Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 12 EE 325 Probability and Random Processes
Lecture Notes 8 August 25, 2014

1 Discrete Random Variables

We now learn an important special case of random-variables, the ones that take discrete random variables. One good thing about these variables is that there is less ambiguity and pathological cases are very rare.

Definition 1 *Consider a countable set E, a function from Ω to E is called a discrete random variable if $\{\omega : X(\omega) = n\} \in \mathcal{F}, \forall n \in E$.*

In other words, the definition ensures that X is measurable with respect to the space $(E, \mathcal{P}(E))$. (Recall that we have to only check measurability for a class of sets which generates the sigma-field of the range-space). Don't be confused by the absence of conventional $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, it is indeed true that for any $B \in \mathcal{B}(\mathbb{R})$,

$$X^{-1}(B \bigcap E) \in \mathcal{F},$$

thus ensuring measurability with respect to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. So by slight abuse of the definition, will define various random-quantities without explicitly stating measurability with respect to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For example, we will call X above as a random variable taking values in E. We will use the notation $\{X = x\}$ to signify $X^{-1}(x) = \{\omega : X(\omega) = x\}$. Consequently, $P(X = x) = P(X^{-1}(x))$, where P is the probability measure associated with the space $(\Omega, \mathcal{F})$. In this way, we avoid referring to the probability measure induced by the random variable X, and conveniently work with one probability measure, i.e. $P(A), \forall A \in \mathcal{F}$.

Proposition 1 *A discrete random variable X is completely specified by its* ***probability distribution function****, denoted as $P(X = x), \forall x \in E$ or $P_X(x)$.*

Our first example of a discrete random variable is perhaps the most popular one.

1.1 Bernoulli Distribution: $Bernoulli(p)$

A binary valued random variable with

$$P(X = 1) = p,$$

is called a **Bernoulli random variable** with parameter p. This is often denoted concisely as $X \sim Bernoulli(p)$.

1.2 Function of a Discrete Random Variable

Functions defined on random-variables allow us to zoom in and out on various aspects of randomness, while preserving all the the measurability properties.

Theorem 1 *Let E_1 and E_2 be countable sets. Let X be a random variable taking values in E_1, and $g : E_1 \to E_2$ be an arbitrary function. Then $Y = g(X)$ is a random variable.*

Proof: We have to show that $\{Y = y\} \in \mathcal{F}$. Observe that

$$\begin{aligned}\{Y = y\} &= \{\omega \in \Omega : X(\omega) = x \text{ and } g(x) = y\} \\ &= \bigcup_{x \in E_1 : g(x) = y} \{X = x\}\end{aligned}$$

Clearly the RHS is the union of events which are already in $\mathcal{F}$, and hence the proof. ■

Example 1 *Let E_1 be RV in $\{1, 2, 3, 4\}$ and consider the function $g(x) = \tan(\frac{\pi i}{8})$. Is $Y = g(X)$ a random variable?*

Solution: If we stick to our original definition, then Y cannot be called a random variable, as its value is undefined when $i = 4$. However, observe that we can still ensure measurability of Y as a function, by taking $\tan^{-1} \infty$ as $\frac{\pi}{2}$. Since measurability is the key property that we need, we can also consider $\bar{\mathbb{R}}-$ valued random variables, i.e. $\mathbb{R} \bigcup \{\pm\infty\}$ is the possible range[1]. Thus Y above can be treated as a random variable. ■

As the last example shows, there is no harm in using $\bar{\mathbb{R}}$ or $\bar{\mathbb{N}}$ as the random variable's range. Also, whenever we talk about the function of a random-variable without any further qualification, we are dealing with an $\bar{\mathbb{R}}$ valued random variable. Let us now propose something which will be of immediate use and simple, however the proof of this will have to wait for the next notes. While this may irk the mathematical purists, it saves the notes from being just a monotonous listing of results. Please accept the following and proceed.

Proposition 2 *If X_1 and X_2 are discrete random variables, so is $X_1 + X_2$.*

The proposition deals with a function of two random variables and will be covered in detail in later sections.

2 Expectation

Expectation of a random variable is a key parameter, which is governed by the distribution of the random variable. In this sense, it is a statistical parameter, something that we can infer from past experiments or modeling experience. This is also known as statistical mean or average, which has close connections to empirical averages of experimental measurements. In fact let us get an intuitive understanding by using the classical frequency interpretation. In particular, the fraction of times any event A occurs in n independent trials of an experiment can be approximated by $P(A)$ for large n, this is the frequency interpretation. Consider repeated draws of a random variable X from the state-space E in an iid fashion. Let us denote the n outcomes as $X_1, X_2, \cdots, X_n$. What is the empirical

[1] $\bar{\mathbb{R}}$ is also known as the two-point compactification of $\mathbb{R}$

average of these quantities.

$$\begin{aligned}\frac{\sum_{i=1}^n X_i}{n} &= \frac{\sum_{i=1}^n \left(\sum_{j\in E} j \mathbb{1}_{\{X_i=j\}}\right)}{n}\\ &= \sum_{j\in E} j \left(\frac{\sum_{i=1}^n \mathbb{1}_{\{X_i=j\}}}{n}\right)\\ &\approx \sum_{j\in E} j\, P(X=j),\end{aligned}$$

where the last step used the frequency interpretation. This is the concept behind the operator called expectation, which we more formally define below.

