Weekly Notes for EE2012 2013/14 – Week 4

T. J. Lim

February 5, 2015

Book sections covered this week: 3.1, 3.2, 3.3.

1 Discrete Random Variables

1.1 Definition and Concept

It is usually inconvenient to work with non-numerical sample spaces, because there is no unified way to address problems phrased in terms of heads and tails, colours, emotions, etc. However, we can easily turn non-numerical outcomes into numerical values through some sensible mapping function, and this is the core concept behind random variables. For instance,

- From the rolling of two dice, we may be interested in the *sum* of the numbers showing on the two top faces, rather than the pair of values obtained, and so we map the outcome (x, y) to $x + y$;
- In a sequence of n Bernoulli trials, we may be concerned about the total number of successes, rather than the exact sequence of trial outcomes, so we map e.g. (110) to 2;
- In a soccer match, we are interested in whether a team wins, loses or draws and we can map the outcomes as "win" $\to 0$, "lose" $\to 1$ and "draw" $\to 2$.

We will see that, with such a mapping, we can use our vast arsenal of algebraic tools to tackle all problems of a probabilistic nature in a unified way.

The mapping of outcomes in the sample space S to numerical values is very much like defining a function

$$X : S \longrightarrow S_X \subset \mathbb{R}. \tag{1}$$

To every element $\zeta \in S$, the function $X(\zeta)$ assigns a real number, much like $f(x) = y$ in a more familiar functional relationship. The number $X(\zeta)$ in an engineering setting will correspond to some physical or logical attribute of ζ.

Example 1: Roll two dice. Let $X(\zeta)$ be the total number of dots showing up on the two faces. The mapping between S and $\mathbb{R}$ can be described as

$$X((a,b)) = a + b \tag{2}$$

where $(a,b) \in S = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$. The range of X (i.e. the set of all values it can take) is denoted S_X, and is given by

$$S_X = \{2,3,\ldots,11,12\}. \tag{3}$$

Example 2: Toss 5 coins. Let $X(\zeta)$ be the total number of heads obtained. Then

$$X(\text{HHHHH}) = 5,\ \ X(\text{HHTTT}) = 2,\ \ X(\text{TTHTH}) = 2,\ \text{etc.} \tag{4}$$

and the range of X is given by $S_X = \{0,1,2,3,4,5\}$. ■

Since a random variable $X(\zeta)$ is nothing more than a function with S as its domain and some subset of $\mathbb{R}$ as its range, we can define any number of random variables from one underlying experiment. For instance, from rolling two dice, we can say that $X(\zeta)$ is the sum of the dots showing up, $Y(\zeta)$ is the larger of the two numbers, and $Z(\zeta)$ is the product of the two numbers.

It is conventional to drop the explicit dependency on ζ when referring to random variables, i.e. we write X rather than $X(\zeta)$ because there is usually no confusion as to the meaning. It is also conventional to use upper-case letters to represent random variables[1], i.e. X, Y and Z are random variables, while x, y and z are particular values that they can take. This should become clearer as we move along.

1.2 Events and Probability Mass Functions

Events involving X are in the form of subsets of S_X e.g. in Example 2 above, we may be interested in $\{X < 2\}$ or $\{X = 4\}$. To be precise, we must translate a set of values in S_X to an equivalent set of outcomes in S, i.e.

$$\{X \in A\} \equiv \{\zeta : X(\zeta) \in A\}. \tag{5}$$

In addition, for X to be a random variable[2], the set $\{X \in (-\infty, x]\}$ must be an event in the underlying probability space for any $x \in \mathbb{R}$, i.e.

$$\{\zeta : X(\zeta) \leq x\} \in \mathcal{F} \tag{6}$$

where $\mathcal{F}$ is the event field. The example below presents a case of a mapping from S to $\mathbb{R}$ that is not a random variable.

Example 3: Pick a student at random, and measure his/her height, weight, age and gender. The sample space will be the set of 4-tuples of all possible values of the

[1]Some authors use boldface but this is rather hard to reproduce by hand.
[2]As opposed to just some other mapping.

four attributes. Let the field of events be defined as all possible events involving height, weight and age, but not gender. In other words, the following are all events:

$$\begin{aligned} A &= \text{"Height larger than 160cm, Weight lower than 60kg";} \quad (7) \\ B &= \text{"Body mass index is in the normal range";} \quad (8) \\ C &= \text{"Weight larger than 60kg, Age below 30".} \quad (9) \end{aligned}$$

But the following is not: $D =$ "Male, taller than 170cm". This means that we do not define $P[D]$.

