Lecture 11: November 4, 2021

Lecturer: Madhur Tulsiani

1 Independence

Recall that two non-zero probability events A and B are said to be independent if $\mathbb{P}[A \mid B] = \mathbb{P}[A]$. One can verify that this is equivalent to $\mathbb{P}[B \mid A] = \mathbb{P}[B]$. In other words, restricting to one event does not change the probability of the other event. Independence is a joint property of events and the probability measure: one cannot make judgment about independence without knowing the probability measure.

Two random variables X and Y defined on the same finite probability space are defined to be independent if $\mathbb{P}[X = x \mid Y = y] = \mathbb{P}[X = x]$ for all non-zero probability events $\{X = x\} := \{\omega : X(\omega) = x\}$ and $\{Y = y\} := \{\omega : Y(\omega) = y\}$.

The notion of independence can also be generalized (in multiple ways) beyond the case of two events or random variables. We say n events $A_1, ..., A_n$ are mutually independent (sometimes we will just say "independent", since this the most commonly used notion of independence for multiple events) if for all subsets $S \subseteq \{1, ..., n\}$ we have:

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i) .$$

We say n random variables $X_1, ..., X_n$ are mutually independent if for all values $x_1, ..., x_n$, the events "$X_1 = x_1$", ..., "$X_n = x_n$" are mutually independent.

There are also weaker notions of independence that are often useful. We say n events are pairwise independent if all *pairs* are independent, and likewise for random variables i.e., we have the above condition only for sets S of size two.

$$\forall S \subseteq \{1, \ldots, n\}, |S| = 2 \quad \mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i) .$$

More generally, the notion of k-wise independence is defined by considering the above condition for all S with $|S| \leq k$.

Exercise 1.1 *Can you think of three events, or three random variables, that are pairwise independent but not mutually independent?*

We saw that for any two random variables X and Y we have $\mathbb{E}[X] + \mathbb{E}[Y] = \mathbb{E}[X+Y]$. However, it is not in general the case that $\mathbb{E}[X] \cdot \mathbb{E}[Y] = \mathbb{E}[X \cdot Y]$ (for example, suppose X and Y are indicator random variables for the same event of probability p; then the LHS is p^2 but the RHS is p). Nonetheless, we *do* get this property when X and Y are independent.

Proposition 1.2 *Let $X, Y : \Omega \to \mathbb{R}$ be two independent random variables. Then*

$$\mathbb{E}[X \cdot Y] \;=\; \mathbb{E}[X] \cdot \mathbb{E}[Y] \,.$$

Proof:

$$\begin{aligned}
\mathbb{E}[X] \cdot \mathbb{E}[Y] &= \left(\sum_a \mathbb{P}(X=a) \cdot a\right) \cdot \left(\sum_b \mathbb{P}(Y=b) \cdot b\right) \\
&= \sum_{a,b} a \cdot b \cdot \mathbb{P}(X=a) \cdot \mathbb{P}(Y=b) \\
&= \sum_{a,b} a \cdot b \cdot \mathbb{P}(X=a \wedge Y=b) \quad \text{(by independence)} \\
&= \sum_c \sum_{(a,b):ab=c} a \cdot b \cdot \mathbb{P}(X=a \wedge Y=b) \quad \text{(grouping)} \\
&= \sum_c c \cdot \mathbb{P}(X \cdot Y = c) = \mathbb{E}[X \cdot Y] \,.
\end{aligned}$$

■

Exercise 1.3 *Check that the converse of the above statement is false i.e., there are random variables X, Y such that $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, but X and Y are not independent.*

1.1 The countably infinite case

The concepts defined in the previous and current lecture for finite probability spaces extend almost verbatim to the the case when the space Ω is countablly infinite i.e., there exists a bijection from Ω to the set $\mathbb{N}$ of natural numbers. However, we need to be careful about the convergence of summations over $\omega \in \Omega$ as these may be inifinite sums, which need to be defined via limits. The extension to the case of uncountably infinite Ω (such as $\Omega = [0,1]$) requires some additional concepts, and we will discuss this in a later lecture.

2 Some important random variables

2.1 Variance

We will now see some very useful random variables. We will also compute the expectation, and another quantity called the *variance* of these random variables, which is a commonly used measure of how "spread" is a random variable. For example a variable X which is always 0, and Y which is ± 1 with probability $1/2$ each, have the same expectation, but the notion of variance can be used to capture the fact that the distribution of Y is spread over more values than that of X (i.e., Y varies more than X).

