Statistics Theory, Self-Study

Selected Solutions for Casell and Berger's Statistical Inference (2nd)
Others to be added on an ad hoc basis

Guilherme Albertini

Contents

Preface

The following briefly covers measure theory that can be used for additional understanding. See some links:

    https://www.youtube.com/watch?v=tBw_OOKEaDA&list=PLBh2i93oe2qtmWsYpxONarbeahMJvhRj0
    https://www.youtube.com/watch?v=FtEmLexUw3Y&list=PLBh2i93oe2quIJS-j1NpbzEvQCmNOOF5o
    https://mbernste.github.io/posts/measure_theory_1/
    https://www.quora.com/How-are-sigma-algebras-different-from-power-sets/answer/Louis-de-Thanhoffer-
    de-Volcsey?ch=15&oid=57710983&share=6b27e662&srid=Yw7L&target_type=answer

The power set of some set S (usually denoted $\mathcal{P}(S)$ or 2^S) is the set of all subsets of this set. This necessarily includes the empty set and it's complement, the set itself. It is the complete enumeration of subsets over the set in question, but comes at the cost of running into the issue of having non-measurable subsets over a field of interest. The sigma algebra rescues us here as it is a clever way to restrict the collection of subsets to those that fit a measure of interest in a translation invariant way. So, to recap, a set $\mathcal{A} \subseteq \mathcal{P}(S)$ is a $\sigma-$algebra if $\emptyset, S$ is in $\mathcal{A}$; all the corresponding complements $\mathcal{A}^c$ are in $\mathcal{A}$; and that the countable union of infinitely many subsets $A_i \in \mathcal{A}$, for index $i \in \mathbb{I}$, are also in $\mathcal{A}$. If these 3 things hold, then $A \in \mathcal{A}$ is called an $\mathcal{A}-$measurable set. The smallest possible $\sigma-$algebra must then be $\mathcal{A} = \{\emptyset, S\}$ and the largest must be the power set.

We can also show that the countable intersection of infinitely many subsets $A_i \in \mathcal{A}$ also form a sigma algebra over $\mathcal{A}$. If we had some set $m \subseteq \mathcal{P}(S)$, there is the smallest sigma algebra that contains this set by

$$\sigma(m) := \bigcup_{\mathcal{A} \supseteq m;\ \mathcal{A}\ \sigma\text{-algebras}} \mathcal{A}$$

and say this is the σ-algebra generated by m. So if we had $m = \{\{a\}, \{b\}\}$ and $S = \{a, b, c\}$, $\sigma(m) = \{\emptyset, S, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{b, c\}, \{c\}\}$. To avoid additional discussion without seeing the lecture clips above, take the Borel sigma algebra $\mathcal{B}(S)$ to be the sigma algebra generated by the open sets (i.e., all the subsets of $]a, b[$).

What is this measure stuff? We say $(S, \mathcal{A})$ form a measurable space, which is the set and its sigma-algebra, or collection of subsets up to the power set. We can also find that $\exists \mu \mid \mathcal{A} \mapsto [-\infty, \infty]$ where this map μ satisfies $\mu(\emptyset) = 0$ and $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ for pairwise disjoint subsets $A_{i,j;\ i \neq j} \in \mathcal{A}$; the latter is called sigma additivity. With this, we can extend the measurable space by selecting this map to form $(S, \mathcal{A}, \mu)$, a measure space.

For our purposes, we seek to find a measure on some set $X = \mathbb{R}^n$ such that $\mu([0, 1]^n) = 1$ and $\mu(x + A) = \mu(A)\ \forall x \in \mathbb{R}^n$. The latter is what is meant by translation invariance; that is, if we take any subset $A \in X$ and any point $x \in A$ then this property of "size persistence after moving" holds. The problem that we run into is that we need the measure μ on $\mathcal{P}(\mathbb{R})$ to behave in such a way that $\mu([a, b]) = b - a,\ b > a$ and $\mu(x + A) = \mu(A)\ A \in \mathcal{P}(\mathbb{R}),\ x \in \mathbb{R}$ BUT we are dealing with an infinitely dense set of real numbers. To show why the measure does not exist using the whole power set (resulting in this problem) –watch video series with shows that the measure $\mu(\mathbb{R}) = 0$– we instead select measurable sets in this power set, namely those having $\mu((0, 1]) < \infty$ and are translation invariant.

