Set theory, and set operations

Sayan Mukherjee

Motivation

It goes without saying that a Bayesian statistician should know probability theory in depth to model.

Measure theory is not needed unless we discuss the probability of two types of events: (1) infinitely repeated operations such as infinite sequences of coin tosses (2) infinitely fine operations such as drawing uniformly from the unit interval. In both these cases we have to worry about infinite events so issues of countability, limits, and measure occur. Measure theoretic probability is needed to understand when and how our intuition about discrete events carries over to both these cases. It also gives us a common framework for both of the above events.

Why need for care ?

One example is order of integration which for Bayesian statistics is very important since we condition and marginalize extensively

$$\begin{aligned} R &= \{0 \leq x \leq 2, 0 \leq y \leq 1\} \\ f(x,y) &= \frac{xy(x^2-y^2)}{(x^2+y^2)^3}, for (x,y) \neq (0,0) \\ ? &= \int\int_R f(x,y)dxdy \end{aligned}$$

If we first integrate y and then x: $\frac{1}{5}$.
If we first integrate x and then y: $-\frac{1}{20}$.

Another example is the Dirichlet process. The following is a simple way of constructing a Dirichlet process:

1. Draw $X_1, ..., X_N \stackrel{iid}{\sim}$ Poi(λ)
2. Normalize $Z_i = \frac{X_i}{\sum_i X_i}$.

What is the distribution underlying this random process or algorithm ? Is there a sense of convergence of the outputs of this algorithm ?

Important ideas

Law of large numbers: Given $X_1, ..., X_n$ of iid random variables we state that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mu,$$

what does convergence ($\rightarrow$) mean ?

Central limit theorem:

$$Z_n = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma} \rightarrow \text{No}(0,1),$$

what does convergence ($\rightarrow$) mean ? What is convergence in distribution ?

Law of the iterated logarithm:

$$\limsup_n \frac{(\sum_{i=1}^{n} X_i - n\mu)}{\sqrt{2n\sigma^2 \log\log n}} = 1,$$

what is $\limsup$?

Notation

Ω: set of possible outcomes of some experiment.

$\omega \in \Omega$: one of the outcomes of the experiment.

$A \subseteq \Omega$ A is a subset of Ω; A is true if $\omega \in A$.

2^Ω: All subsets of Ω called th power set $\mathcal{P}(\Omega)$

Operations

Complement; $A^c = \text{not A} = \{w : w \notin A\}$

Union over arbitrary index set

$$\begin{aligned} \bigcup_\alpha A_\alpha &= \{w : w \in A_\alpha \text{ for at least one } \alpha\}, \\ A \cup B &= A \text{ or } B \text{ (or perhaps both).} \end{aligned}$$

Intersection over arbitrary index set

$$\begin{aligned} \bigcap_\alpha A_\alpha &= \{w : w \in A_\alpha \text{ for all } \alpha\}, \\ A \cap B &= AB = \text{both } A \text{ and } B. \end{aligned}$$

Set difference; symmetric difference

$$\begin{aligned} A \setminus B &= A \cap B^c \\ A \triangle B &= (A \cap B^c) \cup (A^c \cap B). \end{aligned}$$

Relations

Containment: $A \subset B : \omega \in A \Rightarrow \omega \in B$

Disjoint: $A \cap B = \emptyset$

Equality: $A = B : \omega \in A$ iff $\omega \in B$

De Morgan's Law:

$$\begin{aligned} \left(\bigcup_\alpha A_\alpha\right)^c &= \bigcap_\alpha (A_\alpha^c) \\ \left(\bigcap_\alpha A_\alpha\right)^c &= \bigcup_\alpha (A_\alpha^c) \end{aligned}$$

Montone sequence of sets:
$A_n \uparrow A$ means that $A_1 \subset A_2 \subset ....$ and $\lim_{n\to\infty} A_n = \bigcup_{n=1}^\infty A_n$.
$A_n \nearrow A$ means that $A_1 \subseteq A_2 \subseteq ....$ and $\lim_{n\to\infty} A_n = \bigcup_{n=1}^\infty A_n$.
$A_n \downarrow A$ means that $A_1 \supset A_2 \supset ....$ and $\lim_{n\to\infty} A_n = \bigcap_{n=1}^\infty A_n$.
$A_n \searrow A$ means that $A_1 \supseteq A_2 \supseteq ....$ and $\lim_{n\to\infty} A_n = \bigcap_{n=1}^\infty A_n$.