Definition 2 *Let X be a discrete random variable in E, and let $g : E \to \mathbb{R}$ be such that EITHER*

1. *g is non-negative, OR*
2. *$\sum_{x\in E} |g(x)| P(X = x) < \infty$,*

then the expectation of the random variable $g(X)$ is

$$\mathbb{E}[g(X)] \triangleq \sum_{x\in E} g(x)P(X=x) \tag{1}$$

It is common to call $g(X)$ as an *integrable* random variable when $\mathbb{E}[|g(X)|] < \infty$, whereas a more accurate nomenclature is to call it *summable*. Clearly, expectation is a linear operator.

$$\mathbb{E}[\lambda_1 g_1(X) + \lambda_2 g_2(X)] = \lambda_1 \mathbb{E}[g_1(X)] + \lambda_2 \mathbb{E}[g_2(X)].$$

Furthermore, the following monotonicity property is also evident. If $g_1(x) \le g_2(x), \forall x \in E$,

$$\mathbb{E}[g_1(X)] \le \mathbb{E}[g_2(X)].$$

Example 2 *Let S_n represent the number of HEADs in n tosses of a fair coin. The probability law that we choose for this experiment is*

$$P(S_n = k) = \frac{\binom{n}{k}}{2^n}.$$

Find the expectation of S_n, for a given value of n.

Solution: Let $X_i \in \{H, T\}$ be the outcome of toss i. Observe that

$$S_n = \sum_{i=1}^n \mathbb{1}_{\{X_i=H\}}.$$

Thus

$$\begin{aligned}\mathbb{E}[S_n] &= \sum_{i=1}^n \mathbb{E}\, \mathbb{1}_{\{X_i=H\}}\\ &= \sum_{i=1}^n \frac{1}{2}\\ &= \frac{n}{2}.\end{aligned}$$

■

In the last example, we used the indicator function to seamlessly move between the space $\{H, T\}$ and $\{0, 1\}$. In fact this can be done in a more general fashion. Recall that $Y = g(X)$ is a random variable if X is a random variable.

Theorem 2 *Let $g(X)$ be an integrable random variable and take $Y = g(X)$. Then*

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] \tag{2}$$

Proof: At the first look it is too obvious, but observe that $\mathbb{E}[Y] = \sum_y yP(Y = y)$, whereas $\mathbb{E}[g(X)] = \sum_x g(x)P(X = x)$. What guarantees that they are the same? Let $X : E_1 \to E_2$, where E_2 is a countable subset of $\mathbb{R}$.

$$\begin{aligned}
\sum_{y \in E_2} yP(Y = y) &= \sum_{y \in E_2} y \sum_{x:g(x)=y} P(X = x) \\
&= \sum_{y \in E_2} y \sum_{x \in E_1} P(X = x)\mathbb{1}_{\{g(x)=y\}} \\
&= \sum_{x \in E_1} P(X = x) \sum_{y \in E_2} y\mathbb{1}_{\{g(x)=y\}} \\
&= \sum_{x \in E_1} P(X = x)g(x) \\
&= \mathbb{E}[g(X)].
\end{aligned}$$

The integrability assumption allows us to seamlessly interchange the two summations in the second step above.

2.1 Probability is an Expectation

The last example contains an additional lesson, that the probability of an event is nothing but an appropriate expectation. No wonder that both the quantities are specified by the distribution of a random variable.

Theorem 3 *Consider a discrete random variable X in $E \subset \mathbb{R}$. Then for any $A \in \mathcal{B}(\bar{\mathbb{R}})$,*

$$\mathbb{E}[\mathbb{1}_{\{X \in A\}}] = P(A). \tag{3}$$

Proof: Clearly $P(A)$ is the probability induced by the random variable X. Notice that $Y = \mathbb{1}_{\{X \in A\}}$ is indeed a random variable. By using Theorem 2,

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_x \mathbb{1}_{\{x \in A\}} P(X = x) \\
&= \sum_{x \in A} P(X = x) \\
&= P(A).
\end{aligned}$$

Example 3 ***Coupon Collector*** *Let there be m categories of coupons available, with one coupon packed per box. The coupon for each box is randomly chosen. Consider n boxes, and let S_n represent the number of coupons which didn't find its way to any of the boxes. Find $\mathbb{E}[S_n]$.*

