On likelihood function and hypothesis test

Definition 1. Let $f(x|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$, where each of the $\mathbf{X}_i$ denote the outcomes space of the i-th i.i.d. event. Given that $\mathbf{X} = \mathbf{x}$ is observed, the **likelihood function** of θ is defined:

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_i^n f(x_i|\theta) \tag{1}$$

e.g. Suppose an pdf of the form:

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta \end{cases} \tag{2}$$

its likelihood function is then:

$$L(\theta|\mathbf{x}) = \prod_i^n f(x_i|\theta) = \begin{cases} e^{-\sum_i^n x_i + n\theta} & \theta \leq \min_i x_i \\ 0 & \theta > \min_i x_i \end{cases} \tag{3}$$

$\min_i x_i$ is chosen because if $\exists x' \ni x' < \theta$, the likelihood function immediately takes the value of zero. and hence $L(\theta|\mathbf{x})$ is only non zero when each x_i is bounded below by θ

Definition 2. A **hypothesis test** is a rule that specifies:

1. For which sample values the decision is made to accept the null hypothesis as true
2. For which sample value the null hypothesis is rejected and a alternative hypothesis is accepted as true

Definition 3. The **likelihood ratio test** is defined:

$$\lambda(x) = \frac{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{x})} \tag{4}$$

for which the numerator is the maximum probability of the observed sample in null hypothesis parameter space, and the denominator is the one in alternative alternative hypothesis space. If the parameter vector in alternative hypothesis parameter space point toward somewhere where the maximum probability is higher than the one in null hypothesis, then the alternative hypothesis is accepted and the null hypothesis is rejected, and vice versa.

e.g. Suppose the same pdf from equation (2), and consider an null hypothesis: $H_0 : \theta \leq \theta_0$, and where the alternative hypothesis is $H : \theta > \theta_0$, and it can be seen that $L(\theta|\mathbf{x})$ is a strictly increasing w.r.t. θ when

θ is bounded by $\min_i x_i$. and hence the unrestricted maximum of $L(\theta|\mathbf{x})$ is:

$$L(\min_i x_i|\mathbf{x}) = e^{-\sum_i^n x_i + n\theta} \tag{5}$$

and it can be seen that under the definition of the likelihood ratio in equation (4). If $\min_i x_i \leq \theta_0$ the numerator of $\lambda(\mathbf{x})$ is $L(\min_i x_i|\mathbf{x})$, and if $\min_i x_i > \theta$, it is $L(\theta_0|\mathbf{x})$. and hence the likelihood ratio is:

$$\lambda(\mathbf{x}) = \begin{cases} 1 & \min_i x_i \leq \theta_0 \\ e^{-n(\min_i x_i - \theta_0)} & \min_i x_i > \theta_0 \end{cases} \tag{6}$$

then experimenter can choose an arbitrary criteria for $\lambda(x) = c$ so that anything below the criteria is rejected. and the test is the one with a rejection region of $\{\mathbf{x} : \min_i x_i \leq \theta_0 - \frac{lnc}{n}\}$

e.g. Suppose the random sample form a standard normal distribution. and consider a null hypothesis of $\theta = \theta_0$, and an alternative hypothesis of $\theta \neq \theta_0$. the likelihood ratio is written as:

$$\lambda(\mathbf{x}) = \frac{(2\pi)^{-n/2} exp[-\sum_i^n (x_i - \theta_0)^2/2]}{(2\pi)^{-n/2} exp[-\sum_i^n (x_i - \bar{x})^2/2]} \tag{7}$$

conventionally, log-likelihood ratio (L.L.R.) is used instead of the likelihood ratio $\lambda(\mathbf{x})$ when the random variable are of standard normal variable:

$$\text{L.L.R.} = -2\ln\left(\frac{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{x})}\right) \tag{8}$$

and hence the expression for L.L.R. can be simplified to:

$$\text{L.L.R.} = \sum_{i=1}^{n} (x_i - \theta_0)^2 - \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{9}$$

noting that:

$$\sum_{i=1}^{n} (x_i - \theta_0)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2 \tag{10}$$

L.L.R. can be further simplified to:

$$\text{L.L.R.} = n(\bar{x} - \theta_0)^2 \tag{11}$$

and the rejection region with rejection criteria of c can be the found to be:$\{\mathbf{x} : |\bar{x} - \theta_0| \geq \sqrt{c/n}\}$

Definition 4. A **measure** is an extended real-valued set function μ having the following properties:

1. The domain $\mathcal{A}$ of μ is a σ-algebra.

2. μ is nonnegative on $\mathcal{A}$.

3. μ is completely additive on $\mathcal{A}$.

4. $\mu(\emptyset) = 0$

NOTE. The main difference between a probability measure and a general measure is that the codomain of $\mu \in [0, 1]$.

Definition 5. Assume μ and ν are two probability measures on an algebra $\mathcal{A}$ of the subsets of the sample space Ω **total variation distance** (TVD) is defined:

$$\delta(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| \tag{12}$$

It can be seen as a metric defined in sample space, expressing the statistical difference between two measures, or two different probability distribution. It can be seen that TVD is indeed a metric:

Definition 6. A metric is a function: $d : M \times M \to \mathbb{R}$ that is defined in a metric space, (M, d) ,that satisfied the following properties, for $x, y, z \in M$:

1. $d(x, y) = 0 \iff x = y$ (identity of indiscernibles)

2. $d(x, y) = d(y, x)$ (symmetry)

3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)

NOTE. non-negativity is deduced from the above properties, and hence not a direct definition of metric.

proof That TVD is a metric cab be easily proven:

1. (identity of indiscernible) assume $\mu = \nu$, then $\mu(A) = \nu(A)$, and hence $\delta(\mu, \nu) = 0$

2. (summetry) $\delta(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| = \sup_{A \in \mathcal{A}} |\nu(A) - \mu(A)| = \delta(\nu, \mu)$

3. (triangle inequality) $\sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| \leq \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A) + \chi(A) - \chi(A)| = \sup_{A \in \mathcal{A}} |\mu(A) - \chi(A)| + \sup_{A \in \mathcal{A}} |\chi(A) - \nu(A)|$