15 Entropies and hypothesis testing

15.1 Entropies

15.1.1 Notation

Given more than one random variable, X and Y for example,

$$H(X,Y) := S(P_{XY})$$

denotes the "joint entropy" of X and Y. Note that this is simply the entropy of the random variable (X,Y).

If we have states ρ_{ABC} and σ_{ABD} then, unless otherwise specified, $\rho_{\mathsf{A}} := \mathrm{Tr}_{\mathsf{BC}}\rho_{\mathsf{ABC}}$, $\rho_{\mathsf{AC}} := \mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{ABC}}$, $\sigma_{\mathsf{BD}} := \mathrm{Tr}_{\mathsf{A}}\sigma_{\mathsf{ABD}}$, etc. denote the states of the various subsystems. In analogy with our notation for entropies of random variables we write the von Neumann entropies for the states of these systems $H(\mathsf{A})_\rho := S(\rho_{\mathsf{A}})$, $H(\mathsf{D})_\sigma := S(\sigma_{\mathsf{D}})$, $H(\mathsf{ABC})_\rho := S(\rho_{\mathsf{ABC}})$, $H(\mathsf{AB})_\rho := S(\rho_{\mathsf{AB}})$, $H(\mathsf{AB})_\sigma := S(\sigma_{\mathsf{AB}})$, and so on.

15.1.2 Classical information stored in quantum systems

Suppose we have a number of random variables, e.g. X, Y and Z, whose values are stored in quantum systems $\tilde{\mathsf{X}}$, $\tilde{\mathsf{Y}}$ and $\tilde{\mathsf{Z}}$ in the standard way, so that the state of $\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}$ is

$$\rho_{\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}} = \sum_{x,y,z} P_{XYZ}(x,y,z)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |y\rangle\langle y|_{\tilde{\mathsf{Y}}} \otimes |z\rangle\langle z|_{\tilde{\mathsf{Z}}}. \tag{15.1}$$

The quantum entropies of these systems correspond to the classical entropies of the RVs in the obvious way e.g. $H(\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}})_\rho = H(X,Y,Z)$, $H(\tilde{\mathsf{Y}}\tilde{\mathsf{Z}})_\rho = H(Y,Z)$, $H(\tilde{\mathsf{X}})_\rho = H(X)$ etc.

15.1.3 Basic bounds on entropy

Given any distribution p on a finite set $\mathcal{A}$,

$$0 \leq S(p) \leq \log|\mathrm{supp}(p)| \leq \log|\mathcal{A}|, \tag{15.2}$$

where the first equality holds iff $p(x) = 1$ for some $x \in \mathcal{A}$. This result can be established by using the strict concavity of log and Jensen's inequality. Doing so is an exercise on example sheet 3.

Given any state ρ on a finite dimensional Hilbert space $\mathcal{H}$,

$$0 \leq S(\rho) \leq \log\dim(\mathrm{supp}(\rho)) \leq \log\dim(\mathcal{H}), \tag{15.3}$$

where the first equality holds iff ρ is pure. These inequalities follow easily from those in (15.2) by using the fact that if $\rho = \sum_k p(k)|\alpha_k\rangle\langle\alpha_k|$ is an eigendecomposition for ρ then $S(\rho) = S(p)$.

15.2 Conditional entropy and the chain rule

Definition 1 (Classical conditional entropies)**.** Given random variables X and Y and Z:

1. $H(X|Y=y) := S(P_{X|Y=y})$;
2. $H(X|Y) := \sum_{y\in\mathcal{A}_Y} H(X|Y=y)\Pr(Y=y)$;
3. $H(X|Y,Z=z) = \sum_{y\in\mathcal{A}_Y} H(X|Y=y,Z=z)\Pr(Y=y|Z=z)$.

It follows from the results in section 15.1.3 that $H(X|Y)\geq 0$ with equality iff $X=f(Y)$ for some function $f:\mathrm{supp}(P_Y)\to\mathcal{A}_X$. Similarly, $H(X|Y,Z=z)\geq 0$ with equality iff, given $Z=z$, $X=f(Y)$ for some function $f:\mathrm{supp}(P_{Y|Z=z})\to\mathcal{A}_X$.