Solution: Let X_i be the number of boxes with coupon i. Then

$$S_n = \sum_{i=1}^{m} \mathbb{1}_{\{X_i = 0\}}.$$

Taking expectations.

$$\mathbb{E}[S_n] = \sum_{i=1}^{m} \mathbb{E}\mathbb{1}_{\{X_i=0\}}$$
$$= \sum_{i=1}^{m} \left(1 - \frac{1}{m}\right)^n$$
$$= m(1 - \frac{1}{m})^n.$$

3 Independent Random Variables

We have defined independence of events in terms of the probability of events. Similarly, we can define independence of two random variables by demanding the events corresponding to each random-variable be independent. Recall that for a discrete random variable X, the collection of sets of the form $X^{-1}(x), x \in E$ is called the events corresponding to X.

Definition 3 *Two discrete random variables X_1 and X_2 taking respective values in E_1 and E_2 are said to be independent if*

$$P(X_1^{-1}(u), X_2^{-1}(v)) = P(X_1^{-1}(u))P(X_2^{-1}(v)), \ \forall (u,v) \in E_1 \times E_2.$$

i.e. any pair of their respective associated events are independent.

The standard textbook variety of this same statement is repeated below for more clarity.

Definition 4 *Random Variables X_1 and X_2 taking values in E_1 and E_2 respectively are independent if $\forall (i,j) \in (E_1 \times E_2)$*

$$P(X_1 = i, X_2 = j) = P(X_1 = i)P(X_2 = j).$$

The last two definitions are indeed the same.

Example 4 *Consider throwing 2 fair dice, where the probability association is given by*

$$P(X_1 = i, X_2 = j) = \frac{1}{36} \text{ for } 1 \le i,j \le 6.$$

Does this correspond to independent trials.

Solution: Verify that

$$P(X_i = i, X_2 = j) = \frac{1}{36} = P(X_1 = i)P(X_2 = j),$$

since we assumed that the die is a fair one. Thus this probability assignment indeed models an experiment where a fair die is thrown two times, or two fair dice are thrown independently. ■

We can extend the notion of independence to many random variables. (Recall the definition of independence of many events in class).

Definition 5 *The discrete random variables $X_1, X_2, \cdots X_n$, are called independent if all possible tuple of the respective events associated with them are independent.*

Exercise 1 *Show that iff,* $\forall u_i \in E_i$,

$$P(\bigcap_{i=1}^{n} X_i^{-1}(u_i)) = \prod_{i=1}^{n} P(X_i^{-1}(u_i)),$$

then $X_1, \cdots, X_n$ *are independent random variables.*

We can even extend our definition to a countable collection of random variables. In particular, independence of a sequence is defined in terms of finite dimensional probabilities.

Definition 6 *A sequence of discrete random variables* $X_n, n \geq 1$ *are called independent if for all finite set of indices,* $i_1, \cdots, i_k$

$$P(X_{i_1} = x_{i_1}, \cdots, X_{i_k} = x_{i_k}) = \prod_{j=1}^{k} P(X_{i_j} = x_{i_j}).$$

Example 5 *Consider a discrete random variable taking values in* $\{1, 2, \cdots, k\}$ *with* $P(X = i) = p_i, 1 \leq i \leq k$. *For* n *draws of this random variable, say* $\bar{X} = X_1, \cdots, X_n$, *and for any given sequence* $\bar{u} = (u_1, \cdots, u_n)$ *it is specified that*

$$P(\bar{X} = \bar{u}) = \prod_{i=1}^{k} p_i^{N_i(\bar{u})},$$

where $N_i(\bar{u})$ *counts the number of times* i *appears in the sequence* $\bar{u}$. *Are the draws in* $\bar{X}$ *independent.*

Solution: It is easily verified by our definition that these throws are independent. ■

Notice that in the previous example, $N_i(u)$ is a random variable too. This is evident since it is a sum of binary valued random variables. When X is binary, the random variable $N_i(1)$ is called the Binomial distribution.

3.1 Binomial Distribution - Binomial(n, p)

Like the name implies, this is closely related the binomial expansion. Recall that

$$(p + q)^n = \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i}.$$

How does this connect to throwing a coin? Imagine a coin with $P(HEADs) = p$. Let S_n represent the number of HEADs in n independent throws of this coin. Then S_n is Binomial(n, p). To see this, let us denote by X_i the indicator of HEADs in toss i. For any sequence $x_1, \cdots, x_n$ with k HEADs

$$P(X_1 = x_1, \cdots, X_k = x_k) = p^k (1 - p)^{n-k}.$$

Since there are $\binom{n}{k}$ sequences with k heads,

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

We will denote a random variable having Binomial distribution as $X \sim$ Binomial(n, p).

Example 6 *Compute* $\mathbb{E}[X]$ *for* $X \sim \mathrm{Binomial(n, p)}$.