Stats Methods example exam

Phil Uttley

February 2023

You have 2 hours to complete the exam (2.5 hours if you have extra time) and may use a calculator and a ruler if necessary. A total 30 points are available and the exam contributes 25% of the final course grade. In your answers, remember to show your workings and state any assumptions you make.

Note: *unless otherwise stated in the question, when you are asked to give a statistical significance you should provide this in the usual format in terms of integer units of σ, rounded down. E.g. a p-value of 0.01 would be "Better than 2-σ" while 0.003 (to 3 s.f.) could just be stated as "3-σ". Tables of these significance levels for different statistics are provided on the formula sheet.*

1. What is the probability of the following events?

   (a) two heads and one tails (in any order), from three flips of a fair coin?

   (b) at least two heads, from three flips of a fair coin?

   [2 points]

   Solution: First write the list of mutually exclusive exhaustive events:

   HHH, HHT, HTT, HTH, TTT, THT, TTH, THH

   For part (a), we see there are 3 possibilities wherein we can get two heads and one tails from 3 flips of a fair coin, and therefore the probability is 3/8. For part (b), we see there are 4 cases in which we get at least two heads from three flips of the coin, and hence the probability is 1/2.

2. You hear someone say the following: "In the Netherlands there are on average 9 days in April when it rains. That means that it will rain on a given day is 9/30, i.e. 0.3. However, it's extremely unlikely to rain on 10 consecutive days in April, because the probability of that happening is 0.3^{10}, which is less than 1 in 10^5!" Considering the statistical aspects, what is wrong with this statement? **[2 points]**

   Solution: Two main issues I can think of (partial/full credit may be given for others): 1. You need to take account of the number of permutations of 10 days in a row, out of 30 days (this is not that many, so won't make a big difference), 2. most serious is that the event that it rains on a given day is not independent from the weather on previous or following days, i.e. the probability of rain is a conditional probability and cannot be multiplied together as if independent.

3. A specially designed detector searches for a newly predicted dark matter particle over a fixed time interval. The expected background rate in that interval (i.e. due to detected events which are not the predicted particle) is 0.2 counts.

   (a) What number of counts would rule out the null hypothesis that the counts are due to background with > 3-σ significance?

   (b) You want to test between the null hypothesis of background and an alternative hypothesis which predicts that the dark matter particle rate (plus background) in the interval is 2.4 counts. If you set the required significance level at > 3-σ using the minimum counts value calculated in part (a), what is the statistical power of your hypothesis test?

   [**4 points**]

   Solution: Assume the counts are Poisson distributed so we use the Poisson pdf to calculate by hand: $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

   (a) For 3-σ, p-value for 1-cdf must be < 0.0027. For discrete variables the cdf is the probability for $F(X \leq x)$ so for the counts required to falsify the null with 3-σ significance we require the smallest x_α where $F(x_\alpha - 1) > 0.9973$. We can see that $F(x|\lambda) = e^{-\lambda} \sum_{i=0}^{x} \lambda^i/i!$, so for $\lambda = 0.2$, $F(x|\lambda = 0.2) = 0.8187 \times (1 + 0.2 + 0.04/2 + ...)$ and we have $F(1) = 0.982$, $F(2) = 0.9989$. So therefore we need $x_\alpha = 3$ or more counts to rule out the hypothesis that counts are background.

   (b) The power is 1 minus probability of a false negative, i.e. 1 minus the chance of accepting the null if the alternative is true. So we must calculate the cdf for $\lambda = 2.4$ and 2 counts (because with up to 2 counts we would accept the null). The remainder from 1 is the statistical power (1-β): $F(x|\lambda = 2.4) = 0.0907 \times (1 + 2.4 + 2.4 \times 2.4/2) = 0.57$ to 2 s.f., so the power is 0.43.

4. An X-ray telescope observes a newly discovered radio pulsar (a neutron star which shows periodic variations in its radio emission) to determine if it is also an X-ray pulsar. Fig. 1 shows the 'periodogram' of the X-ray time series measured by the telescope, which quantifies the amplitude of variability of the source (the variability 'power') as a function of the Fourier frequency. 2048 frequencies are plotted but a maximum can be seen at a single frequency which corresponds to the observed radio pulsation frequency.

