Project Report

Written as completion requirement of CERN Summer Student Program (NMS)

Written By Ujjwal Agarwal

Project Supervisor: Dr. Andre Tinoco Mendes

Official Supervisor: Dr. Archana Sharma

Asymptotic Likelihood Distribution for Correlated & Constrained Systems

In Particle Physics, search for new signals is done with the help of theory of likelihood tests and their asymptotic approximations. For a recorded data n, likelihood ratio tests are used to find out the likeliness for strength parameter to have a particular value or to find out a range of strength parameter for which confidence level is above a threshold level.

In my project, I had to study the asymptotic likelihood distribution when there are 2 signal processes and these processes are correlated. And thus we have 2 strength parameter namely μ_1 and μ_2. And the covariance matrix is non diagonal. Alternatively, correlation coefficient

$$\rho = \frac{cov(\mu_1, \mu_2)}{\sigma_{\mu 1}\sigma_{\mu 2}}$$

Is non zero. Also, the value of strength parameter are taken to be greater than zero for physical reason that the signal processes can only add to background signal.

Let there be a population K from which n samples are picked up. A variate x is measured for these samples and the distribution function for the x is $f(x, \Theta_1, \Theta_2, \ldots, \Theta_h)$. Mathematically simple hypothesis is a set of values for Θ's .Likelihood tests is a Mathematical theory which is used to test likeliness of a test hypothesis to be true i.e. the value of Θ belongs to test hypothesis. It is also used to test likeliness of a hypothesis in comparison to alternative hypothesis.

Let

Φ = h dimensional space of Θ

ω = test hypothesis

For Θ in ω, we have

$\Theta_i = \Theta_{0i}$ for $(i = 1, 2, \ldots, m)$

Let $X = \{x_1. x_2. \ldots, x_n\}$ = value of x for n samples picked from population

Associated Probability density function will be

$$P(X,\theta) = \prod_{\alpha=1}^{n} f(x_\alpha, \theta_1, \dots, \theta_h)$$

Likelihood ratio is defined as

$$\lambda = P_\omega / P_\varphi$$

$$Where, \quad P_\omega = Sup\big(P(X,\theta)\big) \quad for\ \theta \in \omega$$

$$P_\varphi = Sup\big(P(X,\theta)\big) \quad for\ \theta \in \varphi$$

Wilks assumes existence of maximum likelihood estimators of Θ (Θ which maximize P over complete h dimensional space) and models the probability distribution P around Θ as Gaussian with covariance matrix V and information matrix I.

$$\hat{\theta} = Maximum\ likelihood\ estimators$$

He then proves that the value of λ depends of choice of ω and for Gaussian distribution of probability

$$-2\log(\lambda) = \chi_m^2 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$Where\ \chi_m^2 = \sum_{i,j=1}^{m} c_{ij}\big(\hat{\theta}_i - \theta_{0i}\big)\big(\hat{\theta}_j - \theta_{0j}\big)$$

$$and \quad c_{ij} = elements\ of\ information\ matrix\ I$$

$$Note\colon for\ gaussian\ distribution, I = V^{-1}$$

For large sample, we drop off $O\left(\frac{1}{\sqrt{n}}\right)$ term. Above equation gives the distribution of log-likelihood ratio as a function of test hypothesis.

We now proceed to geometrical interpretation. We can see the distribution as square of distance between maximum likelihood estimator, which would be represented by a point in h dimensional space, and the value of Θ in ω which would maximize P, a point in h dimensional space which lies within region ω. This distance is calculated in a curved Riemannian Geometry in general.

Now, coming to problem statement. We set m = 2. So, for Θ_1 and Θ_2

$\Theta_1 = \Theta_{01}$ and $\Theta_2 = \Theta_{02}$ for region ω

$\Theta_1 \geq 0$ and $\Theta_2 \geq 0$ for region Φ

$\widehat{\theta_1}, \widehat{\theta_2}$ *can be any real number*

We still do not consider correlation between 2 parameters.

