# 1 Quantum probability theory

The fundamental objects in discrete probability theory are distributions, which we can represent as nonnegative vectors $p \in \mathbb{R}_+^n$ with $\sum_{i=1}^n p_i = 1$. One can think of $p$ as describing the state of a statistical system with $n$ possible basic deterministic states. When we observe the system (i.e., sample from $p$), we obtain outcome $i \in \{1, 2, \ldots, n\}$ with probability $p_i$.

Observations of the system described by $p$ correspond to measurable events in the probability space. In the discrete case, these are particularly easy to describe: Consider a partition of the outcomes into $m$ sets:

$$\{1, 2, \ldots, n\} = W_1 \cup W_2 \cup \cdots \cup W_m. \tag{1.1}$$

Then we obtain outcome $W_i$ with probability $p(W_i) := \sum_{x \in W_i} p(x)$ and, upon observing that the outcome is $i$, the state of the system is described by the corresponding conditional probability distribution:

$$p(x \mid W_i) = \frac{p(x)\mathbf{1}_{W_i}(x)}{p(W_i)}.$$

**Density matrices.** The analogous concept in "quantum probability" is that of a *density matrix:* $\rho \in \mathbb{M}_n(\mathbb{C})$ satisfying $\rho \succeq 0$ and $\mathrm{Tr}(\rho) = 1$. In this case, the classical distribution $p$ might correspond to a diagonal matrix $\rho = \mathrm{diag}(p)$.

**Measurements.** To observe the system described by $\rho$, one makes a *measurement* which is specified by a decomposition of the identity (compare (1.1)):

$$\sum_{i=1}^{m} U_i^* U_i = I.$$

This measurement has $m$ outcomes, and outcome $i \in \{1, 2, \ldots, m\}$ occurs with probability $\mathrm{Tr}(U_i \rho U_i^*)$. Conditioned on outcome $i$, the resulting state of the system is described by the density matrix

$$\frac{U_i \rho U_i^*}{\mathrm{Tr}(U_i \rho U_i^*)}.$$

**Pure states.** The set $\mathcal{D} \subseteq \mathbf{H}_n$ of all density matrices is a convex set. The extreme points are called *pure states.* These are precisely the rank-1 matrices $uu^*$ with $u \in \mathbb{C}^n$ and $\|u\|_2 = 1$. Thus density matrices describe mixtures of pure states; the decomposition of a mixed state into pure states is not unique.

It is typical to specify a quantum system by starting with a Hilbert space $\mathcal{H} = \mathbb{C}^n$ of pure states (this is the analog of the set $\{1, 2, \ldots, n\}$ in the classical case), and then considering the set of mixed states over $\mathcal{H}$, which are precisely the linear operators on $\mathcal{H}$ that are positive semidefinite and have unit trace. Define $\mathcal{D}(\mathcal{H}) := \{\rho : \mathcal{H} \to \mathcal{H} \text{ linear} : \rho \succeq 0, \mathrm{Tr}(\rho) = 1\}$.

**Composite systems.** Consider two classical statistical systems $S_1$ and $S_2$ with $m$ and $n$ outcomes, respectively. We can describe the state of $S_1$ by $p_1 \in \mathbb{R}^m$ and the state of $S_2$ by $p_2 \in \mathbb{R}^n$. The joint statistical state is *not* described by the pair $(p_1, p_2) \in \mathbb{R}^m \times \mathbb{R}^n$, unless the systems are independent. Instead, the joint state is given by $q \in \mathbb{R}^m \otimes \mathbb{R}^n$ since we need to assign a probability to all $mn$ pairs of outcomes. If the systems are independent, their state is given by $p_1 \otimes p_2$.

Similarly, if $\mathcal{H}_A = \mathbb{C}^m$ and $\mathcal{H}_B = \mathbb{C}^n$ are two Hilbert spaces, the composite Hilbert space is $\mathcal{H}_A \otimes \mathcal{H}_B$, and the set of mixed states over the composite system is $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \subseteq \mathbb{M}_{mn}(\mathbb{C})$. If the two systems are independent and described by $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ and $\rho^B \in \mathcal{D}(\mathcal{H}_B)$, then the joint state is given by $\rho^{AB} = \rho^A \otimes \rho^B$.

**Marginal distributions and the partial trace.** If $q \in \mathbb{R}^m \otimes \mathbb{R}^n$ describes a joint probability distribution on the space of outcomes $[m] \times [n]$, then we can consider the marginal distributions $q_1$ and $q_2$ induced when we consider only the first or second system, e.g.,

$$q_1(x) = \sum_{y \in [n]} q(x, y).$$

In the quantum setting, there is a similar operation called the *partial trace.* Given a density $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the corresponding marginal density $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ should yield the same outcome for all measurements on the joint system $\mathcal{H}_A \otimes \mathcal{H}_B$ that ignore the $B$ component, i.e.,

$$\operatorname{Tr}\left((U \otimes I)\rho\right) = \operatorname{Tr}\left(U\rho^A\right),$$

and similarly for the marginal density on the $B$-component:

$$\operatorname{Tr}\left((I \otimes V)\rho\right) = \operatorname{Tr}\left(V\rho^B\right).$$