   If the observed variability is only due to photon count variations and not a real signal, the power should be distributed following a χ^2 distribution with 2 degrees of freedom. Estimate the statistical significance (in terms of an integer number of σ) of the X-ray pulsation signal. [**2 points**]

   Solution: Assume power is distributed as χ^2_2. From the table of significance values, the 4-and 5-σ levels for that distribution correspond to powers of 19.33 and 28.74 respectively so the significance of the signal at that frequency is better than 4-σ. Since the radio frequency is known to be at the same frequency, no correction for number of frequencies searched over needs to be made.

5. Suppose that a compact object merger produces both a detectable gamma-ray burst and a detectable gravitational wave event 1 per cent of the time. If the probability of the merger producing a detectable gravitational wave event is 0.3, what is the probability that the merger produces a detectable gamma-ray burst, given that it has been detected as a gravitational wave event? [**1 point**]

   Solution: Define probabilities: detectable GW event $P(GW) = 0.3$, detectable GRB and GW event: $P(GRB \text{ and } GW)$, detectable GRB given there is detectable GW $P(GRB|GW)$ (which we're trying to find out). Now re-arrange addition rule from probability theory (see formula sheet):

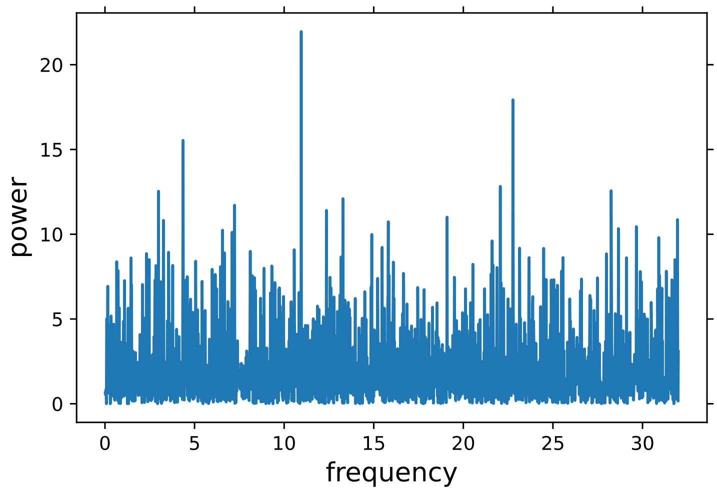

Figure 1: X-ray periodogram of a radio pulsar.

$P(GRB|GW) = \frac{P(GRB \text{ and } GW)}{P(GW)} = 0.01/0.3 = 0.0333...$ recurring

6. A joint distribution has a covariance matrix
$\begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}$
Calculate the correlation coefficient for this distribution. [**1 point**]

Solution: In the covariance matrix the variances σ_x^2, σ_y^2 are the diagonal terms and the off-diagonal is the covariance σ_{xy}, so the correlation coefficient is:

$\rho(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{2}{\sqrt{6 \times 4}} = 0.408$ to 3 s.f.

7. A random variate X is drawn from population following a normal distribution with a mean of 24 and a variance of 49. What is the probability that $X > 38$? [**1 point**]

Solution: We have a normal distribution with $\mu = 24$ and $\sigma = \sqrt{49} = 7$, so 38 is 2 standard deviations away. Thus we use the 2-sigma significance but we must halve it because we are only considering $X > 38$, i.e. $P = 0.0228$ (to 4 s.f.).

8. A continuous random variable x has the cdf shown in Fig. 2. Sketch the pdf for this distribution, including labels showing the values on the x axis where the pdf goes to zero and the y-axis value corresponding to the maximum value of the pdf. [**2 points**]

Solution: See Fig. 6

9. A future particle accelerator experiment measures the cross-section of a newly-discovered particle interaction, with the results and explanation given in Fig. 3. What is wrong with the results shown in the plot? [**1 point**]

Solution: The errors seem to be overestimated. Given the size of the error bars, the data are much too close to the model (for 1-σ we expect 1/3 of the data points to be outside the errors as well as not being so close to the centre of the error bars).

