Appendix A

Quick Review of Probability and Statistics

A.1 Randomness

Randomness plays an essential role in the theory of information–both classical and quantum. It is therefore useful for us to review basic concepts from probability and statistics.

What is randomness anyway? In the classical world randomness is related to ignorance. We lack knowledge of all the conditions and parameters needed to make accurate predictions. For example, flipping a coin and seeing if it lands face up or face down is considered random. But it isn't really random. If Bob watches Alice flip a coin and were able measure exactly how rapidly she made it spin and measure its initial upward velocity, he could predict (using Newton's laws of classical dynamics) how long it will be in the air and whether it will land face up or face down. In more complicated dynamical systems with several interacting degrees of freedom, the motion can be chaotic. Tiny changes in initial conditions (positions and velocities) can lead to large changes in the subsequent trajectory. Thus even though classical mechanics is completely deterministic, motion on long time scales can appear to be random.

Many computer programs rely on so-called random number generators. They are not actually random but rather chaotic iterative maps–they take an initial seed number and compute some complicated function of that seed to produce a new number. That number is then used as the input to the next

round of iteration. The results may look random and may even pass many statistical tests for randomness, but if an observer knows the program that was used and the starting seed, he or she can predict the entire sequence of numbers perfectly.

In quantum mechanics randomness is an ineluctable feature. It is not due to ignorance of initial conditions but rather is an essential part of the theory. Alice can prepare N truly identical copies of a quantum state and Bob can make a measurement of some physical property on each copy of that state and obtain truly random (not pseudo-random) results. The results are not random because of some 'hidden variable' whose value Alice forgot to fix or Bob failed to measure. They are truly random–it is impossible to predict the outcome of the measurement before it is performed.

A.2 Probabilities

The probability p_j of an event j is a non-negative number obeying $0 \leq p_j \leq 1$. The probabilities of all possible events (in a universe of M possible events) obeys the sum rule

$$\sum_{j=1}^{M} p_j = 1. \tag{A.1}$$

This simply means that one of the possible events (from the complete set of all possible mutually exclusive events) definitely happened.

As an example, suppose we have an M-sided die (die is the singular form of dice). On each face of the die is a number. Let x_j denote the number of the jth face of the die, and p_j be the probability that the jth face lands face up so that the number x_j is showing when the die is randomly tossed onto a table. We can define a so-called 'random variable' X to be the number that comes up when we toss the die. X randomly takes on one of the allowed values $x_j; j = 1, \ldots, M$. We can now ask simple questions like, what is the mean (i.e. average) value of X? This is also known as the expectation value of X and is often denoted by an overbar or by double brackets

$$\bar{X} = \langle\langle X \rangle\rangle = \sum_{j=1}^{M} p_j x_j. \tag{A.2}$$

This sum over all possible results weighted by their frequency of occurrence gives the value one would obtain by averaging the results of a huge number of trials of the experiment (of rolling the die).

As an example, suppose we have a standard cube-shaped die with the six faces numbered 1 through 6, that is $x_j = j$. We will take the 'measurement result' to be the number on the top face of the die after it stops rolling. If the die is fair, the probability of result x_j is $p_j = \frac{1}{6}$, and thus the mean value will be

$$\bar{X} = \sum_{j=1}^{6} p_j x_j = \sum_{j=1}^{6} \left(\frac{1}{6}\right) j = 3.5. \tag{A.3}$$

Exercise A.1. Consider a pair of standard six-sided die with faces numbered consecutively from 1 to 6. Assuming the dice are fair

a) What are the unique possible values for the sum of the two numbers showing on the top faces of the dice?

b) What is the probability that each of these unique possible values occurs?

