STATS 310A: Theory of Probability I | Autumn 2016/17

Lecture 8: October 19

Lecturer: Persi Diaconis | *Scribes: Kenneth Tay*

8.1 The Basic Problem of Probability

With the tools developed in the past few weeks, we can state the basic problem of probability more precisely. Let's say we have:

- Some set Ω,
- $\mathcal{A}$ some collection of subsets of Ω,
- $\mu : \mathcal{A} \to [0, 1]$ some probability (as far as it can be).

Basic Problem: For any $B \in \sigma(\mathcal{A})$, compute or approximate $\mu(B)$.

8.2 Measurable Functions & Random Variables

Definition 8.1 *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be 2 measurable spaces. $T : \Omega \to \Omega'$ is* ***measurable*** *if for every $B' \in \mathcal{F}'$, $T^{-1}(B') := \{\omega \in \Omega : T(\omega) \in B'\} \in \mathcal{F}$.*

Example: Let $\Omega =$ all permutations of $[1, \dots, n]$, $\mathcal{F} = 2^\Omega$, and $\Omega' = [1, \dots, n]$, $\mathcal{F}' = 2^{\Omega'}$. Then $T(\pi) = \pi(i)$ is measurable.

The proposition below has 3 useful tricks we can use to manipulate these "inverses":

Proposition 8.2 *Let $\{B_i'\}_{i \in I}$ be a collection of sets in $\mathcal{F}'$.*

1. $[T^{-1}(B')]^c = T^{-1}(B'^c)$.
2. $T^{-1}\left(\bigcup_{i \in I} B_i'\right) = \bigcup_{i \in I} T^{-1}(B_i')$.
3. $T^{-1}\left(\bigcap_{i \in I} B_i'\right) = \bigcap_{i \in I} T^{-1}(B_i')$.

Two easy facts about measurable functions:

1. If $\mathcal{F}' = \sigma(\mathcal{A}')$ for some $\mathcal{A}'$, then T is measurable iff $T^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{A}'$. (This means that we only have to check $T^{-1}(A')$ for generators of $\mathcal{F}'$.)

   Proof: Say $T^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{A}'$. Consider the class of sets $S' \in \Omega'$ such that $T^{-1}(S') \in \mathcal{F}$. This class contains $\mathcal{A}'$, is closed under complements and countable unions, hence it contains $\sigma(\mathcal{A})$. ■

2. If $T : \Omega \to \Omega'$ and $T' : \Omega' \to \Omega''$ are measurable (with respect to the corresponding σ-algebras), then $T' \circ T$ is measurable.

Definition 8.3 *A* ***random variable*** *is a measurable function* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, B(\mathbb{R}))$.

A ***random vector*** Y *is a measurable function* $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, B(\mathbb{R}^k))$.

Proposition 8.4 *Let* $Y_1, \ldots, Y_k$ *be the coordinate functions of random vector* Y*, i.e.* $Y(\omega) = (Y_1(\omega), \ldots, Y_k(\omega))$. *Then* Y *is measurable iff* Y_i *is measurable for all* i.

Proof: Assume that Y_i measurable for all i. Then

$$\{\omega : Y(\omega) \leq (x_1, \ldots x_k) \text{ coordinate-wise}\} = \bigcap_{i=1}^{k} \{\omega : Y_i(\omega) \leq x_i\}.$$

The RHS is an intersection of measurable sets, so the LHS is a measurable set. Since the sets $\{\omega : Y(\omega) \leq (x_1, \ldots x_k) \text{ coordinate-wise}\}$ generates the σ-algebra associated with Ω, Y is measurable.

Conversely, assume that Y is measurable. Then

$$\{\omega : Y_i(\omega) \leq x\} = \bigcup_{n=1}^{\infty} \{\omega : Y(\omega) \leq (n, \ldots, n, x, n, \ldots, n)\},$$

where x is in the i^{th} coordinate. Each of the sets on the RHS is measurable, hence there union is measurable too. ■

The next proposition relates continuity to measurability:

Proposition 8.5 *If* $T : \mathbb{R}^k \to \mathbb{R}^k$ *is continuous, then* T *is measurable.*

The next proposition shows that many of the functions of random variables that we work with are random variables too:

Proposition 8.6

- *The sum,* max *and* min *of random variables are themselves random variables.*
- *The* sup*,* inf*,* lim sup *and* lim inf *of random variables are themselves random variables.*

The set $\{\omega : \lim X_i \text{ exists}\}$ *is measurable.*

But **not everything is measurable!** "Lebesgue's famous mistake": Let $f : [0,1]^2 \to [0,1]$ be a function which is Borel measurable. Consider $g(x) = \sup_y f(x, y)$. Lebesgue claimed that g is measurable, but Lusin showed that this is not always the case! Lusin found a Borel set in $\mathbb{R}^2$ whose projection to one axis is not measurable.

8.2.1 σ-Algebras from Random Variables

If $\{X_i\}_{i \in I}$ are random variables, we can define $\sigma(\{X_i\}_{i \in I})$ as the σ-algebra generated by the sets $X_i^{-1}(a, b]$, where $(a, b]$ ranges over all intervals. This is a sub σ-algebra of $\mathcal{F}$, and is the smallest σ-algebra such that X_i is measurable for all $i \in I$.