Now we want to find a relationship from one measure space $(\Omega_1, \mathcal{A}_1)$ to another measure space $(\Omega_2, \mathcal{A}_2)$. This requires the use of measurable maps (w.r.t. $\mathcal{A}_1, \mathcal{A}_2$). We can say

$$f : \Omega_1 \mapsto \Omega_2 \text{ if } f^{-1}(A_2) \in \mathcal{A}_1 \ \forall A_2 \in \mathcal{A}_2$$

which means that the preimages under f are measurable. This is important because if we have an abstract measure space on the x-axis of a graph, $(\Omega, \mathcal{A}, \mu)$, and have the y-axis represent the corresponding density of the interval (say, 1), then the preimage would represent the x-interval which relates to this density. But if we then want its measure, $\mu(f^{-1}(1))$, then the preimage must be measurable, $f^{-1}(1) \in \mathcal{A}$. Now, this matters when considering measure spaces $(\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and use a characteristic (or indicator) function $\chi_A : \Omega \to \mathbb{R}, \chi_A(\omega) := \mathbb{1}_{\omega \in A}$. For all measurable $A \in \mathcal{A}$, then this function becomes a measurable map. This means that $\chi_A^{-1}(\emptyset) = \emptyset, \chi_A^{-1}(\mathbb{R}) = \Omega$ which are both in $\mathcal{A}$. Similarly, $\chi^{-1}(1) = A, \chi^{-1}(0) = A^c$ which are also in $\mathcal{A}$. If we look back at the two measure spaces from before if we have functions $f, g : \Omega \to \mathbb{R}$ measurable $\implies f + g, f - g$ also measurable.

Now we can explore the Lebesgue integral. Recall measure space $(X, \mathcal{A}, \mu)$ for a set X, a collection of subsets of X called the σ-algebra and a map μ with domain $\mathcal{A}$ and codomain $[0, \infty]$ as its measure. We care for measurable maps $f : X \to \mathbb{R}$ noticing σ-algebra $\mathcal{A}$ is in X and σ-algebra $\mathcal{B}$ is in $\mathbb{R}$. The preimage of all Borel sets $E \subseteq \mathbb{R}$, $f^{-1}(E) \in \mathcal{A}$ is then $\mathcal{A}$-measurable. Now, a simple function (or step function) is one that takes a linear combination of characteristic functions we described before such that you can find $A_1, \ldots, A_n \in \mathcal{A}$ and $c_1, \ldots, c_n \in \mathbb{R}$ so $f(x) = \sum_{i=1}^{\infty} c_i \cdot \chi_{A_i}(x)$ as we know the characteristic function is measurable and that sums of measurable maps are also measurable. The issue we run into is that what if certain measures of subsets A_i are infinite? We can get something like integral $I(f) = \sum_{i=1}^{2} c_i \cdot \mu(A_i) = 3 \cdot \infty - 2 \cdot \infty$. The workaround is then to restrict ourselves either to only have positive c_i or just to focus on subsets A_i with bounded measure. Instead of dropping subsets from consideration, we restrict ourselves to the former.

$$\mathcal{S}^+ := \{f : X \to \mathbb{R} \mid f \text{ simple}, \ f \geq 0\}$$

Thus, we can select a $f \in \mathcal{S}^+$, such that $f(x) = \sum_{i=1}^{n} c_i \cdot \chi_{A_i}(x), \ c_i \geq 0$ without omitting subsets mentioned previously. This is where the Lebesgue integral w.r.t μ is housed, defined by integral

$$I(f) := \int_X f d\mu = \sum_{i=1}^{n} c_i \cdot \mu(A_i) \in [0, \infty]$$

Note that this integral is almost linear (as halfspace gets us only in positive reals for constants) such that $I(\alpha + \beta) = \alpha I(f) + \beta I(f), \ \alpha, \beta \geq 0$. Also note for $f \leq g \implies I(f) \leq I(g)$, aka monotone. By defining a step function $h = \sum_{i=1}^{n} c_i \cdot \chi_{A_i}$, for increasingly finite intervals on a y-axis of c_i for a "fixed" x-axis of the abstract measure space, then we approximate the integral under the curve. Namely,

$$\int_X f d\mu := \sup\{I(h) \mid h \in \mathcal{S}^+, h \leq f\}$$

which defines the Lebesgue integral w.r.t μ. Note that f is called μ-measurable if the integral is bounded.

Chapter 1

Probability Theory

1.1 Priors

Exercise 1.1
r