For later purposes, it will also be useful to consider what happens when we have two independent rolls of the die. Let the random variable X_1 be the number that comes up on the first toss and let X_2 be the number that comes up on the second toss. What is the joint probability distribution for the two results? That is, what is the probability $\mathcal{P}(x_j, x_k)$ that $X_1 = x_j$ and $X_2 = x_k$? Because each result is drawn independently from the same probability distribution we simply have that the probability for a given result of two tosses is just the product of the probabilities of the individual results

$$\mathcal{P}(x_j, x_k) = p_j p_k. \tag{A.4}$$

This is simply the statement that the joint probability distribution *factorizes* into two separate distributions for the individual events (assuming the events

are independent of each other). From this it follows that

$$\begin{aligned}\langle\langle X_1 X_2\rangle\rangle &= \sum_{j=1}^{M}\sum_{k=1}^{M}\mathcal{P}(x_j,x_k)x_j x_k \\ &= \sum_{j=1}^{M}\sum_{k=1}^{M}p_j p_k x_j x_k \\ &= \left\{\sum_{j=1}^{M}p_j x_j\right\}\left\{\sum_{k=1}^{M}p_k x_k\right\} \\ &= \langle\langle X_1\rangle\rangle\langle\langle X_2\rangle\rangle = \bar{X}^2. \end{aligned} \tag{A.5}$$

We conclude that for independent (i.e. uncorrelated) random variables, the mean of the product is equal to the product of the means.

The random variables X_1, X_2 will not be independent if they correspond to the result from the *same* throw of the die. Then of course $X_2 = X_1 = X$ and we have a different result

$$\langle\langle (X)^2\rangle\rangle = \sum_{j=1}^{M} p_j \left(x_j\right)^2. \tag{A.6}$$

This is simply the mean of the square of the number that comes up when we toss the die. Does the mean of the square bear any relation to the square of the mean, $\bar{X}^2$? To find out, let us consider the so-called *variance* of the distribution, the mean of the square of the deviation of the random variable from its own mean

$$\begin{aligned}\sigma^2 \equiv \langle\langle \left(X-\bar{X}\right)^2\rangle\rangle &= \sum_{j=1}^{M} p_j \left(x_j-\bar{X}\right)^2 \\ &= \sum_{j=1}^{M} p_j \left\{(x_j)^2 - 2\bar{X}x_j + \bar{X}^2\right\} \\ &= \langle\langle X^2\rangle\rangle - \langle\langle X\rangle\rangle^2. \end{aligned} \tag{A.7}$$

In deriving this result we have used the fact that

$$\sum_{j=1}^{M} p_j \bar{X}^2 = \bar{X}^2 \sum_{j=1}^{M} p_j = \bar{X}^2, \tag{A.8}$$

and

$$\sum_{j=1}^{M} p_j 2\bar{X} x_j = 2\bar{X} \sum_{j=1}^{M} p_j x_j = 2\bar{X}^2. \tag{A.9}$$

Clearly the variance is non-negative $\sigma^2 \geq 0$ because $(X - \bar{X})^2$ can never be negative. Hence we conclude

$$\langle\langle X^2 \rangle\rangle \geq \langle\langle X \rangle\rangle^2. \tag{A.10}$$

The variance is a measure of the *width* of the probability distribution. Another related quantity is the standard deviation or 'root mean square deviation' of the random variable from its mean

$$\sigma \equiv \langle\langle \big(X - \bar{X}\big)^2 \rangle\rangle^{\frac{1}{2}}. \tag{A.11}$$

To understand better what we mean by the width of the distribution consider the following examples. If a six-sided die has the number 3 painted on every face the probability distribution has zero variance. The only number that ever comes up is 3 so the mean of the distribution is 3 and no result ever deviates from the mean. Similarly, if the die has the standard consecutive numbering of the faces from 1 to 6, then a wide variety of outcomes is possible. If the die is fair then all outcomes are equally likely and the probability distribution is wide as shown in the left panel of Fig. A.1. Suppose however that the die is unfair and yields the number 3 with probability $p_3 = 0.8$, and the other numbers with probability $p_1 = p_2 = p_4 = p_5 = p_6 = 0.04$. This distribution is plotted in the right panel of Fig. A.1. The distribution has wide support (i.e. is non-zero over the full range from 1 to 6) but is still sharply peaked at 3. Hence the variance is small but not zero.

Exercise A.2. A standard six-sided die has faces numbered consecutively from 1 to 6. Find the variance and standard deviation of the probability distribution associated with random throws of the die

a) Assuming the die is fair (as in the left panel of Fig. A.1).

b) Assuming the die is biased with the probabilities given in the right panel of Fig. A.1.

