Testing of Hypothesis

Example 1:

- Assume a coin has a probability p of showing heads each time it is flipped.

- The results of flipping the coin 100 times are viewed as i.i.d. random variables $X_1, X_2, \ldots, X_{100}$

- Suppose $\sum_{n=1}^{100} X_n = 67$

Question: Is the coin fair? ; Is $p = \frac{1}{2}$?

Earlier:

· Estimate p $\equiv$ { Maximum likelihood
Method of moments

· Confidence Interval for p.

Can we device a method to verify / dismiss the conjecture , $p = \frac{1}{2}$?

· find the interval and check if $\frac{1}{2}$ belongs to it

Testing of Hypothesis

Example 1
$X \sim Be(p)$

- $X_1, X_2, \ldots, X_n$ are i.i.d. $\sim X$

and X has a p.d.f. or p.m.f. given by

- $f(x) = f(x \mid p_1, p_2, \ldots p_d) = f(x \mid p)$

which depends on

$p = (p_1, p_2, \ldots, p_d) \in \mathcal{P} \subset \mathbb{R}^d$ for some $d \geq 1$.

$d=1$
$P = [0,1]$

- A conjecture or "null hypothesis" :

$P_0 = \{\frac{1}{2}\}$

$p \in \mathcal{P}_0$, where $\mathcal{P}_0$ is a proper subset of $\mathcal{P}$.

- Our objective will be to devise a procedure.

- One may need to redraft the question into the above format.

(i) Estimate p from $X_1, X_2, \ldots, X_n$ & see if $\hat{p} \in P_0$.

Example : $\hat{p} = 0.67$

(ii) Find a Confidence Interval for p & see if P_0 "$\subseteq$" in the interval

Example 1 : Ex

Testing of Hypothesis

- Construct a "test statistic", which is a function of the data $X_1, X_2, \ldots, X_n$.

- To "test" the conjecture that $p \in \mathcal{P}_0 \subset \mathcal{P}$,
  Check if the observed value of the test statistic could conceivably have come from a value of $p \in \mathcal{P}_0$

- Then quantify the degree of this possibility through a probability that is referred to as the "p-value".

• Define p-value precisely later

p-value should not depend on parameter assuming $p \in P_0$.

• A good test-statistic is one that should not depend on the parameter p if $p \in P_0$.

Example 1: Test-statistic $\bar{X}$

$S_n = X_1 + \ldots + X_n \sim \text{Binomial}(n, p)$

$p \in P_0 = \{\frac{1}{2}\} \rightsquigarrow S_n \sim \text{Binomial}(n, \frac{1}{2})$

Testing of Hypothesis

$X_i \sim X \quad X \overset{d}{=} \text{Ber}(p)$

$0 < p < 1$

Back to Example 1:

- Null Hypothesis: The coin is fair. $p = \frac{1}{2}$

- Test-statistic: $S = \sum_{i=1}^{100} X_i$.

  Under the null $\longleftrightarrow$ $p = \frac{1}{2}$

  $S \sim \text{Binomial}(100, 0.5)$, so $P(S = 67)$ and $P(S \geq 67)$ are given by

```
> dbinom(67, 100, 0.5)

[1] 0.0002324713

> 1 - pbinom(66, 100, 0.5)

[1] 0.0004368599
```

Probabilities are low !

$P(S = 50) \sim 0.08$

$P(S \geq 50) = 0.5$

increase n

check : "$P(S = \frac{n}{2})$"

Testing of Hypothesis

Example 2: Effectiveness of vaccine

- Suppose a medical research team has come up with a potential vaccine for a dangerous disease.

- Choose n individuals from a vulnerable population, give the vaccine to n_1 of them (giving the remaining $n_2 = n - n_1$ individuals a placebo as "control").

  Then, wait for a reasonable period of time (which may depend on the features of the disease) to see how many individuals are affected by the disease in each group.

- Results:

	Affected	Not Affected	Total
Vaccine	X_{11}	X_{12}	n_1
Placebo	X_{21}	X_{22}	n_2

  where X_{11} denote the number of vaccinated individuals who were affected, X_{12} denotes the number of vaccinated individuals who were not affected, ...

Testing of Hypothesis

Example 2:

- The natural independent units are the participants in the study. [Independence]
- Depending on treatment/outcome we place the individuals in one of the the four matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

- Model:

$$p_{11} \sim \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad p_{21} \sim \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad p_{12} \sim \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad p_{22} \sim \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

with the constraints that $0 \leq p_{11}, p_{21}, p_{12}, p_{22} \leq 1$ and $p_{11} + p_{21} + p_{12} + p_{22} = 1$.

- As the n random outcomes are independent, the individuals outcomes M_i can be combined and summarized by their sum, giving the 2×2 table

$$X = \sum_{i=1}^{n} M_i = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

Ex :- Identify distribution of X?

$$X = \sum_{i=1}^{4} M_i \quad = \quad f(\cdot \mid p).$$

M_i are i.i.d $\sim p$

$$\mathcal{P} = \left\{ p = (p_1, p_2, p_3, p_4) : \begin{array}{l} 0 \le p_i \le 1 \\ \sum_{i=1}^{4} p_i = 1 \end{array} \right\}$$

The above model is valid if independence assumption holds

Alternate Approach :

- Each individual is assigned vaccine w.p π
  & assigned placebo w.p. $1-\pi$
- $\mathbb{P}(\text{Affected} \mid \text{vaccine}) = \alpha_{11}$
  $\alpha_{12} = 1 - \alpha_{11}$

- $\mathbb{P}(\text{Affected} \mid \text{placebo}) = a_{21}$

$a_{22} = 1 - a_{21}$

$\Rightarrow \mathbb{P}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = P(V \cap A)$

$= P(A \mid V)\, P(V)$

$= a_{11}\ \pi$

$\mathbb{P}\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = \ldots\ldots\ a_{21}(1-\pi)$

$\mathbb{P}\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \ldots\ a_{12}\,\pi$

$\mathbb{P}\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \ldots\ a_{22}(1-\pi)$

$$X = \sum_{i=1}^{n} M_i = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

Ex :- Identify distribution of X?

One Possible test to "check Effectiveness Vaccine"

lemma: $\{p_{ij}\}$ $i=1,2,$ $j=1,2.$

$$p_{k\bullet} = p_{k1} + p_{k2} \qquad k=1,2$$
$$p_{\star\ell} = p_{1\ell} + p_{2\ell} \qquad \ell=1,2$$

For m^{th} individual let

$$T_m = \begin{cases} 1 & \text{if } m \text{ got vaccine} \\ 0 & \text{otherwise} \end{cases}$$

$$A_m = \begin{cases} 1 & \text{if } m \text{ got disease} \\ 0 & \text{otherwise} \end{cases}$$

T.F.A.E

- T_m & A_m are independent $\forall\, m \geq 1$
- $p_{k\ell} = p_{k\bullet}\, p_{\star\ell}$ $\forall$ $k=1,2$, $\ell=1,2$
- $a_{11} = a_{21}$

$$X = \sum_{i=1}^{n} M_i = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

Under Independence:

$$\hat{p}_{k\bullet} = \frac{X_{k1} + X_{k2}}{n}$$

MLE:

$$\hat{p}_{\bullet\ell} = \frac{X_{1\ell} + X_{2\ell}}{n}$$

$\Leftrightarrow$ $\hat{a}_{12} = \hat{a}_{21} = \frac{1}{n}(X_{11} + X_{21})$