PHYS 8150 - ASTR 8150
Solutions to Problem Set 1

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

a. *Give the CDF of the exponential distribution.*

$$\Pr(T \leq t|\lambda) = \begin{cases} \int_0^t \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x}\right]_0^t = 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}. \tag{1}$$

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

The mode is the location of the maximum of the PDF, so 0. Now we use the integration by part rule from the References.

The mean is $\mu = E[t] = \int_0^{+\infty} \lambda t e^{-\lambda t}\, dt = \left[te^{-\lambda t}\right]_0^{+\infty} - \int_0^{+\infty} e^{-\lambda t}\, dt = \frac{1}{\lambda}$ by posing $u = t$ and $v = e^{\lambda t}$.

The variance is $\sigma^2 = E[t^2] - \mu^2 = \int_0^{+\infty} \lambda t^2 e^{-\lambda t}\, dt - \mu^2 = \left[t^2 e^{-\lambda t}\right]_0^{+\infty} - \int_0^{+\infty} 2te^{-\lambda t}\, dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ by posing $u = t^2$ and $v = e^{\lambda t}$.

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 ?*

If the data D are not independent, we cannot provide an answer. If they are, then:

$$\mathcal{L}(\lambda|D) = \Pr(D|\lambda) = \prod_{i=1}^{n} \Pr(t_i|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i},$$

and:

$$\log \mathcal{L}(\lambda|D) = n \log \lambda - \lambda \sum_{i=1}^{n} t_i.$$

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 ?*

The number of events per unit of time is a random variable governed by Poisson law, with a rate of $\lambda = 1$ per unit of time here (1 supernova per 70 galaxies per year).

$$\Pr(n \geq 2) = 1 - \Pr(n = 1) - \Pr(n = 0) = 1 - \frac{1^0}{0!}e^{-\lambda} - \frac{1^1}{1!}e^{-\lambda} = 1 - 2e^{-\lambda}$$

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 ?*

Here the CDF of the exponential distribution is necessary. One month is t = 1/12 year, and we still have $\lambda = 1$ for 70 galaxies:

$$\Pr(t \leq \frac{1}{12}) = 1 - e^{-\lambda t} = 1 - e^{-1/12} \simeq 0.08$$

For 700 galaxies, we have 10 times the rate, so $\lambda = 10$.

$$\Pr(t \leq \frac{1}{12}) = 1 - e^{-\lambda t} = 1 - e^{-10/12} \simeq 0.57.$$

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.*

   Let pose S="star went supernova" and X="X-rays were detected after an hour". We are looking for $\Pr(\bar{S}|\bar{X})$. We have the prior $\Pr(S) = 0.4$, and so $\Pr(\bar{S}) = 0.6$. Also, if there is no supernova, we won't detect X-rays, and $\Pr(\bar{X}|\bar{S}) = 1$. Finally, using the exponential distribution, $\Pr(\bar{X}|S) = e^{-60/20} = e^{-3} \simeq 0.05$ after 60 minutes. Using Bayes theorem, we find the solution:

$$\Pr(\bar{S}|\bar{X}) = \frac{\Pr(\bar{X}|\bar{S})\Pr(\bar{S})}{\Pr(\bar{X})} = \frac{\Pr(\bar{X}|\bar{S})\Pr(\bar{S})}{\Pr(\bar{X}|\bar{S})\Pr(\bar{S}) + \Pr(\bar{X}|S)\Pr(S)} = \frac{1 \times 0.6}{1 \times 0.6 + 0.05 \times 0.4} \simeq 0.968.$$

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

   We want to solve for the time were the probability $\Pr(\bar{S}|\bar{X}) \geq 0.9$.

$$\begin{aligned} \Pr(\bar{S}|\bar{X}) &\geq 0.9 \\ \frac{1 \times 0.6}{1 \times 0.6 + e^{-t/20} \times 0.4} &\geq 0.9 \\ t &> 20 \log 6 = 35.8 \text{ minutes.} \end{aligned}$$

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 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 ?*

   Let's pose X ="Pluto was one of the two people whose blood was found". Then $\bar{X}$ ="the blood of two unknown people was found". D is the data "O and AB types were found". To "give evidence for" has a particular meaning we saw in class, in this case we search the ratio:

$$\frac{\Pr(X|D)}{\Pr(\bar{X}|D)} = \frac{\Pr(D|X)\Pr(X)}{\Pr(D|\bar{X})\Pr(\bar{X})}$$

   We already know $\Pr(X) = 0.7$ and so $\Pr(\bar{X}) = 1 - 0.7 = 0.3$.

   If X is true, the AB blood came from another person, and $\Pr(D|X) = \Pr(AB) = 0.01$.

