

Lectures of June 2nd, 2006 - AM

Scribe: Mohamed Al Ali, Daniel Beaubien, Silu Chen

1 Testing the Fit of Distribution to Data

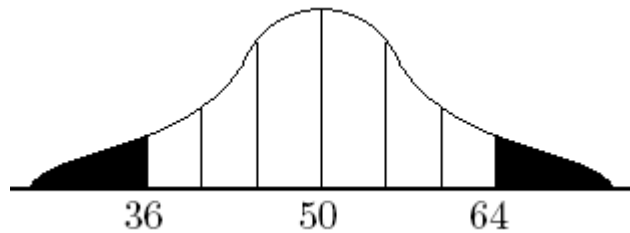
Example 1

Suppose we toss a coin 100 times and we observe 64 "H"'s. Let H_o denote the hypothesis that the coin is fair. [By this, we mean the Bernoulli trial has $p = 0.5$]. Based on the observations, should we reject H_o ?

- Let RV N be the number of "H"'s seen over 100 trials. If H_o is true, we "expect" to see 50 "H"'s. i.e. : $E[N | H_o] = 50$.
- But we observed 14 more "H"'s than expected. Is 64 too large for the coin to be considered fair?

Let us find:

$$P[|N - 50| \geq 14 | H_o]$$



What is the probability of the shaded area?

- If the trials in the shaded areas are large then the coin is not fair.
- If the probabilities are large, then the value is close to the mean
- If the probabilities are small, then the value is far from the mean

$$\begin{aligned} P[|N - 50| \geq 14 | H_o] &= 1 - \sum_{k=37}^{63} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k} \\ &= 0.0093 \end{aligned}$$

If we think 1% probability is small, we should reject H_o .

2 A Formal Procedure for a General Setting (Chi-Square Test)

Setup: Given:

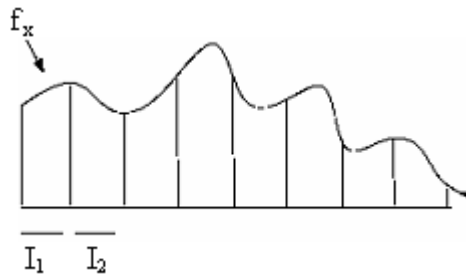
- 1. A hypothesis H_o (which is referred to as the null hypothesis) that describes a distribution function f_X , from which it is postulated that observed data are drawn independently
- 2. Observed data

Objective: Determine whether we should reject H_o based on our observation

3 Procedure

1. Specify a value α , called significance level [where α indicates the probability threshold below which a probability will be considered "small" and the corresponding event will be considered insignificant.] Typical accepted values of α are 1%, 2.5%, 5%.
2. Partition the range S_X into K disjoint intervals I_1, I_2, \dots, I_K and determine according to f_X the probability that the observation is inside the interval

$$P[X \in I_i] = \int_{I_i} f_X(x) dx \quad \forall i \in \{1, 2, \dots, K\}$$



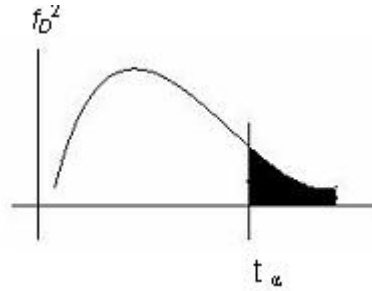
3. Compute

$$D^2 = \frac{\sum_{i=1}^K (N_i - np[X \in I_i])^2}{np[X \in I_i]}$$

→ This tells how far N_i is away from the expected N_i if the hypothesis is true.

where n is the total number of samples in the observation, and N_i is the number of samples in the observation that falls into interval I_i .

The distribution of D^2 is typically considered as chi-square distribution, which we will denote by f_{D^2} . The "degree of freedom" (the parameter) of f_{D^2} is $K-1$.



find D^2 such that the right tail is equal to α .
 we have specified α such that at some point the right tail is equal to α . If D^2 is on the right of α we could reject H_0 but if it is on the left of α then we can not reject it.

4. Find threshold t_α based on α and degree of freedom such that t_α is the value with

$$\int_{-\infty}^{+\infty} f_{D^2}(y) dy = \alpha.$$

This is usually done by a look-up table (table 3.5 in textbook, page 142.)

5. Check whether $D^2 \geq t_\alpha$ or $D^2 < t_\alpha$
- If $D^2 \geq t_\alpha$, we conclude that H_0 is rejected at significance level α .
 - If $D^2 < t_\alpha$, we conclude that H_0 is not rejected at significance level α .

4 Remarks

1. Given K different partitioning of S_X into K intervals leads to somewhat different result.

One should try to use the partitioning such that:

(a)

$$P[X \in I_1] \simeq P[X \in I_2] \simeq \dots \simeq P[X \in I_k]$$

(b)

$$nP[X \in I_i] \geq s$$

Number of intervals shouldn't be too small.

2. If the Postulated distribution f_X is obtained from the observed data, then the Degree Of Freedom for Chi-Square test should be reduced. More Specifically, if f_X is Parameterized by r parameters and every parameter is obtained from the data, then the Degree Of Freedom should be chosen as $(K - 1 - r)$, instead of $(K - 1)$.

5 Characteristic Function

The characteristic function of a RV is defined as:

$$\underbrace{\Phi_X(w)} = E[e^{jwX}]$$

for every fixed w expected value can be calculated

If X is continuous RV, then

$$\Phi_X(w) = \underbrace{E[e^{jwX}]}_{g(x)} = \int_{-\infty}^{\infty} e^{jwX} f_X(x) dx$$

Note: This looks like a Fourier Transform.

If we choose the PDF to be a time signal then the Characteristic Function is its Fourier Transform in the frequency domain. All properties of Fourier Transform hold for characteristic function.

If X is a discrete RV, then

$$\Phi_X(w) = E[e^{jwX}] = \sum_{x_k \in S_X} e^{jwX_k} P_X(x_k) \rightarrow$$

discrete random variables.

If S_X is the set of all integers, then

$$\Phi_X(w) = \sum_{k=-\infty}^{\infty} e^{jwk} P_X(k) \rightarrow$$

integer-valued random variables

note: The $P_X(x_k)$ are simply the coefficients of the Fourier Series of the periodic function $\Phi_X(w)$.

6 Remark

1. The Characteristic Function is essentially the continuous-time Fourier transform of the PDF $f_X(x)$ of the random variable X with x interpreted as continuous time.
2. Every property of Fourier Transform holds for Characteristic Function.

7 Example

If X has PDF f_X , Y has PDF f_Y , and Z has PDF f_Z such that:

$$f_Z(\zeta) = f_X(\zeta) \star f_Y(\zeta)$$

Then

$$\Phi_Z(w) = \Phi_X(w) \cdot \Phi_Y(w)$$

8 Moment Theorem

Since $f_X(x)$ and $\Phi_X(w)$ form a transform pair, we would expect to be able to obtain the moments of X from $\Phi_X(w)$ as follows:

$$E[X^n] = \left[\frac{1}{j^n} \right] \left[\frac{d^n \Phi_X(w)}{dw^n} \right] \Big|_{w=0}$$