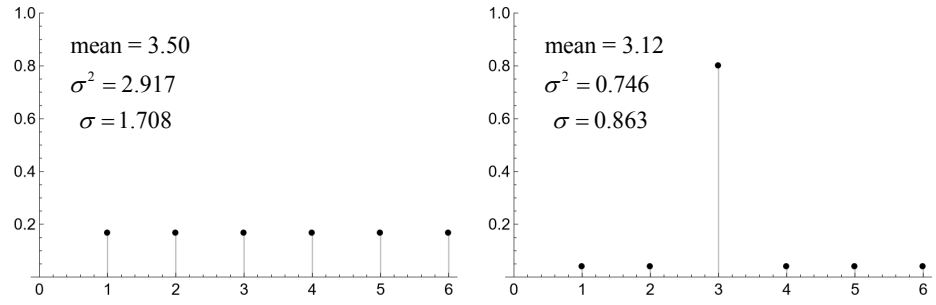

Figure A.1: Left panel: Graph of the probability distribution for the outcome of the throw of a fair die ($p_j = 1/6; j = 1, \ldots, 6$). The variance is large. Right panel: Graph of the probability distribution of an unfair (highly biased) die having $p_3 = 0.8$, and $p_1 = p_2 = p_4 = p_5 = p_6 = 0.04$. The variance is smaller.

A.3 Statistical Estimators

When performing experiments one may wish to attempt to measure the average value of some statistical quantity. Since we can execute only a finite number of runs of the experiment, we cannot be guaranteed to obtain the exact value of the mean, only an estimate.

For N trials we can form a so-called 'estimator' $\tilde{X}$ for $\bar{X}$, the mean value of a random variable via

$$\tilde{X}(N) = \frac{1}{N} \sum_{k=1}^{N} X_k, \tag{A.12}$$

where X_k is the number that came up in the kth run of the experiment (e.g. throw of the die). We use the tilde over the X to indicate that this is an estimator for the mean, not the true mean.

For finite N, this estimator is not likely to be exact, but for large N we expect it to become better and better. Is there a way to determine how accurate this estimator is likely to be? Indeed there is. Let us write

$$\tilde{X}(N) = \bar{X} + \delta_N, \tag{A.13}$$

where the error δ_N is given by

$$\delta_N = \frac{1}{N} \sum_{k=1}^{N} \left[X_k - \bar{X} \right] \tag{A.14}$$

If we were to ensemble average this estimator over a huge (formally infinite) number of experiments consisting of N throws we would obtain

$$\langle\langle \delta_N \rangle\rangle = \frac{1}{N}\sum_{k=1}^{N}\left[\langle\langle X_k\rangle\rangle - \bar{X}\right] = 0. \tag{A.15}$$

Thus the average error vanishes. That is, our estimator is *unbiased* as expected.

We can get a sense of the typical size of the error by considering its variance:

$$\begin{aligned}
\sigma_N^2 &\equiv \langle\langle (\delta_N)^2\rangle\rangle \\
\sigma_N^2 &= \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\langle\langle\left[X_j - \bar{X}\right]\left[X_k - \bar{X}\right]\rangle\rangle \\
&= \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\left\{\langle\langle X_j X_k\rangle\rangle - \bar{X}^2\right\} \\
&= \frac{1}{N^2}\sum_{j\neq k}^{N}\left\{\langle\langle X_j X_k\rangle\rangle - \bar{X}^2\right\} + \frac{1}{N^2}\sum_{j=k}^{N}\left\{\langle\langle X_j X_k\rangle\rangle - \bar{X}^2\right\} \\
&= \frac{1}{N^2}\sum_{j=1}^{N}\left\{\langle\langle (X_j)^2\rangle\rangle - \bar{X}^2\right\} \\
&= \frac{1}{N}\sigma_1^2,
\end{aligned} \tag{A.16}$$

where $\sigma_1 = \sigma$, the standard error for a single throw of the die defined in Eq. (A.11). Thus our estimator has a random error whose variance decreases inversely with N. The standard error thus is

$$\sigma_N = \frac{1}{\sqrt{N}}\sigma_1. \tag{A.17}$$

This is a simple estimate of the size of the error that our estimator of the mean is likely to have.