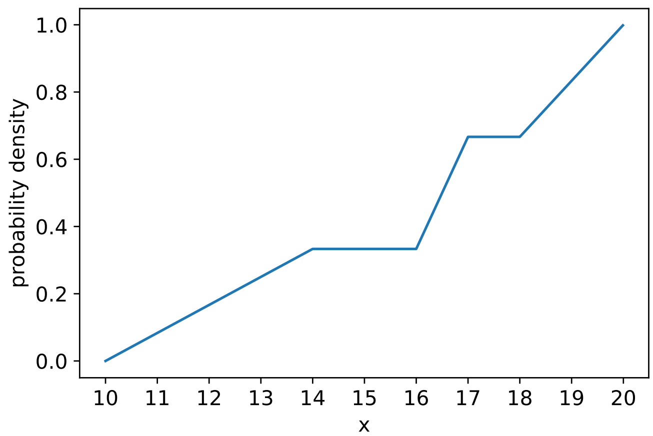

Figure 2: cdf of a variable x.

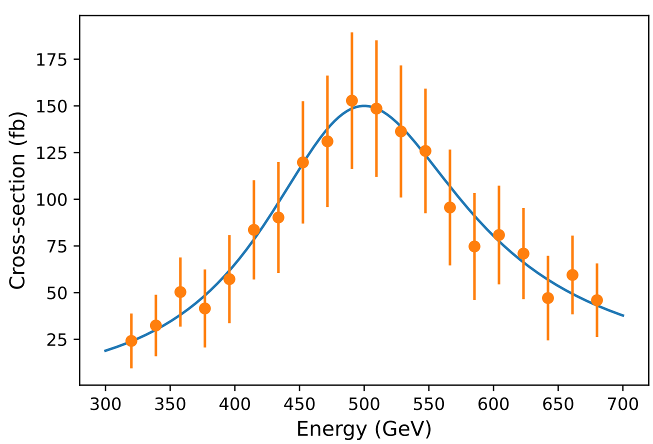

Figure 3: Interaction cross-section vs. beam energy, showing the data and best-fitting model. Errors are 1-σ.

10. A model with 12 free parameters is fitted to a spectrum with 9000 data bins, giving a best fit chi-squared (weighted least squares statistic) of 9390. By making appropriate assumptions, estimate the goodness-of-fit p-value. Is this a good fit? **[2 points]**

Solution: The fit has 8988 d.o.f., so we would need to assume a χ^2_{8988} distribution to calculate g.o.f., which we are not given. However, χ^2_ν is equivalent to ν summed χ^2_1 variables so given ν is very large we can invoke the central limit theorem. For χ^2_ν mean $\mu = E[X \sim \chi^2_\nu] = \nu$, variance $\sigma^2 = V[X \sim \chi^2_\nu] = 2\nu$, so we can assume a normal distribution with $\mu = 8988$ and $\sigma = 134$. The observed χ^2 thus exceeds the mean by 3σ. Note however that since the test is one-tailed, g.o.f. p-value for χ^2 is calculated only from the survival function (not 2 times survival function as for normal distributions), so the p-value is half the 3σ significance for a normal distribution, i.e. $< 1.4 \times 10^{-3}$. Clearly this is a bad fit!

11. A model predicts that the average mass of neutron stars in binary neutron star - black hole mergers should be 2.4 solar masses. Measurements of a sample of 5 of these events with a new gravitational wave detector give a sample mean and standard deviation of 1.9 solar masses and 0.15 solar masses respectively. How significant is this difference from the model prediction? **[1 point]**

Solution: We need to use the 1-sample t-test. From the formula sheet the t-statistic is

$T = \frac{\bar{x}-\mu}{s_x/\sqrt{n}} = \frac{1.9-2.4}{(0.15/\sqrt{5})} = -7.45$ (to 2 s.f.

We should compare with the t-distribution for 4 d.o.f. (since d.o.f for the 1-sample test is $n-1$), so we obtain that the result is significant at better than 3-σ.

12. A fairground artist plays a cup-and-ball game with visitors. The game works as follows: the artist has three cups and places a ball under one of the cups, then shuffles the cups very quickly, so it is impossible for the visitor to follow where the ball is. To win the game the visitor needs to correctly guess which cup has the ball. The game proceeds as follows:

    (a) The visitor guesses which cup has the ball under it.

    (b) The artist then reveals one of the remaining two cups which they know is empty (note: both remaining cups may be empty if the visitor guessed correctly, but then the artist randomly picks one to reveal).

    (c) Given the result of the first 'reveal', the visitor may then choose whether or not to change their guess to remaining cup which the artist did not reveal (note that the artist does not shuffle the cups again!).

The visitor randomly picks cup 1. Then the artist reveals that cup 2 is empty.

