

## Lecture 15: November 17, 2016

Lecturer: Madhur Tulsiani

## 1 Probability over (uncountably) infinite probability spaces

Extending the idea of defining a probability *for each outcome* becomes problematic, when we try to extend it to uncountably infinite spaces. For example, let  $\Omega = [0,1]$ . Let  $\mu : [0,1] \rightarrow [0,1]$  be a function, which we want to think of as a probability distribution. Define the set

$$S_n = \left\{ x \in [0,1] \mid \mu(x) \geq \frac{1}{n} \right\}.$$

Since we want the total probability to add up to 1, we must have  $|S_n| \leq n$ . Also,

$$\text{Supp}(\mu) = \{x \in [0,1] \mid \mu(x) > 0\} \subseteq \bigcup_{n=1}^{\infty} S_n.$$

Since  $\bigcup_{n=1}^{\infty} S_n$  is a countable set,  $\mu(x) > 0$  only for countably many points  $x$ . Hence, it is problematic to think of the probability of the outcome  $x$ , for each  $x \in [0,1]$ . This can be resolved by only talking of probabilities of *events* for an allowed set of events obeying some nice properties. Such a set is known as a  $\sigma$ -algebra or a  $\sigma$ -field.

**Definition 1.1** Let  $2^\Omega$  denote the set of all subsets of  $\Omega$ . A set  $\mathcal{F} \subseteq 2^\Omega$  is called a  $\sigma$ -field (or  $\sigma$ -algebra) if

1.  $\emptyset \in \mathcal{F}$ .
2.  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$  (where  $A^c = \Omega \setminus A$ ).
3. For a (countable) sequence  $A_1, A_2, \dots$  such that each  $A_i \in \mathcal{F}$ , we have  $\bigcup_i A_i \in \mathcal{F}$ .

We then think of the sets in  $\mathcal{F}$  as the allowed events. We can now define probabilities as follows.

**Definition 1.2** Given a  $\sigma$ -field  $\mathcal{F} \subseteq 2^\Omega$ , a function  $\mu : \mathcal{F} \rightarrow [0,1]$  is known as a probability measure if

1.  $\mu(\emptyset) = 0$ .

2.  $\mu(E^c) = 1 - \mu(E)$  for all  $E \in \mathcal{F}$ .
3. For a (countable) sequence of disjoint sets  $E_1, E_2, \dots$  such that all  $E_i \in \mathcal{F}$ , we have

$$\mu(\cup_i E_i) = \sum_i \mu(E_i).$$

Note that the above definition do not say anything about unions of an uncountably infinite collection of sets. We can of course define probability measures on  $\mathcal{F} = 2^\Omega$  and hence define  $\mu(x)$  for all  $x \in \Omega$ . However, as we saw above, such measures will only have  $\mu(x) > 0$  countably many  $x$ . Consider the following example.

**Example 1.3** Let  $\Omega = [0, 1]$  and  $\mathcal{F} = 2^\Omega$ . Let  $T = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ . For each  $S \in \mathcal{F}$ , define

$$\mu(S) = \frac{|S \cap T|}{4}.$$

In the above example,  $\mu(x) > 0$  only for the points in  $T$ . To think of the “uniform distribution” on the the space  $[0, 1]$ , we would like that for  $a, b \in [0, 1]$  with  $a < b$ , the probability measure satisfies  $\mu[a, b] = b - a$ . It is a non-trivial result that such a probability measure indeed exists. This probability measure is known as the *Lebesgue* measure and is defined over a  $\sigma$ -algebra known as the *Borel*  $\sigma$ -algebra. The Borel  $\sigma$ -algebra does not contain all subsets of  $[0, 1]$  but does contain all intervals  $[a, b]$ . In fact, one can use the axiom of choice to show that a measure satisfying the requirement that  $\mu([a, b]) = b - a$  cannot be defined over the  $\sigma$ -algebra  $2^\Omega$ .

## 1.1 Random Variables

To define a random variable, we will also consider a  $\sigma$ -algebra  $\mathcal{F}'$  on the range of the random variable, and allow only functions for which we can correctly define the probability of the random variable taking values in a given set in  $\mathcal{F}'$ .

**Definition 1.4** Let  $\mathcal{F} \subseteq 2^\Omega$  and  $\mathcal{F}' \subseteq 2^\mathcal{R}$  be  $\sigma$ -algebras. A function  $f : \Omega \rightarrow \mathcal{R}$  is said to be measurable (under the  $\sigma$ -algebras  $\mathcal{F}$  and  $\mathcal{F}'$ ) if for all  $E' \in \mathcal{F}'$ , we have  $f^{-1}(E') \in \mathcal{F}$ .

We then define a require (say) a real-valued random variable to be a *measurable* function from  $\Omega$  to  $\mathbb{R}$ . Unless otherwise specified, we consider the  $\sigma$ -algebra on the target space  $\mathbb{R}$  to be the Borel  $\sigma$ -algebra. We will define the expectation of a random variable  $X : \Omega \rightarrow \mathbb{R}$  as the integral with respect to the measure.

