

# STATS310A - Lecture 7

Persi Diaconis  
Scribed by Michael Howes

10/12/21

## Contents

<b>1</b>	<b>Measurable functions and random variables</b>	<b>1</b>
<b>2</b>	<b>Push forwards</b>	<b>2</b>
<b>3</b>	<b>Haar measure</b>	<b>2</b>
3.1	One answer . . . . .	3
3.2	A more mathematical answer . . . . .	3
<b>4</b>	<b>Independence</b>	<b>3</b>
<b>5</b>	<b>Constructing random variables</b>	<b>3</b>
<b>6</b>	<b>Maxima</b>	<b>4</b>

## 1 Measurable functions and random variables

Recall that a function  $T : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$  is *measurable* if  $T^{-1}(A') \in \mathcal{F}$  for all  $A' \in \mathcal{F}'$  where  $T^{-1}(A') = \{\omega \in \Omega : T(\omega) \in A'\}$ .

A *random variable* is a measurable function from  $(\Omega, \mathcal{F})$  to  $(\mathbb{R}, \mathcal{B})$  where  $\mathcal{B}$  is the set of Borel sets.

A *random vector* is a measurable function from  $(\Omega, \mathcal{F})$  to  $(\mathbb{R}^k, \mathcal{B}_k)$  where  $\mathcal{B}_k$  is the set of Borel subsets of  $\mathbb{R}^k$ .

**Lemma 1.** *If  $Y : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}^k$  is a function with coordinates  $Y_i : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ , for  $i = 1, \dots, k$ , then  $Y$  is a random vector if and only if  $Y_i$  is a random variable for each  $i$ .*

*Proof.* Suppose each  $Y_i$  is measurable then

$$\{\omega \in \Omega : Y(\omega) \leq (x_1, \dots, x_k)\} = \bigcap_{i=1}^k \{\omega \in \Omega : Y_i(\omega) \leq x_i\} \in \mathcal{F},$$

since each set  $\{\omega \in \Omega : Y_i(\omega) \leq x_i\}$  is in  $\mathcal{F}$  and  $\mathcal{F}$  is closed under finite intersections. Since sets of the form  $\{y \in \mathbb{R}^k : y \leq x\}$  generate  $\mathcal{B}_k$ , we have that  $Y$  is measurable.

If  $Y$  is measurable, then

$$\{\omega : Y(\omega) \leq x\} = \bigcup_{n=1}^{\infty} \{\omega : Y \leq (n, \dots, x, \dots, n)\} \in \mathcal{F},$$

since  $Y$  is a random vector and  $\mathcal{F}$  is closed under countable unions. The intervals  $(-\infty, x]$  generate  $\mathcal{B}$  and so  $Y_i$  is measurable.  $\square$

**Lemma 2.** If  $T : \mathbb{R}^k \rightarrow \mathbb{R}^j$  is continuous, then  $T$  is Borel-measurable.

*Proof.* Since  $T$  is continuous,  $T^{-1}(U)$  is open for all open sets  $U \subseteq \mathbb{R}^j$  and thus  $T^{-1}(U)$  is Borel for all open sets  $U \subseteq \mathbb{R}^j$ . Since the open sets generate the Borel  $\sigma$ -algebra,  $T$  is measurable.  $\square$

**Corollary 1.** If  $X, Y, (X_n)_{n=1}^\infty$  are random variables, then  $X+Y, XY, \max\{X, Y\}, \sup\{X_n\}, \inf\{X_n\}, \limsup\{X_n\}, \liminf\{X_n\}$  are all random variables. And the set  $\{\omega : \lim X_n(\omega) \text{ exists}\}$  is measurable.

*Proof.* We can write  $X + Y$  as a composition

$$\begin{aligned} \Omega &\rightarrow \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \\ \omega &\mapsto (X(\omega), Y(\omega)) \mapsto X(\omega) + Y(\omega). \end{aligned}$$

From the above lemma,  $X + Y$  is measurable. The others are similar.  $\square$

