Statistical Inference I - Measure Theory and Probability Theory

September 17, 2020

Definition

Sample Set

Ω = Sample Space = Set of all possible outcomes

Example

Sample Set

$\Omega = \mathcal{R}^1$ or $\Omega = \mathcal{R}^2$ or $\Omega = \mathcal{R}^n$

It could be used to representing number of variables

Definition

σ-field

Let $\mathscr{F}$ = a collection of subsets of Ω

$\mathscr{F}$ is a σ-field **iff**

1. $\varnothing \in \mathscr{F}$
2. If $\mathcal{A} \in \mathscr{F} \Rightarrow \mathcal{A}^c \in \mathscr{F}$
3. If $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_1, \cdots, \in \mathscr{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$

Definition

Measure Space

Measure Space $(\Omega, \mathscr{F})$

Elements of $\mathscr{F}$ are called 'measurable' sets

Example

σ-field

Trivial σ-field $\mathscr{F} = \{\Omega, \varnothing\}$

Power set collection of all possinle subsets of Ω. Cardinality of power set $= |P(\mathcal{A})| = 2^n$

Example

σ-field

If $\mathcal{A} \in \Omega \Rightarrow \mathscr{F} = \{\varnothing, \Omega, \mathcal{A}, \mathcal{A}^c\}$ is a σ-field

Example

σ-field

If $\Omega = \{1, 2, 3, \ \ , n\}$ = Natural numbers

$\mathscr{F}$ = Power set $= \{\varnothing, \Omega, \{1\}, \{2\}, \cdots, \{n\}, \{1, 2\}, \cdots, \{n-1, n\}\}$

$|\mathscr{F}| = 2^n$

Example

σ-field

Let $\mathcal{C} = \{\mathcal{A}_1, \mathcal{A}_2\}, \mathcal{A}_1, \mathcal{A}_2 \in \Omega$

The smallest σ-field $= \sigma(\mathcal{C}) = \{\varnothing, \Omega, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_1 \bigcup \mathcal{A}_2, (\mathcal{A}_1 \bigcup \mathcal{A}_2)^c, \mathcal{A}_1 \bigcup \mathcal{A}_2^c \mathcal{A}_1^c \bigcup \mathcal{A}_2, \cdots\}$

Definition

Borel σ-field

Used for most of the statistical applications

Generated by the **open sets**

Example

Borel σ-field

Borel σ-field on $\mathcal{R}^1$ contains all open and closed intervals

Definition

"Measure"

A notion of interval length or volume in a higher dimension

Let $(\Omega, \mathscr{F}$ denote a measurable space, ν = a set function on $\mathscr{F}$ is called a measure **iff**

1. $0 \leq \nu(\mathcal{A}) \leq \infty$
2. $\nu(\varnothing) = 0$
3. If $\mathcal{A}_1, \mathcal{A}_2, \cdots, \in \mathscr{F}$ is disjointed $\Rightarrow \nu(\bigcup_{i=1}^{\infty} \mathcal{A}_i) = \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$

Example

Lebesgue Measure

Length:

A unique measure on $\mathcal{R}^1$ that associates an intervals length to its measure

volume:

Lebesgue measure ~ volume in higher dimension

If $\Omega = [0, 1] \Rightarrow$ Lebesgue measure ~ probability measure

Example

Counting Measure

Let Ω be countable, and $\mathscr{F}$ = power set = collection of all subsets of Ω

$\mathcal{A} \in \mathscr{F} \Rightarrow \nu(\mathcal{A}) = |\mathcal{A}|$ = cardinality of $\mathcal{A}$ = number of elements in $\mathcal{A}$

Proporties

Measure

- Monotoncity:

  If $\mathcal{A} \in \mathcal{B} \Rightarrow \nu(\mathcal{A}) \leq \nu(\mathcal{B})$
- Subadditivity

  $\nu(\bigcup_{n=0}^{\infty} \mathcal{A}_n) \leq \sum_{n=0}^{\infty} \nu((A)_n)$

Proof

Monotoncity:

$\mathcal{B} = \mathcal{A} \bigcup (\mathcal{B} \bigcap \mathcal{A}^c)$ are disjoint union

$\Rightarrow \nu(\mathcal{B} = \nu(\mathcal{A}) + \nu(\mathcal{B} \bigcap \mathcal{A}^c))$

$\quad \geq \nu(\mathcal{A})$

since $\nu(\mathcal{B} \bigcap \mathcal{A}^c) \geq 0$

Subadditivity:

Notes: for 2 events, $\nu(\mathcal{A}_1 \bigcup \mathcal{A}_2) = \nu(\mathcal{A}_1) + \nu(\mathcal{A}_2) - \nu(\mathcal{A}_1 \bigcap \mathcal{A}_2)$

We can find sets disjoint $\mathcal{B}_i$ such that

$\mathcal{B}_1 = \mathcal{A}_1, \mathcal{B}_2 = \mathcal{A}_2 \bigcap \mathcal{A}_1^c, \cdots, \mathcal{B}_n = \mathcal{A}_n \bigcap \mathcal{A}_1^c \bigcap \cdots \bigcap \mathcal{A}_{n-1}$

thus we have

$\nu(\bigcup_{i=1}^{n} \mathcal{A}_i) = \nu(\bigcup_{i=1}^{n} \mathcal{B}_i) = \sum_{i=1}^{n} \nu(\mathcal{B}_i)$.

Since $\mathcal{B}_i \in \mathcal{A}_i \Rightarrow \nu(\bigcup_{i=1}^{n} \mathcal{A}_i) = \sum_{i=1}^{\infty} \nu(\mathcal{B}_\rangle) \leq \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$

If collections is infinite

$\nu(\bigcup_{i=1}^{\infty} \mathcal{A}_i) = \nu(\bigcup_{i=1}^{\infty} \mathcal{B}_i) = \sum_{i=1}^{n} \nu(\mathcal{B}_i)$

$= \lim_{n\to\infty} \sum_{i=1}^{n} \nu(\mathcal{B}_i) = \lim_{n\to\infty} \nu(\bigcup_{i=1}^{n} \mathcal{A}_i) \leq \lim_{n\to\infty} \sum_{i=1}^{n} \nu(\mathcal{A}_i) = \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$

Similarly, since $\mathcal{B}_i \in \mathcal{A}_i \Rightarrow \nu(\bigcup_{i=1}^{n} \mathcal{A}_i) = \sum_{i=1}^{\infty} \nu(\mathcal{B}_\rangle) \leq \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$

Proporties

Probability Measure

- $\Pr(\Omega) = 1, \Pr(\varnothing) = 0$
- $\Pr(\Omega) = 1 = \Pr(\mathcal{A}) + \Pr(\mathcal{A}^c) \Rightarrow \Pr(\mathcal{A}^c) = 1 - \Pr(\mathcal{A})$

Definition

Cumulative Distribution Function (CDF)

P probability measure on $\Omega = \mathbb{R}^1$

CDF = cumulative distribution function denoted $F(x) = \Pr((-\infty, x])$

Facts:

- $F(x = \infty) = 0, F(\infty) = 1$
- F is right continuous, i.e. $\lim_{y \to x} F(y) = F(x)$

Any functions that satisfies these properties are a CDF of some probability measure

Definition

Cumulative Distribution Function (CDF)

We can extend P probability measure on $\Omega = \mathbb{R}^p$

$X = (x_1, x_2, \cdots, x_p)'$

$CDF = F(X) = F(x_1, x_2, \cdots, x_p) = P((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_p]) = \text{joint } CDF$

Definition

Inverse of Function

Let Λ denote some space (other than Ω possibly) (ex $\Lambda = \mathbb{R}^1$)

Let $\mathcal{G} \in \Lambda$

Consider function $f : \Omega \to \Lambda$

Define $f^{-1}(\mathcal{G}) = \{\omega \in \Omega : f(\omega) \in \mathcal{G}\}$

It follows that

- $f^{-1}(\mathcal{G}^c) = (f^{-1}(\mathcal{G}))^c$ for any $\mathcal{G} \subseteq \Lambda$
- $f^{-1}(\bigcup_{i=1}^n \mathcal{G}_i) = \bigcup_{i=1}^n f^{-1}(\mathcal{G}_i)$ for $\mathcal{G}_1, \mathcal{G}_2, \cdots, \in \Lambda$

Note: if $\mathcal{C}$ = collection of sets of $\Lambda \Rightarrow$ define $f^{-1}(\mathcal{C}) = \{f^{-1}(c), c \in \mathcal{C}\}$

Definition

Measurable Functions

Let $(\Omega, \mathscr{F})$ and $(\Lambda, \mathscr{G})$ be measurable space and $f : \Omega \to \Lambda$ is measurable from $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G})$ **iff** $f^{-1}(\mathscr{G}) \subseteq \mathscr{F}$

Often in $\Lambda = \mathbb{R}^p \Rightarrow \mathscr{G}$ = borel σ-field

Definition

Random Variable

In probability, a measurable mapping $X : \Omega \to \mathbb{R}^p$ is called a Random Variable (RV) if $p = 1$ or Random Vector if $p > 1$

We can write $X(\omega), \omega \in \Omega$.