The distribution in this case becomes.

$$-2\log(\lambda) = \begin{cases} \chi_2^2 & for\ \widehat{\theta_1} \geq 0, \widehat{\theta_2} \geq 0 \\ \chi_1^2 & for\ \widehat{\theta_1} < 0, \widehat{\theta_2} \geq 0 \\ \chi_1^2 & for\ \widehat{\theta_1} \geq 0, \widehat{\theta_2} < 0 \\ \chi_0^2 & for\ \widehat{\theta_1} < 0, \widehat{\theta_2} < 0 \end{cases}$$

Now that we have the result for a special case of $\rho = 0$, we can proceed to general case which will involve correlation between the 2 parameters.

For this, we have

$\Theta_1 = \Theta_{01}$ and $\Theta_2 = \Theta_{02}$ for region ω

$\Theta_1 \geq \Theta_{01}$ and $\Theta_2 \geq \Theta_{02}$ for region Φ

$\widehat{\theta_1}, \widehat{\theta_2}\ can\ be\ any\ real\ number$

Zi = $\widehat{\theta_i} - \theta_{0i}$

$I^{ij} = elements\ of\ I^{-1}(\theta_0)$

Now for -2log(λ), we have

$$-2\log(\lambda) = inf_{\theta\in\omega}(Z - (\theta - \theta_0))' I(\theta_0)(Z - (\theta - \theta_0)) \\ - inf_{\theta\in\varphi}(Z - (\theta - \theta_0))' I(\theta_0)(Z - (\theta - \theta_0))$$

In this case, we have to use a transformation T to diagonalize the information matrix and then the distribution becomes

$$for\ Z = P\tilde{Z}$$

$$-2\log(\lambda) = \begin{cases} \widetilde{Z_1^2} + \widetilde{Z_2^2} = \chi_2^2 & for\ \widetilde{Z_1} \geq 0, \widetilde{Z_2} \geq 0 \\ \widetilde{Z_1^2} = \chi_1^2 & for\ \widetilde{Z_1} \geq 0, \widetilde{Z_2} < 0 \\ \widetilde{Z_2^2} = \chi_1^2 & for\ \widetilde{Z_1} < 0, \widetilde{Z_2} \geq 0 \\ 0 = \chi_0^2 & for\ \widetilde{Z_1} < 0, \widetilde{Z_2} < 0 \end{cases}$$

The asymptotic distribution of -2log(λ) is a mixture of $\chi_0^2, \chi_1^2\ and\ \chi_2^2$ with mixing probabilities $1/2 - p,\ 1/2\ and\ p$ respectively. And mixing probability p is given as

$$p = \frac{1}{2\pi}\cos^{-1}\left(I_{12} / \left(\sqrt{I_{11} I_{22}}\right) \right) = \frac{1}{2\pi}\cos^{-1}(-\rho)$$

$$where\ I_{ij}\ are\ (i,j)\ entries\ of\ information\ matrix$$

Above result gives us the asymptotic likelihood distribution for 2 constrained and correlated parameters.

References

1) Cowan, G., Cranmer, K., Gross, E. et al. Eur. Phys. J. C (2011) 71: 1554. doi:10.1140/epjc/s10052-011-1554-0
2) Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann. Math. Statist. 9 (1938), no. 1, 60--62. doi:10.1214/aoms/1177732360. http://projecteuclid.org/euclid.aoms/1177732360.
3) Chernoff, Herman. On the Distribution of the Likelihood Ratio. Ann. Math. Statist. 25 (1954), no. 3, 573--578. doi:10.1214/aoms/1177728725. http://projecteuclid.org/euclid.aoms/1177728725.
4) Self, Steven G., and Liang Kung-Yee. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions." *Journal of the American Statistical Association* 82.398 (1987): 605-10. Web.