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mu.$$

The definition of the integral with respect to a measure with respect to a measure requires some amount of care, though we will not be able to discuss this in much detail. Let  $\mu$  be any probability measure over the space  $\mathbb{R}$  equipped with the Borel  $\sigma$ -algebra. Define the function  $F$  as

$$F(x) := \mu((-\infty, x]),$$

which is well defined since the interval  $(-\infty, x]$  is in the  $\sigma$ -algebra. This can be used to define a random variable  $X$  such that  $\mathbb{P}[X \leq x] = F(x)$ . The function  $F$  is known as the distribution function or the cumulative density function of  $X$ .

When the function  $F$  has the form

$$F(x) = \int_{-\infty}^x f(z) dz,$$

then  $f$  is called the density function of the random variable  $X$ .

## 2 Gaussian Random Variables

A Gaussian random variable  $X$  is defined through the density function

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where  $\mu$  is its mean and  $\sigma^2$  is its variance, and we write  $X \sim \mathcal{N}(\mu, \sigma^2)$ . To see the definition gives a valid probability distribution, we need to show  $\int_{-\infty}^{\infty} \gamma(x) dx = 1$ . It suffices to show for the case that  $\mu = 0$  and  $\sigma^2 = 1$ . First we show the integral is bounded.

**Claim 2.1**  $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$  is bounded.

**Proof:** We see that

$$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx = 2 \int_0^{\infty} e^{-x^2/2} dx \leq 2 \int_0^2 1 dx + 2 \int_2^{\infty} e^{-x} dx = 4 + 2e^{-2},$$

where we use the fact that  $I$  is even and after  $x = 2$ ,  $e^{-x^2/2}$  is upper bounded by  $e^{-x}$ . ■

Next we show that the normalization factor is  $\sqrt{2\pi}$ .

**Claim 2.2**  $I^2 = 2\pi$ .

**Proof:**

$$\begin{aligned}
 I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy \\
 &= \int_0^{\infty} \int_0^{2\pi} e^{-r^2/2} r dr d\theta \quad (\text{let } x = r \cos \theta \text{ and } y = r \sin \theta) \\
 &= 2\pi \int_0^{\infty} e^{-s} ds \quad (\text{let } s = r^2/2) \\
 &= 2\pi.
 \end{aligned}$$

■

This completes the proof that the definition gives a valid probability distribution. We prove a useful lemma for later use.

**Lemma 2.3** For  $X \sim \mathcal{N}(0, 1)$  and  $\lambda \in (0, 1/2)$ ,

$$\mathbb{E} \left[ e^{\lambda \cdot X^2} \right] = \frac{1}{\sqrt{1 - 2\lambda}}.$$

**Proof:**

$$\begin{aligned}
 \mathbb{E} \left[ e^{\lambda \cdot X^2} \right] &= \int_{-\infty}^{\infty} e^{\lambda \cdot x^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2\lambda)x^2/2} dx \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{dy}{\sqrt{1-2\lambda}} \quad (\text{let } y = \sqrt{1-2\lambda}x) \\
 &= \frac{1}{\sqrt{1-2\lambda}}
 \end{aligned}$$

■

### 3 Johnson–Lindenstrauss Lemma

We will use concentration bounds on Gaussian random variables to prove the following important lemma.

**Lemma 3.1 (Johnson–Lindenstrauss [JL84])** Let  $\mathcal{P}$  be a set of  $n$  points in  $\mathbb{R}^d$ . Let  $0 < \varepsilon < 1$ . For  $k = \frac{8 \ln n}{\varepsilon^2/2 - \varepsilon^3/3}$ , there exists a mapping  $\varphi : \mathcal{P} \rightarrow \mathbb{R}^k$  such that for all  $u, v \in \mathcal{P}$

$$(1 - \varepsilon) \|u - v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1 + \varepsilon) \|u - v\|^2.$$

The above lemma is useful for dimensionality reduction, especially when a problem has an exponential dependence on the number of dimensions. We construct the mapping  $\varphi$  as follows. First choose a matrix  $G \in \mathbb{R}^{k \times d}$  such that each  $G_{ij} \sim \mathcal{N}(0, 1)$  is independent. Define

$$\varphi(u) = \frac{Gu}{\sqrt{k}}.$$

Note that by the above construction  $\varphi$  is oblivious, meaning that it doesn't depend on the points in  $\mathcal{P}$ , and it is linear. Before we prove the lemma, we will use the following fact several times.

**Fact 3.2** *Let  $Z = c_1X_1 + c_2X_2$ , where  $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2 \sim \mathcal{N}(0, 1)$  are independent. Then  $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$ .*

The strategy of proving the lemma is to first prove that with high probability the lemma holds for any fixed two points and then apply union bounds to get the result for all pairs of points.