The product rule of probability says that $P_{XY}(x,y)=P_{X|Y=y}(x)P_Y(y)$, so

$$\log\frac{1}{P_{XY}(x,y)} = \log\frac{1}{P_{X|Y=y}(x)} + \log\frac{1}{P_Y(y)}. \tag{15.4}$$

Multiplying by $P_{XY}(x,y)$ and summing over x and y

$$\begin{aligned}\sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_{XY}(x,y)} &= \sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_{X|Y=y}(x)} + \sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_Y(y)}\\ &= \sum_y P_Y(y)\sum_x P_{X|Y}(x|y)\log\frac{1}{P_{X|Y=y}(x)} + \sum_y P_Y(y)\log\frac{1}{P_Y(y)}\end{aligned}$$

which is equivalent to

Proposition 2 (The chain rule)**.** $H(X,Y)=H(X|Y)+H(Y)$.

By treating multiple random variables as a single tuple of random variables, we can extend the chain rule to more variables. For example

$$\begin{aligned}H(X,Y|Z) = H(X,Y,Z) - H(Z) &= (H(X,Y,Z) - H(Y,Z)) + H(Y,Z) - H(Z)\\ &= H(X|Y,Z) + H(Y|Z).\end{aligned}$$

The *quantum conditional entropy* is *defined* to obey the chain rule.

Definition 3. For systems A and B:

1. The **quantum conditional entropy** of A given B is $H(\mathsf{A}|\mathsf{B}) := H(\mathsf{AB}) - H(\mathsf{B})$.
2. The **coherent information** of A given B is $I(\mathsf{A}\rangle\mathsf{B}) := -H(\mathsf{A}|\mathsf{B})$.

Any identity between classical entropies derived using only the chain rule, can be derived in the same way for quantum entropies. For example

$$H(\mathsf{AB}|\mathsf{C}) = H(\mathsf{A}|\mathsf{BC}) + H(\mathsf{B}|\mathsf{C}).$$

A striking difference between classical and quantum conditional entropies is that while any classical conditional entropy is positive, quantum conditional entropy can be negative. For example, if $d_\mathsf{A}=d_\mathsf{B}=d$ and $\rho_\mathsf{AB}=\phi^+_\mathsf{AB}$ then $H(\mathsf{A}|\mathsf{B})_\rho = H(\mathsf{AB})_\rho - H(\mathsf{B})_\rho = S(\phi^+_\mathsf{AB}) - S(\mathbb{1}_\mathsf{B}/d) = 0 - \log d$.

15.3 Mutual information and conditional mutual information

Given two random variables X and Y, the entropy gives us a way to quantify how much Y tells us about X: $H(X)$ is a way of quantifying our uncertainty about X. Prior to learning the value of Y, our expectation for the entropy we will assign to X if we *do* learn the value of Y, is $H(X|Y)$. Therefore, $I(X:Y) := H(X) - H(X|Y)$ is the expectation of the *reduction in entropy* of X which will occur if we learn the value of Y. This is a measure of how much learning Y would *inform* us about X. By the chain rule,

$$I(X:Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X),$$

so $I(X:Y)$ is *symmetric* in the two random variables, and we call it the **mutual information** between X and Y.

Definition 4. For random variables X, Y and Z:

1. The mutual information between X and Y is

$$I(X:Y) := H(X) - H(X|Y). \tag{15.5}$$

2. The conditional mutual information between X and Y given $Z = z$ is

$$I(X:Y|Z=z) := H(X|Z=z) - H(X|Y,Z=z). \tag{15.6}$$

3. The conditional mutual information between X and Y given Z is

$$\begin{aligned} I(X:Y|Z) :=& \sum_z \Pr(Z=z) I(X:Y|Z=z) = H(X|Z) - H(X|Y,Z) \\ =& H(X|Z) + H(Y|Z) - H(X,Y|Z) \\ =& H(Y|Z) - H(Y|X,Z). \end{aligned}$$

Definition 5. For systems A, B and C:

1. The **quantum mutual information** between A and B is

$$I(\mathsf{A}:\mathsf{B}) := H(\mathsf{A}) - H(\mathsf{A}|\mathsf{B}) = H(\mathsf{A}) + H(\mathsf{B}) - H(\mathsf{AB}) = H(\mathsf{B}) - H(\mathsf{B}|\mathsf{A}). \tag{15.7}$$

2. **quantum conditional mutual information** between A and B given C is

$$I(\mathsf{A}:\mathsf{B}|\mathsf{C}) := H(\mathsf{A}|\mathsf{C}) - H(\mathsf{A}|\mathsf{BC}) = H(\mathsf{AC}) + H(\mathsf{BC}) - H(\mathsf{C}) - H(\mathsf{ABC}). \tag{15.8}$$

Ex. sheet 3 asks you to prove the **chain rule for conditional mutual information**:

Proposition 6. $I(X:Y,Z|W) = I(X:Z|W) + I(X:Y|Z,W)$.

Since this can be proved using the chain rule alone, it also holds for quantum entropies:

Proposition 7. $I(\mathsf{A}:\mathsf{BC}|\mathsf{D}) = I(\mathsf{A}:\mathsf{C}|\mathsf{D}) + I(\mathsf{A}:\mathsf{B}|\mathsf{CD})$.

15.4 Hypothesis testing and relative entropy

We now return briefly to a kind of simple state discrimination problem known as "quantum hypothesis testing". Given a quantum system Q, let hypothesis H_0 be that the state of Q is ρ, and hypothesis H_1 be that the state of Q is σ. Suppose we measure a POVM $E : \{0,1\} \to \mathcal{H}_\mathsf{Q}$ obtaining a result $\hat{X}$ which is supposed to identify which hypothesis is true. Since this is a binary POVM, if we set $E(0) = T$ then this determines $E(1) = \mathbb{1} - T$.

$$\text{The "type-I" error probability is } \Pr(\hat{X} = 1|H_0) = \alpha(T,\rho) := 1 - \mathrm{Tr}T\rho, \text{ and} \tag{15.9}$$

$$\text{the "type-II" error probability is } \Pr(\hat{X} = 0|H_1) = \beta(T,\sigma) := \mathrm{Tr}T\sigma. \tag{15.10}$$

In general, there is a trade-off between these two conditional probabilities. We use the following notation for the minimum value of β that can be attained subject to the requirement that $\alpha \leq \epsilon$.

Definition 8. $\beta_\epsilon(\rho\|\sigma) := \min\{\beta(T,\sigma) : \alpha(T,\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\}$.

Proposition 9 (Data processing inequality for β_ϵ)**.** For all $\epsilon \in [0,1]$, states ρ_A and σ_A, and operations $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$, $\beta_\epsilon(\mathcal{N}\rho_\mathsf{A}\|\mathcal{N}\sigma_\mathsf{A}) \geq \beta_\epsilon(\rho_\mathsf{A}\|\sigma_\mathsf{A})$.

Proof. Because the linear map $\mathcal{N}$ is completely positive and trace preserving, its adjoint $\mathcal{N}^\dagger$ is completely positive and unital (identity preserving). So if $0 \leq T \leq \mathbb{1}$ then $0 \leq \mathcal{N}^\dagger T \leq \mathcal{N}^\dagger\mathbb{1} = \mathbb{1}$ and therefore

$$\begin{aligned}\beta_\epsilon(\mathcal{N}\rho\|\mathcal{N}\sigma) &= \min\{\beta(T,\mathcal{N}\sigma) : \alpha(T,\mathcal{N}\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\} &(15.11)\\ &= \min\{\beta(\mathcal{N}^\dagger T,\sigma) : \alpha(\mathcal{N}^\dagger T,\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\} \geq \beta_\epsilon(\rho\|\sigma). &(15.12)\end{aligned}$$

□

Intuitively, if we are given n systems in the state ρ or n systems in the state σ then larger values of n should make it easier to distinguish between the two situations. The next theorem tells us that $\beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$ exhibits exponential decay as n increases, and gives us the rate of the decay.

