Probability & Measures

A measure is an extension of the length, area, or volume of subsets in 1-, 2-, or 3-dimensional Euclidean space. Given a sample space Ω, a measure is a set function defined on the subsets of Ω. Naturally, we define a measure μ on a σ-field as follows:

Definition: Let $\mathcal{F}$ be a collection of subsets of a sample space Ω. $\mathcal{F}$ is called a σ-field (σ-algebra) if and only if (iff) it has the following properties:

(i) The empty set $\phi \in \mathcal{F}$

(ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ [$\mathcal{F}$ is closed under complementation]

(iii) If $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, then $\cup A_i \in \mathcal{F}$ [$\mathcal{F}$ is closed under countable ~~union~~ unions]

Note: If $A_1, A_2, \ldots, \in \mathcal{F}$, then $A_1^c, A_2^c, \ldots, \in \mathcal{F}$ and therefore

$\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$. Using DeMorgan's law, we have

$$\left(\bigcup A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i \text{. That is, } \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Note: The collection of 2 sets $\{\phi, \Omega\}$ is a σ-algebra, called trivial σ-algebra.

Note: The important σ-algebra is the smallest one that contains all of the open sets in Ω.

Example 1: Consider Ω is finite or countable. Define
$$\mathcal{F} = \{\text{all subsets of } \Omega \text{ including } \Omega\}$$

If Ω has n elements, then there are 2^n sets in $\mathcal{F}$.
For example, if $\Omega = \{1,2,3\}$, then $\mathcal{F}$ is a collection of $2^3 = 8$ sets:

$$\begin{array}{ccc} \{1\} & \{1,2\} & \{1,2,3\} = \Omega \\ \{2\} & \{1,3\} & \phi \\ \{3\} & \{2,3\} & \end{array}$$

If Ω is uncountable, it is not an easy task to find $\mathcal{F}$.

Example 2: $\Omega = (-\infty, \infty)$ - Real line. Then $\mathcal{F}$ can be chosen to include all sets of the form:

$$[a,b], \ (a,b], \ (a,b), \ [a,b)$$

~~Then~~ Also from properties of $\mathcal{F}$, $\mathcal{F}$ contains all sets taking unions and intersections of sets of the above forms.

Example 3: Let A be a non-empty subset of Ω $(A \subset \Omega)$.
Then $\mathcal{F} = \{\phi, \Omega, A, A^c\}$ is a σ-algebra containing A (smallest σ-algebra containing A). This σ-algebra is denoted by $\mathcal{F} = \sigma(\{A\}) = \sigma$-algebra induced by A.

Def.: Let $(\Omega, \mathcal{F})$ be a measurable space. A set function μ defined on $\mathcal{F}$ is called a measure iff it has the following

(i) $0 \leq \mu(A) < \infty$ for any $A \in \mathcal{F}$

(ii) $\mu(\phi) = 0$

(iii) If $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, and A_i's are disjoint, then
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space. If $\mu(\Omega) = 1$, then μ is called a probability space. It is denoted by P instead of μ. The triple $(\Omega, \mathcal{F}, P)$ is called a Probability space.

Example (Lebesgue measure): Consider the measure μ on $(\mathbb{R}, \mathcal{F})$ such that
$$\mu([a,b]) = b - a$$ for every finite $[a,b]$. It is called Lebesgue measure. In this case, the integral of f is
$$\int f \, d\mu = \int \cdots \int f(x_1, \ldots, x_n) \, dx_1 \ldots dx_n$$
$\mathcal{F} = \mathcal{B}$ = Borel field.

Example (Counting measure): We may define the measure μ on $(\Omega, \mathcal{F})$ as
$\mu(A)$ = # of elements in the subset A. Here,
$$\int f \, d\mu = \sum_{x \in \Omega} f(x).$$

Example (Probability measure): Let P be a Prob. measure on $(\mathbb{R}, \mathcal{F})$ defined by $F(x) = P((-\infty, x]) = P(X \leq x)$, $x \in \mathbb{R}$

M. Ragab

Measurable functions

* $f: \Omega \longrightarrow \mathbb{R}$, let $B \subset \mathbb{R}$. If
  $f^{-1}(B) = \{x \in \Omega : f(x) \in B\} \in \mathcal{F}$ for every Borel set B.

