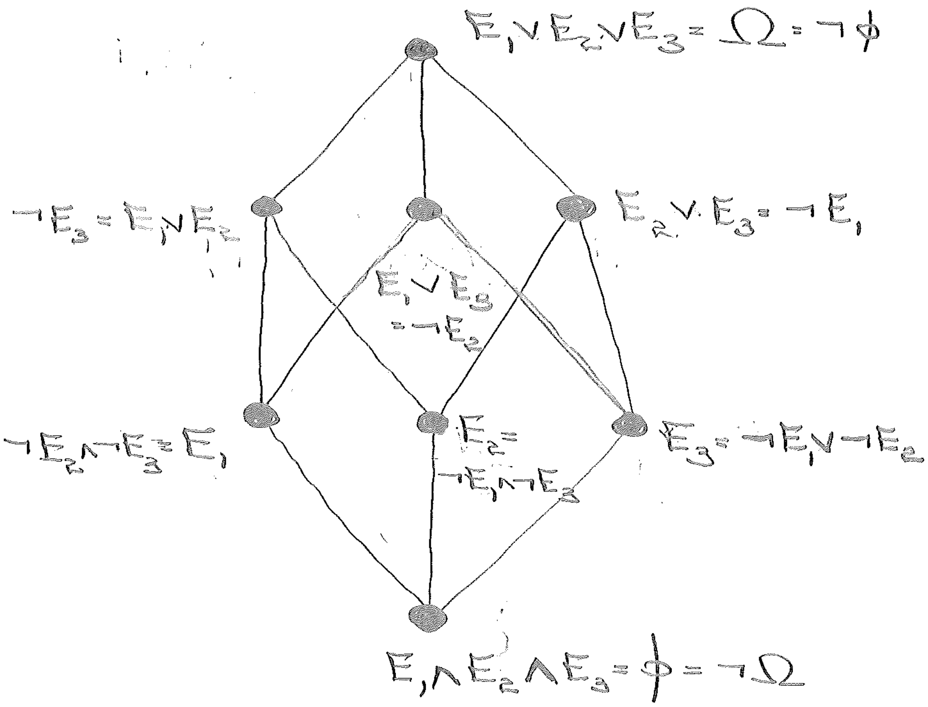
Quantum information theory

Lecture 1

Probabilities and laws of large numbers

Boolean lattice of events / propositions / alternatives

$\vee$	$\cup$
$\wedge$	$\cap$
$\neg$	complement
$\uparrow$	$\uparrow$
logical calculus	set-theoretic operations



$\longleftarrow$ atomic alternatives

Axioms of <u>probabilities</u> $\longleftarrow$ Foundations: See Dutch-book documents on web page.

measure on the event space

① $0 \le p(E) \le 1$

② E is certain $\Longleftrightarrow p(E) = 1$

③ Additivity for mutually <u>exclusive</u> events:

$$p(E_1 \vee E_2) = p(E_1) + p(E_2) \quad \text{if } E_1 \wedge E_2 = \phi$$

④ Bayes's rule: $p(E_1 \wedge E_2) = p(E_1|E_2)\,p(E_2)$

or normalization:

②+③ $\Rightarrow p(E) + p(\neg E) = p(E \vee \neg E) = p(\Omega) = 1$

normalization + ② + ④

$\Longrightarrow E = E_1 \vee E_2$

$$p(E) = \underbrace{[p(E_1|E) + p(E_2|E)]}_{1}\, p(E) = \overset{1}{\overset{\parallel}{p(E|E_1)}}\, p(E_1) + \overset{1}{\overset{\parallel}{p(E|E_2)}}\, p(E_2)$$

$$= p(E_1) + p(E_2)$$

(E)

Random variable X has values $\underbrace{x_1, \dots, x_N}_{\text{atomic alternatives}}$

$P_X(x)$

Two random variables: X, Y

Joint probability: $P_{X,Y}(x,y)$

Marginal probabilities: $P_X(x) = \sum_y P_{X,Y}(x,y)$

$$P_Y(y) = \sum_x P_{X,Y}(x,y)$$

Conditional probabilities:

$$\underbrace{P_{X|Y}(x|y)}_{\substack{\text{probability of } x, \\ \text{given that } Y=y}} P_Y(y) = P_{X,Y}(x,y) = P_{Y|X}(y|x)\, P_X(x)$$

Bayes's rule

We almost always omit the subscripts that indicate which random variable (s) the probability applies to, relying on the argument to provide that information.

$$p \longleftrightarrow P \longleftrightarrow Pr$$

Laws of large numbers

① Chebyshev inequality. For any nonnegative random variable X with finite mean, $P(x>a) \leq \langle x \rangle / a$.