Theorem 10 (Quantum Stein's lemma)**.** For all states ρ, σ of a given system

$$\forall\epsilon \in (0,1), \lim_{n\to\infty} \frac{1}{n} \log \beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = -D(\rho\|\sigma), \tag{15.13}$$

where $D(\rho\|\sigma)$ is the *quantum relative entropy* between ρ and σ...

Definition 11 (Quantum relative entropy)**.**

$$D(\rho\|\sigma) := \begin{cases} -S(\rho) - \mathrm{Tr}\rho\log\sigma & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \tag{15.14}$$

In $\mathrm{Tr}\rho\log\sigma$ we regard both ρ and $\log\sigma$ as operators on $\mathrm{supp}(\sigma)$.

Note that the relative entropy is not symmetric in its arguments and does not obey the triangle inequality.

Proposition 12. Density operators ρ and σ on $\mathcal{H}$ commute if and only if there are eigendecompositions $\rho = \sum_{i=1}^{\dim(\mathcal{H})} p(i)|\alpha_i\rangle\langle\alpha_i|$ and $\sigma = \sum_{i=1}^{\dim(\mathcal{H})} q(i)|\alpha_i\rangle\langle\alpha_i|$. In this case, $D(\rho\|\sigma) = D(p\|q)$, where $D(p\|q)$ is the *classical relative entropy* between the distributions p and q...

Definition 13 (Classical relative entropy)**.**

$$D(p\|q) := \begin{cases} -S(p) - \sum_{x\in\mathrm{supp}(q)} p(x)\log q(x) & \text{if } \mathrm{supp}(p) \subseteq \mathrm{supp}(q), \\ +\infty & \text{otherwise.} \end{cases} \tag{15.15}$$

Theorem 14 (Gibbs' inequality)**.** For any two probability distributions p and q on a finite set $\mathcal{A}$, $D(p\|q) \geq 0$ with equality iff $p = q$.

Proof. That $D(p\|q) \geq 0$ already follows from the operational meaning given to $D(p\|q)$ by Stein's lemma. Here is a direct proof which gives us the equality condition. If $\mathrm{supp}(p) \not\subseteq \mathrm{supp}(q)$ then $D(p\|q) = \infty$ and the result holds, so let's now assume that $\mathrm{supp}(p) \subseteq \mathrm{supp}(q)$. $D(p\|q) = \sum_{x\in\mathrm{supp}(p)} p(x)\log\frac{p(x)}{q(x)} = -\frac{1}{\ln(2)}\sum_{x\in\mathrm{supp}(p)} p(x)\ln\frac{q(x)}{p(x)}$. The line $y = x-1$ is tangent to the graph $y = \ln x$ at $x = 1$ so from the strict concavity of ln it follows that $\ln x \leq x - 1$ with equality iff $x = 1$. Therefore,

$$-D(p\|q) \leq \frac{1}{\ln(2)} \sum_{x\in\mathrm{supp}(p)} p(x)\left(\frac{q(x)}{p(x)} - 1\right) \leq 0.$$

Since the $p(x)$ in the sum are strictly positive, the first inequality is an equality if and only if $q(x)/p(x) = 1$ for all $x \in \mathrm{supp}(p)$, which is true iff $q(x) = p(x)$ for all x, in which case the second inequality is also satisfied. □

Theorem 15 (Data processing inequality for D)**.** For all states ρ_{A} and σ_{A}, and operations $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$, $D(\mathcal{N}\rho_{\mathsf{A}}\|\mathcal{N}\sigma_{\mathsf{A}}) \leq D(\rho_{\mathsf{A}}\|\sigma_{\mathsf{A}})$.

Proof.

$$D(\mathcal{N}\rho\|\mathcal{N}\sigma) = -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon((\mathcal{N}\rho)^{\otimes n}\|(\mathcal{N}\sigma)^{\otimes n}) \tag{15.16}$$

$$= -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon(\mathcal{N}^{\otimes n}\rho^{\otimes n}\|\mathcal{N}^{\otimes n}\sigma^{\otimes n}) \tag{15.17}$$

$$\leq -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma). \tag{15.18}$$

The first and last equalities are by Quantum Stein's lemma. The inequality is by the data processing inequality for β_ϵ, the fact that log is an increasing function, and an elementary fact about limits. □

Theorem 16 (Klein's inequality)**.** $D(\rho\|\sigma) \geq 0$ with equality iff $\rho = \sigma$.