For a (real-valued) random variable X, the variance is defined as

$$\mathsf{Var}[X] \;:=\; \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Note that the inner expectation is a *constant*. Using (say) μ to denote $\mathbb{E}[X]$, we can also write another expression for the variance.

$$\mathsf{Var}[X] \;=\; \mathbb{E}\left[(X-\mu)^2\right] \;=\; \mathbb{E}\left[X^2 - 2\mu \cdot X + \mu^2\right] \;=\; \mathbb{E}\left[X^2\right] - 2\mu^2 + \mu^2 \;=\; \mathbb{E}\left[X^2\right] - \mu^2 .$$

Thus, we can use either of the two expressions below to compute the variance.

$$\mathsf{Var}[X] \;=\; \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \;=\; \mathbb{E}\left[X^2\right] - (\mathbb{E}[X]) .$$

Since the first expression is always non-negative, we also get a proof of the very useful inequality that $\mathbb{E}\left[X^2\right] \geq (\mathbb{E}[X])^2$.

Exercise 2.1 *Can you derive the inequality* $\mathbb{E}\left[X^2\right] \geq (\mathbb{E}[X])^2$ *using the Cauchy-Schwarz-Bunyakovsky inequality?*

2.2 Bernoulli random variables

A $\mathsf{Bernoulli}(p)$ random variable X is defined as taking the value 1 with probability p and the value 0 with probability $1-p$. We can write this as $\mathbb{P}[X = x] = p^x(1-p)^{1-x}$. One may intuitively think of a Bernoulli random variable as the indicator function of "heads" in an outcome space $\Omega = \{\text{tails}, \text{heads}\}$ of a biased coin toss. Alternatively, we simply take the outcome space to be $\Omega = \{0, 1\}$. More generally, indicator functions of events are Bernoulli random variables.

Let X be a $\mathsf{Bernoulli}(p)$ random variable. Then, we have

$$\mathbb{E}[X] \;=\; 1 \cdot p + 0 \cdot (1-p) \;=\; p \;=\; \mathbb{P}[X = 1] .$$

The fact that for a Bernoulli random variable X, $\mathbb{E}[X] = \mathbb{P}[X = 1]$ is extremely useful, particularly when combined with the linearity of expectation, to analyze random variables which can be written as a sum of Bernoulli variables. We can also compute $\mathsf{Var}[X]$, using the fact that $X^2 = X$, since $X \in \{0, 1\}$

$$\mathsf{Var}[X] \;=\; \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \;=\; p - p^2 \;=\; p \cdot (1 - p)\,.$$

2.3 Finite Bernoulli i.i.d. sequences and Binomial random variables

Another important random variable is a sum of (mutually)*indepependent* and indentical Bernoulli random variables. We first define the probability space corresponding to a (finite) collection of Bernoulli variables.

Finite Bernoulli i.i.d. sequence We can also think of a sequence of coin tosses, with

$$X_i = \begin{cases} 1 & \text{if toss i is heads} \\ 0 & \text{if toss i is tails} \end{cases} .$$

being n Bernoulli random variables in the probability space $\Omega_n = \{0, 1\}^n$, i.e., $X_i(\omega) = \omega_i$. Define the product probability measure on this finite space using:

$$\nu_n(\omega) = \prod_{i=1}^{n} p^{\omega_i}(1 - p)^{1-\omega_i}.$$

Note that if $p = \frac{1}{2}$, we have $\nu_n(\omega) = \frac{1}{2^n}$, i.e., $\mathbb{P}_n$ is the uniform distribution over the outcome space, as all outcomes are equally likely.

Exercise 2.2 *For the outcome space defined above, verify that:*

- *For any fixed i, X_i is indeed a Bernoulli(p) random variable, and*
- *If $I \subset [n]$ and $J \subset [n]$ are disjoint, then any function of X_I and any function of X_j are independent random variables.*

As noted in the previous lecture, when the latter point holds, we simply say that $X_1, \cdots, X_n$ are (mutually) independent. Furthermore since all the X_i have the same distribution, we call the sequence i.i.d., meaning independent and identically distributed.

Binomial random variables Let Z_n be a random variable counting the number of heads associated with n independent biased coin tosses. We can model this in Ω_n above as $Z_n = \sum X_i$.

Let us calculate the expectation of Z. By linearity we have $\mathbb{E}[Z_n] = \sum \mathbb{E}[X_i]$. Since $Z_n = \sum X_i$, we have, $\mathbb{E}[Z_n] = \sum \mathbb{E}[X_i]$. Now,

$$\begin{aligned} \mathbb{E}[X_i] &= 1 \cdot \mathbb{P}[X_i = 1] + 0 \cdot \mathbb{P}[X_i = 0] \\ &= \mathbb{P}[X_i = 1] = p \end{aligned}$$

Hence $\mathbb{E}[Z_n] = n \cdot p$. Note that we did not use independence in the above calculations. We just needed that for each i, $\mathbb{E}[X_i] = p$. Let us now compute the variance.

$$\mathsf{Var}[Z_n] = \mathbb{E}\left[Z_n^2\right] - (\mathbb{E}[Z_n])^2 = \mathbb{E}\left[Z_n^2\right] - (n \cdot p)^2 .$$

