APPPHYS225 - Tuesday 23 September 2008

Classical probability review, part 1

1. Discrete random variables as functions on a sample space
2. Probability distribution functions
3. Events
4. Algebras of random variables
5. Expectation, variance, and the notion of state
6. Matrix notation

In this class we will restrict our attenion to finite discrete probability models, in both the classical and quantum contexts. This will greatly simplify the mathematical and notational overhead. In what follows, we make references to the following freely-available documents (links from the course website):

> *Introduction to Probability* - Charles M. Grinstead and J. Laurie Snell
> *ACM217 notes: Stochastic Calculus, Filtering and Stochastic Control* - Ramon van Handel

We make these references not only to attribute quoted material, but also to refer the reader to expanded discussions in the referenced texts.

Textbook discussions of basic probability often start with the example of rolling a six-sided die. Assuming the die and its roller are deemed fair, we may assign equal probabilities to each of the six possible outcomes (number of spots on the side that faces up). Before the die is cast, we can ask simple questions such as:

- What is the probability that the result will be an even number?
- What is the probability that the result will be either 1 or 2?
- What is the probability that the result will not be 6?

All of you already know how to perform the simple calculations required to answer these correctly, so we will here focus instead on using the example of a six-sided die to establish some formal terminology and concepts that will help us eventually to understand the nature of the generalization from classical to quantum probability models. In particular our goal for today will be to understand that a classical probability model comprises a sample space, an algebra of random variables, and a probability state. We will also introduce a matrix notation for classical observables that can naturally be generalized to accommodate quantum probablity models.

Random variables as functions on a sample space

First we note that a single six-sided die, once it has been rolled and come to rest on a level surface, has exactly six possible *configurations* (also called *outcomes*) that are distinct and meaningful for our purposes. We will ignore the exact spatial position of the die and its precise orientation, paying attention only to which face is up. Let us abstractly identify the six possible outcomes resulting from a single die-roll with elements of the set

$$\boxed{\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\},}$$

where the subscript index corresponds to the number of spots on the side that finally faces up. The set of all possible outcomes is known as the *sample space*, and is usually denoted Ω. For reasons that we will discuss below, subsets of Ω are called *events*.

Having established the set of possible outcomes, we can now define *random variables* (also called *observables*) to be functions on Ω. If a given random variable takes values in a set S we call it an S-valued random variable; real- or integer-valued random variables are simply called random variables. For example consider the following random variables, specified first in terms of intuitive definitions and second as explicit functions on Ω :

- $X(\bullet)$: number of spots on the side facing up,

$$\boxed{X(\omega_1) = 1,\ X(\omega_2) = 2,\ X(\omega_3) = 3,\ X(\omega_4) = 4,\ X(\omega_5) = 5,\ X(\omega_6) = 6.}$$

- $Y(\bullet)$: sum of the numbers of spots on the five sides not facing up,

$$\boxed{Y(\omega_1) = 20,\ Y(\omega_2) = 19,\ Y(\omega_3) = 18,\ Y(\omega_4) = 17,\ Y(\omega_5) = 16,\ Y(\omega_6) = 15.}$$

- $Z(\bullet)$: value of the smallest prime number larger than the number of spots on the side facing up,

$$\boxed{Z(\omega_1) = 2,\ Z(\omega_2) = 3,\ Z(\omega_3) = 5,\ Z(\omega_4) = 5,\ Z(\omega_5) = 7,\ Z(\omega_6) = 7.}$$

The notation here is meant to emphasize the view of random variables as functions; note that these functions are not necessarily one-to-one. We will implicitly treat the one-to-one random variable $X(\bullet)$ as a special variable that indicates the numerical value of the die-roll, as this conforms to gambling convention, although in principle $Y(\bullet)$ or any other one-to-one random variable could play the same role.

Note that if we know the exact configuration of our system, we implicitly know the exact value of all observables. Similarly, if we know the exact value of any one-to-one random variable we can infer the configuration and thus the exact value of all other observables.

Probability distribution functions

Another special function on the set of outcomes is the *probability distribution function*, which we will denote $m(\bullet)$. This special function is defined by

$$\boxed{m(\omega_i) = \Pr(\omega_i),\ \forall i.}$$

Since we are assuming that this is a fair die-roll, we have $m(\omega_i) = 1/6$ for all i. Generally speaking, for a probability distribution function in any scenario we require

$$\boxed{0 \leq m(\omega_i) \leq 1, \qquad \sum_{\omega_i \in \Omega} m(\omega_i) = 1.}$$

Occasionally we may have cause to consider unnormalized probability distributions such that $\sum_{\omega_i \in \Omega} m(\omega_i) \neq 1$, and in such cases it is understood that $\Pr(\omega_i) = m(\omega_i) / \sum_{\omega_i \in \Omega} m(\omega_i)$.

