0$. But, in any case, ϕ is differentiable at every point $\alpha \in (-1, 1)$, with the derivative being given by

$$\phi'(\alpha) = \frac{\operatorname{Tr}\bigl(\rho^{1-\alpha} Q^{\alpha} (\ln(\rho) - \ln(Q))\bigr)}{\operatorname{Tr}\bigl(\rho^{1-\alpha} Q^{\alpha}\bigr)}. \tag{5.5}$$

Notice in particular that

$$\phi'(0) = \operatorname{Tr}(\rho \ln(\rho)) - \operatorname{Tr}(\rho \ln(Q)) = \frac{1}{\log(e)} \mathrm{D}(\rho \| Q). \tag{5.6}$$

We also observe that

$$\phi(1/2) = -\ln \operatorname{Tr}\Bigl(\sqrt{\rho}\sqrt{Q}\Bigr) \geq \frac{1}{2\log(e)} \mathrm{D}_{\min}(\rho \| Q) \tag{5.7}$$

with the inequality following from

$$\mathrm{F}(\rho, Q) = \Bigl\| \sqrt{\rho}\sqrt{Q} \Bigr\|_1 \geq \operatorname{Tr}\Bigl(\sqrt{\rho}\sqrt{Q}\Bigr). \tag{5.8}$$

Finally, noting that $\phi(0) = 0$, we see that the theorem will follow from a demonstration that

$$\phi'(0) \geq \frac{\phi(1/2) - \phi(0)}{1/2}. \tag{5.9}$$

This in turn will follow from a demonstration that ϕ is a concave function.

To prove that ϕ is concave, it suffices to compute its second derivative and observe that its value is non-positive. To make this as simple as possible, and to avoid a messy calculation, let us use the spectral theorem to write

$$\rho = \sum_{a \in \Sigma} p(a) x_a x_a^* \quad \text{and} \quad Q = \sum_{b \in \Gamma} q(b) y_b y_b^* \tag{5.10}$$

for alphabets Σ and Γ, orthonormal sets $\{x_a : a \in \Sigma\}$ and $\{y_b : b \in \Gamma\}$, a probability vector $p \in \mathcal{P}(\Sigma)$, and $q \in (0, \infty)^{\Gamma}$ being a vector of positive real numbers. Let us also define a function

$$r_{a,b}(\alpha) = \frac{|\langle x_a, y_b \rangle|^2 p(a)^{1-\alpha} q(b)^{\alpha}}{\operatorname{Tr}(\rho^{1-\alpha} Q^{\alpha})} \tag{5.11}$$

for every $(a, b) \in \Sigma \times \Gamma$ and $\alpha \in (-1, 1)$, so that

$$\phi'(\alpha) = \sum_{(a,b) \in \Sigma \times \Gamma} r_{a,b}(\alpha) \bigl(\ln(p(a)) - \ln(q(b))\bigr). \tag{5.12}$$

We may then express the second derivative of ϕ as

$$\phi''(\alpha) = \Bigg(\sum_{(a,b)\in\Sigma\times\Gamma} r_{a,b}(\alpha) \big(\ln(p(a)) - \ln(q(b))\big) \Bigg)^2 - \sum_{(a,b)\in\Sigma\times\Gamma} r_{a,b}(\alpha) \big(\ln(p(a) - \ln(q(b))\big)^2. \tag{5.13}$$

Observing that $r_{a,b}(\alpha) \geq 0$ and

$$\sum_{(a,b)\in\Sigma\times\Gamma} r_{a,b}(\alpha) = 1 \tag{5.14}$$

for every $\alpha \in (-1, 1)$, we find that $\phi''(\alpha)$ is non-positive by Jensen's inequality, which completes the proof. □

5.2 Conditional max-entropy

Next we will discuss the conditional max-entropy, which is defined through the min-relative entropy in precisely the same way that the conditional min-entropy is defined through the max-relative entropy.