Now let $X(\zeta) = 1$ if ζ is male, and $X(\zeta) = 0$ if ζ is female. Then X is not a random variable, because $\{X = 1\} = \{\zeta : \text{"}\zeta \text{ is male"}\}$ is not an element of $\mathcal{F}$ and hence is not an event. This corresponds to the case where either the gender distribution is not known and we do not want to make a guess, or it is not of interest. However, we can still define X, though we cannot find its probability distribution because we do not have a probability value for the elementary events defined on S_X. ■

In Example 2 above,

$$\{X < 2\} \equiv \{\text{TTTTT,TTTTH,TTTHT,TTHTT,THTTT,HTTTT}\},$$

and

$$\{X = 4\} \equiv \{\text{THHHH,HTHHH,HHTHH,HHHTH,HHHHT}\}.$$

Therefore, if the underlying probability space is completely specified, then we can find the probability of an event involving X by finding the probability of the equivalent event.

If the range S_X of a random variable X is countable, we say that X is a discrete random variable. For instance, S_X may be $\{1, 2, 3\}$, $\{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$, etc. If S_X is uncountable, then X may be a continuous or a mixed random variable, depending on its probability distribution. We will defer discussion of continuous and mixed random variables to a future lesson.

If we have a discrete X, with $S_X = \{x_1, x_2, \ldots, x_n\}$, then to compute the probability of any event in X requires only the *probability mass function*, defined as

$$p_X(x_k) = P[X = x_k] = P[\{\zeta : X(\zeta) = x_k\}], \quad x_k \in S_X. \tag{10}$$

Note the shorthand convention that $P[X = x_k]$ is the probability of the event $\{\zeta : X(\zeta) = x_k\}$.

The pmf of X completely specifies or describes the random variable X, in the sense that all events involving X have probabilities that can be expressed in terms of $p_X(x_k)$. This is because any event $X \in A_X \subset S_X$ can be seen as the union of $|A_X|$ single-element sets, i.e.

$$A_X = \bigcup_{x \in A_X} \{x\}.$$

There is a set of outcomes in S that are mapped to each value $x \in A_X$, i.e. $X : A_x \to x$. The sets A_x, $x \in A_X$, are mutually exclusive, because each outcome can only be mapped to one value in S_X. Therefore, by Axiom III, we have

$$\begin{aligned} P[X \in A_X] &= P\left[\bigcup_{x \in A_X} \{x\}\right] && (11) \\ &= \sum_{x \in A_X} P[\zeta \in A_x] && (12) \\ &= \sum_{x \in A_X} p_X(x). && (13) \end{aligned}$$

The third line is obtained from the second using the definition of PMF, equation (10).

Example 4: To find the pmf of X in Example 2, we note that $\{X = k\}$ is equivalent to "k heads in 5 coin tosses". Assuming the probability of a head is p, and recalling our previous discussion on the binomial probability law, we see that

$$p_X(k) = \binom{5}{k} p^k (1-p)^{5-k}, \quad k = 0, 1, 2, 3, 4, 5. \tag{14}$$

To obtain $P[X < 2]$, we use (13) with $A = \{0, 1\}$:

$$P[X < 2] = p_X(0) + p_X(1) = (1-p)^5 + 5p(1-p)^4. \tag{15}$$

Exercise: Find the pmf of X in Example 1, then find the probability of an even-valued X. ■

Given that $p_X(x)$ is a probability value, it must of course satisfy the following properties:

1. $0 \le p_X(x) \le 1$ for all $x \in S_X$.
2. $\sum_{x \in S_X} p_X(x) = 1$ because $X \in S_X$ is equivalent to the certain event S.

These properties often provide good "sanity checks" on results.