   If $\bar{X}$ is true, both blood types came from unknown persons, $\Pr(D|\bar{X}) = 2 \times \Pr(AB) \times \Pr(O) = 0.012$, where the factor 2 comes from the two ways we can pick the AB and O persons (the first person is type O and the second is type AB, and the first is type AB and the second type O). We see at this stage that the likelihoods support $\bar{X}$ as slightly more probable than X. The end result is $\frac{\Pr(X|D)}{\Pr(\bar{X}|D)} = \frac{0.01 \times 0.7}{0.012 \times 0.3} \simeq 1.94$. According to the table in our slides, this corresponds to only anecdotal evidence for Pluto's presence at the scene. Note that the same problem, but with Pluto being of AB type, would result in much stronger evidence for his presence.

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) ?*

   First we note that in all this problem, N is not a random variable, but instead is set by the experimenter. We also note that the prior information on p is that it follows a uniform distribution on the interval $[0, 1]$ and is

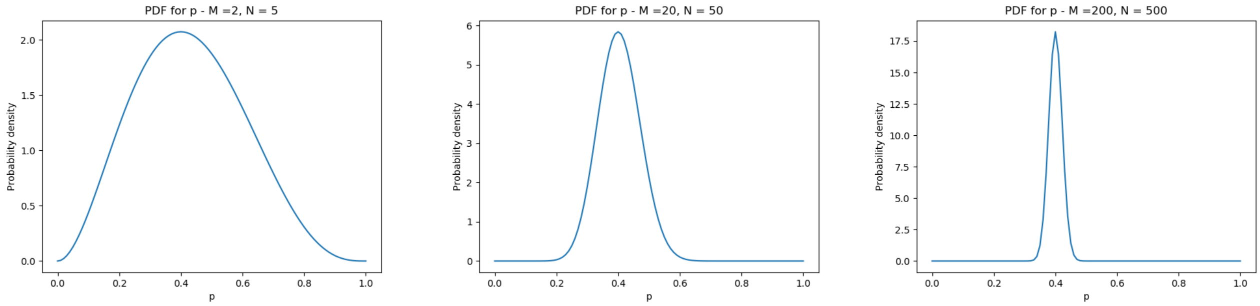

Figure 1: The probability density function for p.

independent of N, so $\Pr(p|N) = \Pr(p) = \text{constant} = 1$. We further note that $\Pr(M|p, N)$ follows a binomial distribution, i.e. :

$$\Pr(M|p, N) = \frac{N!}{M!(N-M)!} p^M (1-p)^{N-M}$$

Using Bayes' theorem:

$$\Pr(p|M, N) = \frac{\Pr(M|p, N)\Pr(p|N)}{\int_0^1 \Pr(M|p, N)\Pr(p|N)\,dp} = \frac{\cancel{\frac{N!}{M!(N-M)!}} p^M (1-p)^{N-M}}{\int_0^1 \cancel{\frac{N!}{M!(N-M)!}} p^M (1-p)^{N-M}\,dp} = \frac{p^M (1-p)^{N-M}}{\int_0^1 p^M (1-p)^{N-M}\,dp}$$

Then using the integral provided in the references, $\int_0^1 p^M (1-p)^{N-M}\,dp = \frac{M!(N-M)!}{(N+1)!}$, and:

$$\Pr(p|M, N) = \frac{(N+1)!}{M!(N-M)!} p^M (1-p)^{N-M}$$

The figure above demonstrate how the probability distribution behaves for $N = 5, M = 2$, then $N = 50, M = 20$, and finally $N = 500, M = 200$, which all would seem to entail that $p = 0.4$ but with various degrees of confidence.

What is the probability that the next try will result in another failure ?

The question is maybe a bit tricky. Let's first underline that we are not looking for the probability of getting M successes given $N+1$ trials, without any other knowledge. Instead we are given the background information that (*"After repeating the experiment N times, she finds she detected the particle M times."*). If we were using binomial with known p, the probability that the next try would result in a failure would be $1-p$, which follows the probability distribution with PDF equal to $1 - \Pr(p|M, N)$. In fact, stopping there was a valid answer. But we can go one step beyond for extra bonus points. What is the most probable probability of failure for the next try ? To answer, we can look for the mode or the expected value of $\Pr(p|M, N)$. The mode is the easiest since the value that maximizes $p^N(1-p)^{N-M}$ is the same that maximizes the binomial distribution, i.e. $p = \frac{M}{N}$, and so the probability of failure is $1 - \frac{M}{N}$. If you chose to use the expected value:

$$\begin{aligned} E\left[\Pr(p|M, N)\right] &= \int_0^1 \frac{(N+1)!}{M!(N-M)!} p^M (1-p)^{N-M} \times p\,dp = \frac{(N+1)!}{M!(N-M)!} \int_0^1 p^{M+1}(1-p)^{N-M}\,dp \\ &= \frac{(N+1)!}{M!\cancel{(N-M)!}} \times \frac{\cancel{(N-M)!}(M+1)!}{(N+2)!} = \frac{M+1}{N+2} \end{aligned}$$

And in this case, we have $1 - \frac{M+1}{N+2}$.