Definition 5.4 (Conditional max-entropy)**.** Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of a pair of registers (X, Y). The *conditional max-entropy* of X given Y for the state ρ is defined as

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = - \inf_{\sigma\in\mathrm{D}(\mathcal{Y})} \mathrm{D}_{\min}(\rho \| \mathbb{1}_{\mathcal{X}} \otimes \sigma). \tag{5.15}$$

Equivalently,

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = \sup_{\sigma\in\mathrm{D}(\mathcal{Y})} \log\big(\mathrm{F}(\rho, \mathbb{1}_{\mathcal{X}} \otimes \sigma)^2\big). \tag{5.16}$$

We obtain from Theorem 5.3 that $\mathrm{H}(\mathsf{X}|\mathsf{Y})_\rho \leq \mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho$, and so

$$\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})_\rho \leq \mathrm{H}(\mathsf{X}|\mathsf{Y})_\rho \leq \mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho. \tag{5.17}$$

Writing $n = \dim(\mathcal{X})$ and $\omega = \mathbb{1}_{\mathcal{X}}/n$, we see from the definition of the conditional max-entropy that

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = \log(n) + \sup_{\sigma\in\mathrm{D}(\mathcal{Y})} \log\big(\mathrm{F}(\rho, \omega \otimes \sigma)^2\big). \tag{5.18}$$

Semidefinite program for conditional max-entropy

One way to compute the value $\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho$ is to use the semidefinite program for the fidelity function that was discussed in CS 766/QIC 820, obtaining the following semidefinite program.

Optimization Problem 5.5 (SDP for conditional max-entropy)

Primal problem

$$\begin{aligned} \text{maximize:} \quad & \frac{1}{2}\operatorname{Tr}(X) + \frac{1}{2}\operatorname{Tr}(X^*) \\ \text{subject to:} \quad & \begin{pmatrix} \rho & X \\ X^* & \mathbb{1}_{\mathcal{X}} \otimes \sigma \end{pmatrix} \geq 0, \\ & X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y}), \\ & \sigma \in \mathrm{D}(\mathcal{Y}) \end{aligned}$$

Dual problem

$$\begin{aligned} \text{minimize:} \quad & \frac{1}{2}\langle \rho, Y \rangle + \frac{1}{2}\lambda_1(\operatorname{Tr}_{\mathcal{X}}(Z)) \\ \text{subject to:} \quad & \begin{pmatrix} Y & -\mathbb{1}_{\mathcal{X}\otimes\mathcal{Y}} \\ -\mathbb{1}_{\mathcal{X}\otimes\mathcal{Y}} & Z \end{pmatrix} \geq 0, \\ & Y, Z \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}). \end{aligned}$$

It is the case that

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = 2\log(\alpha), \tag{5.19}$$

for

$$\alpha = \sup_{\sigma \in \mathrm{D}(\mathcal{Y})} \mathrm{F}(\rho, \mathbb{1}_{\mathcal{X}} \otimes \sigma) \tag{5.20}$$

being the optimal value of this semidefinite program.

Remark 5.6. The dual problem may be simplified to obtain the expression

$$\alpha = \inf_{Z>0} \left(\frac{\langle \rho, Z \rangle}{2} + \frac{\| \operatorname{Tr}_{\mathcal{X}}(Z^{-1}) \|}{2} \right) \tag{5.21}$$

for the optimal value of Optimization Problem 5.5. Using the arithmetic-geometric mean inequality, one may conclude that

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = \inf_{Z>0} \big(\log\langle \rho, Z \rangle + \log\| \operatorname{Tr}_{\mathcal{X}}(Z^{-1}) \| \big) \tag{5.22}$$

Examples

We may again consider a few examples of classes of states, to gain some intuition on the conditional max-entropy.

Example 5.7. For any choice of $\sigma \in \mathrm{D}(\mathcal{X})$ and $\xi \in \mathrm{D}(\mathcal{Y})$, it is the case that

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_{\sigma\otimes\xi} = 2\log \operatorname{Tr}\sqrt{\sigma}. \tag{5.23}$$

Example 5.8. Calculating the conditional max-entropy $\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho$ for a classical-quantum state ρ of the form

$$\rho = \sum_{k=1}^{n} p_k |k\rangle\langle k| \otimes \xi_k \tag{5.24}$$

yields

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = \log \sup_{\sigma\in\mathrm{D}(\mathcal{Y})} \left(\sum_{k=1}^{n} \sqrt{p_k}\,\mathrm{F}(\xi_k,\sigma) \right)^2. \tag{5.25}$$