Using Bayes' theorem, calculate the probability that cup 1 contains the ball, given that the artist reveals cup 2 and state whether or not it would be better for the visitor to stick with cup 1, or switch their guess to cup 3. **[4 points]**

Solution: *Aside: this is effectively the same problem as the famous 'Monty Hall problem' in Bayesian statistics.*

Define B_1 for cup 1 having the ball, R_2 for cup 2 being revealed (and similar notation for other situations). According to Bayes' formula:

$P(B_1|R_2) = \frac{P(R_2|B_1)P(B_1)}{\sum_{i=1}^{3} P(R_2|B_i)P(B_i)} = \frac{1/2\times1/3}{1/2\times1/3+0\times1/3+1\times1/3} = \frac{1}{3}$

So sticking with cup 1 offers only a 1/3 chance of success with the complement (switching) being 2/3, thus it is better to switch!

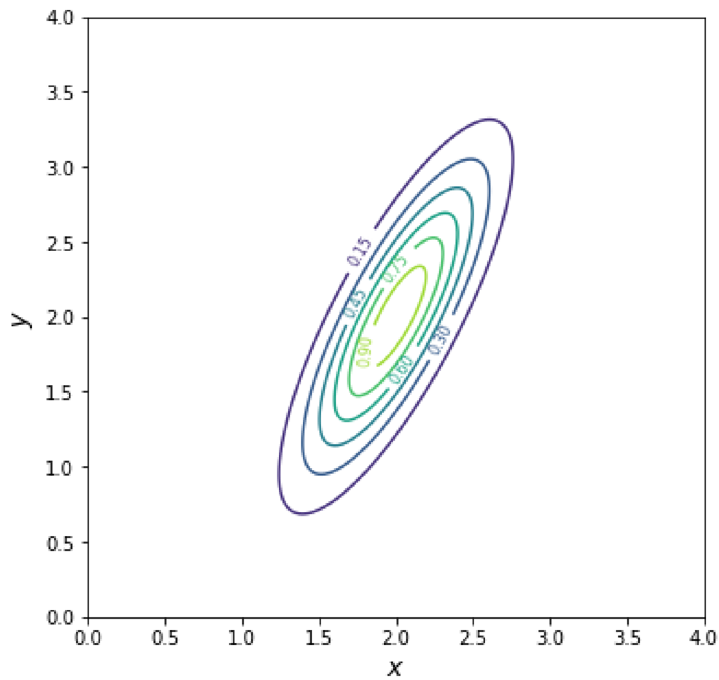

Figure 4: Joint pdf of x and y.

13. Fig. 4 shows the joint pdf of two random variables $p(x, y)$. Sketch **on the same plot** and with appropriate estimates of values on the horizontal and vertical axes, the conditional pdfs $p(y|x_0 = 2)$ and $p(x|y_0 = 1.5)$. **[2 points]**

    Solution: See Fig. 7

14. Fig. 5 shows data for the masses and diameters of a large sample of widgets, which were published in a paper. The accompanying explanation states: "The spearman ρ correlation coefficient is 0.77 for 769 data points, so that the correlation between mass and diameter is highly significant $(p, 10^{-100})$". Comment (with explanation) on whether this interpretation is correct or not. **[1 point]**

    Solution: The interpretation is incorrect because the data are clearly clustered into three groups, indicating 3 separate populations. Thus the assumption of independent and identically distributed data is clearly incorrect since the data are not drawn from a single population.

15. Five different instruments are used to measure the speed of light in vaccuum and the measured values compared to the true value. Table 1 below shows the mean difference from the true value (Δ) and the standard deviation for each experiment, calculated using samples of 10000 measurements.

    Which instrument is the most precise? Which is the most accurate? **[1 point]**

    Solution: The most precise is the one with the smallest standard deviation, i.e. experiment 5. The most accurate has the least bias, i.e. deviation from the true value, so it is experiment 2.