1.3 Discrete Random Variables of Practical Importance

1.3.1 Discrete Uniform Random Variable

This r.v. has equal probability of taking on any value within its range S_X. In other words, if $S_X = \{x_1, x_2, \ldots, x_M\}$, then

$$p_X(x_k) = \frac{1}{M}, \quad k = 1, 2, \ldots, M. \tag{16}$$

The uniform distribution is used for instance to model a problem wherein only the possible values of X are known but nothing else. It expresses the maximum uncertainty about the value of X, over the range S_X. This concept of a measure of uncertainty can be quantified using information theory.

1.3.2 Bernoulli Random Variable

We can indicate the occurrence of an event A of interest using the *indicator function*, defined as

$$I_A(\zeta) = \begin{cases} 1 & \text{if } \zeta \in A \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Then I_A is a random variable with range $\{0, 1\}$, and pmf

$$p_{I_A}(0) = P(A^c) = 1 - p, \quad \text{and} \quad p_{I_A}(1) = p. \tag{18}$$

1.3.3 Binomial Random Variable

We previously encountered the binomial probability law, which dictates the probability of obtaining k successess in n independent identical Bernoulli trials. Now we define the number of successes in n trials as a random variable X with pmf

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{19}$$

This notation is more concise and allows us to write events such as "3 or more heads in 5 coin tosses" as $\{X \geq 3\}$, whose probability can be computed as

$$P(X \geq 3) = \sum_{k=3}^{5} p_X(k) = \sum_{k=3}^{5} \binom{5}{k} p^k (1-p)^{5-k}.$$

1.3.4 Geometric Random Variable

The geometric distribution law was earlier introduced to tackle the event "m Bernoulli trials required before seeing the first success". We can conveniently define X as the number of trials required before the first success, which will have the pmf

$$p_X(k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots \tag{20}$$

The probability that we need to wait more than N trials before the first success is then

$$P(X > N) = \sum_{k=N+1}^{\infty} (1-p)^{k-1} p = p \frac{(1-p)^N}{1-(1-p)} = (1-p)^N \tag{21}$$

as previously derived.

1.3.5 Other Distributions

There are several other discrete distributions of practical importance, in particular the Poisson distribution, but we will defer discussion of the Poisson random variable to a later class.

2 Expected Value of Discrete Random Variables

2.1 Definition

Let X be a random variable defined from a repeatable experiment. If we perform the experiment N times independently, and observe the outcome $\{X = x(i)\}$ in the i-th trial, then it is obvious that the average or mean value of X over these N trials is

$$\langle X \rangle_N = \frac{\sum_{i=1}^{N} x(i)}{N}. \tag{22}$$

This value is known as the *sample mean* because it is the mean of N "samples" of the r.v. X. If the sample space of X is $S_X = \{x_1, \ldots, x_n\}$, then N_1 of the elements of the set of observations $\{x(1), \ldots, x(N)\}$ will take the value x_1, N_2 will take the value x_2 and so on. By grouping the samples $x(1), \ldots, x(N)$ according to their values, we then have

$$\langle X \rangle_N = \frac{\sum_{k=1}^{n} N_k x_k}{N} = \sum_{k=1}^{n} x_k \frac{N_k}{N}. \tag{23}$$

Since N_k/N is the relative frequency of x_k within the collection of N samples, as N grows large, it converges to $P[X = x_k] = p_X(x_k)$. In other words,

$$\lim_{N\to\infty} \langle X \rangle_N = \sum_{k=1}^{n} x_k p_X(x_k) = \sum_{x \in S_X} x \cdot p_X(x). \tag{24}$$

By extension to random variables arising from non-repeatable experiments, we *define* the **expected or mean value of** X for discrete X as

$$E[X] = \sum_{x \in S_X} x \cdot p_X(x), \tag{25}$$

if it exists. (For some distributions, the sum on the RHS does not converge. For these random variables, we say that the mean does not exist.)

Note that:

1. $E[X]$ is to be read as "the expected value of X". It is not a *function* of X, even though the notation makes it look as if it is. $E[X]$ is computed using the pmf of X, and is a value, not another random variable as a function of a random variable will be (more on this later).

2. $E[X]$ is the centre of gravity of a uniform rod with weights attached at points $x = x_k$ that are proportional to $p_X(x_k)$. Therefore, $E[X]$ lies between the smallest and largest values in S_X.