Example 5.9. Let $n = \dim(\mathcal{X})$, let

$$\tau = \frac{1}{n} \sum_{a,b=1}^{n} |a\rangle\langle b| \otimes |a\rangle\langle b| \tag{5.26}$$

and suppose that τ can be recovered perfectly by applying a channel locally to Y for the state $\rho \in \mathrm{D}(\mathcal{X}\otimes\mathcal{Y})$. This is equivalent to ρ taking the form

$$\rho = (\mathbb{1}_{\mathcal{X}} \otimes V)(\tau\otimes\xi)(\mathbb{1}_{\mathcal{X}} \otimes V)^* \tag{5.27}$$

for some choice of a density operator $\xi \in \mathrm{D}(\mathcal{Z})$ and an isometry $V \in \mathrm{U}(\mathcal{X}\otimes\mathcal{Z},\mathcal{Y})$. Then we have

$$\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_\rho = -\log(n), \tag{5.28}$$

just like the conditional min-entropy and conditional quantum entropy.

A relationship between conditional min- and max-entropy

We will now prove a fundamental relationship between conditional min- and max-entropy, which is stated in the following theorem.

Theorem 5.10. *Let* X, Y, *and* Z *be registers and assume the triple* $(\mathsf{X},\mathsf{Y},\mathsf{Z})$ *is in a pure state* uu^*, *for* $u \in \mathcal{X}\otimes\mathcal{Y}\otimes\mathcal{Z}$ *a unit vector. It is the case that*

$$\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y}) + \mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z}) = 0. \tag{5.29}$$

Proof. First let us prove

$$2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})} \leq 2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})}. \tag{5.30}$$

Let

$$\rho = \mathrm{Tr}_{\mathcal{Z}}(uu^*) \tag{5.31}$$

and choose $\Phi \in \mathrm{C}(\mathcal{Y},\mathcal{X})$ to be a channel for which

$$2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})} = \mathrm{F}\big((\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\rho), \mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*\big)^2. \tag{5.32}$$

Among the many nice properties that the fidelity function possesses is the fact that if $Q_0, Q_1 \in \mathrm{Pos}(\mathcal{U})$ and $P_0 \in \mathrm{Pos}(\mathcal{U} \otimes \mathcal{V})$ satisfies $\mathrm{Tr}_{\mathcal{V}}(P_0) = Q_0$, then

$$\mathrm{F}(Q_0, Q_1) = \max\bigl\{\mathrm{F}(P_0, P_1) \,:\, P_1 \in \mathrm{Pos}(\mathcal{U} \otimes \mathcal{V}),\ \mathrm{Tr}_{\mathcal{V}}(P_1) = Q_1\bigr\}. \tag{5.33}$$

From this fact we find that

$$2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})} = \mathrm{F}\bigl(\bigl(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(uu^*), \mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes \sigma\bigr)^2 \tag{5.34}$$

for some state $\sigma \in \mathrm{D}(\mathcal{Z})$, as operators of the form $\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes \sigma$ are the only operators that leave $\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*$ when $\mathcal{Z}$ is traced out. The fidelity is nondecreasing under the partial trace on the second tensor factor of $\mathcal{X}$, and therefore

$$2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})} \leq \mathrm{F}\bigl(\mathrm{Tr}_{\mathcal{Y}}(uu^*), \mathbb{1}_{\mathcal{X}} \otimes \sigma\bigr)^2 \leq 2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})}. \tag{5.35}$$

Now let us prove the reverse inequality

$$2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})} \leq 2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})}, \tag{5.36}$$

which is based on a similar idea. Choose a density operator $\sigma \in \mathrm{D}(\mathcal{Z})$ so that

$$2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})} = \mathrm{F}\bigl(\mathrm{Tr}_{\mathcal{Y}}(uu^*), \mathbb{1}_{\mathcal{X}} \otimes \sigma\bigr)^2. \tag{5.37}$$

The operator $\mathbb{1}_{\mathcal{X}} \otimes \sigma$ can be purified as

$$\mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}} \otimes \sqrt{\sigma}\bigr)\,\mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}} \otimes \sqrt{\sigma}\bigr)^* \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{X} \otimes \mathcal{Z}). \tag{5.38}$$