Limits of sets

Infinitely often (io): $\limsup_{n\to\infty} A_n \equiv \bigcap_{m\geq 1} \bigcup_{n\geq m} A_n$. Also called infinitely often as A_n occurs infinitely often as $n \to \infty$. The event $\limsup A_n$ occurs if and only if infinitely many of the A_n occur.

Almost all (aa): $\liminf_{n\to\infty} A_n \equiv \bigcup_{m\geq 1} \bigcap_{n\geq m} A_n$. Also called almost all since A_n occurs with at most finitely many exceptions.

$\{\liminf A_n\} \subseteq \{\limsup A_n\}$

$$\begin{aligned} A_n = \{n, n+1, ...\} &\qquad \limsup A_n = \liminf A_n = \emptyset \\ A_n = \{1, ..., n\} &\qquad \limsup A_n = \liminf A_n = \mathbb{N} \\ A_{2n} = \{2, 4, 6, ...\} &\qquad \limsup A_n = \mathbb{N}, \liminf A_n = \emptyset \end{aligned}$$

Convergence: $A_n \to A \quad \Rightarrow \quad \liminf A_n = \limsup A_n = A.$

Indicator functions

Indicator functions allow for Boolean algebra to be turned into ordinary algebra.

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

The operations of unions and intersections correspond to pointwise maxima ($\vee$) or minima ($\wedge$)

$$\mathbb{I}_{\cup A_i}(x) = \bigvee_i \mathbb{I}_{A_i}(x) \quad \mathbb{I}_{\cap A_i}(x) = \bigwedge_i \mathbb{I}_{A_i}(x).$$

$$\begin{aligned} \mathbb{I}_{A^c}(x) &= 1 - \mathbb{I}_A(x) \\ \mathbb{I}_{A \setminus B}(x) &= \max(0, \mathbb{I}_A(x) - \mathbb{I}_B(x)) \\ \mathbb{I}_{A \triangle B}(x) &= |\mathbb{I}_A(x) - \mathbb{I}_B(x)|. \end{aligned}$$

$$\left(\bigcap_{i=1}^n A_i\right) \triangle \left(\bigcap_{i=1}^n B_i\right) \subseteq \bigcup_{i=1}^n (A_i \triangle B_i) \Rightarrow \left|\prod_i \mathbb{I}_{A_i} - \prod_i \mathbb{I}_{B_i}\right| \leq \max_i |\mathbb{I}_{A_i} - \mathbb{I}_{B_i}|.$$

Revisit $\limsup$: $\limsup$ of a set is really much like $\limsup$ of a function

$$\mathbb{I}_{\limsup_n A_n} = \limsup_n \mathbb{I}_{A_n}.$$

$$\begin{aligned} \limsup_{n\to\infty} &\equiv \lim_{n\to\infty} \left(\sup_{m \geq n} x_m\right) \\ \liminf_{n\to\infty} &\equiv \lim_{n\to\infty} \left(\inf_{m \geq n} x_m\right) \end{aligned}$$

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n.$$

$x_n \to x \quad \Rightarrow \quad \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = x.$

Some sets are bigger than others

The cardinality of a set $|\Omega|$ is the number of elements in the set.

Theorem 0.0.1 (Cantor) *For any set Ω and power set $\mathcal{P}(\Omega)$, $|\Omega| < |\mathcal{P}(\Omega)|$.*

Example 0.0.1 $\Omega = \mathbb{N}$ *– an infinite but countable set*
$\mathcal{P}(\Omega)$ *– uncountable*
$\mathbb{R}$ *– uncountable*
$\mathbb{Q}$ *– the rationals are countable.*

Fields and algebras

Definition 0.0.1 (algebra) *Let $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is called a field (algebra) if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and finite union:*

1. *(i) $\Omega \in \mathcal{F}$*
2. *(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$*
3. *(iii) $A_1, ..., A_n \in \mathcal{F} \Rightarrow \bigcup_{j=1}^n A_j \in \mathcal{F}$.*