Her colleague Goofy is certain he knows that $p = 0.3$. 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 ?

First let's express the probability of a current situation for Daisy, who is using the beta distribution. For M

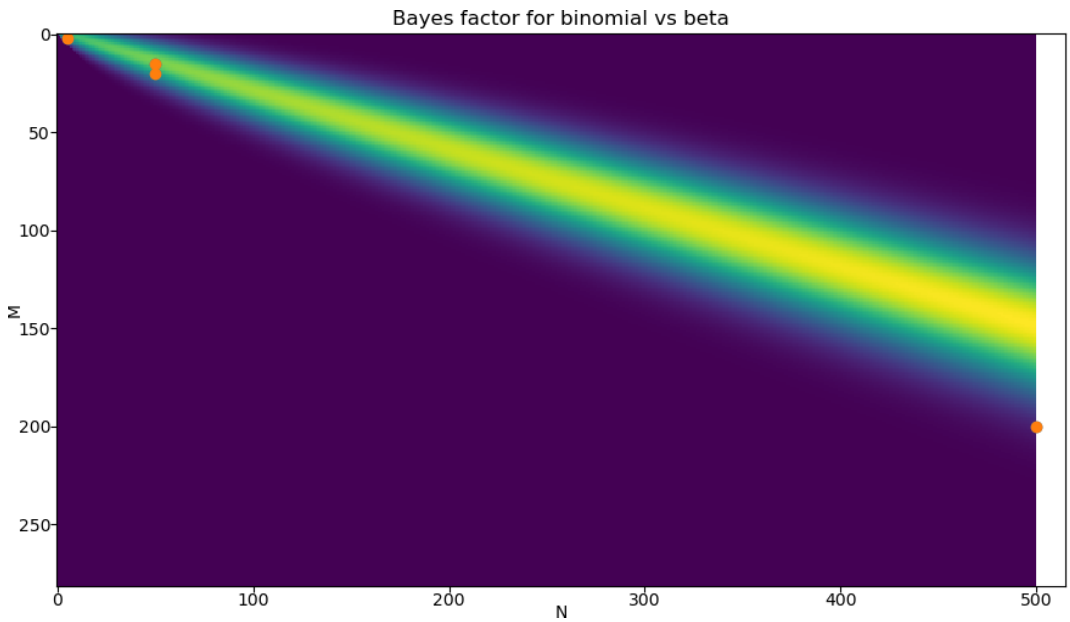

Figure 2: Bayes factor K for different values of N and M. Orange points represents (N, M) pairs in the problem.

successes given N trials, assuming Daisy's distribution:

$$\Pr(M|N) = \int_0^1 \Pr(M|p, N) \Pr(p|N)\, dp = \frac{N!}{M!(N-M)!} \times \int_0^1 p^M (1-p)^{N-M}\, dp$$
$$= \frac{N!}{\cancel{M!(N-M)!}} \times \frac{\cancel{M!(N-M)!}}{(N+1)!} = \frac{1}{N+1}.$$

This doesn't depend on M (since draws are independent) nor p (since we do not know it). In addition we know that M can take $N+1$ values $(0, 1, \ldots N)$. What we found is that M is uniformly distributed if we do not know p.

Meanwhile, Goofy is using the binomial distribution with $p = 0.3$. Which hypothesis H_{binom} or H_{beta} is the best is determined by model selection, given the data (i.e. M, since N is constant for a given dataset):

$$K = \frac{\Pr(H_{\text{binom}}|M, N)}{\Pr(H_{\text{beta}}|M, N)} = \frac{\Pr(M|H_{\text{binom}}, N)}{\Pr(M|H_{\text{beta}}, N)} \frac{\Pr(H_{\text{binom}})}{\Pr(H_{\text{beta}})} = \frac{\Pr(M|H_{\text{binom}}, N)}{\Pr(M|H_{\text{beta}}, N)} = \frac{\frac{N!}{M!(N-M)!} 0.3^M 0.7^{N-M}}{1/(N+1)}$$

where we chose $\Pr(H_{\text{beta}}) = \Pr(H_{\text{binom}})$ since we don't have an *a priori* reason to prefer one of the distributions. For $M = 2$, $N = 5$, $K \simeq 1.85$ so the binomial distribution is more probable but by a marginal amount. For $M = 20$, $N = 50$, $K \simeq 1.89$, still marginally more probable. For $M = 15$ and $N = 50$ this would rise to $K \simeq 6.2$, giving more evidence for $p = 0.3$ as expected since $15 = 50 \times 0.3$. Conversely, if we get $M = 200$, $N = 500$, $K \simeq 0.0002$, there is now enough data to make the possibility of $p = 0.3$ improbable compared to alternatives (such as, in this case, $p = 0.2$).