Every extension of $\mathrm{Tr}_{\mathcal{Y}}(uu^*)$ to an element of $\mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{X} \otimes \mathcal{Z})$ can be expressed as

$$\bigl(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Xi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(uu^*) \tag{5.39}$$

for a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z} \otimes \mathcal{X})$, meaning that these are exactly the operators that leave $\mathrm{Tr}_{\mathcal{Y}}(uu^*)$ when the first tensor factor of $\mathcal{Z}$ and the second tensor factor of $\mathcal{X}$ are traced out. By the same fact regarding the fidelity function from before, we find that

$$2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})} = \mathrm{F}\bigl(\bigl(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Xi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(uu^*), \mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}} \otimes \sqrt{\sigma}\bigr)\,\mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}} \otimes \sqrt{\sigma}\bigr)^*\bigr)^2 \tag{5.40}$$

for some choice of a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z} \otimes \mathcal{X})$. Because the fidelity is nondecreasing under the partial trace on both copies of $\mathcal{Z}$, we obtain

$$2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})} \leq \mathrm{F}\bigl(\bigl(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi\bigr)(\mathrm{Tr}_{\mathcal{Z}}(uu^*)), \mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}}\bigr)\,\mathrm{vec}\bigl(\mathbb{1}_{\mathcal{X}}\bigr)^*\bigr)^2 \tag{5.41}$$

for $\Phi = \mathrm{Tr}_{\mathcal{Z}} \circ \Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{X})$. This implies

$$2^{\mathrm{H}_{\max}(\mathsf{X}|\mathsf{Z})} \leq 2^{-\mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})}, \tag{5.42}$$

as required. □

5.3 Hypothesis-testing relative entropy

We will define the *hypothesis-testing relative entropy* as follows.

Definition 5.11. Let $\rho \in \mathrm{D}(\mathcal{X})$, $Q \in \mathrm{Pos}(\mathcal{X})$, and $\varepsilon \in [0,1]$. The *ε-hypothesis-testing relative entropy* of ρ with respect to Q is defined as

$$\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q) = -\inf\bigl\{\log\langle Q, X\rangle \,:\, X \in \mathrm{Pos}(\mathcal{X}),\, \langle \rho, X\rangle \geq 1,\, \varepsilon X \leq \mathbb{1}\bigr\}. \tag{5.43}$$

Elementary observations

Before we try to understand the intuitive meaning of this quantity, let us note a few simple things about it.

First, we see that $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q) = \infty$ is possible:

1. $\mathrm{D}_{\mathrm{H}}^{0}(\rho\|Q) = \infty$ if and only if $\mathrm{im}(\rho) \not\subseteq \mathrm{im}(Q)$.
2. For $\varepsilon \in (0,1)$ we have $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q) = \infty$ if and only if
$$\bigl\langle \Pi_{\ker(Q)}, \rho \bigr\rangle \geq \varepsilon. \tag{5.44}$$
3. $\mathrm{D}_{\mathrm{H}}^{1}(\rho\|Q) = \infty$ if and only if $\mathrm{im}(\rho) \perp \mathrm{im}(Q)$.

Next, notice that as ε decreases, the infimum decreases, because decreasing ε means relaxing the constraint $\varepsilon X \leq \mathbb{1}$, and therefore the ε-hypothesis-testing relative entropy increases: $\delta \leq \varepsilon$ implies $\mathrm{D}_{\mathrm{H}}^{\delta}(\rho\|Q) \geq \mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q)$. Stated another way, taking ε to be smaller means taking the ε-hypothesis-testing relative entropy to be a stronger notion of divergence. (The same may be said about the ε-smoothed max-relative entropy.)