16. Consider a model with two parameters, θ, ϕ, which you are using to explain your data D. Given the likelihood $p(D|\phi, \theta)$ and priors for θ and ϕ, show how to calculate the marginal posterior probability distribution for ϕ. You can assume that the priors for θ and ϕ are independent of one another. **[2 points]**

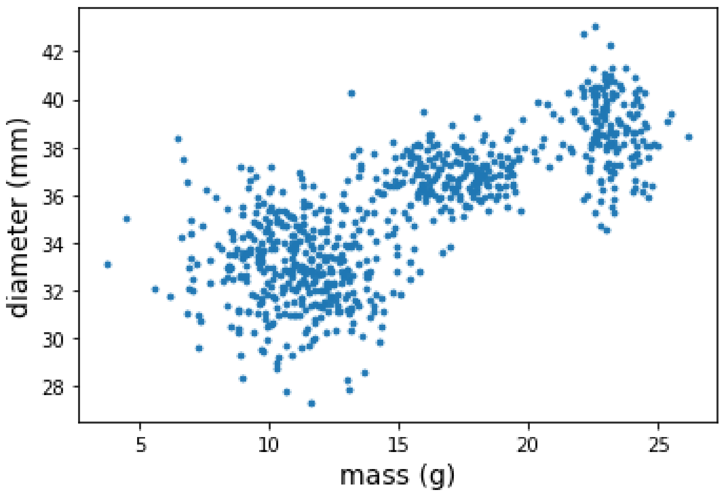

Figure 5: Masses and diameters for widgets.

Experiment	Δ (m s^{-1})	σ (m s^{-1})
1	+30	50
2	-11	42
3	+23	21
4	+14	75
5	-35	16

Table 1: Speed of light experimental results.

Solution: *A full calculation is not possible here, instead we need to show* **how** *to do the calculation given the information we already have (the likelihood and the priors). See Ep. 4 for how to work with joint pdfs, and remember to keep the variables on the same side of the conditional term.*

Start by writing Bayes' theorem in terms of the joint probability distribution for ϕ, θ:

$$p(\phi,\theta|D) = \frac{p(D|\phi,\theta)p(\phi,\theta)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(D|\phi,\theta)p(\phi,\theta)\mathrm{d}\phi\mathrm{d}\theta}$$

Then replace the joint prior with the product of $p(\phi)$, $p(\theta)$ (since the priors are independent of one another), and integrate both sides w.r.t. θ to get the final result:

$$p(\phi|D) = \int_{-\infty}^{\infty} p(\phi,\theta|D)\mathrm{d}\theta = \frac{\int_{-\infty}^{\infty} p(D|\phi,\theta)p(\phi)p(\theta)\mathrm{d}\theta}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(D|\phi,\theta)p(\phi)p(\theta)\mathrm{d}\phi\mathrm{d}\theta}$$

17. An X-ray spectrum from an accreting black hole can be well-fitted with a blackbody spectrum with a chi-squared statistic of 44 for 47 degrees of freedom. You see some weak residuals which suggest an additional power-law component. When you add the power-law component and find the best fit, the chi-squared improves to 34.8. What is the statistical significance of this improvement? **[1 point]**

    Solution: The power-law has 2 free parameters so the improvement is 9.2 for 2 additional free parameters and is distributed as χ^2_2 if the simpler model is correct. The resulting significance (from the table on the formula sheet) is better than 2-σ.

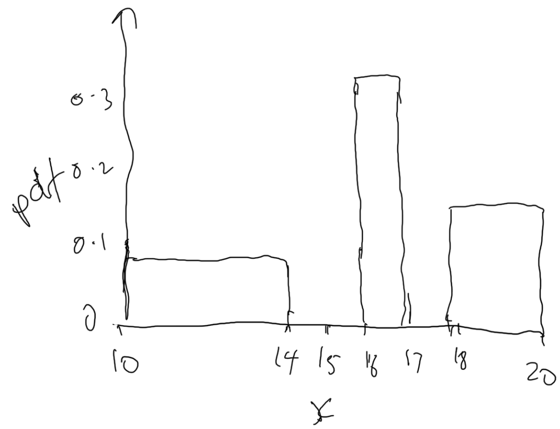

Figure 6: Question 7 solution.

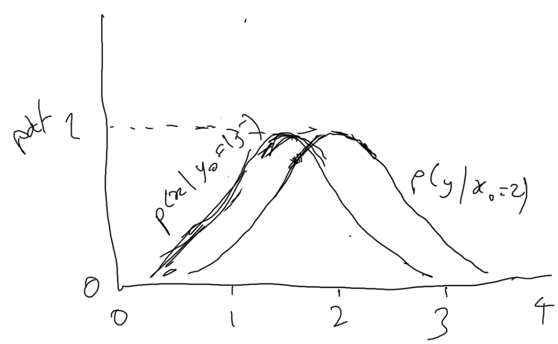

Figure 7: Question 11 solution.