We could replace closure under finite unions with closure under finite intersections (due to de Morgan's law)

$$A_1, ..., A_n \in \mathcal{F} \Rightarrow \bigcap_{j=1}^n A_j \in \mathcal{F}.$$

Definition 0.0.2 (σ-algebra) *Let $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is called a field (algebra) if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and countable unions:*

1. *(i)* $\Omega \in \mathcal{F}$
2. *(ii)* $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. *(iii)* $A_1, ..., A_n \in \mathcal{F} \Rightarrow \bigcup_{j=1}^n A_j \in \mathcal{F}$
4. *(iv)* $A_1, ..., A_n, ... \in \mathcal{F} \Rightarrow \bigcup_{j=1}^\infty A_j \in \mathcal{F}$

An algebra is a σ-algebra but not vice versa.

Example 0.0.2

$$\Omega = \mathbb{R}, \quad A_n = (a, b].$$

$\mathcal{F}$ is the collection of all finite disjoint unions of $(a, b]$.
$\mathcal{F}$ is an algebra.
$\mathcal{F}$ is not a σ-algebra.

$$\bigcup_{n=1}^{\infty} (0, n^{-1}] = (0, 1).$$

Definition 0.0.3 (Borel set) *Let $\Omega = \mathbb{R}$ and $\mathcal{B}_0$ be the field of right-semiclosed intervals. Then $\sigma(\mathcal{B}_0)$ is the Borel σ-algebra of $\mathbb{R}$.*

This idea can be extended to metric spaces such as $\mathbb{R}^d$ or the space of continuous functions.

Probability spaces

The idea of a probability space is to assign a measure or nonnegative number to subsets of of Ω.

Definition 0.0.4 (Probability space) *A probability space is a triple $(\Omega, \mathcal{F}, \mathbf{P})$ where $\mathbf{P}$ is a function (probability measure) $\mathbf{P} : \mathcal{F} \to [0, 1]$ such that*

1. *(i)* $\mathbf{P}(A) \geq 0, \quad \forall A \in \mathcal{F}$
2. *ii* $\mathbf{P}(\Omega) = 1$
3. *iii if $A_j \in \mathcal{F}$ are disjoint*

$$\mathbf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbf{P}(A_j).$$

Properties of probability spaces

1. (i) monotonicity: if $A \subset B$ then $\mathbf{P}(A) < \mathbf{P}(B)$
2. (ii) σ-subadditive: for A_n, $n \geq 1$

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n)$$

3. (iii) continuity:
   if $A_n \uparrow A$ and $A_n \in \mathcal{F}$ then $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$
   if $A_n \downarrow A$ and $A_n \in \mathcal{F}$ then $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$
4. (iv) inclusion-exclusion: for $A_1, ..., A_n$ (related to Euler characteristic in topology)

$$\mathbf{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbf{P}(A_j) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i A_j) + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i A_j A_k) - \cdots (-1)^{n+1} \mathbf{P}(A_1 \cdots A_n)$$

Example 0.0.3 *Two events that are disjoint but the sum of the probability of the union is equal to the probability of the sum. $A = (0, .5]$ and $B = [.5, 1)$ with $\mathbf{P}(A) = .5$ and $\mathbf{P}(B) = .5$. $\mathbf{P}(A \cup B) = 1 = \mathbf{P}(A) + \mathbf{P}(B)$ and $A \cap B = \{.5\} \neq \emptyset$.*

Lemma 0.0.1 (Fatou) *For $A_n \in \mathcal{F}$ for $n \geq 1$*

1. *(i)*

$$\begin{aligned} \mathbf{P}(\liminf A_n) &\leq \liminf \mathbf{P}(A_n) \\ &\leq \limsup \mathbf{P}(A_n) \leq \mathbf{P}(\limsup A_n) \end{aligned}$$

2. *(ii) If $A_n \to A$ then $\mathbf{P}(A_n) \to \mathbf{P}(A)$.*

For a probability measure μ and a sequence of functions f_n (think $f_n = \mathbb{I}_{A_n}$) *(i)* is sometimes written with functional notation

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu.$$

A typical use of Fatous lemma is the following. Suppose we have $f_n \to f$ and $\sup_{n\geq 1} |f_n| \leq K < \infty$. Then $|f_n| \to |f|$, and by Fatou's lemma $\int |f| \leq K$. (What does the above have to do with any problem of inference ?)