Thus, we need to compute the first term $\mathbb{E}\left[Z_n^2\right]$ to understant the variance. We can write

$$\begin{aligned} \mathbb{E}\left[Z_n^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i,j} X_i \cdot X_j\right)\right] \\ &= \sum_{i,j} \mathbb{E}\left[X_i \cdot X_j\right] \\ &= \sum_{i} \mathbb{E}\left[X_i^2\right] + \sum_{i \neq j} \mathbb{E}\left[X_i \cdot X_j\right] \\ &= n \cdot p + n(n-1) \cdot p^2 , \end{aligned}$$

where we used the fact that $\mathbb{E}\left[X_i \cdot X_j\right] = \mathbb{E}[X_i] \cdot \mathbb{E}\left[X_j\right] = p^2$ using independence, when $i \neq j$. Using the above, we get that

$$\mathsf{Var}[Z_n] = n \cdot p + n(n-1) \cdot p^2 - n^2 \cdot p^2 = n \cdot p - n \cdot p^2 = n \cdot p(1-p) = \sum_i \mathsf{Var}[X_i] .$$

Exercise 2.3 *Check that for any collection of* pairwise independent *(and not necessarily identical) random variables $X_1, \ldots, X_n$, we still have that for $Z = \sum_i X_i$*

$$\mathsf{Var}[Z] = \sum_i \mathsf{Var}[X_i] .$$

We do need independence, and namely the product probability measure, to calculate $\mathbb{P}(Z_n = k)$ for $k \in [n]$ (this is often called the probability mass function. First note that the shorthand $(Z_n = k)$ simply means $\{\omega \in \Omega : Z_n(\omega) = k\}$. Since all ω that have the same number (in this case k) of 1's have the same probability, we simply need to count how many such ω's there are, and multiply by this individual probability.

Exercise 2.4 *Verify that* $\mathbb{P}_n(Z_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Z_n is called a $\mathsf{Binomial}(n, p)$ random variable.

2.4 Infinite Bernoulli i.i.d. sequence and Geometric random variables

We would like to generalize the Bernoulli sequence probability space to an infinite sequence. We would like to choose $\Omega = \{0,1\}^{\mathbb{N}}$ as our outcome space, but this is not a countable set. We will come back to the issue of properly defining the probability space with this uncountable Ω.

For now, if we still consider the mental experiment of infinite i.i.d. Bernoulli(p) sequence of random variables $X_1, X_2, \cdots$, which we interpret once more as coin tosses. We define Y be the number of tosses till the first heads. If we are just interested in Y (the first heads rather than all outcomes of all tosses), we can take Ω to be $\mathbb{N}$.

Exercise 2.5 *Although we cannot define a countable probability space for the infinite i.i.d. Bernoulli sequence, show that if we just want define a space for* Y*, we can take* $\Omega = \mathbb{N}$ *and* $\mathbb{P}(i) = (1-p)^{i-1} \cdot p$ *for* $i \geq 1$.

Y is known as a $\mathsf{Geometric}(p)$ random variable.

Let us calculate $\mathbb{E}[Y]$, in a somewhat creative way. Let E be the event that the first toss is heads. Then by total expectation we have,

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[Y|E] \cdot \mathbb{P}[E] + \mathbb{E}[Y|E^c] \cdot \mathbb{P}[E^c] \\ &= 1 \cdot \mathbb{P}[E] + (1 + \mathbb{E}[Y]) \cdot (1-p) \end{aligned}$$

Thus we have, $\mathbb{E}[Y] = \frac{1}{p}$. The main observation that we used here is that, thanks to independence, when the first toss is *not* heads, then the problem resets (with the hindsight of one consumed toss).

Exercise 2.6 *Compute* $\mathsf{Var}[Y]$ *for a* $\mathsf{Geometric}(p)$ *random variable* Y.

3 Coupon Collection

Consider the following problem: There are n kinds of items/coupons and at each time step we get one coupon chosen to be from one of the n types at random. All types are equally likely at each step and the choices at different time steps are independent. We define a random variable, T which is the time when we first have all the n types of coupons. Find $\mathbb{E}[T]$.

We can make the following claim:

$$T = \sum_{i=1}^{n} X_i \,,$$

where X_i is the time to get from the $i-1$ to the i types of coupons. Thus we have,

$$\mathbb{E}\left[T\right] = \sum_{i} \mathbb{E}\left[X_i\right]$$

Note that X_i is a geometric random variable with parameter $\frac{n-i+1}{n}$, since if we have $i-1$ type of coupons, X_i represents the time till we receive a coupon belonging to any one of the remaining $n-i+1$ types. Thus,

$$\mathbb{E}\left[X_i\right] \;=\; \frac{n}{n-i+1}\,.$$

Therefore,

$$\mathbb{E}\left[T\right] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} = n \cdot H(n)$$

where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is the n^{th} *harmonic number*. It is known (see Wikipedia for example) that $H_n = \ln n + \Theta(1)$. Thus, we have that $\mathbb{E}\left[T\right] = n \ln n + \Theta(n)$.