Continuing on, although our primary focus will be on the range of values $\varepsilon \in (0,1)$, we will take a moment to consider the extreme cases $\varepsilon = 0$ and $\varepsilon = 1$. If it is the case that $\varepsilon = 0$, the constrain $\varepsilon X \leq \mathbb{1}$ is trivially satisfied. We obtain precisely the max-relative entropy, as an examination of the dual form of Optimization Problem 2.4 reveals:

$$\mathrm{D}_{\mathrm{H}}^{0}(\rho\|Q) = \mathrm{D}_{\max}(\rho\|Q). \tag{5.45}$$

At the other extreme, we may consider $\varepsilon = 1$, so that the constraint $\varepsilon X \leq \mathbb{1}$ becomes $X \leq \mathbb{1}$. The infimum is evidently achieved for $X = \Pi_{\mathrm{im}(\rho)}$, and so we obtain

$$\mathrm{D}_{\mathrm{H}}^{1}(\rho\|Q) = -\log\bigl\langle Q, \Pi_{\mathrm{im}(\rho)}\bigr\rangle. \tag{5.46}$$

This quantity has also been called the min-relative entropy by some, but obviously we will not use this name given that we have already used it for something else—the name *1-hypothesis-testing relative entropy* will do just fine.

Semidefinite programming characterization

The definition of the hypothesis-testing relative entropy immediately suggests a semidefinite programming characterization. Specifically, the value $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q)$ is the negative logarithm of the optimal value of the following semidefinite program.

Optimization Problem 5.12 (SDP for hypothesis-testing relative entropy)

Primal problem		*Dual problem*	
minimize:	$\langle Q, X\rangle$	maximize:	$\lambda - \mathrm{Tr}(Y)$
subject to:	$\langle \rho, X\rangle \geq 1,$	subject to:	$\lambda\rho \leq Q + \varepsilon Y,$
	$\varepsilon X \leq \mathbb{1},$		$Y \in \mathrm{Pos}(\mathcal{X}),$
	$X \in \mathrm{Pos}(\mathcal{X}).$		$\lambda \geq 0.$

The primal problem is strictly feasible provided that $\varepsilon \in [0,1)$. In particular,

$$X = \begin{cases} \frac{(1+\varepsilon)\mathbb{1}}{2\varepsilon} & \text{if } \varepsilon \in (0,1) \\ 2\mathbb{1} & \text{if } \varepsilon = 0 \end{cases} \tag{5.47}$$

is strictly primal feasible. Strict primal feasibility is, on the other hand, impossible when $\varepsilon = 1$. The dual problem is strictly feasible when $\varepsilon \in (0,1]$, by $\lambda = 1$ and $Y = 2\mathbb{1}/\varepsilon$ for instance. In the case $\varepsilon = 0$, the dual problem is strictly feasible if and only if $\mathrm{im}(\rho) \subseteq \mathrm{im}(Q)$.

Therefore, strong duality holds and the optimal values are achieved for all choices of ρ and Q when $\varepsilon \in (0,1)$, by Slater's theorem. We also have strong duality and an optimal value achieved in the primal problem when $\varepsilon = 1$, again by Slater's theorem, as $X = \mathbb{1}$ is primal feasible (although not strictly so); and strong duality also holds in the case $\varepsilon = 0$, as our examination of Optimization Problem 2.4 has already revealed.

Interpretation

One way to interpret the ε-hypothesis-testing relative entropy, at least in the case $\varepsilon > 0$, begins with observation that

$$\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q) = -\inf\left\{\log\frac{\langle Q, P\rangle}{\varepsilon} \,:\, 0 \leq P \leq \mathbb{1},\ \langle \rho, P\rangle \geq \varepsilon\right\}. \tag{5.48}$$

For the sake of the discussion that follows, let us consider the case that $Q = \sigma$ is a density operator.

We can now consider a test being performed that aims to distinguish between the states ρ and σ. Think of σ as representing an idealized model for a state, corresponding to the null-hypothesis of the test being performed, whereas ρ is the actual state of the system being tested. The operator P may be associated with a measurement operator, the outcome corresponding to which is to be seen as a signal supporting an alternative hypothesis. If we were to measure ρ with respect to such a measurement, the alternative hypothesis may not be signaled with high probability, but the probability is at least ε. Think of the probability ε as representing how small of a signal one is willing to tolerate in support of an alternative hypothesis. The value $2^{-\mathrm{D}_{\mathrm{H}}^{\varepsilon}}(\rho\|Q)$ is then equal to the smallest possible value for $\langle Q, P\rangle/\varepsilon$ that we could achieve by selecting P optimally. Informally speaking, $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|\sigma)$ is a measure of how surprising it would be to obtain the outcome corresponding to P in the idealized case represented by σ.

