Stats Methods Week 1 example exam questions

1. 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):

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

2. The estimated rate for core-collapse supernovae (CCSN) to occur in our Galaxy is 1.63 CCSN per 100 years. A neutrino detector can detect the neutrino pulse produced by CCSN in our Galaxy. What is the probability that the detector will detect one **or more** CCSN in our Galaxy in 20 years of operation? [**2 points**]

   Solution: Since CCSN are independent randomly occuring events we can assume that the number of Galactic CCSN in a given time interval is a Poisson distributed variable, with rate in 20 years $\lambda = 1.63/5 = 0.326$. We use the Poisson distribution:

   $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

   The probability of getting one or more is $1 - p(0|\lambda = 0.326) = 1 - e^{-0.326} = 0.278$ (to 3 s.f.)

3. A continuous random variable x has the pdf shown in Fig. 1. Sketch the cdf for this distribution, including labels showing the values on the x and y axis at the beginning and end of each scale and the x and y values where inflection points (i.e. significant changes in gradient) in the function are seen. [**2 points**]

   Solution: See Fig. 2

4. A particular theory predicts that a neutrino detector should detect neutrinos from a distant collider experiment in fixed time intervals at a rate of 21.35 neutrinos per interval. From 4000 intervals you measure a mean rate of 21.01 neutrinos per interval. Stating your assumptions and any calculations, perform a significance test and determine whether the theory is ruled out at the 5-σ significance level. [**3 points**]

   Solution: For a constant underlying rate of $\lambda = 21.35$ the observed rate in each interval should be Poisson distributed, with population mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$. We can calculate the Z-statistic (see formula sheet) for the sample mean $\bar{x} = 21.01$ and size $n = 4000$:

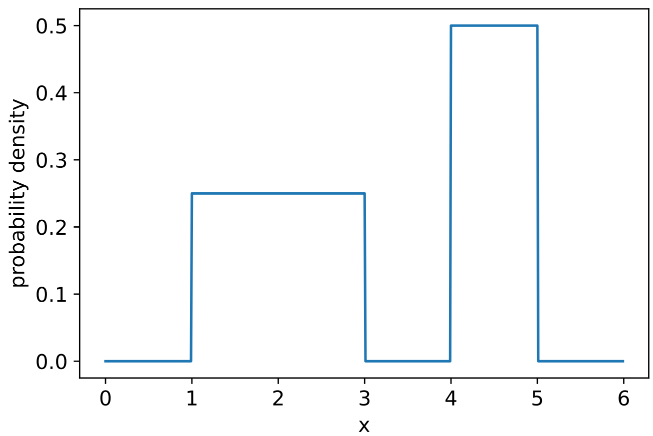

Figure 1: pdf for variable x.

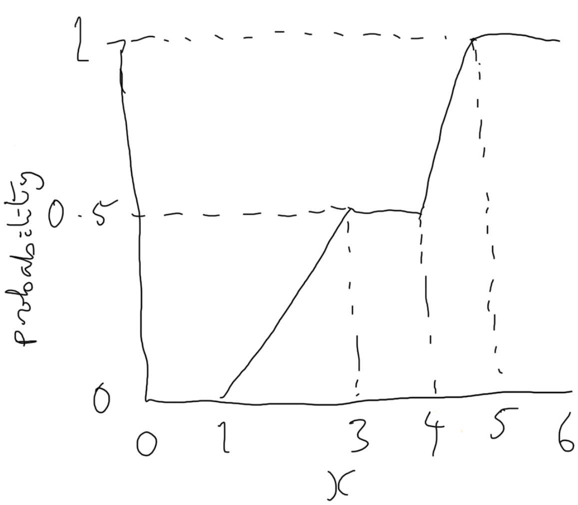

Figure 2: cdf sketch for the given pdf.

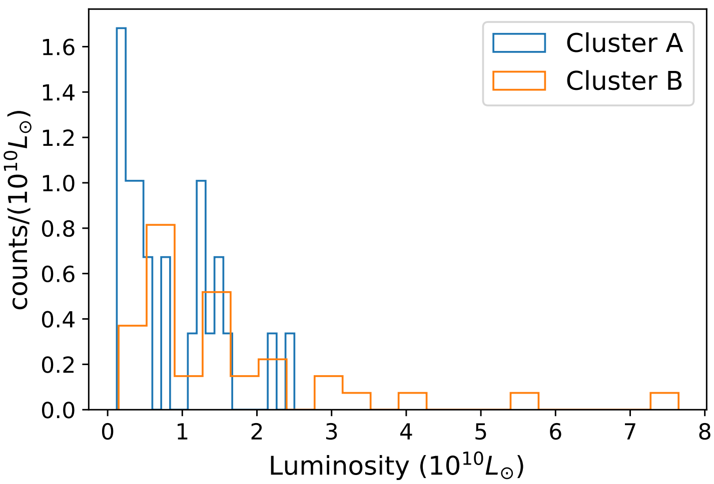

Figure 3: Galaxy cluster luminosity distributions.

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{21.01-21.35}{\sqrt{21.35/4000}} = -4.65$ to 3 s.f.

Assuming that for this count rate (>20 counts per bin!) and 4000 intervals, the central limit holds out to the tails of the distribution and the sample mean is normally distributed around the true mean, the Z-statistic should be distributed as a standard normal. The observed deviation is large but still <5-σ below the theory prediction. Therefore the theory is **not** ruled out at the 5-σ significance level.

5. A research group publishes some results which include the following figure (see Fig. 3) and statement: *"The figure shows histograms of the measured galaxy luminosities for galaxy cluster A and galaxy cluster B. A 2-sample t-test shows that the mean luminosities of each cluster are significantly different at the 3-σ level."*

   Why is a t-test the wrong approach here?" [**1 point**]

   Solution: A t-test assumes that the sample mean is normally distributed, which requires either that the samples are large (hundreds of measurements) or the data are normally distributed. Neither is clearly the case here.