PHYS 8150 - ASTR 8150
Statistics & Probabilities
Problem set 1

Problem 1: Exponential distribution (characterizing a distribution, analytic work)

a. Give the CDF of the exponential distribution.

b. Give the mode of the exponential distribution, as well as its mean and its variance.

c. Daisy the Physicist knows that t follows an exponential distribution with parameter λ, and got experimental data with elapsed times $t_1, t_2, \ldots t_n$. What is the likelihood of λ ? The log-likelihood ?

Problem 2: Supernova monitoring (requires solving problem 1a, applying PDF/CDF)

Supernovae are rare stellar explosions, so rare that the last one to occur in our galaxy was observed by Kepler in 1604. Astronomers consequently have to monitor other galaxies to have a chance detecting supernovae. It is estimated they need to observe around 70 galaxies for a year in order to witness one supernova.

a. What is the probability of observing 2 or more supernovae in a given year while monitoring 70 galaxies ?

b. Last year, Donald the astronomer saw exactly two supernova events. How likely was it that these events were less than a month apart if he was monitoring 70 galaxies ? 700 galaxies ?

Problem 3: Supernova goes boom (Bayes theorem)

Minnie the astronomer thinks she has found that the star Beeblebrox-5 has 40% chance to transform in a supernova today ! She points her X-ray detector toward it. If the star goes supernova, she will have to wait t minutes before detecting X-rays, where t has an exponential distribution with $1/\lambda = 20$ minutes. If the star does not go supernova, no X-ray will be detected.

a. Find the conditional probability that the star did not go supernova given that no X-ray has been detected after an hour.

b. How long will Minnie have to wait without detecting X-ray to be 90% sure the star did no go supernova ?

Problem 4: The hound of the Mickeyvilles (Bayesian model selection)

a. Two people have left traces of their own blood at the scene of a crime. The blood groups of the two traces are found to be of type O (a common type in the local population, having frequency 60%) and of type AB (a rare type, with frequency 1%). A suspect, Pluto, is tested and found to have type O blood. We also know that based on other evidence (including his schedule) there is a only 70% chance he could be present at the crime scene. Do the data (the blood types found at the scene) and the background information on his schedule give evidence for the proposition that Pluto was one of the two people whose blood was found at the scene ?

b. Daisy the physicist is attempting to detect a new type of exotic particle. After careful study she is now convinced that this detector can detect these particles with probability p during each experiment, with all possible values of p equiprobable. After repeating the experiment N times independently, she finds she detected the particle M times. What is the probability distribution for p as a function of N and M (this is known as the beta distribution) ? What is the probability that the next try will result in another failure ? Her colleague Goofy is certain he knows that $p = .3$. Based on data that found $M = 2$ for $N = 5$, can you tell who is likely right, Daisy or Goofy ? Same question for $N = 50$ and $M = 20$, $N = 50$ and $M = 15$, and $N = 500$ and $M = 200$. What is the interpretation ?

Problem 5: **Nonsense sensors (maximum likelihood, maximum a posteriori)**

Dr. McDuck has two photo-sensors with fixed integration time that he wants to use to study planets around a star.

a. He is pointing the first sensor toward the star, where he's expecting around 1000 counts, and the second toward the much fainter planet, expecting 5 counts. He actually measures 1010 counts in the first sensors, and 2 counts in the second. What was the probability of such an event (assuming photon shot noise) ?

b. He's now pointing both sensors toward the yet another star, i.e. he should be getting the same value on both detectors, yet he measures 1150 and 1270 due to noise. Express the log-likelihood, what is the most probable value of the flux according to maximum likelihood ?

c. He's now got a new type of sensor. It is a single-pixel sensor more adapted to planet work, but the detected flux is subject to strong read noise, i.e. it follows a Gaussian distribution with standard deviation of about 2 flux units. Pointing to the planet, he measures a flux value of 7. He also thinks higher flux values are quite improbable, with the probability $\Pr(\text{flux} = f) \propto e^{-f^2}$. According to MAP, what is the most probable value of the planet flux ?

References

Poisson and exponential distributions

In a Poisson process, events occur continuously and independently at a constant average rate λ. The Poisson probability mass function is given by:

$$\Pr(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

While the number of Poisson events happening during a given lapse of time is modeled by the Poisson distribution, the probability distribution that describes the amount of time t elapsed between two events is modeled by the exponential distribution. The exponential PDF for the random variable t is:

$$\Pr(t|\lambda) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

Integration by parts

$$\int u\, dv = uv - \int v\, du$$

Stirling's formula

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

Beta integral

$$\int_0^1 t^a (1-t)^b dt = \frac{a!b!}{(a+b+1)!}$$