PHY 4105: Quantum Information Theory
Lecture 2

Anil Shaji
School of Physics, IISER Thiruvananthapuram
(Dated: August 6, 2013)

A. Laws of large numbers

The ability to estimate the probabilities $p(x)$ for a random variable taken on each of its allowed values depend on the various forms of the laws of large numbers. It tells you how the probability vector - whether it is an artifact of your mind or otherwise - can take shape based on a finite number of observational data on events, atomic or otherwise. Take care not to confuse laws of large numbers with the central limit theorem, which in many cases end up saying the same thing. In particular the weak law says that the mean taken over several instances of a random variable will converge to the expected value. It guarantees stable averages even for random events.

Chebyshev inequality: For any non-negative random variable X with finite mean,

$$P(x > a) \leq \frac{\langle x \rangle}{a}.$$

Proof

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

For a random variable X with a finite mean and variance

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

where the variance of X is

$$\langle (\Delta x)^2 \rangle = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2.$$

Let us consider N tosses of a six sided die. Each toss may be considered as a random variable, X_j which takes values 1 through 6. Alternatively one can think of N identical dies being tossed and the results of each being taken as random variable. What matters is that in both cases, the probability distribution for the outcomes is identical, either from trial to trial or from die to die as the case may be. Implicit in this statement is the fact that the next toss of the die does not depend on the previous one or the toss of one die does not depend on the other ones as the case may be. In other words X_1, $X_2, \ldots, X_N$ are *Independent, Identically Distributed* (iid) random variables. Furthermore let these random variables have a finite mean $\langle x \rangle$ and finite variance $\langle (\Delta x)^2 \rangle$.

We can define the a new variable; the sample mean,

$$S = \frac{1}{N} \sum_{j=1}^{N} X_j.$$

The sample mean has mean,

$$\langle s\rangle = \frac{1}{N}\sum_{j=1}^{N}\langle X_l\rangle = \langle x\rangle,$$

and second moment

$$\langle s^2\rangle = \frac{1}{N^2}\sum_{j,k}\langle X_j X_k\rangle = \langle x\rangle^2 + \frac{1}{N}\langle(\Delta x)^2\rangle,$$

using

$$\sum_{jk}\langle X_j X_k\rangle = \langle(\langle x\rangle + \Delta X_j)(\langle x\rangle + \Delta X_k)\rangle = \langle x\rangle^2 + \langle \Delta X_j \Delta X_k\rangle = \langle x\rangle^2 + \delta_{jk}\langle(\Delta x)^2\rangle.$$

So the variance of the sample mean is

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

This is the first form of the weak law of large numbers that we are going to learn.

A different form for the weak law is

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

The central limit theorem tells you a bit more than what the weak law of large numbers does. It tells you not only the mean and the variance of the sample mean but it tells that the sample mean is a random variable distributed according to the Normal distribution. Let us push things a bit forward and ask how the measured frequencies $f_i = n_i/N$ obtained from the number of times n_i that a random variable X took on a value x_i out of N trials are connected to the probabilities P_i associated with that outcome. Let the possible values taken by the random variables X be labeled $x_1, x_2, \ldots, x_D$. In N trials we get a sequence $X_1, \ldots X_N$ of the D possible outcomes in which the value x_j occurs n_j times, x_k occurs n_k times and so on. We have

$$\sum_{j=1}^{N} n_j = N.$$

Since we are worried about the frequencies $f_j = n_j/N$ of the occurrence of each of the D values, let us look at the probability that the sequence we obtain has occurrence numbers $n_1, \ldots, n_D$:

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

Normalization of the probability distribution means that

$$\sum_{n_1,\ldots n_D} p(n_1, \ldots, n_D) = (p_1 + \cdots + p_D)^N = 1.$$

The mean values of the occurrence number can be computed as

$$\langle n_j\rangle = \sum_{n_1,\ldots n_D} n_j p(n_1, \ldots, n_D) = p_j \frac{\partial}{\partial p_j}\Big(\sum_{n_1,\ldots n_D} p(n_1, \ldots, n_D)\Big) = N p_j.$$

Similarly for the second moment matrix,

$$\begin{aligned}\langle n_j n_k\rangle &= \sum_{n_1,\ldots n_D} n_j n_k p(n_1,\ldots,n_D)\\ &= p_j\frac{\partial}{\partial p_j}\Big(p_k\frac{\partial}{\partial p_k}\Big(\sum_{n_1,\ldots n_D} p(n_1,\ldots,n_D)\Big)\Big)\\ &= p_j\frac{\partial}{\partial p_j}\Big(Np_k(p_1+\ldots+p_D)^{N-1}\Big)\\ &= Np_j\delta_{jk}(p_1+\ldots+p_D)^{N-1}+N(N-1)p_jp_k(p_1+\ldots+p_D)^{N-2}\\ &= N^2p_jp_k+N(p_j\delta_{jk}-p_jp_k).\end{aligned}$$

The correlation matrix:

$$\begin{aligned}\langle\Delta n_j\Delta n_k\rangle &= \langle(n_j-\langle n_j\rangle)(n_k-\langle n_k\rangle)\rangle\\ &= \langle n_jn_k\rangle-\langle n_j\rangle\langle n_k\rangle\\ &= N(p_j\delta_{jk}-p_jp_k).\end{aligned}$$

So the variances are

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

Now looking at the frequency of occurrence of each alternative, $f_j = n_j/N$,

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

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

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

So we have a third form of the weak law of large numbers,

$$P(|f_j-p_j|\le\epsilon,\ j=1,\ldots,D)\ge 1-\frac{1-\sum_{j=1}^D p_j^2}{N\epsilon^2}.$$

The probability that *all* the frequencies lie within ϵ of the corresponding probability is now bounded by the expression on the right side of the above inequality. This form of the weak law says that

$$\lim_{N\to\infty}P(|f_j-p_j|\le\epsilon,\ j=1,\ldots,D)=1,$$

for any $\epsilon>0$.

Proof

$$\begin{aligned}P(|f_j-p_j|\le\epsilon,\ j=1,\ldots,D) &\ge P(\sum_j(f_j-p_j)^2\le\epsilon^2)\\ &= 1-P(\sum_j(f_j-p_j)^2>\epsilon^2)\\ &\ge 1-\frac{\sum_j\langle(\Delta f_j)^2\rangle}{\epsilon^2}=\frac{1-\sum_j p_j^2}{N\epsilon^2}.\end{aligned}$$

The last inequality comes form Chebyshev inequality.

The strong law of large numbers deals directly with limits of a sequence of frequencies and is stated in the form

$$P(\lim_{N\to\infty} f_j^{(N)} = p_j) = 1,$$

while the weak laws refer to a limit of a sequence of probabilities (see the third form above). Typically we are interested in limits of probabilities and so we are interested in the weak laws and not on the strong laws.