Proof:

$$P(x>a) = \int_a^\infty dx\, p(x) \leq \int_a^\infty dx\, \frac{x}{a} p(x) \leq \int_0^\infty dx\, \frac{x}{a} p(x) = \frac{\langle x \rangle}{a}$$

$\Rightarrow$ ② For any random variable X with finite mean and variance,

$$P(|x-\langle x \rangle| > a) = P\left((x-\langle x \rangle)^2 > a^2\right) \leq \frac{\langle (\Delta x)^2 \rangle}{a^2}$$

$$\begin{pmatrix}\text{variance}\\ \text{of } X\end{pmatrix} = \langle (\Delta x)^2 \rangle = \langle (x-\langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \sigma_x^2$$

alternate notation we won't use much

Now let $X_1, \ldots, X_N$ be independent, identically distributed (i.i.d.) random variables with finite mean $\langle x \rangle$ and variance $\langle (\Delta x)^2 \rangle$. The sample mean

$$S = \frac{1}{N} \sum_{\ell=1}^{N} X_\ell$$

has mean

$$\langle s \rangle = \frac{1}{N} \sum_{\ell=1}^{N} \underbrace{\langle x_\ell \rangle}_{= \langle x \rangle} = \langle x \rangle$$

and second moment

$$\langle S^2\rangle = \frac{1}{N^2}\sum_{l,m}\underbrace{\langle x_l x_m\rangle}_{} = \langle x\rangle^2 + \frac{1}{N}\langle(\Delta x)^2\rangle$$

$$= \langle(\langle x\rangle + \Delta x_l)(\langle x\rangle + \Delta x_m)\rangle$$

$$= \langle x\rangle^2 + \underbrace{\langle \Delta x_l \Delta x_m\rangle}_{= \delta_{lm}\langle(\Delta x)^2\rangle} \longleftarrow \text{i.i.d.}$$

$$= \langle x\rangle^2 + \delta_{lm}\langle(\Delta x)^2\rangle$$

Thus the variance of the sample mean is

$$\langle(\Delta S)^2\rangle = \langle S^2\rangle - \langle S\rangle^2 = \frac{1}{N}\langle(\Delta x)^2\rangle$$

<u>Weak law of large numbers I.</u>

③ <u>Weak law of large numbers. II.</u>

$$P\big(|S-\langle x\rangle| > \epsilon\big) \underset{\text{Chebyshev inequality}}{\leq} \frac{\langle(\Delta S)^2\rangle}{\epsilon^2} = \frac{\langle(\Delta x)^2\rangle}{N\epsilon^2}$$

ⓐ This weak law says that

$$\lim_{N\to\infty} P\big(|S_N-\langle x\rangle| \leq \epsilon\big) = 1 \text{ for any } \epsilon > 0.$$

To prove the <u>asymptotic equipartition property</u> (AEP), we will use this form of the weak law with a random variable that takes on values

$p_l = -\log p(x_l)$.

ⓑ Relation to <u>central-limit theorem</u>?

Now let X be a discrete random variable that can take on D values $x_1, \dots, x_D$. Let $p_j = p(x_j)$ be the probability of the jth value. In N trials, the probability of a sequence $x_1, \dots, x_N$ is

$$p(x_1, \dots, x_N) = p(x_1) \dots p(x_D).$$

Let n_j be the number of occurrences of j $\left(\sum_{j=1}^{D} n_j = N \right)$. The probability of occurrence numbers $n_1, \dots, n_D$ is

$$p(n_1, \dots, n_D) = \frac{N!}{n_1! \dots n_D!} p_1^{n_1} \dots p_D^{n_D}$$

(i) Normalization: $\displaystyle\sum_{n_1, \dots, n_D} p(n_1, \dots, n_D) = (p_1 + \dots p_D)^N = 1$

(ii) Means:

$$\langle n_j \rangle = \sum_{n_1, \dots, n_D} n_j \, p(n_1, \dots, n_D) = \underbrace{p_j \frac{\partial}{\partial p_j} \left(\sum_{n_1, \dots, n_D} p(n_1, \dots, n_D) \right)} = N p_j$$

$$= p_j \frac{\partial}{\partial p_j} (p_1 + \dots + p_D)^N$$

$$= N p_j (p_1 + \dots + p_D)^{N-1}$$

$$= N p_j$$

(iii) Correlation matrix:

$$\langle n_j n_k \rangle = \sum_{n_1,\dots,n_D} n_j n_k\, p(n_1,\dots,n_D) = p_j \frac{\partial}{\partial p_j}\left(p_k \frac{\partial}{\partial p_k}\left(\sum_{n_1,\dots,n_D} p(n_1,\dots,n_D)\right)\right)$$