Proof. Suppose that $X = 0$ with probability $1/2$ and $X = 1$ with probability $1/2$, and suppose that the state of Q is ρ when $X = 0$ and the state is σ when $X = 1$. Suppose we measure a POVM E to obtain an estimate $\hat{X}$ of X and then prepare Q in the computational basis state $|\hat{X}\rangle\langle\hat{X}|$. This corresponds to applying the operation

$$\mathcal{N} : \tau \mapsto |0\rangle\langle 0|\mathrm{Tr}E(0)\tau + |1\rangle\langle 1|\mathrm{Tr}E(1)\tau. \tag{15.19}$$

We choose the POVM which maximises $\Pr(\hat{X} = X)$.

Letting $p(i) = \Pr(\hat{X} = i|X = 0)$ and $q(i) = \Pr(\hat{X} = i|X = 1)$, $\mathcal{N}\rho = p(0)|0\rangle\langle 0| + p(1)|1\rangle\langle 1|$ and $\mathcal{N}\sigma = q(0)|0\rangle\langle 0| + q(1)|1\rangle\langle 1|$. By the data processing inequality for quantum relative entropy, Proposition 12, and Gibbs' inequality

$$D(\rho\|\sigma) \geq D(\mathcal{N}\rho\|\mathcal{N}\sigma) = D(p\|q) \geq 0 \tag{15.20}$$

so we require $p = q$ for $D(\rho\|\sigma) = 0$ to hold. The Holevo-Helstrom theorem tells us that

$$\Pr(\hat{X} = X) = \frac{1}{2}\left(p(0) + q(1)\right) = \frac{1}{2}\left(1 + q(1) - p(1)\right) = \frac{1}{2}\left(1 + \frac{1}{2}\|\sigma - \rho\|_1\right) \tag{15.21}$$

so $q(1) - p(1) = \frac{1}{2}\|\sigma - \rho\|_1$, and $q = p$ iff $\|\sigma - \rho\|_1 = 0$ iff $\sigma = \rho$. □

15.5 Entropic inequalities

Example sheet 3 asks you to show

Proposition 17. $I(X : Y) = D(P_{XY}\|P_X P_Y)$.

Therefore, Gibb's inequality tells us that $I(X : Y) \geq 0$ with equality iff $P_{XY} = P_X P_Y$. From this and the definitions of the classical quantities in section 15.3 we obtain

Proposition 18. For any random variables X, Y and Z (taking values in finite sets):

1. $I(X : Y) \geq 0$ with equality iff X and Y are independent, which means

$$\forall x, y : P_{XY}(x, y) = P_X(x)P_Y(y).$$

2. $I(X : Y|Z = z) \geq 0$ with equality iff X and Y are conditionally independent given $Z = z$, which means

$$\forall x, y : P_{XY|Z=z}(x, y) = P_{X|Z=z}(x)P_{Y|Z=z}(y).$$

3. $I(X : Y|Z) \geq 0$ with equality iff X and Y are conditionally independent given Z, which means

$$\forall z \in \mathrm{supp}(P_Z), x, y : P_{XY|Z}(x, y|z) = P_{X|Z}(x|z)P_{Y|Z}(y|z).$$

Example sheet 3 also asks you to show

Proposition 19. $I(\mathsf{A} : \mathsf{B})_\rho := D(\rho_{\mathsf{AB}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}})$.

So Klein's inequality gives us

Proposition 20. For any state ρ_{AB}, $I(\mathsf{A} : \mathsf{B})_\rho \geq 0$ with equality iff $\rho_{\mathsf{AB}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$.

Quantum *conditional* mutual information is also positive, a fact known as the **strong subadditivity** of von Neumann entropy. Proving this from scratch is much more involved than the classical case, but it is a fairly easy consequence of the data processing inequality for quantum relative entropy.