Example 5: Consider the pmf $p_X(0) = \frac{1}{6}$, $p_X(1) = \frac{1}{3}$ and $p_X(2) = \frac{1}{2}$. The mean value would be

$$E[X] = 0 \times \frac{1}{6} + 1 \times \frac{1}{3} + 2 \times \frac{1}{2} = 1\frac{1}{3}.$$

Since the probability mass is "heavier" for the larger values in S_X, the centre of gravity $E[X]$ is closer to the largest value in S_X than the smallest. ■

Example 6: Symmetry of the pmf of X about a value $x = a$ means that $E[X] = a$. Consider $S_X = \{1, 2, \ldots, M\}$ and $p_X(k) = 1/M$ for all $k \in S_X$. Then

$$E[X] = \frac{1}{M}(1 + 2 + \cdots + M) = \frac{M+1}{2}$$

which is the mid-point of the real numbers in the set S_X. Or consider the binomial distribution with $n = 3$ and $p = 0.5$ (i.e. the number of heads in 3 fair coin tosses). If we call this r.v. Y, then

$$p_Y(k) = \binom{3}{k} 0.5^3, \quad k = 0, 1, 2, 3,$$

which has even symmetry about $x = 1.5$. The mean of Y is by definition

$$\begin{aligned} E[Y] &= \frac{1}{8}\sum_{k=0}^{3} k\binom{3}{k} \\ &= \frac{1}{8}(3 + 6 + 3) \\ &= 1\frac{1}{2} \end{aligned}$$

which is exactly the mid-point of the distribution, as expected. ■

2.2 Functions of a Random Variable

Suppose a random variable R represents the radius of a circle, then the area of the circle is πR^2, which will itself be a random variable. Or if V represents a random voltage applied across a resistance of r Ohms, then the power dissipated V^2/r will be random too. The area of the circle and the power dissipated in the resistor are examples of *functions* of a random variable.

In general, one can define a r.v. $Z = g(X)$ where $g(x)$ is some function with a domain that includes S_X. For every outcome ζ, we have $Z(\zeta) = g(X(\zeta))$, and the range of Z is S_Z, containing all unique values of $g(x_i)$, $x_i \in S_X$. The pmf of Z is then given by

$$p_Z(z_k) = P[g(X) = z_k] = P[\{\zeta : g(X(\zeta)) = z_k\}]. \tag{26}$$

This is (sometimes) not as difficult to obtain as it looks.

Example 7: A die is tossed. Define X as the number of dots on the top face, and $Y = X - 3$ as the number of dollars won in a game involving rolling this die. Then $S_Y = \{-2, -1, 0, 1, 2, 3\}$, and the pmf of Y is

$$p_Y(y) = p_X(y + 3) = \frac{1}{6} \tag{27}$$

for all $y \in S_Y$. ■

Example 8: The voltage V applied across a 1-Ohm resistor has a binomial distribution with $n = 3$ and $p = 0.5$. The power $W = V^2$ will have the range $S_W = \{0, 1, 4, 9\}$ and pmf

$$p_W(x) = \begin{cases} \frac{1}{8} & x = 0, 9 \\ \\ \frac{3}{8} & x = 1, 4 \end{cases} \tag{28}$$

since $P[W = x] = P[V = \sqrt{x}] = p_V(\sqrt{x})$. ■

The distribution of a function of a discrete random variable is often possible to obtain easily, and hence the mean of that function is also obtainable, at least using numerical methods. However instead of first finding the pmf of Z and then applying the expected value definition, we can also compute

$$E[Z] = E[g(X)] = \sum_{x \in S_X} g(x) p_X(x) \tag{29}$$

using only the pmf of the original r.v. X.

