Practical Statistics Problems, Sept 2024

1) An experiment is searching for quarks of charge 2/3, which are expected to produce 4/9 the ionisation I_0 of unit charged particles. In an exposure in which 10^5 cosmic ray tracks are observed, 1 track has its ionisation measured as $0.44I_0$. The detector is such that ionisation measurements are Gaussian distributed about their true values with standard deviation σ. Calculate the probability that this could be a statistical fluctuation on the ionisation of a unit charged particle for the following different assumptions:
   a) $\sigma = 0.07I_0$ for all 10^5 track,
   b) For 99% of the tracks $\sigma = 0.07I_0$, while for the remainder it is $0.14I_0$.

2) An experiment is determining the decay rate λ for a new particle X, whose probability density for decay at time t is proportional to exp(-λt). A total of nine decays are observed at decay times 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 picoseconds. Calculate the likelihood function L(λ) at suitable values of λ (most easily done by a simple computer program), and draw a graph of the results. Find the best estimate of λ from the maximum of the likelihood curve, and a "±σ" range for λ by finding the values of λ where the logarithm to the base e of the likelihood function decreases by 0.5 units from its maximum value.

3) i) A tracker has detector 6 elements at x = -11, -10, -9, +9, +10 and +11 cms, which each measure a track's y-coordinate to an accuracy of ±1 cm. A straight line y = a + bx is fitted (for example by chi-squared) to the data from the 3 elements at positive x (L1); a second (L2) for the data at negative x; and a third (L3) to all 6 detector elements. The inverse covariance matrix for a and b has elements $M_{aa} = \Sigma 1/\sigma_i^2$ $M_{bb} = \Sigma x_i^2/\sigma_i^2$ $M_{ab} = \Sigma x_i/\sigma_i^2$, where the measurements are $y_i \pm \sigma_i$ at x_i. Evaluate the covariance matrix for a and b for each of the 3 fits. How do the uncertainties and correlations compare with what you expect?
   ii) When two measurements for a pair of quantities are combined optimally, the uncertainties on the combined parameters are such that $M_c = M_1 + M_2$, where M_c is the inverse covariance matrix for the combination, and M_1 and M_2 are those for the separate measurements. Determine the covariance matrix for the combination of the parameters of L1 and L2. Explain why the uncertainties for the combination are considerably smaller than those for L1 and L2 separately.

4) The coverage C(μ) is a property of a statistical technique for estimating a range for a parameter μ at a confidence level α (e.g. 68%, 90% or whatever). It is the fraction of times that, in repetitions of the procedure with different data each with its own statistical fluctuations, the estimated range contains the true value μ.
   In a Poisson counting experiment with n observed events, one method of estimating a range for the Poisson parameter μ uses the estimate n +- sqrt(n) i.e. from n - sqrt(n) to n + sqrt(n). This is supposed to have 68% coverage. Determine the actual coverage C(μ) at μ = 3.41 and 3.42 as

follows:
Determine for which measured values of n the nominal range from the "n+-sqrt(n)" procedure includes the specified true value, and then add up the Poisson probabilities for obtaining these measured values, again assuming the specified value of the Poisson parameter.
Explain why a plot of the coverage $C(\mu)$ as a function of the Poisson parameter value μ has discontinuities.
The difference in the coverage C at the two values of μ is very similar to a specific Posson probability $P_{Poisson}(n|\mu)$. What are the values of n and μ?

5) (a) Explain briefly the Bayesian and Frequentists approaches to 'probability'.
(b) Outline how Bayesians and how Frequentists would obtain 90% upper limits on the Poisson parameter μ for a counting experiment in which N events are observed.

6) An experiment is searching for a SUSY particle. With no such particle production, 100 events are expected; if the SUSY particle is produced, 110 events are expected. The experiment observes 130 events, which is 3σ above the 'No SUSY' prediction, so the p-value for the null hypothesis is 0.1%. The Lab Publicity Officer announces that we now are 99.9% certain that SUSY has been discovered.
Comment.

TAKE AWAY POINT FROM PROBLEMS:

1) EFFECT OF MISMODELLING
The distribution of a measurement is assumed to be exactly Gaussian, whereas in fact there is a 1% tail. This has an enormous effect on the result.

2) If you have never calculated a likelihood, this shows you how amazingly simple it can be. And the expected result and an approximate value of its uncertainty are easily known, so you can see whether your values are reasonable
.

3) This is a simple example of combining two measurements of the same quantities. The 2 separate estimates of the gradient have large uncertainties but the gradient for the combination is very much smaller. This problem should help you understand why.

4) People think they understand what coverage is until they see the highly structured plots for discrete data. By actually calculating the coverage for Poission data at the two specified values of the Poisson parameter, you should understand the origin of the jumps in coverage, and to get a better feeling for what coverage is.