The precise scaling in the definition above has been selected so that a variant of Klein's inequality holds, provided $\varepsilon \in [0,1)$. That is, for all $\varepsilon \in [0,1)$ it is the case that $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|\sigma) \geq 0$, with equality if and only if $\rho = \sigma$. Indeed, if $\sigma \neq \rho$, we may choose a unit vector u for which $\langle uu^*, \rho\rangle > \langle uu^*, \sigma\rangle$, and then consider

$$X = (1-\delta)\mathbb{1} + \frac{\delta}{\langle uu^*, \rho\rangle}\, uu^* \tag{5.49}$$

in the primal problem, for $\delta \in (0,1)$. The operator X is clearly positive semidefinite, and it is the case that $\langle \sigma, X\rangle < \langle \rho, X\rangle = 1$. The constraint $\varepsilon X \leq \mathbb{1}$ is satisfied so long as

$$1 - \delta + \frac{\delta}{\langle uu^*, \rho\rangle} \leq \frac{1}{\varepsilon}, \tag{5.50}$$

which is so for all sufficiently small δ, as $1/\varepsilon$ is strictly larger than 1 by the assumption $\varepsilon < 1$. By selecting any such δ, one obtains a primal feasible X having objective value strictly smaller than 1, which implies that $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|\sigma)$ is positive. In the case $\varepsilon = 1$, one has $\mathrm{D}_{\mathrm{H}}^{1}(\rho\|\sigma) = 0$ if and only if $\mathrm{im}(\sigma) \subseteq \mathrm{im}(\rho)$.

Monotonicity under the action of channels

Suppose that Φ is a trace-preserving and positive map. Consider the dual form of Optimization Problem 5.12. If it is the case that $\lambda \geq 0$ and $Y \in \mathrm{Pos}(\mathcal{X})$ satisfy the constraint

$$\lambda\rho \leq Q + \varepsilon Y, \tag{5.51}$$

then by the positivity of Φ it must also hold that

$$\lambda\Phi(\rho) \leq \Phi(Q) + \varepsilon\Phi(Y), \tag{5.52}$$

and therefore $(\lambda, \Phi(Y))$ is dual-feasible for the instance of this problem corresponding to $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\Phi(\rho)\|\Phi(Q))$, with the same objective value being achieved by the assumption that Φ preserves trace. It follows that

$$2^{-\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\Phi(\rho)\|\Phi(Q))} \geq 2^{-\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q)}, \tag{5.53}$$

or, equivalently,

$$\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\Phi(\rho)\|\Phi(Q)) \leq \mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q). \tag{5.54}$$

Relationship to smoothed max-relative entropy

We will conclude our discussion of the hypothesis-testing relative entropy by observing its close relationship to the smoothed max-relative entropy. The two theorems that follow reveal this close relationship.

Recall that $\mathrm{D}_{\max}^{\varepsilon}(\rho\|Q)$ is the logarithm of the optimal value of this conic program:

Optimization Problem 5.13 (SDP for smoothed max-relative entropy)

Primal problem		*Dual problem*	
minimize:	η	maximize:	$\psi_{\rho}^{\varepsilon}(Z)$
subject to:	$\xi \leq \eta Q$	subject to:	$\langle Q, Z\rangle \leq 1$
	$\xi \in \mathcal{B}_{\varepsilon}(\rho)$,		$Z \in \mathrm{Pos}(\mathcal{X})$
	$\eta \geq 0$		

where

$$\psi_{\rho}^{\varepsilon}(Z) = \inf_{\xi \in \mathcal{B}_{\varepsilon}(\rho)} \langle \xi, Z\rangle. \tag{5.55}$$

In the theorems that follow, we will again use trace-distance smoothing, as we did in Lecture 4:

$$\mathcal{B}_{\varepsilon}(\rho) = \bigl\{\xi \in \mathrm{D}(\mathcal{X}) \,:\, \tfrac{1}{2}\|\rho - \xi\|_1 \leq \varepsilon\bigr\}. \tag{5.56}$$

