Feynman Liang
STAT 210A
Lecture #1
8/25/2016

1 Measure theory basics

Given a set X, a *measure* maps subsets $A \subset X$ to $\mu(A)[0, +\infty]$.

Example 1.1. X contable. The *counting measure* $\#(A) = \#$ elements in A

Example 1.2. $X = \mathbb{R}^n$. The *Lebesgue measure* $\lambda(A) = \int_A dx_1 \cdots dx_n$ ("volume of A")

Definition 1.3. A *σ-field* $\mathcal{F}$ is a collection of subsets for which μ is defined.

Example 1.4. X countable, $\mathcal{F} = \mathcal{P}(X)$.

Example 1.5. $X = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n) = \sigma\{\text{open sets of } \mathbb{R}^n\}$ is the *Borel σ-field*

Definition 1.6. $(X\mathcal{F})$ is a *measurable space*. A *measure* $\mu : \mathcal{F} \to [0, +\infty]$ satisfies

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1.1}$$

for disjoint $A_i \in \mathcal{F}$.

Definition 1.7. A *probability measure* $\mathbb{P}$ is a measure where $\mathbb{P}(X) = 1$.

We notate an *integral* wrt measure μ as $\int_X f(X)\mu(dx)$.

Example 1.8. Counting: $\int_X f(x)\mu(dx) = \sum_{x \in X} f(x)$
Lebesgue: $\int f(x)\lambda(dx) = \int_{\mathbb{R}^n} f(x) dx_1 \cdots dx_n$

2 Densities

Given $(X, \mathcal{F})$ measurable space and two measures P, μ

Definition 2.1. P is *dominated by* μ (denoted $P \ll \mu$) if $\mu(A) = 0 \implies P(A) = 0$. If $\mu = \lambda$, we say P is *absolutely continuous*.

If $P \ll \mu$, then by the Radon-Nikodym theorem the *density* $p(x) = \frac{dP}{d\mu}(x)$ exists, is essentially unique, and satisfies

$$P(A) = \int_A p(x) d\mu(x) \tag{2.1}$$

$$\int_X f(x) dP(x) = \int_X f(x) p(x) d\mu(x) \tag{2.2}$$

$\frac{dP}{d\mu}$ is essentially unique: If $p_0(x)$ and $p_1(x)$ are both densities for $\frac{dP}{d\mu}$, then $p_0 \overset{\text{a.s.}}{=} p_1$ i.e. $P(\{p_0(x) \neq p_1(x)\}) = 0$.

Example 2.2. $\mathcal{N}(0,1)$, $p(x) = (2\pi)^{-1/2} e^{-x^2/2}$. $P((a,b)) = \int_a^b p(x) dx$.

3 Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

- $\omega \in \Omega$ is an *outcome*
- $A \in \mathcal{F}$ is an *event*
- $\mathbb{P}(A)$ is the probability of event A

Definition 3.1. A *random variable* (vector, matrix) is a measurable function $X : \Omega \to \mathbb{R}$ ($\mathbb{R}^n$). That is, $X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{B}$.

We say $X \sim Q$ if $\mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) = Q(B)$.

Definition 3.2. An *expectation* is an integral wrt $\mathbb{P}$:

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} x dQ(x) \tag{3.1}$$

Corollary 3.3. *Q is a probability measure, since $B = \mathbb{R} \implies Q(\mathbb{R}) = P(X \in \mathbb{R}) = 1$.*

4 Estimation

Statisticl model: a family of candidate probability distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ for observed data $X \sim P_\theta$.

Goal: observe X, estimate $g(\theta)$ (*estimand*)

Example 4.1. Flip a coin n times. θ = probability of heads. X = # heads in n trials $\sim \text{Binom}(n, \theta) = \theta^x (1-\theta)^{n-x} \binom{n}{x}$.

Definition 4.2. A *statistic* $\Gamma(X)$ is a function of the data X. Statistics are generally used as intermediate quantities to summarize data within a statistical procedure.

An *estimator* $\Gamma(X)$ is also a function $\delta(X)$, usually chosen to be "close" to a function of the parameters $g(\theta)$.

Example 4.3. For the binomial example, $\delta_0(X) = X/n$ is an estimator of θ.

5 Loss and Risk

Loss function: $L(\theta, \delta)$ measures how good an estimator δ is.

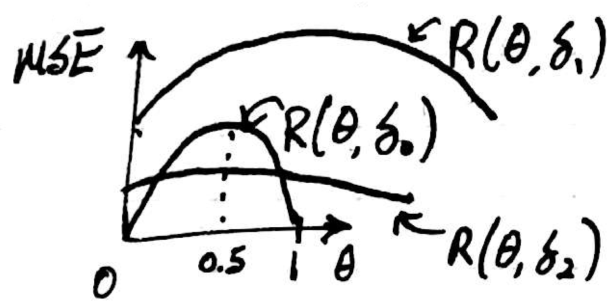
Example 5.1. Squared error loss $L(\theta, \delta) = (\delta - g(\theta))^2$

Definition 5.2. The *risk*

$$R(\theta, \delta(\cdot)) = \mathbb{E}_\theta[L(\theta, \delta(X))] = \mathbb{E}_{X \sim P_\theta}[L(\theta, \delta(X))] \tag{5.1}$$

Notation: $\mathbb{E}_\theta$ denotes an expectation where the subscript θ is fixed and all other random quantities are integrated over.

Example 5.3. $R(\theta, \delta_0) = \frac{1}{n}\theta(1-\theta)$, $L =$ squared error, $R =$ MSE.
Consider $\delta_1 = \frac{X+3}{n}$ and $\delta_2 = \frac{X+3}{n+6}$.



δ_1 is *inadmissibile*: $\forall\theta : R(\theta, \delta_0) < R(\theta, \delta_1)$.
Comparing δ_0 to δ_2 shows that different estimators may perform better for different values of θ.