Section 8 – Solutions

Review

- **Law of Total Probability (partition based on value of a r.v.):** If X is a discrete random variable, then

$$\mathbb{P}(A) = \sum_{x \in \Omega_X} \mathbb{P}(A|X = x)p_X(x)$$

If X is a continuous random variable, then

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x)f_X(x)\ dx$$

- **Conditional Expectation**: Let X and Y be random variables. Then, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y) \qquad X \text{ discrete}$$

and for any event A,

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|A) \qquad X \text{ discrete}$$

Note that linearity of expectation still applies to conditional expectation: $\mathbb{E}[X + Y|A] = \mathbb{E}[X|A] + \mathbb{E}[Y|A]$

- **Law of Total Expectation (Event Version)**: Let X be a random variable, and let events $A_1, ..., A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

- **Law of Total Expectation (RV Version)**: Suppose X and Y are random variables. Then,

$$\mathbb{E}[X] = \sum_{y} \mathbb{E}[X|Y = y]p_Y(y) \qquad Y \text{ discrete r.v..}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y]f_Y(y)dy \qquad Y \text{ continuous r.v.}$$

Maximum Likelihood Estimation

1) **Realization/Sample**: A realization/sample x of a random variable X is the value that is actually observed.

2) **Likelihood**: Let $x_1, \ldots x_n$ be iid realizations from probability mass function $p_X(\mathsf{x}\ ; \theta)$ (if X discrete) or density $f_X(\mathsf{x}\ ; \theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data.

   If X is discrete:

$$L(x_1, \ldots, x_n\ ;\ \theta) = \prod_{i=1}^{n} p_X(x_i\ ;\ \theta)$$

   If X is continuous:

$$L(x_1, \ldots, x_n\ ;\ \theta) = \prod_{i=1}^{n} f_X(x_i\ ;\ \theta)$$

3) Maximum Likelihood Estimator (MLE): We denote the MLE of θ as $\hat{\theta}_{\text{MLE}}$ or simply $\hat{\theta}$, the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(x_1, \ldots, x_n \ ; \ \theta) = \arg\max_{\theta} \ln L(x_1, \ldots, x_n \ ; \ \theta)$$

4) Log-Likelihood: We define the log-likelihood as the natural logarithm of the likelihood function. Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

If X is discrete:

$$\ln L(x_1, \ldots, x_n \ ; \ \theta) = \sum_{i=1}^{n} \ln p_X(x_i \ ; \ \theta)$$

If X is continuous:

$$\ln L(x_1, \ldots, x_n \ ; \ \theta) = \sum_{i=1}^{n} \ln f_X(x_i \ ; \ \theta)$$

5) Steps to find the maximum likelihood estimator, $\hat{\theta}$:

(a) Find the likelihood and log-likelihood of the data.

(b) Take the derivative of the log-likelihood and set it to 0 to find a candidate for the MLE, $\hat{\theta}$.

(c) Take the second derivative and show that $\hat{\theta}$ indeed is a maximizer, that $\frac{\partial^2 L}{\partial \theta^2} < 0$ at $\hat{\theta}$. Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

(d) If we are finding the MLE for a set of parameters, then we set up the system of equations obtained by taking the partial derivative of the log-likelihood function with respect to each of the parameters and setting it equal to 0. We then solve this system to get the MLEs. (And again, second order conditions need to be checked.)

6) An estimator $\hat{\theta}$ for a parameter θ of a probability distribution is **unbiased** iff $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$

Task 1 – Content Review

a) True or False: The Log-Likelihood gives a slightly different estimate, but because it is close enough and easier to compute we use it for our estimate of θ.

False: Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

b) True or False: $\hat{\theta}$ is the true parameter and θ is our estimate.

False: It is the other way around. Remember to switch to $\hat{\theta}$ when you set your equation to zero!

c) True or False: An estimator is unbiased if Bias$(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta = 0$ or equivalently $\mathbb{E}[\hat{\theta}] = \theta$

True by definition of

d) You flip a coin 10 times and observe HHHTHHTHHH (8 heads, 2 tails). What is the MLE of θ, where θ is the true probability of seeing tails?

◯ $\hat{\theta} = .2$

◯ $\hat{\theta} = .25$

◯ $\hat{\theta} = .8$

◯ $\hat{\theta} = .3$

Option 1: $\hat{\theta} = .2$

Task 2 – Mystery Dish!

A fancy new restaurant has opened up that features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability θ, dish C with probability 2θ, and dish D with probability $0.5 - 3\theta$. Each diner is served a dish independently. Let x_A be the number of people who received dish A, x_B the number of people who received dish B, etc, where $x_A + x_B + x_C + x_D = n$. Find the MLE $\hat{\theta}$ for θ.

The data tells us, for each diner in the restaurant, what their dish was. We begin by computing the likelihood of seeing the given data given our parameter θ. Because each diner is assigned a dish independently, the likelihood is equal to the product over diners of the chance they got the particular dish they got, which gives us:

$$\mathcal{L}(x;\theta) = \prod_{i=1}^{n} P(x_i;\theta) = 0.5^{x_A}\theta^{x_B}(2\theta)^{x_C}(0.5-3\theta)^{x_D}$$

From there, we just use the MLE process to get the log-likelihood, take the first derivative, set it equal to 0, and solve for $\hat{\theta}$.

$$\ln \mathcal{L}(x;\theta) = x_A \ln(0.5) + x_B \ln(\theta) + x_C \ln(2\theta) + x_D \ln(0.5 - 3\theta)$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(x;\theta) = \frac{x_B}{\theta} + \frac{x_C}{\theta} - \frac{3x_D}{0.5 - 3\theta}$$

$$\frac{x_B}{\hat{\theta}} + \frac{x_C}{\hat{\theta}} - \frac{3x_D}{0.5 - 3\hat{\theta}} = 0$$

Solving yields $\hat{\theta} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$.