Theorem 21. For any state ρ_{ABC}, $I(\mathsf{A}:\mathsf{B}|\mathsf{C}) \geq 0$.

Proof. By the chain rule and Proposition 19,

$$\begin{aligned} I(\mathsf{A}:\mathsf{B}|\mathsf{C}) =& I(\mathsf{A}:\mathsf{BC}) - I(\mathsf{A}:\mathsf{C}) \\ =& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}}\otimes\rho_{\mathsf{BC}}) - D(\rho_{\mathsf{AC}}\|\rho_{\mathsf{A}}\otimes\rho_{\mathsf{C}}) \\ =& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}}\otimes\rho_{\mathsf{BC}}) - D(\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{ABC}}\|\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{A}}\otimes\rho_{\mathsf{BC}}) \\ \geq& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}}\otimes\rho_{\mathsf{BC}}) - D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}}\otimes\rho_{\mathsf{BC}}) = 0, \end{aligned}$$

where the inequality is the data processing inequality for quantum relative entropy. □

Proposition 22. If ρ_{AB} is a pure state, then $H(\mathsf{A})_\rho = H(\mathsf{B})_\rho$.

Proof. From the existence of a Schmidt decomposition for ρ_{AB} we know that ρ_{A} and ρ_{B} have the same non-zero eigenvalues with the same multiplicities, and therefore have the same von Neumann entropy. □

Theorem 23. Data processing inequality for the coherent information:
If $\sigma_{\mathsf{AC}} = \mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$ for some operation $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}$ then $I(\mathsf{A}\rangle\mathsf{B})_\rho \geq I(\mathsf{A}\rangle\mathsf{C})_\sigma$.

Proof. Let ρ_{RAB} be a purification of ρ_{AB}, and let $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}X_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{E}}VX_{\mathsf{B}}V^\dagger$ be a Stinespring representation for $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}$ (so V is an isometry in $\mathcal{L}(\mathcal{H}_{\mathsf{B}}, \mathcal{H}_{\mathsf{C}}\otimes\mathcal{H}_{\mathsf{E}})$). Defining

$$\psi_{\mathsf{RACE}} = \mathbb{1}_{\mathsf{RA}}\otimes V\rho_{\mathsf{RAB}}\mathbb{1}_{\mathsf{RA}}\otimes V^\dagger,$$

we have $\mathrm{Tr}_{\mathsf{RE}}\psi_{\mathsf{RACE}} = \sigma_{\mathsf{AC}}$, so let's write $\sigma_{\mathsf{RACE}} := \psi_{\mathsf{RACE}}$. Note that σ_{RACE} is a pure state. Now, the coherent information before the operation (the 'data processing') is

$$I(\mathsf{A}\rangle\mathsf{B})_\rho = H(\mathsf{B})_\rho - H(\mathsf{AB})_\rho \overset{(a)}{=} H(\mathsf{RA})_\rho - H(\mathsf{R})_\rho \overset{(b)}{=} H(\mathsf{RA})_\sigma - H(\mathsf{R})_\sigma = H(\mathsf{A}|\mathsf{R})_\sigma, \quad (15.22)$$

where (a) is by two applications of Proposition 22 to the pure state ρ_{RAB}, (b) is because $\rho_{\mathsf{RA}} = \sigma_{\mathsf{RA}}$, and the other equalities are true by definition. After the operation,

$$\begin{aligned} I(\mathsf{A}\rangle\mathsf{C})_\sigma =& H(\mathsf{C})_\sigma - H(\mathsf{AC})_\sigma \overset{(c)}{=} H(\mathsf{RAE})_\sigma - H(\mathsf{RE})_\sigma \\ =& H(\mathsf{AE}|\mathsf{R})_\sigma + H(\mathsf{R})_\sigma - (H(\mathsf{E}|\mathsf{R})_\sigma + H(\mathsf{R})_\sigma) = H(\mathsf{AE}|\mathsf{R})_\sigma - H(\mathsf{E}|\mathsf{R})_\sigma, \end{aligned} \quad (15.23)$$

