Examine for Statistic Model.

• $\{X_k\}_1^n$ i.i.d from cdf. $F(\cdot|\theta)$, we will examine that whether the assumption $F(\cdot|\theta)$ is reasonable.

Extend: $\mathcal{N}_0 = \{F(\cdot|\theta) | \theta \in \Theta\}$ to

$$\mathcal{N} = \mathcal{N}_0 \cup \{\text{other Asst's}\}.$$

(1) **Procedure:**

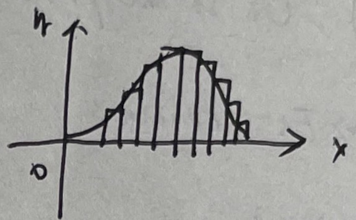
① Observe $X_1, X_2 \cdots X_n$

② Separate $[X_{(1)}, X_{(n)}]$ into $\{(t_{i-1}, t_i]\}_1^m$;

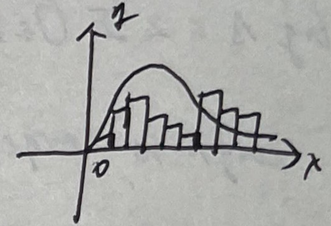
O_i is the number of $\{X_k\}_1^n$ falling into $(t_{i-1}, t_i]$. (indept of $\{X_k\}_1^n$)

③ We obtain: $O_1 \cdots O_m \sim \text{Multi}(n, p_1 \cdots p_m)$ and Histogram

④ Check whether the graph of $F(\cdot|\hat{\theta}_0)$ fit for the histogram.



or



(2) **Examples:**

① **GLR test for goodness of fit:**

LR statistic:

$$\Lambda = \frac{\max_{\vec{p} \in \mathcal{N}_0} \left[\frac{n!}{O_1! \cdots O_m!} p_1^{O_1} \cdots p_m^{O_m} \right]}{\max_{\vec{p} \in \mathcal{N}} \left[\frac{n!}{O_1! \cdots O_m!} p_1^{O_1} \cdots p_m^{O_m} \right]} = \prod_1^m \left(\frac{p_i(\hat{\Theta})}{\hat{p}_i} \right)^{O_i}$$

$$\mathcal{N}_0 = \{p(\Theta) : p(\Theta) = (p_1(\Theta), \cdots p_m(\Theta))\}.$$

$$p_i(\Theta) = F(t_i|\Theta) - F(t_{i-1}|\Theta)$$

$$\Rightarrow -2\log\Lambda = 2\sum_1^m n\hat{p}_i \log\left(\frac{n\hat{p}_i}{np_i(\hat{\Theta})}\right)$$

$$\begin{cases} O_i = n\hat{p}_i : \text{the observed count} \\ E_i = np_i(\hat{\omega}) : \text{the expected count.} \end{cases}$$

$\therefore -2\log\Lambda = 2\sum O_i \log\left(\frac{O_i}{E_i}\right)$. O_i is the promotion

When $-2\log\Lambda$ is large $\Rightarrow O_i \gg E_i$. reject!

$\therefore$ Reject Region: $\{-2\log\Lambda \geq c\}$.

② <u>Pearson's Chi-square Test:</u>

· Test Statistic: $X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

E_i is for the proposition!

$\Rightarrow$ Reject Region: $X^2 \geq c$

<u>Remark:</u> i) From Taylor Expansion of $x\log\frac{x}{x_0}$ at x_0:

$$-2\log\Lambda \simeq X^2 \longrightarrow \chi^2_t.$$

ii) It's easy to calculate that GLR!

③ <u>Loss of Information:</u>

$$\{X_k\}_1^n \underset{①}{\Longleftrightarrow} \{X_{(k)}\}_1^n \underset{②}{\Longleftrightarrow} \{O_i\}_1^m$$

① : Need i.i.d

② : Need discrete.

(3) Application of GLR

— Poisson Dispersion Test:

- We wanna to examine whether the samples from a Poisson model are sampled in a constant rate.

① Suppose $X_i \sim Poi(\lambda_i)$.

$\Omega = \{(\lambda_1, \lambda_2 \cdots \lambda_n) \mid \lambda_i > 0\}$. $\dim \Omega = n$.

$\Omega_0 = \{(\lambda_1 \cdots \lambda_n) \mid \lambda_i = \lambda_j > 0\}$. $\dim \Omega_0 = 1$

$H_0: \vec{\lambda} \in \Omega_0$. $H_A: \vec{\lambda} \in \Omega / \Omega_0 \triangleq \Omega_A$

If under Ω. MLE of $\{\lambda_i\}$ is $\{\tilde{\lambda}_i\} = \{X_i\}$

since under Ω_0. MLE is $\bar{X}$

$$\therefore \Lambda = \prod_{i}^{n} \left(\frac{\bar{X}}{x_i}\right)^{x_i} e^{x_i - \bar{X}}$$

$$= \frac{\prod \tilde{\lambda}^{x_i} e^{-\tilde{\lambda}} / x_i!}{\prod \bar{\lambda}_i^{x_i} e^{-\bar{\lambda}_i} / x_i!}$$

$$-2\log \Lambda = 2\sum \left[x_i \log\left(\frac{x_i}{\bar{x}}\right) + x_i - \bar{x}\right]$$

$$\simeq \frac{1}{\bar{x}} \sum (x_i - \bar{x})^2 = \frac{n\hat{\sigma}^2}{\bar{x}}$$. by Taylor expansion

Remark: i) It's reasonable. Since for $X \sim P(\lambda)$

$Var(X) = E(X) = \lambda$.

So we compare: $\frac{\hat{\sigma}^2}{\bar{X}}$ to decide whether the rate λ is const!

ii) It's different with general GLR test. Since it only check $Var \simeq E$? But not check the model is Poisson! (If there's a dist that $Var \simeq E$. Then it won't be rejected!)

But it's efficient than GLR test when the model is Poisson!