Statistics 235A / Math 235A
Probability & Measure
UC Davis, Fall 2016
Instructor: Ethan Anderes

Lets start the class with a paradox.

Suppose X, Y are two independent positive random variables (each with a density).

Question:

What is the distribution of X given the event $\{X = Y\}$?

Two possible solutions:

1. Take $X|X=Y$ to mean $X|Z=0$ where $Z = X - Y$.

   By "change of variables" $X|Z=0$ has density
   $$f_{X|Z=0}(x) = \frac{f_{X,Z}(x,0)}{f_Z(0)} \propto f_{X,Y}(x, x-0)\underbrace{\left|\frac{d(X,Y)}{d(X,Z)}\right|}_{=1}$$
   since $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} X \\ Z \end{pmatrix}$

2. Take $X|X=Y$ to mean $X|W=1$ where $W = \frac{Y}{X}$.

   Again by "change of variables" $X|W=1$ has density
   $$f_{X|W=1}(x) = \frac{f_{X,W}(x,1)}{f_W(1)} = f_{X,Y}(x, x \cdot 1)\underbrace{\left|\frac{d(X,Y)}{d(X,W)}\right|}_{= \left|\begin{matrix} 1 & 0 \\ w & x \end{matrix}\right|}$$
   since $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X \\ XW \end{pmatrix}$.

Paradox:

1. says $X|X=Y$ has density $\propto f_{X,Y}(x,x)$
2. says $X|X=Y$ has density $\propto x\, f_{X,Y}(x,x)$

why are they different? which one is correct?

Course Content

- Rigorous measure theoretic underpinnings of Probability theory:
  - measure & integration
  - σ-fields
  - conditional probability.
  - ⋮
- Advanced topics using the general theory.
  - Martingales
  - Markov chains
  - Convergence of prob measures
  - Borel cantelli lemmas
  - LIL
  - ⋮

Remark: Although I rarely (never) do research in measure theory, I find it extremely useful to know about it and not be afraid of it.

Course Material

- I have a collection of Notes (still a work in progress) which you can find on my github page "github.com/EthanAnderes/ProbabilityAndMeasureNotes"
- I'll also put p.d.f copies of my lecture notes on github
- Book Ref: "Prob & Measure" by Billingsley.
  "Probability, Theory & examples" by Durrett

Course grading

- weekly HWks. (80% of total grade)
- Exams (20% of total grade)
  I'll announce each one with at least 1 week notice

(3)

Borel's Normal number theorem

In modern terms this thm is basically the SLLN for coin flips.

Thm: If $X_1, X_2, \ldots$ are indep R.V.s s.t.

$$X_i = \begin{cases} 1 & \text{w.p. } 1/2 \\ 0 & \text{w.p. } 1/2 \end{cases}$$

then $P\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \frac{1}{2}\right) = 1.$

Here is how one would prove it in modern times.

Let $S_n = 2n\left[\frac{X_1+\cdots+X_n}{n} - 1/2\right]$

$$= \sum_{i=1}^{n} \underbrace{(2X_i - 1)}_{= R_i = \pm 1}$$

radamacher coin flip

Note $E(R_i) = 0$, $\mathrm{var}(R_i) = E(R_i^2) - 0 = 1$

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = 2P\left(\frac{S_n}{n} \geq \varepsilon\right)$$

$$= 2P(\varepsilon S_n \geq n\varepsilon^2)$$

$$= 2P(e^{\varepsilon S_n} \geq e^{n\varepsilon^2})$$

want to apply Markov on the MGF of S_n since it is compactly supported & ∴ very smooth at $\varepsilon = 0$

$$\leq 2\frac{E(e^{\varepsilon S_n})}{e^{n\varepsilon^2}}$$

$$= 2e^{-n\varepsilon^2}\left[E(e^{\varepsilon R_1})\right]^n$$

where

$$E(e^{\varepsilon R_1}) = \frac{1}{2}(e^{-\varepsilon} + e^{\varepsilon})$$

$$\leq e^{\varepsilon^2/2}, \quad \text{use Taylor series}$$

$$\therefore P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq 2e^{-n\varepsilon^2} e^{n\varepsilon^2/2}$$

$$= 2e^{-n\varepsilon^2/2}$$

(4)