Events

Our next step is to discuss the association of events (subsets of Ω) with yes-or-no questions about the outcome. Suppose I roll the six-sided die but do not show you the result. Any relevant yes-or-no question you could ask me regarding the outcome of the die-roll can be associated with a subset $E \subset \Omega$, such that I will say yes if and only if the actual outcome is in $E.$ For example:

- Was the result 1? $E = \{\omega_1\}$
- Was the result an even number? $E = \{\omega_2, \omega_4, \omega_6\}$
- Was the result 1 or 2? $E = \{\omega_1, \omega_2\}$
- Was the result not 6? $E = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$

Of course there is more than one way to formulate a yes-or-no question corresponding to a given set of elements:

- $E = \{\omega_1, \omega_2\}$: Was the result less than 3?
- $E = \{\omega_1\}$: Was the result not an even number, and less than 3?

Clearly the information content of the answer to a yes-or-no question regarding membership in a given subset depends only on the subset, and is independent of the precise way that the question is worded.

Note that knowledge of membership in a subset E implies knowledge of membership in the complementary subset $E^C.$ In words, this corresponds to the fact that the answer to a yes-or-no question implies the answer to the negation of this question. Likewise, if we have knowledge of membership in two different subsets E_1 and $E_2,$ then we can infer membership in the combined subsets $E_1 \cup E_2$ and $E_1 \cap E_2$ (exercise: write out the corresponding truth tables). It is useful to note in this context that

$$\boxed{E_1 \cap E_2 = \left(E_1^C \cup E_2^C\right)^C,}$$

which means that complementation and union are really the essential operations in this type of inference game. In any case we note that, if I allow you to ask about membership in a "starter collection" of subsets $\{E_i\},$ you can actually infer membership in a larger collection of subsets generated by complementation and union *(for explicit definitions of set complement, union, intersection and difference see Grinstead and Snell, p. 21-22)*.

Note that random variables can also serve to define events:

- Was the result of the die-roll such that $X = 5$? $E = \{\omega_5\}$
- Was the result of the die-roll such that $Y = 15$ or $Y = 16$? $E = \{\omega_5, \omega_6\}$
- Was the result of the die-roll such that $Z = 5$? $E = \{\omega_3, \omega_4\}$

Knowing the value of a random variable does not necessarily allow you to determine the exact configuration, but you can narrow it down to a subset of $\Omega.$ The term *level set* is commonly used to refer to the event that contains all configurations for which a random variable assumes a given value. Note that the level sets of a random variable are non-overlapping, and that the union of all level sets of a random variable is $\Omega.$

It is natural to extend the probability distribution function $m(\bullet)$ so that it is defined not only on elementary outcomes but also on events. Explicitly,

$$\boxed{m(E) = \sum_{\omega_i \in E} m(\omega_i).}$$

When viewed as a function from subsets to the reals, $m(\bullet)$ is often referred to as a *probability measure* (especially in scenarios with continuous random variables). It is easy to show that the following properties hold [Grinstead and Snell, Theorem 1.1]:

1. $m(E) \geq 0$ for every $E \subset \Omega.$
2. $m(\Omega) = 1.$

3. If $E \subset F \subset \Omega$ then $m(E) \leq m(F)$.

4. If A and B are disjoint subsets of $\Omega,$ then $m(A \cup B) = m(A) + m(B)$.

5. $m(A^C) = 1 - m(A)$ for every $A \subset \Omega$.

Here A^C indicates the complement of A in $\Omega,$ as in our above discussion of events.

Note that the probability distribution function thus induces probabilities for the values of random variables. If we define

$$\boxed{E_{A,a} = \{\omega_i : A(\omega_i) = a\}}$$

as the event $A = a$, then

$$\boxed{\Pr(A = a) = m(E_{A,a}) = \sum_{\omega_i \in E_{A,a}} m(\omega_i).}$$

For example $X = 5$ occurs only for $\{\omega_5\},$ so $\Pr(X = 5) = m(\omega_5) = 1/6.$ On the other hand, $\Pr(Z = 5) = m(\omega_3) + m(\omega_4) = 1/3.$

Algebras of random variables

Once we have defined some random variables, such as X, Y, Z, it is very easy to generate more (here we will assume that all random variables can be viewed as taking real values). Note that sums and products of random variables are themselves random variables, as are the products of random variables with real numbers. Hence, random variables have a natural algebraic structure. For example, if we define

$$\boxed{R(\bullet) = \alpha X(\bullet) + \beta Z(\bullet),}$$

with α, β real numbers, then

$$\begin{array}{|lll|} \hline R(\omega_1) = \alpha + 2\beta, & R(\omega_2) = 2\alpha + 3\beta, & R(\omega_3) = 3\alpha + 5\beta, \\ \hline R(\omega_4) = 4\alpha + 5\beta, & R(\omega_5) = 5\alpha + 7\beta, & R(\omega_6) = 6\alpha + 7\beta. \\ \hline \end{array}$$