Random Variable are measurable functions

Note: if $f : \Omega \to \Lambda$ is measurable, $(\Omega, \mathscr{F})$ to $(\Lambda, \mathscr{G}) \Rightarrow f^{-1}(\mathscr{G})$ is a sub σ-field of $\mathscr{F}$

Notation

Random Variable

$\sigma(X) = \sigma$-field generated by X

Let $\mathcal{A} \subseteq \Omega$ and define

$I_A(\omega) =$ indicator function such that

$$I_A(\omega) = \begin{cases} 0, \text{ if } \omega \notin \mathcal{A} \\ 1, \text{ if } \omega \in \mathcal{A} \end{cases}$$

Notation

Inverse Function and sub σ-field

$(\Omega, \mathscr{F}), \mathcal{A} \in \mathscr{F}$

$$I_A(\omega) = \begin{cases} 0, \text{ if } \omega \notin \mathcal{A} \\ 1, \text{ if } \omega \in \mathcal{A} \end{cases}$$

$\mathcal{B} \subseteq \mathbb{R}^1$

$$I_A^{-1}(\mathcal{B}) = \begin{cases} \varnothing, \text{ if } 0, 1 \notin \mathcal{B} \\ \mathcal{A}, \text{ if } 1 \in \mathcal{B}, 0 \notin \mathcal{B} \\ \mathcal{A}^c, \text{ if } 1 \notin \mathcal{B}, 0 \in \mathcal{B} \\ \Omega, \text{ if } 0, 1 \in \mathcal{B} \end{cases}$$

The sub σ-field generated by function I_A is $\sigma(I_A) = \{\varnothing, \Omega, \mathcal{A}, \mathcal{A}^c\}$

Definition

Simple Function

Let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ be measurable sets in $\mathscr{F}$, and $a_1, a_2, \cdots, a_n, a_i \in \mathbb{R}^1$

$\varphi(\omega) = \sum_{i=1}^n a_i I_{A_i}(\omega)$ is a simple function (building blocks to generate functions)

To prove results, often interest to show result for simple functions and take limits

If $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ form a partition of $\omega \Rightarrow$ sub σ-field generated by φ is $\sigma(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n)$

Proposition

Borel Measurable Function

1. f is borel measurable function **iff** $f^{-1}((a,\infty)) \in \mathscr{F}$ for all $a \in \mathbb{R}^1$
2. f, g are borel measurable functions $\Rightarrow f+g, f\cdot g$ are also borel measurable functions
3. $\{f_i\}$ are borel measurable functions $\Rightarrow \sup f_n, \inf f_n, \liminf f_n, \limsup f_n$ are borel measurable functions
4. continuous functions are borel measurable functions

Theorem

Let $f \geq 0$ be borel measurable function $f : \Omega \to \mathbb{R}^1$

Then $\exists$ a sequence of simple functions $\varphi_1 \leq \varphi_2 \leq \varphi_3 \leq \cdots \leq \varphi_j \leq f$ and $\lim \varphi_n = f$

Proof

for any $n = 1, 2, \cdots, 0 \leq k \leq 2^{2n-1}$

Let $\mathcal{A}_n = f^{-1}(k2^{-n}, (k+1)2^{-n}), \mathcal{B}_n = f^{-1}(2^n, \infty)$

$\Rightarrow \varphi_n(\omega) = \sum_{k=0}^{2^{(2n-1)}} k2^{-n} I_{A_n}(\omega) + 2^n I_{B_n}(\omega)$ satisfies theorem.