Let us define partial trace operators $\operatorname{Tr}_A : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_B)$ and $\operatorname{Tr}_B : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_A)$ that "trace out" the corresponding component so that $\rho^A = \operatorname{Tr}_B(\rho)$ and $\rho^B = \operatorname{Tr}_A(\rho)$. These should satisfy: For all $U, V$:

$$\operatorname{Tr}_B(U \otimes V) = U\operatorname{Tr}(V), \quad \operatorname{Tr}_A(U \otimes V) = \operatorname{Tr}(U)V.$$

The name "partial trace" comes from thinking of an element $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as a block matrix. Let us use greek letters $\alpha, \beta$ to index the $A$-system and arabic letters $i, j$ for the $B$-system. Then it is possible to think of $\rho$ as an $m \times m$ block matrix where $\rho_{\alpha\beta} \in \mathcal{D}(\mathcal{H}_B)$ for $\alpha, \beta \in [m]$. In this case,

$$\operatorname{Tr}_A(\rho) = \sum_{\alpha=1}^{m} \rho_{\alpha\alpha} \in \mathcal{D}(\mathcal{H}_B).$$

Note that

$$\operatorname{Tr}(\operatorname{Tr}_A(\rho)) = \sum_{i=1}^{n} \left(\operatorname{Tr}_A(\rho)\right)_{ii} = \sum_{i=1}^{n} \left(\sum_{\alpha=1}^{n} (\rho_{\alpha\alpha})_{ii}\right) = \operatorname{Tr}(\rho).$$

If we instead represent $\rho$ as a block matrix where $\rho_{ij} \in \mathcal{D}(\mathcal{H}_A)$ for each $i, j \in [n]$, then

$$\operatorname{Tr}_B(\rho) = \sum_{i=1}^{n} \rho_{ii}.$$

**Entanglement.** A classical probability distribution on $[m] \times [n]$ is independent between the two subsystems if the joint distribution is a tensor: $q = p_1 \otimes p_2$. Equivalently, if $q(x, y) = p_1(x)p_2(y)$ for all $x \in [m], y \in [n]$. One can easily check that the set of probability distributions on $[m] \times [n]$ is precisely the convex hull of the set of independent distributions $p_1 \otimes p_2$ as $p_1$ and $p_2$ range over distribution on $[m]$ and $[n]$, respectively.

Similarly, we can consider the collection $\mathscr{S} \subseteq \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ of operators of the form $\sum_i \lambda_i (U_i \otimes V_i)$ where $\lambda_i \geqslant 0$, $\sum_i \lambda_i = 1$, and $U_i \in \mathcal{D}(\mathcal{H}_A)$, $V_i \in \mathcal{D}(\mathcal{H}_B)$ for all $i$. These are called *separable states*, and they represent states that are *classically* correlated across the $A$-$B$ partition.

The biggest novelty of quantum information is that there are states whose correlations on non-classical: The elements of $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \setminus \mathscr{S}$ are called *entangled states*. As a simple example, consider the unit vector $v \in \mathbb{C}^2 \otimes \mathbb{C}^2$ given by

$$v = \frac{1}{\sqrt{2}} \left( e_1 \otimes e_2 - e_2 \otimes e_1 \right),$$

and the corresponding density matrix $\rho = vv^*$. (This is a pure entangled state.) Another example of a *maximally entangled state* in $\mathcal{D}(\mathbb{C}^n \otimes \mathbb{C}^n)$:

$$\rho = \frac{1}{n^2} \sum_{i,j=1}^{n} e_{ij} \otimes e_{ij}.$$

**State transformations.** Suppose we have a density $\rho \in \mathcal{D}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. This state could then be made to interact with some external system that is itself represented by a density $\rho^{\text{env}} \in \mathcal{D}(\mathcal{H}_{\text{env}})$. Quantum dynamics are described by unitary matrices, hence the joint state after interaction and "processing" will be of the form

$$U(\rho \otimes \rho^{\text{env}})U^*,$$

for some unitary $U$ acting on $\mathcal{H} \otimes \mathcal{H}_{\text{env}}$. Finally, we can look at the marginal state of our system after this, which corresponds to tracing out the environment

$$\tilde{\rho} = \text{Tr}_{\text{env}}(U(\rho \otimes \rho^{\text{env}})U^*) \in \mathcal{D}(\mathcal{H}).$$

This is called a "state transformation."