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. 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.

a. *What was the probability of such an event (assuming photon shot noise) ?*

We have $\lambda_1 = 1000$ per unit of time (integration time) for the first star, $\lambda_2 = 5$ for the second, $n_1 = 1010$ and $n_2 = 2$. Since the events are independent:

$$\Pr(n_1, n_2|\lambda_1, \lambda_2) = \Pr(n_1|\lambda_1) \times \Pr(n_2|\lambda_2)$$

We can easily calculate $\Pr(n_2|\lambda_2) = \frac{5^2}{2!}e^{-5} \simeq 0.084$. We can use Stirling's formula from the References to compute $\Pr(n_1|\lambda_1) = \left(\frac{\lambda_1}{n_1}\right)^{n_1} \frac{e^{n_1-\lambda_1}}{\sqrt{2\pi n_1}} \simeq 0.012$. This shows that Stirling's formula is handy to compute Poisson probabilities even for large numbers. The final answer is $\Pr(n_1, n_2|\lambda_1, \lambda_2) \simeq 0.001$.

b. *He's now pointing both sensors toward the exact same spot at the center of the star, i.e. he's expecting 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 ?*

The log-likelihood is:

$$\mathcal{L}(\lambda|n) = \log \Pr(n|\lambda) = n \log \lambda - \lambda - \log n! \simeq n \log \lambda - n \log n + n - \lambda - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log n$$

So for two data points:

$$\mathcal{L}(\lambda|n_1, n_2) = \log \Pr(n_1|\lambda) + \log \Pr(n_2|\lambda) = (n_1+n_2)\log\lambda - n_1 \log n_1 - n_2 \log n_2 + n_1 + n_2 - 2\lambda - \log(2\pi) - \frac{1}{2}(\log n_1 + \log n_2)$$

Looking at the expression, we can recognize the characteristic $\sum n \log n$ of entropy. The most probable value of the flux is the estimated rate from the two data points, i.e. the rate that maximizes the likelihood:

$$\frac{\partial \mathcal{L}(\lambda|n_1, n_2)}{\partial \lambda} = \frac{n1 + n2}{\lambda} - 2 = 0$$

and so $\lambda = (n_1 + n_2)/2$, the average.

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* $\Pr(\textit{flux} = f) \propto e^{-f^2}$. *According to MAP, what is the most probable value of the planet flux ?*

This is a regularization problem. The data are the measurement $m = 7$ and the uncertainty/standard deviation of the noise $\sigma = 2$. The likelihood is:

$$\mathcal{L}(f|m) = \Pr(m|f) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(f-m)^2}{2\sigma^2}}$$

The MAP solution is:

$$\hat{f} = \underset{f\in\mathbb{R}^+}{\text{argmax}} \left[\frac{\Pr(m|f)\Pr(f)}{\Pr(m,\sigma)}\right] \propto \underset{f\in\mathbb{R}^+}{\text{argmax}} \left[\mathcal{L}(f|m)\Pr(f)\right] = \underset{f\in\mathbb{R}^+}{\text{argmin}} \left[-\log \Pr(m|f) - \log \Pr(f)\right]$$
$$\propto \underset{f\in\mathbb{R}^+}{\text{argmin}} \left[\frac{(f-7)^2}{8} + f^2\right]$$

To solve for the mimimum we set the gradient to zero:

$$\frac{\partial(9f^2 - 14f + 49)}{\partial f} = 0 \implies f = \frac{7}{9}$$

The most probable value for the planet flux given the data and prior knowledge is $\hat{f} = \frac{14}{18}$, which is quite far from the measured value ($f = 7$). Why ? Because the chosen prior is very strongly penalizing high fluxes. In fact we may be over-regularizing. To prevent this, we saw the Tikhonov regularization has an adjustable scalar weight λ, so that $\Pr(\text{flux} = f) \propto e^{-\lambda f^2}$, and we would find $\hat{f} = \frac{14}{2+16\lambda}$. As seen in class, the L-curve can be used to select the best λ value.