where (c) is by two applications of Proposition 22 to the pure state ρ_{RACE} and the other equalities are definitions. Subtracting equation (15.23) from (15.22) yields

$$I(\mathsf{A}\rangle\mathsf{B})_\rho - I(\mathsf{A}\rangle\mathsf{C})_\sigma = H(\mathsf{A}|\mathsf{R})_\sigma + H(\mathsf{E}|\mathsf{R})_\sigma - H(\mathsf{AE}|\mathsf{R})_\sigma = I(\mathsf{A}:\mathsf{E}|\mathsf{R}) \geq 0$$

by definition and positivity of QCMI. □

Theorem 24. Data processing inequality for the quantum mutual information:
If $\sigma_{\mathsf{A'B'}} = \mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}}\otimes\mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$ where $\mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}}$ and $\mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}$ are operations then

$$I(\mathsf{A}:\mathsf{B})_\rho \geq I(\mathsf{A'}:\mathsf{B'})_\sigma.$$

Proof. With $\rho'_{\mathsf{AB'}} := \mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}}\otimes\mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$, $\sigma_{\mathsf{A'B'}} = \mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}}\otimes\mathbf{id}^{\mathsf{B'}\leftarrow\mathsf{B'}}\rho'_{\mathsf{AB'}}$, Using $\rho'_{\mathsf{A}} = \rho_{\mathsf{A}}$ and $\rho'_{\mathsf{B'}} = \sigma_{\mathsf{B'}}$, the relationships between mutual information and coherent information, and applying the DPI for coherent information twice

$$\begin{aligned} & I(\mathsf{A}:\mathsf{B})_\rho = H(\mathsf{A})_\rho + I(\mathsf{A}\rangle\mathsf{B})_\rho \geq H(\mathsf{A})_{\rho'} + I(\mathsf{A}\rangle\mathsf{B'})_{\rho'} \\ =& I(\mathsf{A}:\mathsf{B'})_{\rho'} = H(\mathsf{B'})_{\rho'} + I(\mathsf{B'}\rangle\mathsf{A})_{\rho'} \geq H(\mathsf{B'})_\sigma + I(\mathsf{B'}\rangle\mathsf{A})_\sigma = I(\mathsf{A'}:\mathsf{B'})_\sigma. \end{aligned}$$

□

15.6 Fano's inequality

If $\mathcal{A}_X = \{0, 1\}$ then $H(X) = h(P_X(1))$ where h is the **binary entropy function**:

Definition 25. $h(\lambda) := (1-\lambda)\log\frac{1}{1-\lambda} + \lambda\log\frac{1}{\lambda}$.

Proposition 26 (Fano's inequality)**.** Given two random variables M and $\hat{M}$ which both take values in a finite set $\mathcal{A}$, we have

$$H(M|\hat{M}) \leq \Pr(\hat{M} \neq M)\log(|\mathcal{A}| - 1) + h(\Pr(\hat{M} \neq M)).$$

Proof. Let E be a binary RV which is 1 if $\hat{M} \neq M$ and 0 if $\hat{M} = M$. Using the chain rule and $H(E|M, \hat{M}) = 0$ (because E is a function of M and $\hat{M}$):

$$\begin{aligned} H(M|\hat{M}) =& H(M|\hat{M}, E) + H(E|\hat{M}) - H(E|M, \hat{M}) \\ =& H(M|\hat{M}, E = 1)\Pr(E = 1) + H(M|\hat{M}, E = 0)\Pr(E = 0) + H(E|\hat{M}) \end{aligned}$$

$H(M|\hat{M}, E = 0) = 0$ because, conditioned on $E = 0$ (i.e. $\hat{M} = M$) M is a function of $\hat{M}$. $H(M|\hat{M}, E = 1) \leq \log(|\mathcal{A}| - 1)$ because, for all $a \in \mathcal{A}$, the support of $P_{M|\hat{M}=a,E=1}$ is no larger than $|\mathcal{A}| - 1$. Finally, by positivity of mutual information, $H(E|\hat{M}) \leq H(E) = h(\Pr(\hat{M} \neq M))$. □