Let us examine the general form of such a transformation. Suppose that $\mathcal{H} = \mathbb{C}^n$ and $\mathcal{H}_{\text{env}} = \mathbb{C}^m$. Since the mapping $\rho \mapsto \tilde{\rho}$ is linear in $\rho^{\text{env}}$, we can assume that $\rho^{\text{env}} = zz^*$ for some unit vector $z \in \mathbb{C}^m$. Let us write the unitary $U$ in block matrix form where $U_{ij} \in \mathbb{M}_n(\mathbb{C})$. (We will write all operators on $\mathbb{C}^n \otimes \mathbb{C}^m$ in this form.) Then we have:

$$\begin{aligned}
\tilde{\rho} &= \text{Tr}_{\text{env}} \left( U(\rho^{\text{env}} \otimes \rho)U^* \right) = \sum_i \left( U(\rho \otimes \rho^{\text{env}})U^* \right)_{ii} \\
&= \sum_{i,k,\ell} U_{ik} (\rho \otimes \rho^{\text{env}})_{k\ell} (U^*)_{\ell i} = \sum_{i,k,\ell} U_{ik} (z_k \bar{z}_\ell) \rho (U_{i\ell})^* = \sum_i \left( \sum_k z_k U_{ik} \right) \rho \left( \sum_\ell z_\ell U_{i\ell} \right)^*.
\end{aligned}$$

If we denote $A_i := \sum_k z_k U_{ik}$, then we have written

$$\tilde{\rho} = \sum_i A_i \rho A_i^*,$$

and these operators satisfy

$$\sum_i A_i^* A_i = \sum_{i,k,\ell} \bar{z}_k z_\ell U_{ik}^* U_{i\ell} = \sum_{k,\ell} \bar{z}_k z_\ell \sum_i U_{ik}^* U_{i\ell} = \sum_{k,\ell} \bar{z}_k z_\ell \mathbf{1}_{\{k=\ell\}} I_n = \left( \sum_k |z_k|^2 \right) I_n = I_n,$$

where we used the fact that $U$ is unitary so that $U^*U = I$, hence

$$\sum_i U_{ik}^* U_{i\ell} = \sum_i (U_{ki})^* U_{i\ell} = (U^*U)_{k\ell} = \begin{cases} I_n & k = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

We have thus proved the difficult part of the next theorem; we leave the converse as an exercise.

**Theorem 1.1.** *Every state transformation $\rho \mapsto \mathcal{E}(\rho)$ as above can be written as*

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^*$$

*with $\sum_i A_i^* A_i = I$. And, conversely, all such $\mathcal{E}$ are state transformations.*

## 1.1 The von Neumann entropy

The von Neumann entropy of a density $\rho \in \mathcal{D}(\mathbb{C}^n)$ is defined as

$$\mathcal{S}(\rho) = -\mathrm{Tr}(\rho \log \rho).$$

This is precisely the Shannon entropy of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\rho$:

$$H(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^n \lambda_i \log \frac{1}{\lambda_i},$$

with our standard convention that $0 \log 0 = 0$. It is straightforward to verify that $H$ is nonnegative and strongly concave, with the unique maximum occurring at the uniform distribution $H(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}) = \log n$.

**Purification and monotonicity of entropy.** In the classical world, we have monotonicity of entropy: Suppose that $q \in \mathbb{R}^m \otimes \mathbb{R}^n$ is a joint distribution on $[m] \times [n]$ and that $q_1$ and $q_2$ are the marginals. Then:

$$H(q_1) \leqslant H(q). \tag{1.2}$$

To see this, let us define the conditional distribution $q_2^x$ on $[n]$ by

$$q_2^x(y) = \frac{q(x,y)}{\sum_{y \in [n]} q(x,y)}.$$

One first verifies the chain rule:

$$H(q) = H(q_1) + \sum_{x \in [m]} q_1(x) H(q_2^x), \tag{1.3}$$

and then uses that all the terms in the latter sum are nonnegative, yielding (1.2). Note that (1.3) is usually written more simply as

$$H(X, Y) = H(X) + H(Y \mid X),$$

where $X, Y$ are random variables on $[m]$ and $[n]$, respectively.

This monotonicity property fails resoundingly in the quantum setting. Note that if $\rho = uu^*$ is a pure state, then $\mathcal{S}(\rho) = 0$. On the other hand, one has the following fact.

**Lemma 1.2** (Purification). *For every state $\rho^A \in \mathcal{D}(\mathcal{H}_A)$, there is a Hilbert space $\mathcal{H}_B$ and a* **pure state** *$\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\rho^A = \mathrm{Tr}_B(\rho^{AB})$.*

In other words, every state is the partial trace (the "marginal") of a pure state. In particular the analogous monotonicity $\mathcal{S}(\rho^A) \leqslant \mathcal{S}(\rho^{AB})$ fails to hold.

To construct the purification, decompose $\rho = \sum_{i=1}^n \lambda_i v_i v_i^*$ in its eigenbasis, and let $\mathcal{H}_B$ be the Hilbert space spanned by an orthonormal basis $\{e_1, \dots, e_n\}$. Define

$$u^{AB} := \sum_i \sqrt{\lambda_i} v_i \otimes e_i, \qquad \rho^{AB} = u^{AB}(u^{AB})^* = \sum_{i,j} \sqrt{\lambda_i \lambda_j} (v_i \otimes e_i)(v_j \otimes e_j)^*.$$

Note that $\rho^{AB}$ can be interpreted as a block matrix where $\rho^{AB}_{ij}$ is the marix $\sqrt{\lambda_i \lambda_j} v_i v_j^*$. Therefore we have

$$\mathrm{Tr}_B(\rho_{AB}) = \sum_i (\rho^{AB})_{ii} = \sum_i \lambda_i v_i v_i^* = \rho.$$