  Then f is called a measurable function.

In probability theory, a measurable function is called a random element and denoted by $X, Y, \ldots$

* $X \sim P_X$ = absolutely continuous with density p. Then
  $$F_X(x) = P(X \le x) = P\left((-\infty, x]\right) = \int_{-\infty}^{x} p(u)\,du$$
  $$\Longrightarrow \quad p(x) = F_X'(x)$$

* If X is a random variable with prob. measure P, then
  $$EX = \int x \, dP = \int x\, p(x)\,dx$$

Inferential Theory

Let Θ = parameter space $[\mathbb{R}^q$ - q-dimensional vectors of real$]$

- $g(\theta)$ = quantity to be estimated, $g(\theta) \in \mathbb{R}$
- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ - collection of Prob. dist.'s
- X = observed data taking values in Ω with $P_\theta \in \mathcal{P}$
- $\delta : \Omega \longrightarrow \mathbb{R}$ is an estimator
- $L : \Theta \times \delta \longrightarrow [0, \infty)$ loss function. It is denoted by $L(\theta, \delta)$ or $L(\theta, d)$ with $L(\theta, g(\theta)) = 0$

Examples: squared error loss (SEL): $L(\theta, d) = (g(\theta) - d)^2$

Absolute = = (AEL): $L(\theta, d) = |g(\theta) - d|$

Relative = = (REL): $L(\theta, d) = \left|\frac{d}{g(\theta)} - 1\right|^p, \; p > 1$

For $K(\theta) > 0$, $L(\theta, d) = K(\theta)\,|g(\theta) - d|^p$ includes all others.

- Let $R(\theta, d) = E_\theta \mathcal{L}(\theta, \delta(X))$ — Risk function. We say that estimator δ_1 dominates δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all $\theta \in \Theta$ with strict inequality for some $\theta \in \Theta$.

- Estimators are chosen based on the following criteria:
  - Best within unbiased or invariant class of estimators
  - Best on the average (Bayes estimator) i.e. minimizes $\int R(\theta, d) w(\theta) d\theta$

    weighting average or prior dist.
  - Best under worst circumstances i.e. minimizes $\max_{\theta \in \Theta} R(\theta, \delta)$

Convexity

Def.: A subset $A \in \mathbb{R}^n$ is convex if $\underset{\sim}{x}, \underset{\sim}{y} \in A \Rightarrow \gamma \underset{\sim}{x} + (1-\gamma) \underset{\sim}{y} \in A$ for $0 < \gamma < 1$ [A, B convex $\Rightarrow A \cap B$ convex].

Def.: Let A be convex and $f: A \longrightarrow \mathbb{R}$. f is convex if

$$f(\gamma \underset{\sim}{x} + (1-\gamma) \underset{\sim}{y}) \leq \gamma f(\underset{\sim}{x}) + (1-\gamma) f(\underset{\sim}{y})$$

$\forall \ \underset{\sim}{x}, \underset{\sim}{y} \in A \ \& \ 0 < \gamma < 1$.

Theorem (Holder Inequality): For positive r.v.'s X and Y with finite means, $E(X^\alpha Y^{1-\alpha}) \leq (EX)^\alpha (EY)^{1-\alpha}$.

Theorem (Jensen's Inequality): Let $f: A \longrightarrow \mathbb{R}$ be convex, then

$$f(EX) \leq E(f(X))$$

Example: $f(x) = x^2$, convex $\Rightarrow EX^2 \geq (EX)^2 = \mu^2 \Rightarrow \sigma^2 \geq 0$.

Exponential family:

Def.: A parametric family $\{P_\theta : \theta \in \Theta\}$ is called an exponential family iff $P(x|\eta) = P_\eta(x) = e^{\eta' T(x) - A(\eta)} \cdot h(x)$ where $T \in \mathbb{R}^s$, with

$$\eta' T = \sum_{i=1}^{s} \eta_i T_i \text{ and } A(\eta) = \log \int e^{\eta' T} h(x)\, d\mu(x) < \infty$$

$$\Xi = \left\{ \eta \,\middle|\, \int e^{\eta' T} h(x)\, d\mu < \infty \right\} = \text{natural parameter space}$$

$$\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_s \end{pmatrix}$$

or $\Xi = \{ \eta \mid A(\eta) < \infty \}$.

