Statistical Methods Exercises 10 Autumn 2020

Return your solutions by 12.00 Finnish time on Thursday 3.12.2020 to Moodle course page: `https://moodle.helsinki.fi/course/view.php?id=30207`

1. **Acceptance-rejection/"Hit-or-miss" method and MC estimate**. (i) Estimate the volume of a sphere with radius $R = 1$ (arbitrary units) using Monte Carlo (MC) methods without using formulas for the volume, any explicit integration or anything requiring information about the value of π. Describe the method in detail in your answer and add any computer code used. Hint: generate 10000 random space points in a cube allowing the x-, y- and z to vary within the range [-1,1] and then use the acceptance-rejection method on the generated space points.

   (ii) The volume of a sphere is $4\pi R^3/3$. Use the result of (i) to estimate the value of π. Estimate also the uncertainty on the π determination.

2. **Significance of a signal & confidence intervals from likelihood ratio**. The number of observed events in a new physics search is n. This number can be treated as a Poisson variable with a mean of $s+b$, where s is the expected number of events for the signal process (= the new physics phenomenon) and b is the number of expected background events (= known physics). The likelihood function is therefore

   $$L(s,b) = (s+b)^n e^{-(s+b)}/n!$$

   Suppose that $b = 2.8$ (known exactly !) and we observe $n = 15$ events.

   (i) Compute the P-value for the hypothesis that $s = 0$, i.e. there is no new physics. To sum the Poisson probabilities, you can use the relation

   $$\sum_{n=0}^{m} P(n\,;\nu) = 1 - F_{\chi^2}(2\nu\,;n_{dof}) ,$$

   where $P(n\,;\nu)$ is the Poisson probability for n given a mean value ν, and $F_{\chi^2}(2\nu\,;n_{dof})$ is the cumulative χ^2 distribution for $n_{dof} = 2(m+1)$ degrees of freedom. Can be computed using mathematical packages or by web-based applets (google for "probability distribution applet"). NB! make sure the tool used calculates the P-value with sufficient precision.

   (ii) Compute the corresponding significance i.e. how many standard deviations of a standard Gaussian does the P-value correspond to? Is it a discovery i.e. the significance larger than 5 standard deviations?

   (iii) Assume we have seen signal of a new physics phenomenon, what is the confidence interval for the number of signal events at 68.3 % confidence level (one Gaussian standard deviation)? One way of estimating the confidence interval is using the likelihood ratio [1]:

$$\lambda(\hat{b}\,;n) \quad = \quad L(s, b_{exact}\,;n)/L(\hat{s}, b_{exact}\,;n) \ ,$$

where $L(\hat{s}, b_{exact}\,;n)$ is likelihood value when b is assumed exactly known and $\hat{s}$ estimated by maximizing $L\,(s, b)$ i.e. $\partial L/db = 0$, and $L(s, b_{exact}\,;n)$ likelihood value when b is assumed exactly known and s allowed to vary freely (irrespective of n). The 68.3 % CL interval are the s values satisfying $-2\,ln\lambda(\hat{b}\,;n) \leq -2\,ln\lambda|_{min} + 1$. Hint: plot $-2\,ln\lambda(\hat{b}\,;n)$ vs s.

iv) In reality b is affected by systematic uncertainties. Systematic uncertainties usually don't follow a Gaussian distribution. However, a uniform probability distribution at e.g. 95 % confidence level can often be defined for them. Assume b to be uniformly distributed in the range $[b - \sigma_b,\ b + \sigma_b]$, where $\sigma_b = 0.5$. Calculate now the confidence interval for the number of signal events at 68.3 % CL. The likelihood function stays the same but now the new b is uniformily distributed variable in the range defined above. Find s values satisfying $\lambda(\hat{b}\,;n) \leq \langle -2ln\lambda|_{min,\sigma_b}\rangle + 1$. NB! remember to avoid "undercoverage".

Hint: Test whether a certain s satisfies the condition by making sure that at least 68.3 % of the allowed b-values satisfies the condition. Simplest done by making pseudoexperiments for the new b (for example drawn 10000 b-values from a uniform probability density function in the defined range) and see whether sufficient fraction satisfy condition. Any change of the confidence interval? Exercise gives max 12 points.

THAT'S ALL EXERCISES FOR THIS COURSE, FOLKS !!

[1] e.g. W.A. Rolke, A.M. Lopez and J. Conrad: Limits and Confidence Intervals in the Presence of Nuisance Parameters, *Nucl. Instr. & Meth.*, **A 511** (2005) 493-503; arXiv:physics/0403059.