Advanced statistical methods

Frédéric Pascal

CentraleSupélec, Laboratory of Signals and Systems (L2S), France
frederic.pascal@centralesupelec.fr
http://fredericpascal.blogspot.fr

MSc in Data Sciences & Business Analytics
CentraleSupélec / ESSEC
Oct. 2nd - Dec. 20th, 2017

CentraleSupélec

Part B

On the use of classical hypothesis techniques

Part C: Contents

I. Simple Hypothesis Testing

II. χ^2 test

Simple Hypothesis Testing

Context

We are interesting to test the assumption that dying persons can delay their death until very important (for them!) events. For that purpose, Phillips D. P. and King E. W. have collected in the article *"Death takes a holiday: mortality surrounding major social occasions"; Lancet, 1988, pp 728–732* several observations of deaths around a Jewish religious holiday. On 1919 deaths, 922 (respectively 997) occurred the previous while the others the week after.

Statistical modelling

1. By denoting p the probability of dying before the Jewish religious holiday, write a statistical model.
2. Let H_0 the hypothesis "dying persons can delay their death", establish the binary hypothesis test.

ML estimation with PDF

Discussion on the hypothesis test

Theoretical analysis

3. Built a convergent asymptotic test at the level α. What is the conclusion of this test for $\alpha = 5\%$?

4. Built the UMP test at the level α.

5. For a binomial-distributed random varaible X, with parameters $n = 1919$ and $p = 1/2$, one has $\mathrm{P}(X \leq 922) = 0.0456$ and $\mathrm{P}(X = 923) = 0,0045$. Deduce the UMP test at the level $\alpha = 0,05$ and draw conclusions.

I. Simple Hypothesis Testing

II. χ^2 test

χ^2 test

Context

In order to analyze the activity of a call center, one count during 200 consecutive observations (i.e. during 200 seconds), the number of incoming calls per second. Results are reported in the following table:

Number of incoming calls per second	0	1	2	3	4	5	6	7	8	9	10	11
Observations	6	15	40	42	37	30	10	9	5	3	2	1

Discussion: What are we interested in?

$\Rightarrow$ Statistical modelling, estimation problem, decision making...

χ^2 test

1. Does the empirical distribution fit to a Poisson distribution? Conclude for $\alpha = 5\%$.
2. What would give the distribution fitting if it was a Poisson distribution with a parameter $\lambda_0 = 3,7$?
3. Previously, it has been decided that the number of incoming calls per second follows a Poisson distribution with a parameter $\lambda_1 = 4$. Test the stability of the behavior on the observed data.

k	0	1	2	3	4	5	6	7
$P_{3,7}(k)$	0,0247	0,0915	0,1692	0,2087	0,1931	0,1429	0,0881	0,0466
$P_4(k)$	0,0183	0,0733	0,1465	0,1954	0,1954	0,1563	0,1042	0,0595
k	8	9	10	11	12	13	14	15
$P_{3,7}(k)$	0,0215	0,0089	0,0033	0,0011	0,0003	0,0001	$< 0,0001$	$< 0,0001$
$P_4(k)$	0,0298	0,0132	0,0053	0,0019	0,0006	0,0002	0,0001	$< 0,0001$

Table: Poisson distribution: for all $k \in \mathbb{N}$, $P_\lambda(k) = e^{-\lambda}\lambda^k / k!$.

q	75%	80%	85%	90%	95%	97.5%	99%
χ^2_7	9.04	9.80	10.75	12.02	14.07	16.01	18.48
χ^2_8	10.22	11.03	12.03	13.36	15.51	17.53	20.09

Table: Useful quantiles.