Theorem 5.14. *Let $\rho \in \mathrm{D}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{X})$ be operators satisfying* $\mathrm{im}(\rho) \subseteq \mathrm{im}(Q)$ *and let $\varepsilon \in (0,1)$. It is the case that*

$$\mathrm{D}_{\max}^{\sqrt{\varepsilon}}(\rho\|Q) \leq \mathrm{D}_{H}^{\varepsilon}(\rho\|Q) + \log\Bigl(\frac{1}{\varepsilon(1-\varepsilon)}\Bigr). \tag{5.57}$$

Proof. Consider any operator $Z \in \mathrm{Pos}(\mathcal{X})$ that satisfies $\langle Q, Z\rangle \leq 1$, and let

$$Z = \sum_{k=1}^{n} \lambda_k(Z)\, z_k z_k^* \tag{5.58}$$

be a spectral decomposition of Z. We will use the basis $\{z_1, \ldots, z_n\}$ to construct a family of feasible solutions X to the primal form of Optimization Problem 5.12.

In particular, for each real number $\lambda \in \mathbb{R}$, define a subset $S_\lambda \subseteq \{1, \ldots, n\}$ as

$$S_\lambda = \bigl\{k \in \{1, \ldots, n\} \,:\, \bigl\langle z_k z_k^*, \rho - 2^\lambda Q \bigr\rangle > 0 \bigr\}, \tag{5.59}$$

and define

$$\Pi_\lambda = \sum_{k \in S_\lambda} z_k z_k^*. \tag{5.60}$$

Observe that

$$\bigl\langle \Pi_\lambda, \rho \bigr\rangle \geq 2^\lambda \bigl\langle \Pi_\lambda, Q \bigr\rangle, \tag{5.61}$$

with the equality being strict so long as Π_λ is nonzero. The operator

$$X = \frac{\Pi_\lambda}{\varepsilon} \tag{5.62}$$

is therefore feasible for primal form of Optimization Problem 5.12 provided that $\langle \Pi_\lambda, \rho \rangle \geq \varepsilon$, and with this observation in mind it is informative to consider the supremum over all such values of λ:

$$\gamma = \sup\bigl\{\lambda \in \mathbb{R} \,:\, \langle \Pi_\lambda, \rho \rangle \geq \varepsilon \bigr\}. \tag{5.63}$$

In particular, for every $\delta > 0$ we see that the operator

$$X = \frac{\Pi_{\gamma - \delta}}{\varepsilon} \tag{5.64}$$

is primal feasible, and so we obtain the inequalities

$$2^{-\,\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho \| Q)} \leq \frac{\langle Q, \Pi_{\gamma-\delta} \rangle}{\varepsilon} \leq \frac{2^{-\gamma+\delta}}{\varepsilon} \langle \rho, \Pi_{\gamma-\delta} \rangle \leq \frac{2^{-\gamma+\delta}}{\varepsilon}. \tag{5.65}$$

As these inequalities hold for every $\delta > 0$, it follows that

$$\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho \| Q) \geq \gamma + \log(\varepsilon). \tag{5.66}$$

On the other hand, for any choice of $\delta > 0$, it must be that $\langle \Pi_{\gamma+\delta}, \rho \rangle < \varepsilon$. The density operator

$$\xi = \frac{(\mathbb{1} - \Pi_{\gamma+\delta})\rho(\mathbb{1} - \Pi_{\gamma+\delta})}{\langle \mathbb{1} - \Pi_{\gamma+\delta}, \rho \rangle} \tag{5.67}$$

therefore satisfies

$$\mathrm{F}(\rho, \xi) \geq \sqrt{1 - \varepsilon} \tag{5.68}$$

by Winter's gentle measurement lemma, which implies that

$$\frac{1}{2}\|\rho - \xi\|_1 \leq \sqrt{\varepsilon} \tag{5.69}$$

by one of the Fuchs–van de Graaf inequalities. Consequently,

$$\begin{aligned} \psi_\rho^{\sqrt{\varepsilon}}(Z) &\leq \langle \xi, Z\rangle = \frac{1}{\langle \mathbb{1} - \Pi_{\gamma+\delta}, \rho\rangle} \sum_{k \notin S_{\gamma+\delta}} \lambda_k(Z) \langle z_k z_k^*, \rho\rangle \\ &\leq \frac{2^{\gamma+\delta}}{1-\varepsilon} \sum_{k \notin S_{\gamma+\delta}} \lambda_k(Z)\langle z_k z_k^*, Q\rangle \leq \frac{2^{\gamma+\delta}}{1-\varepsilon}\langle Z, Q\rangle \leq \frac{2^{\gamma+\delta}}{1-\varepsilon}. \end{aligned} \tag{5.70}$$