Note: $0 < f - \varphi_n \leq 2^{-n}$

Definition

$X \sim R.V.$ on $(\Omega, \mathscr{F}, P)$

Then $\sigma(X) = \sigma$-field generated by X is the smallest σ-field with respect to which X is measurable

Typically $\sigma(X)$ smallest than $\mathscr{F}$

Extension 1

$X_1, X_2 \sim R.V.$ on $(\Omega, \mathscr{F}, P) \Rightarrow \sigma(X_1, X_2) = \sigma$-field generated by X_1, X_2 =smallest σ-field with respect to which each X_1, X_2 measurable

Extension 2

$\sigma(X_1, X_2, \cdots X_n)$ or $\sigma(X_1, X_2, \cdots)$

$\mathcal{A} \in sigma(X_1, X_2, \cdots X_n)$ if $A = \{\omega \in \Omega : (X_1(\omega), \cdots, X_n(\omega)) \in \mathcal{B}^n\}$ where $\mathcal{B}^n$ is borel σ-field on $\mathbb{R}^n$

Theorem

Theorem 20.1 in Billingslely Book

$Y = R.V.$ is measurable with respect to $\sigma(X_1, X_2, \cdots X_n)$ **iff**

$Y = f(X_1, X_2, \cdots X_n), f \sim$measurbale $f : \mathbb{R}^n \to \mathbb{R}^1$

Very useful result for **conditional expectation**

Definition

Integration

If $\varphi(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega) \sim$ simple function $a_i \geq 0$

$\int_\Omega \varphi(\omega) d\nu(\omega) = \int \varphi = \sum_{i=1}^{k} a_i \nu(\mathcal{A}_i)$

Note: convention $a_i \nu(\mathcal{A}_i) = 0$ if $a_i = 0, \mu(\mathcal{A}_i) = \infty$

Let $f \geq 0$ and $0 \leq \varphi_1 \leq \varphi_2 \leq \varphi_3 \leq \cdots$ so that $\lim_{n\to\infty} \varphi_n(X) = f(X)$

Then $\int f d\nu = \lim_{n\to\infty} \int \varphi_n d\nu$

Alternatively $\int f d\nu = \sup\{\int \varphi d\nu, 0 \leq \varphi \leq f\}$ where φ is simple function

Definition

Integrable

A non-negative measurable function f, is called 'integrable' with respect to ν

$\int f d\nu < \infty$

For arbitrary measurable f write $f(x) = f^+(x) - f^-(x)$ where

$f^+(x) = \max(0, f(x))$

$f^-(x) = -\min(0, f(x))$

Define $\int f d\nu = \int f^+ d\nu - \int f^- d\nu$

f is integrable if f^+ and f^- are both integrable

Note: $\nu =$ counting measure

$\int_\Omega f d\nu = \sum_\Omega f(x_i)$

Note: In most continuous setting, our integral here coincides with Reimann integral

Example

Not Integrable

Expectation of Cauchy Distribution (special case of t-distribution with degree of freedom 1) does not exist

$PDF = f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, -\infty < x < \infty$

To compute expectation $\to$ integrate $f9x)$ from $-\infty$ to ∞

$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{\infty} xf(x)dx = -\infty + \infty$

Example

Integration

$\varphi \sim$ simple function, $\varphi \geq 0$

$(\Omega, \mathscr{F}, \nu), \mathcal{A} \in \mathscr{F}$

Define $\nu^*(\mathcal{A}) = \int_{\mathcal{A}} \varphi d\nu \Rightarrow \nu^*$ is also a measure

Notation

Intergrable

$f \in L^p(\mu) \Rightarrow \int |f|^p d\nu < \infty$

Norm of f in $L^p = (\int |f|^p d\nu)^{\frac{1}{p}} = ||f||_p$

Often use $p = 2$ which is ofer refered as Hilbert spaces

Example

Intergrable $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$ where $\mathbb{Q}$ =rationals

$\int f d\mu = 0$ but f not Reimann-integrable

$\mu =$ Lebesgue measure

$\mu(\mathbb{Q}) = 0$ where $\mathbb{Q}$ is countable

Example

Cantor Set

Let $\mathcal{C} =$ Cantor set, $\mu(\mathcal{C}) = 0$ where $\mathcal{C}$ is uncountable

$x \in \mathcal{C}$

$\sum_{i=1}^{\infty} a_i (\frac{1}{3})^i, a_i = 0, 1, 2, \cdots$

Let $a_i \in \{0, 2\}$. we can map $\mathcal{C}$ 1-1 to $[0, 1]$

Change base from $\frac{1}{3}$ to $\frac{1}{2}$, and map a_i to $[0, 1]$

Facts

Integration is Linear Functions

Functions $f_1, f_2, \cdots, f_n$

constants $a_1, a_2, \cdots, a_n$

$\int \sum_{i=1}^{n} a_i f_i \ \ \sum_{i=1}^{n} a_i \int f_i$

Notation

Almost Everywhere (a.e.)