Now let $\varepsilon_n \to 0$ slow enough that $n\varepsilon_n^2 \to \infty$ & $\sum_{n=1}^{\infty} e^{-n\varepsilon_n^2/2} < \infty$

$$\therefore \sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon_n\right) < \infty$$

$$\therefore P\left(\left|\frac{S_n}{n}\right| < \varepsilon_n \text{ for all but finite } n\right) = 1$$

by the first Borell-Cantelli lemma

$$\therefore P\left(\lim_{n\to\infty} \frac{S_n}{n} = 0\right) = 1$$

$$\therefore P\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \frac{1}{2}\right) = 1.$$

QED.

But if your Borel in 1909 you...

- don't know what a prob measure P is
- don't know the meaning of "length(A)" for non-standard sets $A \subset \mathbb{R}$.
- don't even know if there exists an infinite sequence of R.V.s $X_1, X_2, \ldots$ s.t. $X_i = \begin{cases} 0 & \text{w.p. } 1/2 \\ 1 & \text{w.p. } 1/2 \end{cases}$
- don't know Markov's thm or first Borel-Cantelli lemma.

(side Note: Borel's original proof in 1909 had a mistake which was fixed in 1914 by Hausdorff).

Lets prove this from first principles to...

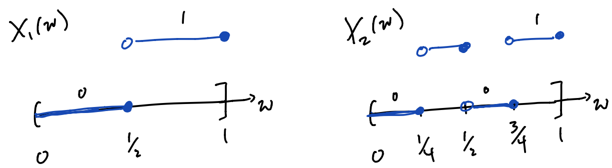
- Motivate the definition of a probability space.
- show the existance of an infinite coin flip sequence of independent random variables (which will be used again later in the course).
- In case you time traveled back to 1900 you could write a proof others would understand.
- Show what one can do with simple finite addivity.

For $k=1,2,\dots$ let $X_k:(0,1]\to\{0,1\}$

be the function:

$$X_k(w) = k^{th}\ \underbrace{\text{binary digit}}_{\substack{\text{non-terminating,}\\ \text{i.e. use } 0.0111111\cdots \\ \text{instead of } 0.1000\cdots}} \text{ of } w$$

e.g.



e.g. $X_2(3/4) = X_2(0.1\underset{\uparrow}{0}11111\cdots) = 0$

Let $N = \{w\in(0,1] : \frac{1}{n}\sum_{k=1}^{n} X_k(w) \to \frac{1}{2}\}$

$=$ Normal numbers in $(0,1]$

$A = N^c$

$=$ Abnormal numbers in $(0,1]$

Borel's question: is length$(N)=1$?

is length$(A)=0$?

what does "length" mean ?

Def: $(\Omega,\mathcal{F},P)$ is a finitely additive probability (FAP) model if

1. Ω is a set
2. $\mathcal{F}$ is a collection of subsets of Ω s.t.
   - 2a. $\Omega\in\mathcal{F}$
   - 2b. $A\in\mathcal{F} \Rightarrow A^c\in\mathcal{F}$
   - 2c. $A,B\in\mathcal{F} \Rightarrow A\cup B\in\mathcal{F}$

   i.e. $\mathcal{F}$ is a field
3. $P:\mathcal{F}\to[0,1]$ s.t.
   - 3a: $P(\Omega)=1$
   - 3b: $P(\underbrace{A\cup B}_{\substack{\text{disjoint}\\ \text{members of } \mathcal{F}.}}) = P(A)+P(B)$

In our case:

$\Omega = (0,1]$

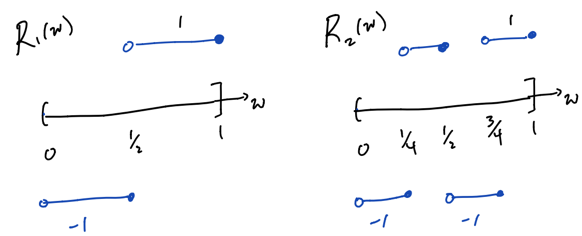
$\mathcal{F} =$ Finite disjoint unions of intervals $(a,b]\subset\Omega$ ↳ denote $\mathcal{B}_0^{(0,1]}$ later

$$P\Big(\underbrace{\bigcup_{i=1}^{n}(a_i,b_i]}_{\text{disjoint}}\Big) = \sum_{i=1}^{n}(b_i-a_i).$$

forms a FAP.