Box A.1. Sample variance vs. true variance Care must be exercised when estimating the variance σ_1^2 of an unknown probability distribution from a finite sample size drawn from the distribution. If (somehow) we know the true mean of the distribution, then we can simply use as our estimator of the variance

$$\tilde{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N}(X_j - \bar{X})^2. \tag{A.18}$$

It is straightforward to show that this is an unbiased estimator since

$$\langle\langle\tilde{\sigma}^2\rangle\rangle = \sigma_1^2. \tag{A.19}$$

The situation is not so simple when we do not know the mean of the unknown distribution and are forced to estimate it using Eq. (A.12). While this estimator of the mean is unbiased it still will have some small error and that error is positively correlated with the values of X_j in our sample. This means that if we use as our estimator of the variance

$$\tilde{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N}(X_j - \tilde{X}(N))^2, \tag{A.20}$$

it will be biased because it is too small. As an extreme example, consider the case $N = 1$. Our estimate of the mean is $\tilde{X}(N) = X_1$. If we substitute this for the mean in Eq. (A.20) we always obtain $\tilde{\sigma}^2 = (X_1 - X_1)^2 = 0$. A straightforward calculation shows that for general N

$$\langle\langle\tilde{\sigma}^2\rangle\rangle = \frac{N-1}{N}\sigma_1^2, \tag{A.21}$$

which is consistent with our result that it vanishes for $N = 1$. Therefore if we want to have an unbiased estimate of the variance we should use

$$\tilde{\sigma}_S^2 = \frac{N}{N-1}\tilde{\sigma}^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_j - \tilde{X}(N))^2.. \tag{A.22}$$

We will refer to this as the unbiased sample variance.

Let us return now to the question of estimating the mean of a distribution from N samples. We have an unbiased estimator and have already computed the variance of our estimator in Eq. (A.17). The question arises as to whether or not we could say something about the probability distribution of the error. For large N, the error δ_N in Eq. (A.14) is the sum of a large number of small random terms. By the *central limit theorem*, the error will be, to a good approximation, Gaussian distributed. That is, the probability distribution for the error is well approximated by a *continuous* distribution having *probability density*

$$P(\delta_N) = \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{1}{2\sigma_N^2}\delta_N^2}. \tag{A.23}$$

The interpretation of probability density for a continuous variable x is the following. $P(x)dx$ is the probability that the value of the random variable X lies between x and $x + dx$. The normalization condition on probability becomes an integral (see the discussion on Gaussian integrals in Box A.2)

$$\int_{-\infty}^{+\infty} dx\, P(x) = 1. \tag{A.24}$$

Summations over random variables such as in Eq. (A.14) can be interpreted as random walks. Suppose that Δ is a random variable with equal probability of being $\pm\epsilon$, where ϵ is a fixed step length. Then a random walk of N steps ends up a position

$$x = \sum_{j=1}^{N} \Delta_j, \tag{A.25}$$

where Δ_j is the value of the random variable Δ for the jth step.

In Fig. A.2 we see plots of the probability distribution for x for different values of N, together with the Gaussian approximation to it. We see that the Gaussian approximation is quite good even for modest values of N. This is the essence of the central limit theorem. The sum of a large number of random variables (with bounded variance) is well-approximated by a Gaussian distribution.

Exercise A.3 provides an opportunity to prove the central limit theorem for the particular case of a random walk. To see how this works, let us derive the *exact* probability distribution for a random walk of N steps, each

of size $\epsilon = 1/2$. We take N even so that the final position m of the walker is always an integer and lies in the interval $m \in [-N/2, +N/2]$. Let $p_\pm$ be the probability of stepping to the right or left respectively. Let the number of steps to the right be R and the number to the left be L. We have $N = R + L$ and the final position is given by $m = (R - L)\epsilon$. The probability of any given sequence of steps is

$$P = p_+^R p_-^L. \tag{A.26}$$

The number of different walks with R steps to the right and L steps to the left can be determined from combinatorics. Think of a string of N symbols, ℓ and r denoting the direction of each step in the random walk. There are altogether $N!$ permutations of the order of these symbols. However $L!$ permutations merely swap ℓ's with other ℓ's and so should not be counted as distinct walks. Similarly there are $R!$ permutations of the r's that should not be counted. The number of distinct walks $M(R, L)$ is therefore

$$M(R, L) \;=\; \frac{N!}{R!L!}. \tag{A.27}$$

The probability of ending up at $m = (R - L)\epsilon$ after $N = R + L$ (even) steps is therefore

$$P(N, m) = p_+^R p_-^L \frac{N!}{R!L!} = p_+^R p_-^{(N-R)} \left(\begin{array}{c} N \\ R \end{array} \right) \tag{A.28}$$

where the last expression is the binomial coefficient

$$\left(\begin{array}{c} N \\ R \end{array} \right) = \frac{N!}{R!(N - R)!}. \tag{A.29}$$