Condition holds except possibly on set of measure zero

Example

Almost Everywhere

$\{x_i\}$ are $R.V.$s

$\lim_{n \to \infty} X_n(\omega) = X(\omega), \omega \in \Omega$

Theorem

Continuity from Below

$(\Omega, \mathscr{F}, \nu), \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}_3 \subseteq \cdots$ all in $\mathscr{F}$

Then $\nu(\bigcup_{j=1}^{\infty} \mathcal{A}_j) = \lim_{n\to\infty} \nu(\mathcal{A}_n)$

Proof

Set $\mathcal{A}_0 = \varnothing \Rightarrow \nu(\bigcup_{j=1}^{\infty} \mathcal{A}_j) = \sum_{j=1}^{\infty} \nu(\mathcal{A}_j \backslash \mathcal{A}_{j-1}) = \lim_{n\to\infty} \sum_{j=1}^{n} \nu(\mathcal{A}_j \backslash \mathcal{A}_{j-1})$

Theorem

Let φ and ψ simple functions, $\varphi, \psi \geq 0$

If $\varphi \leq \psi \Rightarrow \int \varphi \leq \int]\psi$

Proof

Let $\varphi = \sum_{j=1}^{n} a_j I_{\mathcal{A}_j}, \psi = \sum_{k=1}^{m} b_k I_{\mathcal{B}_k}$ where $\mathcal{A}_j, \mathcal{B}_k$ are disjoint sets where

$$\begin{aligned}\int \varphi = \sum_{j=1}^{n} a_j \nu(\mathcal{A}_j) &= \sum_{j=1}^{n} a_j \sum_{k=1}^{m} \nu(\mathcal{A}_J \bigcap \mathcal{B}_j) \\ &\leq \sum_{j=1}^{n} b_j \sum_{k=1}^{m} \nu(\mathcal{A}_J \bigcap \mathcal{B}_j) = \sum_{k=1}^{m} b_j \nu(\mathcal{B}_k) = \int \psi\end{aligned}$$

Theorem

Monotone Convergence Theorem

$\Omega, \mathscr{F}, \nu$

$\{f_n\}$ are sequence of functions, $f_n \geq 0$ s.t. $f_1 \leq f_2 \leq f_3 \leq \cdots$ and $\lim_{n\to\infty} f_n(\omega) = f(\omega)$

Then $\lim_{n\to\infty} \int f_n = \int \lim f_n = \int f$

Proof

Since $f_n \leq f \Rightarrow \int f_n \leq \int f \Rightarrow \lim_{n\to\infty} \int f_n \leq \int f$

Fix $\varepsilon \in (0, 1)$

Let φ simple functions $0 \leq \varphi \leq f$

Let $\mathcal{A}_n = \{\omega \in \Omega : f_n(\omega) \geq \varepsilon\varphi(\omega)\} \Rightarrow \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}_3 \subseteq \cdots$ and $\bigcup_{n=1}^{\infty} \mathcal{A}_n = \Omega$ since f_n approach f from below

Now $\int_{\Omega} f_n \geq \int_{\mathcal{A}_n} f_n \geq \varepsilon \int_{\mathcal{A}_n} \varphi \Rightarrow \lim_{n\to\infty} \int_{\Omega} f_n \geq \varepsilon \int_{\Omega} \varphi$

And we need to make sure that the above inequality stand for all $\varepsilon \in (0, 1)$

Recall definition that $\int f = \lim_{j\to\infty} \int \varphi_j, \varphi_1 \leq \varphi_2 \leq \cdots \leq f$

OR $\int f = \sup\{\int \varphi : 0 \leq \varphi \leq f\} \Rightarrow \lim_{n\to\infty} f_n \geq \varepsilon \sup\{\int \varphi : 0 \leq \varphi \leq f\} \Rightarrow \lim_{n\to\infty} \int_{\Omega} f_n \geq \varepsilon \int_{\Omega} \varphi$ for all $\varepsilon \in (0, 1)$

Theorem

Fatou's Lemma

integrable $f_n \geq 0 \Rightarrow \int \liminf f_n \leq \liminf \int f_n$

Proof

For each k, $\inf_{n\geq k} f_n \leq f_j$ for $j \geq k$

$\Rightarrow \int \inf_{n\geq k} f_n \leq \int f_j, j \geq k$

$\Rightarrow \int \inf_{n\geq k} f_n \leq \inf_{j\geq k} \int f_j$

Now, lets $k \to \infty$

by monotone convergence theorem

$\lim_{k\to\infty} \int \inf_{n\geq k} f_n = \int \liminf f_n \leq \lim_{k\to\infty} \inf_{n\geq k} \int f_n$

Fatou's Lemma equivalent to monotone convergence theorem

Theorem

Dominated Convergence Theorem

$f_n, n = 1, 2, \cdots$, integrable, s.t. $f_n \to f$ a.e.