Similarly,

$$\boxed{Z^2(\bullet) \equiv [Z(\bullet)]^2}$$

has values

$$\boxed{Z^2(\omega_1) = 4, \quad Z^2(\omega_2) = 9, \quad Z^2(\omega_3) = 25, \quad Z^2(\omega_4) = 25, \quad Z^2(\omega_5) = 49, \quad Z^2(\omega_6) = 49,}$$

and

$$\boxed{XZ(\bullet) = X(\bullet)Z(\bullet)}$$

has values

$$\boxed{XZ(\omega_1) = 2, \quad XZ(\omega_2) = 6, \quad XZ(\omega_3) = 15, \quad XZ(\omega_4) = 20, \quad XZ(\omega_5) = 35, \quad XZ(\omega_6) = 42.}$$

The probability distribution function on Ω clearly provides probability distribution functions for such random variables as well.

An *indicator function* $\chi_E(\bullet)$ of an event (subset) E is a random variable such that

$$\begin{array}{|rl|} \hline \chi_E(\omega_i) = 1, & \omega_i \in E, \\ \hline = 0, & \omega_i \notin E. \\ \hline \end{array}$$

Technically speaking, any random variable can be expressed in terms of indicator functions on its level sets:

$$R(\bullet) = \sum_i r_i \, \chi_{\Omega_{r_i}}(\bullet),$$

where $R(\bullet)$ takes values in the set $\{r_i\}$ and Ω_{r_i} is the level set corresponding to the value r_i. For example,

$$Z(\bullet) = 2\chi_{\{\omega_1\}}(\bullet) + 3\chi_{\{\omega_2\}}(\bullet) + 5\chi_{\{\omega_3,\omega_4\}}(\bullet) + 7\chi_{\{\omega_5,\omega_6\}}(\bullet).$$

It thus appears that indicator functions are like ‘basis functions’ for random variables. Note that for two events A and $B,$

$$\chi_{A\cap B}(\omega_i) = \chi_A(\omega_i)\chi_B(\omega_i).$$

Hence for a pair of random variables $R(\bullet)$ and $T(\bullet),$

$$R(\bullet)T(\bullet) = \left[\sum_i r_i \, \chi_{\Omega_{r_i}}(\bullet)\right]\left[\sum_j t_j \, \chi_{\Omega_{t_j}}(\bullet)\right] = \sum_{i,j} r_i t_j \chi_{\Omega_{r_i}\cap\Omega_{t_j}}(\bullet) = T(\bullet)R(\bullet).$$

Expectation, variance, and the notion of state

The *expectation* of a random variable $R(\bullet),$ which we will write $\langle R\rangle,$ is defined as

$$\langle R\rangle \equiv \sum_{\omega_i\in\Omega} R(\omega_i)m(\omega_i).$$

This is the average, or mean value of R with respect to the probability distribution function $m(\bullet).$ Note that for indicator functions,

$$\langle \chi_E\rangle = m(E).$$

Similarly, the *variance* of $R(\bullet)$ is defined as

$$\mathrm{var}[R] \equiv \langle R^2\rangle = \sum_{\omega_i\in\Omega} R^2(\omega_i)m(\omega_i) = \sum_{\omega_i\in\Omega} [R(\omega_i)]^2 m(\omega_i).$$

It is common also to define the *standard deviation* of $R(\bullet),$ also called the *uncertainty* of $R(\bullet),$ as

$$\mathrm{std}[R] \equiv \sqrt{\langle R^2\rangle - \langle R\rangle^2} = \sqrt{\sum_{\omega_i\in\Omega} [R(\omega_i)]^2 m(\omega_i) - \left[\sum_{\omega_i\in\Omega} R(\omega_i)m(\omega_i)\right]^2}.$$

It is common also to define the *covariance* of two random variables $A(\bullet)$ and $B(\bullet)$ as

$$\begin{aligned}\mathrm{cov}[A,B] &\equiv \langle (A-\langle A\rangle)(B-\langle B\rangle)\rangle = \langle AB\rangle - \langle A\rangle\langle B\rangle \\ &= \sum_{\omega_i\in\Omega} A(\omega_i)B(\omega_i)m(\omega_i) - \left[\sum_{\omega_i\in\Omega} A(\omega_i)m(\omega_i)\right]\left[\sum_{\omega_i\in\Omega} B(\omega_i)m(\omega_i)\right].\end{aligned}$$

It should be clear from these definitions that, in general, $\langle R^2\rangle \neq \langle R\rangle^2$ and $\langle AB\rangle \neq \langle A\rangle\langle B\rangle.$ If $\mathrm{cov}[A,B] = 0$ we say that $A(\bullet)$ and $B(\bullet)$ are *independent* random variables.