For this reason, this probability distribution is known as the binomial distribution. Notice that this expression is correctly normalized because

$$\sum_{m=-N}^{+N} P(N, m) = \sum_{R=0}^{N} p_+^R p_-^{(N-R)} \frac{N!}{R!(N - R)!} = (p_+ + p_-)^N = (1)^N = 1, \tag{A.30}$$

where we have used the binomial theorem to evaluate the sum.

If the final position is $m = (R - L)\epsilon$, we have (for $\epsilon = 1/2$) $R = (N + m)/2$ and $L = (N - m)/2$ so Eq. (A.28) can also be written

$$P(N, m) = p_+^{(N+m)/2} p_-^{N-m)/2} \frac{N!}{[(N + m)/2]!\,[(N - m)/2]!} \tag{A.31}$$

Exercise A.3. A random walker moving in one dimension takes steps of length $\epsilon = 1/2$ to the right ($x \to x + \epsilon$) with probability p and to the left ($x \to x - \epsilon$) with probability $q = 1 - p$. The walker starts at $x = 0$ and the position after N steps is simply the algebraic sum of all the steps. The final position lies in the interval $[-N\epsilon, +N\epsilon]$. If $\epsilon = 1/2$ and N is even, only integer positions are possible.

a) *Derive* exact expressions for the mean and variance of the position m after N steps by two different methods. Hint: You can either derive the appropriate properties of the binomial distribution, or you can use the fact that the individual walk steps have a mean and variance and are statistically independent. For the former it is useful to notice that $p_+ \frac{\partial}{\partial p_+}(p_+)^R = R(p_+)^R$.

b) Discuss and explain what happens to the variance as p approaches either 0 or 1.

c) Take the logarithm of $P(N, m)$ in Eq. (A.31) and use Stirling's (asymptotic*) expansion

$$\ln Z! \sim Z \ln Z - Z + \frac{1}{2} \ln(2\pi Z) \tag{A.32}$$

for the factorials in the binomial coefficients to show that for large N the binomial distribution with $q = p = 1/2$ approaches the Gaussian distribution

$$P_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}, \tag{A.33}$$

where the standard deviation is $\sigma = \sqrt{N}\epsilon = \sqrt{N}/2$. Hint: In evaluating the logarithm of the probability, keep terms of order x^2/N but neglect higher-order terms in $1/N$ such as $(x/N)^2$.

This constitutes a proof of the central limit theorem for this special case.

*Note that an asymptotic expansion does not mean that the *difference* between the LHS and RHS of Eq. (A.32) becomes arbitrarily close to zero for large N. It means that the *ratio* of the RHS to the LHS approaches unity for large N. Since both quantities are diverging, their difference can still be large even if their ratio approaches unity.

Box A.2. Gaussian Integrals A general Gaussian integral over a single variable has the form

$$I = \int_{-\infty}^{+\infty} dz\, e^{-az^2+bz}. \tag{A.34}$$

It turns out that integrals of this form (and its multi-dimensional generalization) are ubiquitous in physics, so it is handy to know how to carry them out. We begin by 'completing the square' by writing

$$I = \int_{-\infty}^{+\infty} dz\, e^{-a[z-\frac{b}{2a}]^2+\frac{b^2}{4a}}. \tag{A.35}$$

Shifting the dummy variable of integration to $x = z - \frac{b}{2a}$ we have

$$I = e^{\frac{b^2}{4a}} \int_{-\infty}^{+\infty} dx\, e^{-ax^2}. \tag{A.36}$$

It turns out to be easier to consider the square of the integral which we can write as

$$I^2 = \left[e^{\frac{b^2}{4a}}\right]^2 \int_{-\infty}^{+\infty} dxdy\, e^{-a[x^2+y^2]}. \tag{A.37}$$

This two-dimensional integral is readily carried out using polar coordinates r, θ with $r = \sqrt{x^2+y^2}$

$$I^2 = \left[e^{\frac{b^2}{4a}}\right]^2 \int_0^{\infty} 2\pi r dr\, e^{-ar^2}. \tag{A.38}$$

Defining $u = ar^2$ and using $du = 2ardr$ we have

$$I^2 = \frac{\pi}{a}\left[e^{\frac{b^2}{4a}}\right]^2 \int_0^{\infty} du\, e^{-u} \tag{A.39}$$

which yields the final result

$$I = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}}. \tag{A.40}$$

From this we see that the Gaussian probability distribution in Eq. (A.23) is properly normalized to unity.