**Claim 3.3** *Fix  $u, v \in \mathcal{P}$ . Let  $w = u - v$ . With probability greater than  $1 - 1/n^3$ , the following inequality holds,*

$$(1 - \varepsilon) \cdot \|w\|^2 \leq \|\varphi(w)\|^2 \leq (1 + \varepsilon) \cdot \|w\|^2.$$

**Proof:** Recall that  $\varphi(u) = \frac{Gu}{\sqrt{k}}$ . Let

$$Z = \frac{k\|\varphi(w)\|^2}{\|w\|^2} = \frac{\sum_{i=1}^k (Gw)_i^2}{\|w\|^2}.$$

We need to show  $(1 - \varepsilon)k \leq Z \leq (1 + \varepsilon)k$ . We know that the sum of Gaussian random variables is still a Gaussian random variable, so  $(Gw)_i = G_i w = \sum_{j=1}^n G_{ij} w_j$  is a Gaussian variable. Besides,  $\text{Var} \left[ \sum_{j=1}^n G_{ij} w_j \right] = \sum_j w_j^2 = \|w\|^2$  according to Fact 3.2. In other words,  $G_i w \sim \mathcal{N}(0, \|w\|^2)$ . As a result,  $Z = \sum_{i=1}^k \frac{(Gw)_i^2}{\|w\|^2} = \sum_{i=1}^k X_i^2$ , where  $X_i \sim \mathcal{N}(0, 1)$ . The expectation of each individual element in  $Gw$  is

$$\mathbb{E} [(Gw)_i^2] = \mathbb{E} [(G_i w)^2] = \mathbb{E} \left[ \left( \sum_{j=1}^n G_{ij} w_j \right)^2 \right] = \text{Var} \left[ \sum_{j=1}^n G_{ij} w_j \right] = \|w\|^2.$$

In addition,

$$\mathbb{E} [Z] = \frac{\sum_{i=1}^k \mathbb{E} [(Gw)_i^2]}{\|w\|^2} = k.$$

Now we prove the concentration bound for  $Z$ . The proof is almost identical to Chernoff bound.

$$\begin{aligned}
\mathbb{P}[Z \geq (1 + \varepsilon)k] &\leq \mathbb{P}\left[e^{tZ} \geq e^{\lambda \cdot (1 + \varepsilon)k}\right] \\
&\leq \frac{\mathbb{E}\left[e^{\lambda \cdot Z}\right]}{e^{\lambda \cdot (1 + \varepsilon)k}} && \text{(by Markov's inequality)} \\
&= \frac{\mathbb{E}\left[e^{\lambda \cdot \sum_{i=1}^k X_i^2}\right]}{e^{\lambda \cdot (1 + \varepsilon)k}} = \frac{\prod_{i=1}^k \mathbb{E}\left[e^{\lambda \cdot X_i^2}\right]}{e^{\lambda \cdot (1 + \varepsilon)k}} && \text{(by the independence of } X_1, \dots, X_k) \\
&= \frac{\prod_{i=1}^k \frac{1}{\sqrt{1 - 2\lambda}}}{e^{\lambda \cdot (1 + \varepsilon)k}} && \text{(by Lemma 2.3)} \\
&\leq \left(\frac{e^{-2(1 + \varepsilon)\lambda}}{1 - 2\lambda}\right)^{k/2} && \text{(assume } \lambda < 1/2) \\
&\leq (e^{-\varepsilon}(1 + \varepsilon))^{k/2} && \text{(let } \lambda = \frac{\varepsilon}{2(1 + \varepsilon)}) \\
&\leq \left(\left(1 - \varepsilon + \frac{\varepsilon^2}{2}\right)(1 + \varepsilon)\right)^{k/2} && \text{(by Taylor expansion of } e^{-x}) \\
&\leq e^{-\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{2}\right)\frac{k}{2}} && \text{(by } 1 + x \leq e^x)
\end{aligned}$$

We can derive the other side of the inequality in an analogous way. Thus, we have

$$\mathbb{P}[|Z - k| \geq \varepsilon k] \leq 2 \cdot \exp\left(-\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{2}\right)\frac{k}{2}\right) \leq 2 \cdot \exp(-3 \ln n) = \frac{2}{n^3},$$

where we choose

$$k = \left\lceil \frac{6 \ln n}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{2}} \right\rceil.$$

■

To prove Johnson–Lindenstrauss Lemma, we apply the union bound and get the desired result

$$\begin{aligned}
\mathbb{P}\left[\forall u, v \in \mathcal{P}, (1 - \varepsilon)\|u - v\|^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1 + \varepsilon)\|u - v\|^2\right] &\geq 1 - \binom{n}{2} \frac{2}{n^3} \\
&\geq 1 - \frac{1}{n}.
\end{aligned}$$

## References

[JL84] W Johnson and J Lindenstrauss, *Extensions of Lipschitz maps into a Hilbert space*, Contemporary Math **26** (1984). 4