Example 0.0.4 *Discrete probability measure on the real numbers whose support is the real line.*

The support of a discrete probability mesure is not necessarily just the set of points at which it places mass, two points really really close may not be individually distinguished. Let $r_1, r_2, ..$ be an enumeration of the rational numbers (how would one enumerate them ?) and let $\mathbf{P}$ place mass $1/2^n$ at r_n. $\mathbf{P}$ is discrete but has support on the line since for any x and any $\varepsilon > 0$

$$\mathbf{P}(\{x - \varepsilon, x + \varepsilon\}) > 0,$$

since the interval contains at leasnt one rational number.

Does this example have any uses in non-parametric statistical models ?

Definition 0.0.5 (Distribution function) *A function $F : \mathbb{R} \to [0, 1]$ is a probability distribution function (df) with*

$$F(x) = \mathbf{P}((-\infty, x]), \quad x \in \mathbb{R},$$

if F

1. *is continuous*
2. *is monotone and non-decreasing*
3. *has limits at $\pm\infty$*

$$\begin{aligned} F(\infty) &:= \lim_{x\uparrow\infty} F(x) = 1 \\ F(-\infty) &:= \lim_{x\downarrow\infty} F(x) = 0. \end{aligned}$$

Example 0.0.5 *Distributions that are neither discrete nor continuous.*

1. *Define a distribution F that is χ-squared with two degrees of freedom.*
   1. *Observe K from a Poisson distribution with mean λ.*
   2. *Sum K observations from F and report this value, if $K = 0$ report 0.*
2. *Define a distribution F that is standard normal.*
   1. *Observe X from a Bernoulli distribution with paramater p.*
   2. *If $X = 0$ report 0 otherwise report a draw from F.*

Are these distributions interesting priors ?

Dynkin's $\pi - \lambda$ theorem

We will soon need to define probability measures on infinite and possible uncountable sets, like the power set of the naturals. This is hard. It is easier to define the measure on a much smaller collection of the power set and state consistency conditions that allow us to *extend* the measure to the the power set. This is what Dynkin's $\pi - \lambda$ theorem was designed to do.

Definition 0.0.6 (π-system) *Given a set Ω a π system is a collection of subsets $\mathscr{P}$ that are closed under finite intersections.*
1) $\mathscr{P}$ is non-empty; 2) $A \cap B \in \mathscr{P}$ whenever $A,\ B \in \mathscr{P}$.

Definition 0.0.7 (λ-system) *Given a set Ω a λ system is a collection of subsets $\mathscr{L}$ that contains Ω and is closed under complementation and disjoint countable unions.*
1) $\Omega \in \mathscr{L}$; 2) $A \in \mathscr{L} \Rightarrow A^c \in \mathscr{L}$; 3) $A_n \in \mathscr{L},\ n \geq 1$ with $A_i \cap A_j = \emptyset\ \forall i \neq j \quad \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{L}$.

Definition 0.0.8 (σ-algebra) *Let $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is called a field (algebra) if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and countable unions,*
1) $\Omega \in \mathcal{F}$; 2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$; 3) $A_1, ..., A_n \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{n} A_j \in \mathcal{F}$; 4) $A_1, ..., A_n, ... \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.

A σ-algebra is a π system.
A σ-algebra is a λ system but a λ system need not be a σ algebra, a λ system is a weaker system.

Example 0.0.6 $\Omega = \{a, b, c, d\}$ *and* $\mathscr{L} = \{\Omega, \emptyset, \{a, b\}, \{b, c\}, \{c, d\}, \{a, d\}, \{b, d\}, \{a, c\}\}$ *$\mathscr{L}$ is closed under disjoint unions but not unions.*

However:

Lemma 0.0.2 *A class that is both a π system and a λ system is a σ-algebra.*

Next lecture we will start constructing probability measures on infinite sets and we will need to push the idea of extension of measure. In this the following theorem is central.

Theorem 0.0.2 (Dynkin) *If $\mathscr{P}$ is a π system and $\mathscr{L}$ is a λ system, then $\mathscr{P} \subset \mathscr{L}$ implies $\sigma(\mathscr{P}) \subset \mathscr{L}$.*