Example: Let P_θ be the binomial dist. $B(n, \theta)$

$$P(x|\theta) = P_\theta(x) = e^{x \log \frac{\theta}{1-\theta} + n \log(1-\theta)} \binom{n}{x} I_{\{0,1,\dots,n\}}(x)$$

$T(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$, $A(\theta) = -n \log(1-\theta)$ and $h(x) = \binom{n}{x} I_{\{0,1,\dots,n\}}(x)$,

$$\Rightarrow P(x|\eta) = e^{\eta x - n \log(1 + e^{\eta})} \binom{n}{x} I_{\{0,1,\dots,n\}}(x) \quad \text{– Natural exp. family}$$

$$\Rightarrow \Xi = \mathbb{R}$$

Example: Suppose μ is Lebesgue measure on $\mathbb{R}$; $s=1$, $T_1(x) = x$, $h = 1 I_{(0,\infty)}(x)$

Then
$$A(\eta) = \log \int_0^\infty e^{\eta x} dx = \begin{cases} \log\left(-\frac{1}{\eta}\right) & \text{if } \eta < 0 \\ \infty & \text{if } \eta \geq 0 \end{cases}$$

$$\Rightarrow P(x|\eta) = e^{\eta x - \log\left(-\frac{1}{\eta}\right)} I_{(0,\infty)}(x)$$

where $\eta \in \Xi = (-\infty, 0)$.

M. Rafab

Example: Normal dist. $N(\mu, \sigma^2)$. Consider $\underset{\sim}{\theta} = (\mu, \sigma^2)$

$$P(x|\underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \left(\frac{\mu^2}{2\sigma^2} + \log\sigma\right)}$$

2-Parameter exp. form.

$T_1 = x$, $T_2 = -x^2$, $\eta_1(\underset{\sim}{\theta}) = \frac{\mu}{\sigma^2}$, $\eta_2(\underset{\sim}{\theta}) = +\frac{1}{2\sigma^2}$, $h(x) = \frac{1}{\sqrt{2\pi}}$

$$\Rightarrow P(x|\underset{\sim}{\eta}) = e^{\eta_1 x + \eta_2 x^2 - A(\eta_1, \eta_2)}, \quad A(\eta_1, \eta_2) = \frac{\eta_1^2}{4\eta_2} + \log(1/\sqrt{2\eta_2})$$

This leads to the fact that $P(x|\underset{\sim}{\eta})$ is a natural exp. family of full rank with $\underset{\sim}{\Xi} = \mathbb{R} \times (0, \infty)$

Example (Multinomial family). $(X_0, X_1, \ldots, X_k) \sim \text{Mult.}(n; p_0, \ldots, p_k)$

$$f_\theta(x_0, \ldots, x_k) = \frac{n!}{x_0!\, x_1! \cdots x_k!} p_0^{x_0} p_1^{x_1} \cdots p_k^{x_k} I_B(x_0, \ldots, x_k)$$

where $B = \{(x_0, x_1, \ldots, x_k) : x_i \geq 0, \sum_{i=0}^{k} x_i = n\}$

Consider $\eta = (\log p_0, \log p_1, \ldots, \log p_k)$, $h(\underset{\sim}{x}) = \frac{n!}{x_0!\, x_1! \cdots x_k!} I_B(\underset{\sim}{x})$

$$f_\theta(\underset{\sim}{x}) = e^{\eta' x}\, h(\underset{\sim}{x}), \quad \underset{\sim}{x} \in \mathbb{R}^{k+1}.$$

$\Rightarrow$ The above form does not give an Exp. family of full rank, since there is no open set of $\mathbb{R}^{k+1}$ contained in the natural parameter space. Reparameterization leads to Exp. family of full rank.