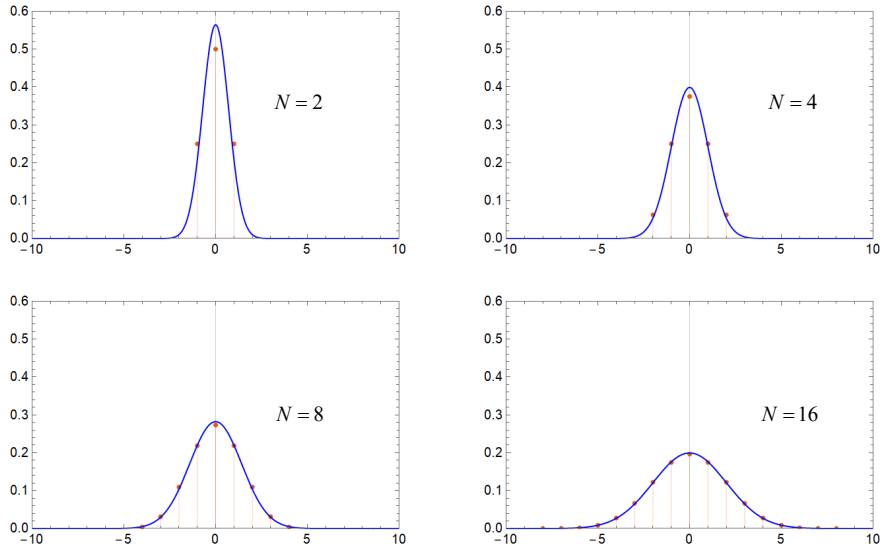

Figure A.2: Dots: Probability distribution for the ending position of random walks of N with step length $\epsilon = 1/2$. The smooth curve is the Gaussian distribution with standard deviation $\sqrt{N}/2$. The Gaussian approximation rapidly becomes very accurate as N increases.

A.4 Joint Probability Distributions

This section provides background for the discussion of multi-qubit measurements in Sec. 4.2.

Suppose that we have a probability distribution $P(A, B)$ for two random variables, A and B. A and B might for example be the measured values of the Pauli Z operator for each two qubits in a quantum computer. Thus for example $P(+1, -1)$ is the probability that the measured value of Z for q_0 is -1 and the measured value of Z for q_1 is $+1$. Altogether there are four possible measurement results for the combined measurement of the two qubits and P obeys the usual sum rule

$$\sum_{A=\pm 1} \sum_{B=\pm 1} P(A, B) = 1. \tag{A.41}$$

The probability distribution for measurement results on q_1 is simply

$$P_1(A) = \sum_{B=\pm 1} P(A, B). \tag{A.42}$$

That is, it is the sum of the probabilities of all the independent ways that q_0 can be found with measured value A independent of what B is. Notice that it follows immediately from the definition in Eq. (A.42) and the sum rule in Eq. (A.41) that this distribution is properly normalized

$$\sum_{A=\pm 1} P_1(A) = \sum_{A=\pm 1} \sum_{B=\pm 1} P(A,B) = 1. \tag{A.43}$$

Similarly, the probability distribution for measurement results on q_0 is

$$P_0(B) = \sum_{A=\pm 1} P(A,B), \tag{A.44}$$

and this too is properly normalized.