## 2 Push forwards

**Definition 1.** Suppose that  $(\Omega, \mathcal{F}, \mu)$  is a measure space and  $T : (\Omega, \mathcal{F}, \mu) \rightarrow (\Omega', \mathcal{F}')$  is measurable. We define the *push forward* of  $\mu$  along  $T$ , to be the measure  $\mu^{T^{-1}}$  on  $(\Omega', \mathcal{F}')$  defined by

$$\mu^{T^{-1}}(A) := \mu(T^{-1}(A)) = \mu(\{\omega : T(\omega) \in A\}).$$

Note  $\mu^{T^{-1}}$  is a measure. It is well defined because  $T$  is measurable. And

$$\mu^{T^{-1}}(\emptyset) = \mu(\emptyset) = 0.$$

If  $A \subseteq B$ , then  $T^{-1}(A) \subseteq T^{-1}(B)$  and so

$$\mu^{T^{-1}}(A) = \mu(T^{-1}(A)) \leq \mu(T^{-1}(B)) = \mu^{T^{-1}}(B).$$

If  $\{A_i\}_{i=1}^\infty$  are disjoint, then  $\{T^{-1}(A_i)\}_{i=1}^\infty$  are disjoint and so

$$\begin{aligned} \mu^{T^{-1}}\left(\bigcup_{i=1}^\infty A_i\right) &= \mu\left(T^{-1}\left(\bigcup_{i=1}^\infty A_i\right)\right) \\ &= \mu\left(\bigcup_{i=1}^\infty T^{-1}(A_i)\right) \\ &= \sum_{i=1}^\infty \mu(T^{-1}(A_i)) \\ &= \sum_{i=1}^\infty \mu^{T^{-1}}(A_i). \end{aligned}$$

Lebesgue's mistake/a warning: If  $U \subseteq \mathbb{R}^2$  is a Borel set, then the projections of  $U$  are not necessarily Borel sets.

## 3 Haar measure

Let  $O_n = \{M \in \mathbb{R}^{n^2} : M^T M = I_n\}$  be the orthogonal group. The group  $O_n$  has an invariant probability  $\nu$  which we call Haar measure. That is for all measurable  $A \subseteq O_n$  and  $m \in O_n$ ,  $\nu(m \cdot A) = \nu(A)$ . What is this measure?

### 3.1 One answer

We will give a recipe for drawing  $M \in O_n$  from  $\nu$ . To start let  $Z_{i,j} \sim N(0,1)$  be independent for  $1 \leq i, j \leq n$ . Let  $Z = (Z_{i,j})_{i,j=1}^n$  and apply Gram-Schmidt to  $Z$  to get a matrix  $M \in O_n$ .

### 3.2 A more mathematical answer

We know that  $\Phi(x) = \int_{-\infty}^x \exp(-t^2/2)dt$  is a distribution. Define on  $\mathbb{R}^{n^2}$

$$F(x_{1,1}, x_{1,2}, \dots, x_{n,n}) = \prod_{i,j=1}^n \Phi(x_{i,j}).$$

One can check that this defines a probabilities distribution  $\mu$  on  $\mathbb{R}^{n^2}$ . Define a function  $T : \mathbb{R}^{n^2} \rightarrow O_n$  given by given a matrix  $Z$ , apply Gram-Schmidt to  $Z$  to get  $M \in O_n$ . Finally define  $\nu := \mu^{T^{-1}}$  to be the push forward of  $\mu$  along  $T$ .

## 4 Independence

**Definition 2.** If  $\{X_i\}_{i \in I}$  is a collection of random variables, then we define the  $\sigma$ -algebra generated by  $\{X_i\}_{i \in I}$  to be

$$\sigma(X_i, i \in I) := \sigma(\{X_i^{-1}((a, b]) : i \in I, a, b \in \mathbb{R}\}).$$

**Definition 3.** Two random variables  $X, Y$  are *independent* if  $\sigma(X)$  and  $\sigma(Y)$  are independent. That is equivalently,

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y),$$

for all  $x, y \in \mathbb{R}$ . Yet another equivalent statement is that for all  $A, B \subseteq \mathbb{R}$  Borel

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

## 5 Constructing random variables

How do we pick from  $F$  where  $F$  is a univariate probability distribution? We first pick  $U$  which is uniformly distributed on  $[0, 1]$  and then we define  $T : [0, 1] \rightarrow \mathbb{R}$  by

$$T(u) = \inf\{x \in \mathbb{R} : T(x) \geq u\}.$$

Then  $\mathbb{P}(T(U) \leq x) = F(x)$ .