$\exists \quad g \sim$ integrable s.t. $|f_n| \leq g \ \forall \quad n$

Then: f integrable and $\lim_{n\to\infty} \int f_n = \int f$

Proof

$f_n \to f, f_n \sim$ measurable $\Rightarrow f \sim$ measurable

$(A) \Rightarrow f + f_n \geq 0 \quad a.e.$

$$\int (g+f) = \int g + \int f = \overbrace{\int (g + \liminf f_n) \leq \int g + \liminf \int f_n}^{\text{Fatou's Lemma}}$$

$\Rightarrow \int f \leq \liminf \int f_n$

$(B) \Rightarrow g - f_n \geq 0 \quad a.e.$

$$\begin{aligned}\int (g-f) = \int g - \int f = \int (g - \liminf f_n) = &\overbrace{\int \liminf (g - f_n) \leq \liminf \int (g - f_n)}^{\text{Fatou's Lemma}} \\ &= \int g - \limsup \int f_n\end{aligned}$$

$\Rightarrow \int f \leq \liminf \int f_n$

However, we have $\limsup \geq \liminf$

Thus, we have $\liminf \int f_n = \limsup \int f_n \Rightarrow \lim_{n\to\infty} \int_\Omega f(\omega) d\nu(\omega)$ QED

Example

Maximum Likelihood Estimation

$(\Omega, \mathscr{F}, P)$ probability space

$f(\omega, \theta) \sim$ Borel function. Typical setup $\theta \in \Theta \quad e.g. \Theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$

Suppose $\frac{\partial f}{\partial \theta} \sim$ exists and $\exists \quad g$ s.t. $|\frac{\partial f}{\partial \theta}| \leq g$

To solve the MLE, we need to evaluate

$\frac{\partial}{\partial \theta} \int f(\omega, \theta) dP(\omega) = \int \frac{\partial}{\partial \theta} f(\omega, \theta) dP(\omega)$ by Dominated Convergence Theorem

Definition

Absolutely Continuous

$(\Omega, \mathscr{F}, \mu)$ measure space

Let $\mathcal{A} \in \mathscr{F}$ define $\nu(\mathcal{A}) = \int_{\mathcal{A}} f d\mu$ for some measurable $f \geq 0$

$\Rightarrow \nu$ is a measure on $(\Omega, \mathscr{F})$, we write: $d\nu = f d\mu$

In general, we say ν is absolutely continuous w.r.t. μ if whenever $\mu(A) = 0 \Rightarrow \nu(A) = 0$

We write $\nu << \mu$

Theorem

Radon–Nikodym Theorem

ν, μ two measures on $(\Omega, \mathscr{F})$ with $\nu << \mu$

$\Rightarrow \exists$ measurable $f \geq 0$ s.t.

$\nu(\mathcal{A}) = \int f(\omega) d\mu(\omega)$ and f is unique a.e.

$f =$ Radon–Nikodym derivative

Typically write $f = \frac{d\nu}{d\mu}$

Proof

Proof uses monotone convergence theorem, see Billingsley book

Example

Radon–Nikodym Theorem

Suppose $X \sim$ random variable (R.V.)

$CDF \quad F(x) = P(\omega \in \Omega : X(\omega) \leq x)$ or $P(X < x)$

so $P(X \in \mathcal{A}) = \int_{\mathcal{A}} dF(x) \qquad e.g. \mathcal{A} = (a, b)$

If X has a R-N derivative with respect to Lebesgue measure, say $f(x)$

$f =$ Probability Density Function (pdf) $=$ R-N derivative

So, $P(X \in \mathcal{A}) = \int_{\mathcal{A}} dF(x) = \int_{\mathcal{A}} f(x) dx$

Theorem

Fubini's Theorem Inter-change order of integration

$f \sim$ Borel measure on $\Omega_1 \times \Omega_2$

$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d\nu_1 \times d\nu_2$

from Fubini's, for a.e. $\omega_1 \in \Omega_1, f_2(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) d\nu_1 \sim$ measruable function on Ω_2 and integrable

It's also called marginal density function