A.4.1 Conditional Probabilities and Bayes Rule

Another useful quantity is the *conditional probability distribution* of A given a fixed value of B which is denoted $P(A|B)$. You might naively think that this is simply given by

$$P(A|B) = P(A,B), \tag{A.45}$$

with B held fixed. This however is not properly normalized. The correctly normalized expression is

$$P(A|B) = \frac{P(A,B)}{\sum_{A=\pm 1} P(A,B)} \tag{A.46}$$

$$= \frac{P(A,B)}{P_0(B)}. \tag{A.47}$$

From this it follows that

$$P(A,B) = P(A|B)P_0(B). \tag{A.48}$$

Similarly, the probability distribution of B given a fixed value of A is

$$P(B|A) = \frac{P(A,B)}{\sum_{B=\pm 1} P(A,B)} \tag{A.49}$$

$$= \frac{P(A,B)}{P_1(A)}. \tag{A.50}$$

From this it follows that

$$P(A,B) = P(B|A)P_1(A). \tag{A.51}$$

Equating the above two expressions for $P(A,B)$ yields the very important Bayes rule:

$$P(B|A) = P(A|B)\frac{P_0(B)}{P_1(A)}. \tag{A.52}$$

This is especially powerful in situations (which occur frequently) in which $P(B|A)$ is hard to compute but $P(A|B)$ is easy to compute.

To understand Bayes rule better, let us consider the example discussed in Box 4.6 in which we given the problem of quantifying the information gain from making a measurement. To recap the setup of the problem: Alice randomly gives Bob one copy of one of the following four states

$$|\psi_0\rangle = |0\rangle \tag{A.53}$$
$$|\psi_1\rangle = |1\rangle \tag{A.54}$$
$$|\psi_2\rangle = |+\rangle \tag{A.55}$$
$$|\psi_3\rangle = |-\rangle \tag{A.56}$$

selected with equal probability. Here of course the Shannon entropy of the distribution prior to the measurement is $S_{\text{prior}} = 2\,\text{bits}$. Suppose Bob measures Alice's selected qubit in the Z basis and obtains the result $z = +1$. With this new knowledge he must update the probability distribution from the one prior to the measurement to the new one conditioned on the measurement result. The Born rule tells us immediately what the probability is for the measurement result being the observed $z = +1$ conditioned on the state being $|\psi_j\rangle$

$$P(z=+1|\psi_0) = 1 \tag{A.57}$$
$$P(z=+1|\psi_1) = 0 \tag{A.58}$$
$$P(z=+1|\psi_2) = \frac{1}{2} \tag{A.59}$$
$$P(z=+1|\psi_0) = \frac{1}{2}. \tag{A.60}$$

What we are after however is the reverse, namely the probability that the observed result $z = +1$ was obtained from state ψ_j, $P(\psi_j|z=+1)$. This is

precisely the situation where Bayes Rule is useful. We have from Eq. (A.52)

$$P(\psi_j|z=+1) = P(z=+1|\psi_j)\frac{P_\psi(\psi_j)}{P_z(z=+1)}, \tag{A.61}$$

where $P_\psi(\psi_j)$ is the prior probability that Alice selects state ψ_j. In this case, since Alice chooses each state with equal probability, we have $P_\psi(\psi_j) = \frac{1}{4}$ for all four values of j. $P_z(z)$ is the prior probability of obtaining the measurement result z. In this case, we find using Eqs. (A.57-A.60)

$$P_z(z=+1) = \sum_{j=0}^{3} P(z=+1|\psi_j)P_\psi(\psi_j) = \left[1+0+\frac{1}{2}+\frac{1}{2}\right]\left(\frac{1}{4}\right) = \frac{1}{2}. \tag{A.62}$$

We see that $P_z(z=+1)$ on the RHS of Eq. (A.62) is just a constant factor needed to fix the normalization of the posterior probability distribution on the LHS.

Combining all these results with Eq. (A.61) yields the results given in Box 4.6

$$P(\psi_0|z=+1) = \frac{1}{2} \tag{A.63}$$
$$P(\psi_1|z=+1) = 0 \tag{A.64}$$
$$P(\psi_2|z=+1) = \frac{1}{4} \tag{A.65}$$
$$P(\psi_3|z=+1) = \frac{1}{4}. \tag{A.66}$$
$$\tag{A.67}$$

As noted in Box 4.6, this probability distribution has Shannon entropy of (3/2) bits. Thus the information gain Bob receives from his measurement is

$$I = S_{\text{prior}} - S_{\text{post}} = \left(2-\frac{3}{2}\right)\text{ bits} = \frac{1}{2}\,\text{bit}.$$