**Example 1.** Consider the case when

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 - e^{-x} & \text{if } x > 0. \end{cases}$$

Let  $u = 1 - e^{-x}$ , then  $x = -\log(1 - u)$ . Define  $T : (0, 1) \rightarrow \mathbb{R}$  by  $T(u) = -\log(1 - u)$  and  $X = T(U)$  where  $U$  is uniform on  $(0, 1)$ . Then if  $x > 0$ ,

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(-\log(1 - U) \leq x) \\ &= \mathbb{P}(-x \leq \log(1 - U)) \\ &= \mathbb{P}(e^{-1} \leq 1 - U) \\ &= \mathbb{P}(U \leq 1 - e^{-x}) \\ &= 1 - e^{-x}. \end{aligned}$$

Another good example if when  $X$  is discrete. Say  $X = a_i$  with probability  $p_i$ . Then the above construction divides  $[0, 1]$  into intervals  $A_i$  of length  $p_i$ . Then if  $U$  lies in  $A_i$ , then we set  $T(U)$  to be  $a_i$ . Thus  $T(U)$  and  $X$  have the same distribution.

## 6 Maxima

Let  $X_1, \dots, X_n$  be independent random variables with distribution

$$\mathbb{P}(X_i \leq x) = F(x).$$

Define  $M_n = \max\{X_i : i = 1, \dots, n\}$ . Then

$$\mathbb{P}(M_n \leq x) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \leq x) = F(x)^n.$$

What happens as  $n \rightarrow \infty$ ? Suppose that  $F(x) = 1 - e^{-x}$ . Then  $\mathbb{P}(M_n < x) = (1 - e^{-x})^n$ . We are interested in what happens when  $n \rightarrow \infty$ . Let  $x = \log(n) + y$ , then

$$\mathbb{P}(M_n \leq x) = \left(1 - \frac{e^{-y}}{n}\right)^n \sim e^{-e^{-y}}.$$

Then function  $F(y) = e^{-e^{-y}}$ ,  $y \in \mathbb{R}$  is a distribution function and is called the standard Gumble distribution.

**Definition 4.** We say that a sequence of distributions  $F_n$  converges in distribution to a distribution  $F$  if

$$F_n(x) \rightarrow F(x),$$

for all  $x$  such that  $F$  is continuous at  $x$ .

Why do we only restrict to  $x$  at which  $F$  is continuous? Consider the following example:  $X_n$  is a point mass at  $1 + \frac{1}{n}$  and  $X$  is a point mass at 1. Then

$$F_n(x) = \begin{cases} 0 & \text{if } x < 1 + 1/n, \\ 1 & \text{if } x \geq 1 + 1/n, \end{cases}$$

and

$$F(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

Thus  $F_n(x) \rightarrow F(x)$  if and only if  $x \neq 1$ . Thus in the definition of convergence in distribution we do not worry about the points at which  $F$  is not continuous.

We can now say that  $M_n - \log(n)$  converges in distribution to a Gumble distribution.

Now let's consider the maximum of Gaussians. Let  $X_1, \dots, X_n \sim N(0, 1)$ . We know that

$$\mathbb{P}(M_n \leq x) = (\Phi(x))^n = e^{n \log(\Phi(x))} = e^{n \log(1 - (1 - \Phi(x)))}.$$

We will use the approximation  $\log(1 - y) \sim -y$  as  $y \rightarrow 0$ . We also have (homework problem)

$$\frac{x}{1+x^2} \exp^{-x^2/2} \leq \int_x^\infty \exp(-t^2/2) \leq \frac{1}{x} \exp^{-x^2/2}.$$

Thus we can say  $1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x}$ . Thus for  $n$  large

$$\mathbb{P}(M_n \leq x) \sim e^{-n \frac{e^{-x^2/2}}{\sqrt{2\pi}x}}.$$

Let  $x = \sqrt{2 \log(n) - \log(\log(n))} + y$ , so  $x \sim \sqrt{2 \log(n) + y}$ , then

$$\mathbb{P}(M_n \leq \sqrt{2 \log(n) - \log(\log(n))} + y) \sim e^{-\frac{e^{-y/2}}{\sqrt{2\pi}}},$$

another Gumble distribution. We can not always perform these sorts of calculations. There are distributions such that  $\lim \mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right)$  does not exist for any choice of  $a_n, b_n$ . Discrete distributions such as the geometric or Poisson distributions tend to show this behaviour. Be careful when looking at the limiting behaviour of the maxima of discrete random variables.