We can derive this result through a relative frequency argument as follows. Suppose we perform the experiment N times independently, and obtain N samples of X, denoted $x(1)$ through $x(N)$. By applying the function $g(\cdot)$ on every one of these N samples, we obtain the N values $z(1)$ through $z(N)$ where $z(k) = g(x(k))$. The sample mean of $z(k)$, $k = 1, \ldots, N$ is then

$$\langle Z \rangle_N = \frac{1}{N} \sum_{i=1}^{N} z(i) = \frac{1}{N} \sum_{i=1}^{N} g(x(i)). \tag{30}$$

By grouping terms in the summation according to the values in $S_X = \{x_1, \ldots, x_n\}$ as we did earlier, we then have

$$\langle Z \rangle_N = \frac{1}{N} \sum_{k=1}^{n} N_k g(x_k) = \sum_{k=1}^{n} g(x_k) \frac{N_k}{N}. \tag{31}$$

Since in the limit $N \to \infty$ relative frequency is equivalent to probability, we then obtain (29).

Using (29), we can derive the following important results:

- $E[aX] = aE[X]$;
- $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$;
- $E[X + c] = E[X] + c$, where c is a deterministic constant;
- $E[c] = c$, where c is a deterministic constant.

2.3 Variance

The variance of a discrete r.v. is a measure of how "spread out" the pmf is, e.g. if the ranges of two random variables X and Y are both equal to $S_X = \{0, 1, 2\}$, and $p_X(k) = \frac{1}{3}$ while $p_Y(0) = p_Y(2) = 0.1$ and $p_Y(1) = 0.8$, it should be clear from a plot of the two pmf's that the Y values are less dispersed than the X ones. We would like to quantify this **deviation from the mean** precisely, using the concept of variance.

While there are many ways to characterize deviation from the mean μ_X, e.g. $|X - \mu_X|$, $(X - \mu_X)^{20}$, etc., the mean squared deviation or **variance** is easiest to work with:

$$\sigma_X^2 = \mathrm{var}(X) = E[(X - \mu_X)^2]. \tag{32}$$

Its positive square root is known as the **standard deviation** of X, and is denoted σ_X. Note that σ_X has the same units as X (metres, kilograms, seconds) whereas σ_X^2 is in terms of the square of the units of X (m^2, kg^2).

By expanding the square in the definition (32) and using the results at the end of the last section, we have

$$\begin{aligned}
\sigma_X^2 &= E[X^2 - 2\mu_X X + \mu_X^2] & (33)\\
&= E[X^2] - 2\mu_X E[X] + E[\mu_X^2] & (34)\\
&= E[X^2] - 2\mu_X^2 + \mu_X^2 & (35)\\
&= E[X^2] - \mu_X^2. & (36)
\end{aligned}$$

The last expression is often most convenient for the computation of variances.

Example 9: Let $p_X(k) = 1/3$, $k = 0, 1, 2$ and $p_Y(0) = p_Y(2) = 0.1$, $p_Y(1) = 0.8$. Note that $E[X] = E[Y] = 1$ due to symmetry. We can compute the variances of X and of Y using (32) as:

$$\begin{aligned}
\sigma_X^2 &= (1^2 + 0^2 + 1^2)/3 = 0.67 & (37)\\
\sigma_Y^2 &= (1^2 + 1^2) \times 0.1 + 0^2 \times 0.8 = 0.2 & (38)
\end{aligned}$$

As expected, $\sigma_Y < \sigma_X$. ■

2.4 Importance of Mean and Variance

The mean and variance are important *partial descriptors* of a random variable. They give us a rough idea of the likely values of a random variable, e.g. we will see in the next class that for a binomial r.v., $\mu_X = np$ and $\sigma_X^2 = npq$. So in 100 independent Bernoulli trials with probability of success 0.2, we expect 20 successes, with a standard deviation of $\sqrt{16} = 4$. As a rule of thumb, we then expect that with very high probability, X lies in $\mu_X \pm 3\sigma_X = \{8, \ldots, 32\}$. Indeed, the exact probability of X lying within that range can be computed as 0.9982.[3]

[3] In Matlab, use the function `binopdf` in the Statistics toolbox.

3 Diagnostic Questions

1. Identify three examples of discrete random variables that you have encountered in your daily life. Describe the underlying probability space of each example.

2. Explain how the parameter p impacts the binomial PMF (with fixed n) and the geometric PMF.

3. The PMF of two random variables X and Y are respectively $p_X(k) = 0.4\delta(k) + 0.6\delta(k-1)$ and $p_Y(k) = p_X(k-1)$. Find $E[X]$, $E[Y]$, σ_X^2 and σ_Y^2.