Definition 8.7 *Two random variables* X *and* Y *are independent if* $\sigma(X)$ *is independent of* $\sigma(Y)$*, i.e. for all* x*,* y*,*

$$\mu\{\omega : X(\omega) \leq x, Y(\omega) \leq y\} = \mu\{\omega : X(\omega) \leq x\} \cdot \mu\{\omega : Y(\omega) \leq y\}.$$

8.3 New Measures from Old

Definition 8.8 *Suppose we have $(\Omega, \mathcal{F}, \mu)$ a measure space, $(\Omega', \mathcal{F}')$ a measurable space, $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ measurable.*

Define, $\mu^{T^{-1}}$ by

$$\mu^{T^{-1}}(B') := \mu(T^{-1}(B'))$$

*for $B' \in \mathcal{F}'$. This is called the **push-forward** of μ by T.*

Note that $\mu^{T^{-1}}$ is a measure. **Proof:**

- $\mu^{T^{-1}}(\emptyset') = \mu(\emptyset) = 0$.
- If $A' \subseteq B'$, then $T^{-1}(A') \subseteq T^{-1}(B')$. Hence, $\mu^{T^{-1}}(A') \leq \mu^{T^{-1}}(B')$.
- For any sets $B'_i \in \mathcal{F}'$, $i = 1, 2, \ldots$

  $$\mu^{T^{-1}}\left(\bigcup_i B'_i\right) = \mu\left(T^{-1}\left(\bigcup_i B'_i\right)\right).$$

  Note that $T^{-1}\left(\bigcup_i B'_i\right) = \bigcup_i T^{-1}(B'_i)$. If the B'_i's are disjoint, then the $T^{-1}(B'_i)$'s are disjoint, so

  $$\begin{aligned}\mu^{T^{-1}}\left(\bigcup_i B'_i\right) &= \sum_i \mu(T^{-1}(B'_i)) \\ &= \sum \mu^{T^{-1}}(B'_i).\end{aligned}$$

■

Example ("Pick $m \in \mathcal{O}_n$ from Haar measure"). Consider the following algorithm:

1. Pick X_{ij} iid from $\mathcal{N}(0,1)$, i.e. on $\mathbb{R}^{n^2}$, build a measure with $F(x_{11}, x_{12}, \ldots, x_{nn}) = \prod \Phi(x_{ij})$, where $\Phi(x) = \displaystyle\int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$.
2. Map the X_{ij}'s deterministically by the Gram-Schmidt algorithm $T : \mathbb{R}^{n^2} \to \mathcal{O}_n$.

Then $F^{T^{-1}}$ is Haar measure.

8.4 How to Pick a Random Value from a Distribution F

8.4.1 Case 1: F continuous, strictly increasing CDF

Because F is continuous and strictly increasing, given any point y, there is x such that $F(x) = y$. (In fact, $F(F^{-1}(y)) = y$ and $F^{-1}(F(x)) = x$.)

Now, let U be a uniform random variable, i.e. $P(U \leq x) = x$ for $0 \leq x \leq 1$. Consider $Y = F^{-1}(U)$. We have

$$\begin{aligned} P(Y \leq x) &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

Thus, if you know how to pick from a uniform distribution, you can use that to pick from F.

Example (Exponential random variable): Let $F(x) = (1 - e^{-x})_+$. We wish to pick y from F.

$$\begin{aligned} y &= 1 - e^{-x}, \\ e^{-x} &= 1 - y, \\ x &= -\log(1 - y). \end{aligned}$$

Thus, $-\log(1 - U)$ will be distributed like an exponential random variable:

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

8.4.2 Case 2: F corresponds to a discrete measure

In this case, let's say that there are outcomes $A_1, A_2, \ldots$ with outcome probabilities $P(A_1), P(A_2), \ldots$. Split the unit interval into parts of length $P(A_1), P(A_2), \ldots$.

Pick from the uniform distribution and see which interval it lands in. If it lands in the part of length $P(A_i)$, assign the value i.

8.4.3 General CDFs

In general, for $F : \mathbb{R} \to [0, 1]$, we can define $F^{-1}(y) = \inf\{x : y \leq F(x)\}$. Then we have the following:

Proposition 8.9 *For all x,*

$$P\{F^{-1}(U) \leq x\} = F(x).$$

Example: Let $X_1, \ldots, X_n$ be iid, exponential variables. Let $M_n = \max X_i$. Find the distribution function (and limit as $n \to \infty$) of M.

The distribution function of M_n is given by

$$\begin{aligned} P\{M_n \leq x\} &= P\{X_i \leq x \text{ for all } i\} \\ &= P\{X_1 \leq x\}^n \\ &= (1 - e^{-x})^n. \end{aligned}$$

Take $x = \log n + c$. Then for fixed $c \in (-\infty, \infty)$,

$$\begin{aligned} P\{M_n \le \log n + c\} &= (1 - e^{-\log n - c})^n \\ &= \left(1 - \frac{e^{-c}}{n}\right)^n \\ &\to e^{-e^{-c}} \end{aligned}$$

as $n \to \infty$.

Thus, $F(c) = e^{-e^{-c}}$ is a distribution function, and it is called the extreme value distribution, or Gumbel distribution.