Using $\sum_{i=0}^{n} x_i = n$ & $\sum_{i=0}^{n} p_i = 1 \Rightarrow x_0 = n - \sum_{i=1}^{n} x_i$, $p_0 = 1 - \sum_{i=1}^{n} p_i$

$$\Rightarrow f_\theta(\underset{\sim}{x}) = e^{\log p_0 \left(n - \sum_{i=1}^{n} x_i\right) + \sum_{i=1}^{n} \log p_i x_i} = e^{\sum_{i=1}^{n} \log \frac{p_i}{p_0} x_i + n \log p_0}$$

$$= e^{\eta_*' x_* - A(\eta_*)} h(\underset{\sim}{x}), \quad x_* = (x_1, \ldots, x_k), \quad \eta_* = \left(\log \frac{p_1}{p_0}, \ldots, \log \frac{p_k}{p_0}\right)$$

$A(\underset{\sim}{\eta_*}) = -n \log p_0 \Rightarrow$ Exp. family with full rank, $\underset{\sim}{\Xi} = \mathbb{R}^k$.

M. Rajab

Moments of T_i's

Let $g(\underset{\sim}{\eta}) = \int e^{\underset{\sim}{\eta}'\underset{\sim}{T}} h(x)\, d\mu = e^{A(\underset{\sim}{\eta})}$, $\underset{\sim}{\eta} \in \mathbb{R}^s$

Differentiating w.r.t η_j, gives

$$e^{A(\underset{\sim}{\eta})} \frac{\partial A(\underset{\sim}{\eta})}{\partial \eta_j} = \int \frac{\partial}{\partial \eta_j} e^{\sum^s \eta_i T_i} h(\underset{\sim}{x})\, d\mu$$

$$= \int T_j(\underset{\sim}{x})\, e^{\sum^s \eta_i T_i} h(\underset{\sim}{x})\, d\mu$$

$$\Rightarrow \frac{\partial A(\underset{\sim}{\eta})}{\partial \eta_j} = \int T_j(\underset{\sim}{x})\, e^{\sum^s \eta_i T_i - A(\underset{\sim}{\eta})}$$

$$= \int T_j(\underset{\sim}{x})\, p(\underset{\sim}{x} | \underset{\sim}{\eta})$$

$$\boxed{\frac{\partial A(\underset{\sim}{\eta})}{\partial \eta_j} = E(T_j)}$$

Now, $ET_i = \dfrac{\partial A(\underset{\sim}{\eta})}{\partial \eta_i} = \displaystyle\int T_i\, e^{\underset{\sim}{\eta}'\underset{\sim}{T} - A(\underset{\sim}{\eta})} h(\underset{\sim}{x})\, d\mu$

$$\frac{\partial^2 A(\underset{\sim}{\eta})}{\partial \eta_j \partial \eta_i} = \int T_i\, e^{\underset{\sim}{\eta}'\underset{\sim}{T} - A(\underset{\sim}{\eta})} \left(T_j - \frac{\partial A(\eta)}{\partial \eta_j}\right) h(\underset{\sim}{x})\, d\mu$$

$$= E(T_i T_j) - E(T_i)\, E(T_j)$$

$$= Cov(T_i, T_j)$$

The moment generating function (MGF) of $\underset{\sim}{T} = (T_1, \ldots, T_s)'$ is

$$M_T(\underset{\sim}{u}) = E\, e^{\underset{\sim}{u}'\underset{\sim}{T}} = \int_\Omega e^{\underset{\sim}{u}'\underset{\sim}{T}}\, e^{\underset{\sim}{\eta}'\underset{\sim}{T} - A(\underset{\sim}{\eta})} h(\underset{\sim}{x})\, d\mu(\underset{\sim}{x})$$

$$= \int_\Omega e^{(\underset{\sim}{u}+\underset{\sim}{\eta})'\underset{\sim}{T}} h(\underset{\sim}{x})\, d\mu(\underset{\sim}{x}) \Big/ e^{A(\underset{\sim}{\eta})}$$

$$= e^{A(\underset{\sim}{u}+\underset{\sim}{\eta})} \big/ e^{A(\underset{\sim}{\eta})} = e^{A(\underset{\sim}{u}+\underset{\sim}{\eta}) - A(\underset{\sim}{\eta})}$$