As this is so for all $\delta > 0$, it follows that

$$\log \psi_\rho^{\sqrt{\varepsilon}}(Z) \leq \gamma - \log(1-\varepsilon), \tag{5.71}$$

and therefore

$$\log \psi_\rho^{\sqrt{\varepsilon}}(Z) \leq \mathrm{D}_{\mathrm{H}}^{\varepsilon}(\rho\|Q) + \log\Big(\frac{1}{\varepsilon(1-\varepsilon)}\Big). \tag{5.72}$$

By optimizing over all operators $Z \in \mathrm{Pos}(\mathcal{X})$ that satisfy $\langle Q, Z\rangle \leq 1$ we obtain the required inequality. □

Theorem 5.15. *Let $\rho \in \mathrm{D}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{X})$ be operators satisfying $\mathrm{im}(\rho) \subseteq \mathrm{im}(Q)$, let $\varepsilon \in (0,1)$, and let $\delta \in (0, 1-\varepsilon)$. It is the case that*

$$\mathrm{D}_{H}^{\varepsilon+\delta}(\rho\|Q) \leq \mathrm{D}_{\max}^{\varepsilon}(\rho\|Q) - \log\Big(\frac{\delta}{\varepsilon+\delta}\Big). \tag{5.73}$$

Proof. Suppose that $\xi \in \mathcal{B}_\varepsilon(\rho)$ satisfies

$$\xi \leq 2^{\mathrm{D}_{\max}^{\varepsilon}(\rho\|Q)} Q. \tag{5.74}$$

Given that $\xi \in \mathcal{B}_\varepsilon(\rho)$ we have that

$$\rho \leq \xi + R \tag{5.75}$$

for $R \in \mathrm{Pos}(\mathcal{X})$ satisfying $\mathrm{Tr}(R) \leq \varepsilon$.

Now consider the quantity $\mathrm{D}_{\mathrm{H}}^{\varepsilon+\delta}(\rho\|Q)$, and in particular consider the choices

$$\lambda = 2^{-\mathrm{D}_{\max}^{\varepsilon}(\rho\|Q)} \quad \text{and} \quad Y = \frac{2^{-\mathrm{D}_{\max}^{\varepsilon}(\rho\|Q)} R}{\varepsilon+\delta} \tag{5.76}$$

in the dual form of Optimization Problem 5.12 (with ε being replaced by $\varepsilon + \delta$). These choices represent a feasible solution to this conic program, and the objective value is at least

$$2^{-\mathrm{D}_{\max}^{\varepsilon}(\rho\|Q)}\left(1 - \frac{\varepsilon}{\varepsilon+\delta}\right) \tag{5.77}$$

It follows that

$$\mathrm{D}_{\mathrm{H}}^{\varepsilon+\delta}(\rho\,\|\,Q) \leq \mathrm{D}_{\max}^{\varepsilon}(\rho\,\|\,Q) - \log\Bigl(\frac{\delta}{\varepsilon+\delta}\Bigr), \tag{5.78}$$

which completes the proof. □

Corollary 5.16. *Let* $\rho \in \mathrm{D}(\mathcal{X})$ *and* $Q \in \mathrm{Pos}(\mathcal{X})$ *be operators. For every* $\varepsilon \in (0,1)$, *we have*

$$\lim_{n\to\infty} \frac{\mathrm{D}_{\mathrm{H}}^{\varepsilon}\bigl(\rho^{\otimes n}\,\|\,Q^{\otimes n}\bigr)}{n} = \mathrm{D}(\rho\,\|\,Q). \tag{5.79}$$