2nd-moment matrix

$$p_k \frac{\partial}{\partial p_k}\left(\sum_{n_1,\dots,n_D} p(n_1,\dots,n_D)\right) = N p_k (p_1 + \dots + p_D)^{N-1}$$

$$p_j \frac{\partial}{\partial p_j}\left(N p_k (p_1 + \dots + p_D)^{N-1}\right) = N p_j \delta_{jk} (\quad)^{N-1} + N(N-1) p_j p_k (\quad)^{N-2}$$

$$= N^2 p_j p_k + N(p_j \delta_{jk} - p_j p_k)$$

$$\langle n_j n_k \rangle = N^2 p_j p_k + N(p_j \delta_{jk} - p_j p_k)$$

The correlation matrix is

$$\langle \Delta n_j \Delta n_k \rangle = \langle (n_j - \langle n_j \rangle)(n_k - \langle n_k \rangle) \rangle$$

$$= \langle n_j n_k \rangle - \langle n_j \rangle \langle n_k \rangle$$

$$= N(p_j \delta_{jk} - p_j p_k)$$

Notice that for $j \neq k$,

$$\langle \Delta n_j \Delta n_k \rangle = - N p_j p_k$$

anticorrelation because occurrence numbers must add to N.

The variances are

$$\langle (\Delta n_j)^2 \rangle = N p_j (1 - p_j)$$

We want to work with the <u>frequency</u> of occurrence of each alternative:

$$f_j = \frac{n_j}{N}$$

$$\langle f_j \rangle = \frac{\langle n_j \rangle}{N} = p_j$$

$$\langle f_j f_k \rangle = \frac{\langle n_j n_k \rangle}{N^2} = p_j p_k + \frac{1}{N}(p_j \delta_{jk} - p_j p_k)$$

$$\langle \Delta f_j \Delta f_k \rangle = \frac{1}{N}(p_j \delta_{jk} - p_j p_k)$$

$$\langle (\Delta f_j)^2 \rangle = \frac{p_j(1 - p_j)}{N}$$

(4) Weak law of large numbers. III.

$$\underbrace{P\left(|f_j - p_j| \le \epsilon,\ j=1,\dots,D\right)}_{} \ge 1 - \frac{1-\sum_{j=1}^{D} p_j^2}{N\epsilon^2}$$

probability that all the frequencies lie within ϵ of the corresponding probability

We could do the frequency of each alternative separately but here we ask a question about all frequencies.

This weak law says that

$$\lim_{N\to\infty} P\left(|f_j^N - p_j| \le \epsilon,\ j=1,\dots,D\right) = 1 \text{ for any } \epsilon > 0.$$

Proof:

$$\underbrace{P\left(|f_j - p_j| \le \epsilon,\ j=1,\dots,D\right)}_{\text{hypercube with sides of length } 2\epsilon} \ge P\Big(\overbrace{\sum_j (f_j - p_j)^2 \le \epsilon^2}^{\text{Sphere of radius } \epsilon \text{ inside the hypercube}}\Big)$$

$$= 1 - \underbrace{P\left(\sum_j (f_j - p_j)^2 > \epsilon^2\right)}_{\le \frac{\sum_j \langle(\Delta f_j)^2\rangle}{\epsilon^2} = \frac{1-\sum p_j^2}{N\epsilon^2}}$$

Chebyshev inequality

$$\ge 1 - \frac{1-\sum_{j=1}^{D} p_j^2}{N\epsilon^2}$$

The weak laws refer to a limit of a sequence of probabilities. The strong law of large numbers deals with limits of the frequencies themselves. To illustrate this, let's consider the case of binary alternatives, 0 and 1. A sequence $x_1, \ldots, x_N$, has a frequency of occurrence of 1 given by

$$f_N = \frac{1}{N} \sum_{n=1}^{N} x_n \,.$$

The strong law deals directly with the limit of this frequency sequence and says that

$$P\Big(\lim_{N \to \infty} f_N = \underset{\substack{\uparrow \\ \text{probability for } 1}}{p}\Big) = 1 \,.$$

Many sequences don't have a frequency limit at all, so you have to show that there are sequences that do have a frequency limit and that these have all the probability (this latter task requires formulating a probability measure directly on the space of infinite sequences).

We never seem to use the strong law because ultimately we're interested in limits of probabilities on finite sets, which we can understand, not probabilities of limits, which have only a formal significance.