Formally, a *state* is a consistent assignment of an expectation value to every random variable in an algebra. It should be clear from the above that a state specifies variances and covariances by virtue of the fact that if $A(\bullet)$ and $B(\bullet)$ are random variables in our algebra, then so are $A^2(\bullet),$ $B^2(\bullet)$ and $AB(\bullet).$ The probability measure $m(\bullet)$ is a compact way of summarizing the state on an algebra of random variables. Note that state and configuration are quite different in our useage of the terms

- classically we assume that there exists an ‘actual’ configuration of the system in question (the actual disposition of the die after it has been rolled), which may or may not be known to anyone, but we also have a ‘state’ of knowledge/belief that summarizes the information we use to make predictions within a probabilistic framework.

Matrix notation

In a finite discrete setting, for which the sample space Ω contains N elements, it is natural to associate random variables with $N \times N$ real matrices. For an arbitrary random variable $R(\bullet)$, we simply place the values $R(\omega_i)$ along the diagonal and put zeros everywhere else. Hence, continuing with our example of the six-sided die:

$$X(\bullet) \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix}, \quad Z(\bullet) \leftrightarrow \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix}.$$

We use (X) to denote the matrix representation of a random variable $X(\bullet)$. With a bit of thought you can convince yourself that with this matrix representation, we can use the usual rules of matrix arithmetic and multiplication to carry out algebraic manipulations among random variables. For example,

$$\begin{aligned} R(\bullet) \equiv \alpha X(\bullet) + \beta Z(\bullet) \leftrightarrow \alpha &\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix} + \beta \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix} \\ &= \mathrm{diag}(\alpha + 2\beta, 2\alpha + 3\beta, 3\alpha + 5\beta, 4\alpha + 5\beta, 5\alpha + 7\beta, 6\alpha + 7\beta), \end{aligned}$$

where the **diag**$(\ldots)$ notation hopefully is obvious. Note that because of the fact that all matrices we use in this classical probability setting are diagonal, the matrix representations of an algebra of random variables form a commutative matrix algebra.

We note that the probability distribution can be written in exactly the same matrix notation, and that we thus arrive with the convenient expressions such as

$$
\begin{aligned}
\langle X \rangle &\equiv \sum_{\omega_i \in \Omega} X(\omega_i) m(\omega_i) \\
&\leftrightarrow \mathrm{Tr}\left[\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1/6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/6 \end{pmatrix} \right].
\end{aligned}
$$

We will use the suggestive notation $\rho \equiv \mathrm{diag}\,(m(\omega_1), \ldots, m(\omega_N))$ for the matrix representing the probability distribution function. Hence, in general, the expectation $\langle R \rangle$ of an arbitrary random variable R can be computed by taking the trace of the product of ρ with (R). The matrix ρ provides a convenient representation of a state for our algebra of random variables.

Indicator functions have a somewhat special appearance in this matrix notation, as they correspond to matrices with zeros and ones on the diagonal. Viewed as linear operators, they are therefore projection (idempotent) operators. For example, the indicator function $\chi_E(\cdot)$ for the event $E = \{\omega_1, \omega_2\}$ has matrix representation

$$
\chi_E(\cdot) \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},
$$

where clearly $(\chi_E)^2 = (\chi_E)$. It should be evident that the matrix representations of the indicator functions on all of the individual outcomes $\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_N\}$ provide a linear basis for the commutative matrix algebra representing all possible random variables on Ω. In particular,

$$
(R) = \sum_{i=1}^{N} R(\omega_i)(\chi_{\{\omega_i\}}).
$$

Hopefully, this perspective also highlights the fact that we can easily identify sub-algebras. For example if we think about the linear span of the matrix representations of indicator functions on $\{\omega_1, \omega_3, \omega_5\}$ and $\{\omega_2, \omega_4, \omega_6\}$, we obtain a closed matrix algebra for which the first, third and fifth diagonal elements are always the same, as are the second, fourth and sixth. It is only really two-dimensional. Exercise: determine the sample space for a single roll of two six-sided dice. Of what dimension are the matrix representations of random variables on this sample space?

Note that once we have obtained the matrix representations for the observables that we care about, and for the state, we can actually forget about Ω and the underlying configurations! Our original notion of random variables as functions on a sample space dictated the dimension of the matrix representations and their diagonality (required for multiplication to be commutative).

Summary points

1. Classical observables (random variables) can be viewed as functions on a sample space
2. Classical observables form commutative algebras with indicator-function bases
3. Classical observables can be represented by diagonal matrices (indicator functions by projectors)
4. Expectations are obtained by trace with a matrix representation of the state (probability measure)
5. Configuration refers to “physical reality”; state refers to knowledge/belief