Stats Methods Week 3 example exam questions

1. A sentence in a research paper states: *"Our model fit resulted in a chi-squared of 40 for 100 degrees of freedom."*

   What should you conclude from this? [**1 point**]

   Solution: The $\chi^2/d.o.f.$ is 0.4, which for this many degrees of freedom is highly unlikely to happen by chance *Aside: the previous sentence is good enough but if you want to fully justify it:* χ^2_{100}*-distributed variates are approximately normal with standard deviation* $\sqrt{(}200) \simeq 14$.

   Therefore it seems most likely that the errors have been overestimated in this case.

2. The posterior probability distribution of a parameter θ is a chi-squared distribution with 4 degrees of freedom. Calculate the MLE for θ. [**2 points**]

   Solution: Using the formula sheet, substituting $x = \theta$ and $\nu = 4$:

   $p(\theta|\nu = 4) = \frac{(1/2)^2}{\Gamma(2)}\theta e^{-\theta/2} = \frac{1}{8}\theta e^{-\theta/2}$

   to find the MLE we calculate the 1st derivative:
   $\frac{\mathrm{d}p(\theta|\nu=4)}{\mathrm{d}\theta} = \frac{1}{8}\left(1 - \theta/2\right) e^{-\theta/2}$

   So the MLE $\hat{\theta}$ occurs at the zero-gradient point $0 = 1 - \hat{\theta}/2 \rightarrow \hat{\theta} = 2$.

   Note: for completeness we could also calculate the 2nd derivative to show that the value corresponds to a maximum, but since we know the chi squared distribution does not contain minima we don't need to do that here to get full marks.

3. A dark matter search experiment detects 103 candidate dark matter particles in its first run. Making reasonable assumptions, what is the 2-σ lower limit on the true rate of candidate dark matter particles? [**1 point**]

   Solution: The observed rate should be Poisson distributed, with MLE equal to the observed rate λ_{obs} and variance also λ_{obs}. Due to the central limit theorem and large number of counts, we can approximate the distribution as normal, out to at least the 2-σ part of the tail of the distribution, so the 2-σ lower limit is $(103 - 2 \times \sqrt{103}) \simeq 83$ candidate dark matter particles.

4. Fig. 1 shows the log-likelihood distribution for a model parameter θ. Use the figure to estimate $\hat{\theta}$ and it's 1-sigma error, stating any assumptions you make. [**2 points**]

   Solution: We assume the MLE is normally distributed so the 1-σ error corresponds to a $\Delta L(\theta) = 0.5$. For the resulting estimate, see Fig. 2

5. You measure spectra from two objects which you fit with a blackbody (i.e. Planck) spectrum, which has two parameters (temperature and normalisation). When you fit both spectra with the same temperature (and allow the normalisations to be different) you obtain a weighted least squares (chi-squared) statistic of 212 for 200 degrees of freedom. When you also allow the temperature to be

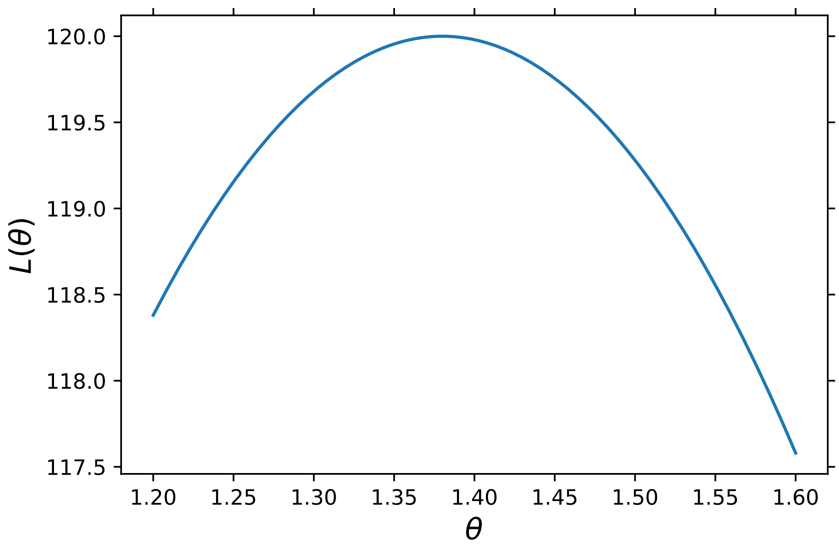

Figure 1: Log-likelihood distribution of parameter θ.

different, you obtain a weighted least squares statistic of 202. Assuming that the MLEs are normally distributed, what can you infer from this result? **[1 point]**

Solution: Improvement in chi-squared statistic is 10 for 1 additional free parameter. According to Wilks' theorem if the more constrained null hypothesis is true the expected change would be distributed as χ^2_1, for which the 1, 2, 3 sigma significances (i.e. for cdf values corresponding to those significances) are 1, 4, 9, ... i.e. for Z-sigma the value is Z^2. Thus the resulting improvement shows that the temperatures are different at $>$3-σ significance.

6. You obtain a gamma-ray spectrum of a gamma-ray burst which has 5 data points (with normally distributed errors), which you fit with a power-law model using the weighted least-squares method, to obtain a minimum of the chi-squared statistic of 14.2. What is the goodness-of-fit for this model? **[1 point]**

   Solution: The model has 2 free parameters (slope and normalisation), so we should compare the fitted chi-squared with the survival function for $5-2=3$ d.o.f. Significance levels (i.e. s.f.) values are given in the formula sheet table, with 14.156 corresponding to 3-σ for χ^2_3, so g.o.f. is approximately 0.3%, i.e. a bad fit.

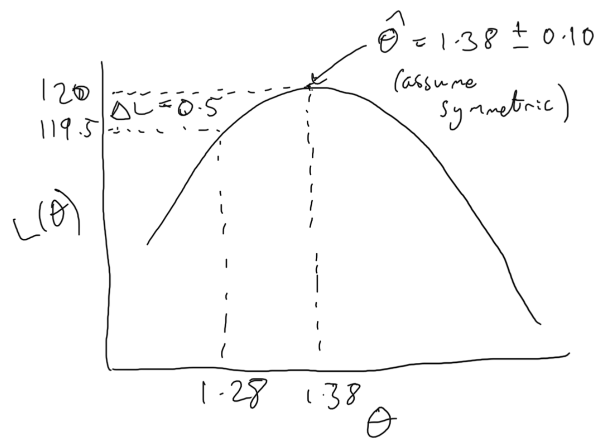

Figure 2: Sketch of log-likelihood estimate of MLE and 1-sigma errors.