Note: $N, A\notin\mathcal{F}$ so it doesn't make sense to ask if $P(N)=1$ & $P(A)=0$.

As before let $R_k(w) = 2X_k(w)-1$



and set $S_n(w) := \sum_{k=1}^{n} R_k(w)$

Even though $N, A\notin\mathcal{F}$ we do have

$$\{\tfrac{S_n}{n}\geq\varepsilon\} := \Big\{w\in(0,1] : \Big|\frac{S_n(w)}{n}\Big| \geq \varepsilon\Big\} \in\mathcal{F}$$

Thm (WLLN): $\forall\varepsilon>0$

$$P\Big(\frac{S_n}{n}\geq\varepsilon\Big) \longrightarrow 0$$

as $n\to\infty$.

Proof: Note that $R_1, R_2,\dots$ are orthonormal functions in $L^2(0,1]$

i.e. $\langle R_i, R_j\rangle = \int_0^1 R_i(w)R_j(w)\,dw = \delta_{ij}$

$$\therefore\ n = \int_0^1 \Big(\sum_{i=1}^n R_i(w)\Big)^2 dw$$

$$= \int_0^1 S_n^2(w)\,dw$$

$$\geq \int_{|S_n/n| \geq \varepsilon} S_n^2(w)\,dw$$

on this set of w we have $S_n^2 \geq n^2\varepsilon^2$

$$\geq n^2\varepsilon^2 \int_{|S_n/n|\geq\varepsilon} 1\,dw$$

$$= n^2\varepsilon^2 \underbrace{P(|S_n/n| \geq \varepsilon)}_{\text{defined since this is simply length}\left(\left\{\left|\frac{S_n}{n}\right|\geq\varepsilon\right\}\right)}$$

$$\therefore\ P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2} \to 0 \text{ as } n\to\infty.$$

basically chebyshev

QED.

One of your exercises will show one can get better bounds:

$$P\left(\left|\frac{S_n}{n}\right|\geq\varepsilon\right) \leq 2e^{-n\varepsilon^2/2}$$

From first principles.

Now we just need to do a Borel-cantelli argument by hand.

Thm (SLLN, Borel's normal number thm)

For all $\varepsilon > 0$, $\exists\, F_1, F_2, \ldots \in \mathcal{F}$ s.t.

$$A \subset \bigcup_{k=1}^\infty F_k$$

$$\sum_{k=1}^\infty P(F_k) \leq \varepsilon.$$

The combined "length" of the F_k's $\leq \varepsilon$.

i.e. A is negligible.

Proof:

For any $\varepsilon_n \to 0$ we have

$$\left\{\left|\frac{S_n}{n}\right| < \varepsilon_n\ \forall \text{ large } n\right\} \subset \left\{\lim_{n\to\infty}\frac{S_n}{n} = 0\right\} = N$$

$$\therefore\ A = N^c \subset \left\{\left|\frac{S_n}{n}\right| \geq \varepsilon_n \text{ for infinitely many } n\text{'s}\right\}$$

$$\subset \underbrace{\bigcup_{n=m}^\infty \left\{\left|\frac{S_n}{n}\right|\geq\varepsilon_n\right\}}_{\text{This is the cover for some appropriately chosen } m \,\&\, \varepsilon_n}, \ \forall m$$

where

$$\sum_{n=m}^\infty P\left(\left|\frac{S_n}{n}\right|\geq\varepsilon_n\right) \leq \sum_{n=m}^\infty 2e^{-n\varepsilon_n^2/2}$$

can be made to be $\leq$ any given ε by choosing $\varepsilon_n \to 0$ so that $e^{-n\varepsilon_n^2/2}$ is summable, then choosing m large enough.

QED.