Task 3 – A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. In other words, they follow the distributions $\mathbb{P}(k;\theta) = \theta^k e^{-\theta}/k!$, where $k \in \mathbb{N}$ and $\theta > 0$ is a positive real number.

Find the MLE of θ.

We follow the recipe given in class:
First define likelihood. Then compute the natural log of that likelihood. Then get the derivative of

the log. Finally set it to 0 and solve for $\hat{\theta}$.

$$
\begin{aligned}
\mathcal{L}(x_1, \ldots, x_n \;;\; \theta) &= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} \\
\ln \mathcal{L}(x_1, \ldots, x_n \;;\; \theta) &= \sum_{i=1}^{n} \left[-\theta - \ln(x_i!) + x_i \ln(\theta)\right] \\
\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(x_1, \ldots, x_n \;;\; \theta) &= \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\theta}\right] \\
0 &= \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\hat{\theta}}\right] \\
0 &= -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\theta}} \\
\hat{\theta} &= \frac{\sum_{i=1}^{n} x_i}{n}
\end{aligned}
$$

Task 4 – A biased estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2)\right)$$

where $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

By the definition of an unbiased estimator, $\hat{\theta}_2$ is an unbiased estimator of σ^2 iff $\mathbb{E}[\hat{\theta}_2] = \sigma^2$. Let $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\mathbb{E}\left[\hat{\theta}_2\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right]$$

which by linearity of expectation (and distributing the sum) is

$$
\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\frac{2}{n} \overline{X} \sum_{i=1}^{n} X_i\right] + \mathbb{E}\left[\overline{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - 2\mathbb{E}\left[\overline{X} \frac{1}{n} \sum_{i=1}^{n} X_i\right] + \mathbb{E}\left[\overline{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - 2\mathbb{E}\left[\overline{X}^2\right] + \mathbb{E}\left[\overline{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\overline{X}^2\right] . \qquad (**)
\end{aligned}
$$

We know that for any random variable Y, since $\mathrm{Var}(Y) = \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2$ it holds that

$$\mathbb{E}\left[Y^2\right] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2.$$

Also, we have $\mathbb{E}\left[X_i\right] = \mu,$ Var $(X_i) = \sigma^2\ \forall i$ and $\mathbb{E}\left[\overline{X}\right] = \mu,$ Var $\left(\overline{X}\right) = \frac{\sigma^2}{n}$. Combining these facts, we get

$$\mathbb{E}\left[X_i^2\right] = \sigma^2 + \mu^2 \quad \forall i \qquad \text{and} \qquad \mathbb{E}\left[\overline{X}^2\right] = \frac{\sigma^2}{n} + \mu^2.$$

Substituting these equations into (**) we get

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2)\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\overline{X}^2\right] = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) \\ &= \left(1 - \frac{1}{n}\right)\sigma^2.\end{aligned}$$

Thus $\hat{\theta}_2$ is not unbiased.

Task 5 – Weather Forecast

A weather forecaster predicts sun with probability θ_1, clouds with probability $\theta_2 - \theta_1$, rain with probability $\frac{1}{2}$ and snow with probability $\frac{1}{2} - \theta_2$. This year, there have been 55 sunny days, 100 cloudy days, 160 rainy days and 50 snowy days. What is the maximum likelihood estimator for θ_1 and θ_2?

We want to find the likelihood of the data samples given the parameters θ_1 and θ_2. To do this, we take the following product over all the data points.

$$\mathcal{L}(x_1, ..., x_{365}; \theta_1, \theta_2) = \theta_1^{55}(\theta_2 - \theta_1)^{100}\left(\frac{1}{2}\right)^{160}\left(\frac{1}{2} - \theta_2\right)^{50}$$

Then, we use this to determine the log likelihood.

$$\begin{aligned}\ln \mathcal{L}(x_1, ..., x_{365}; \theta_1, \theta_2) &= \ln(\theta_1^{55}(\theta_2 - \theta_1)^{100}\left(\frac{1}{2}\right)^{160}\left(\frac{1}{2} - \theta_2\right)^{50}) \\ &= \ln \theta_1^{55} + \ln(\theta_2 - \theta_1)^{100} + \ln\left(\frac{1}{2}\right)^{160} + \ln\left(\frac{1}{2} - \theta_2\right)^{50} \\ &= 55\ln\theta_1 + 100\ln(\theta_2 - \theta_1) + 160\ln\left(\frac{1}{2}\right) + 50\ln\left(\frac{1}{2} - \theta_2\right)\end{aligned}$$

Then, we take the derivative of the log likelihood with respect to θ_1.

$$\frac{\partial}{\partial\theta_1}\ln\mathcal{L}(x_1, ..., x_{365}; \theta_1, \theta_2) = \frac{55}{\theta_1} - \frac{100}{\theta_2 - \theta_1}$$

Setting this equal to 0, we solve for $\hat{\theta_1}$:

$$\frac{55}{\hat{\theta_1}} - \frac{100}{\hat{\theta_2} - \hat{\theta_1}} = 0$$

$$55(\hat{\theta_2} - \hat{\theta_1}) - 100\ \hat{\theta_1} = 0$$

$$55\ \hat{\theta_2} = 155\ \hat{\theta_1}$$

$$\hat{\theta_1} = \frac{11}{31}\ \hat{\theta_2}$$

Then, we take the derivative of the log likelihood with respect to θ_2.

$$\frac{\partial}{\partial \theta_2} \ln \mathcal{L}(x_1, ..., x_{365}; \theta_1, \theta_2) = \frac{100}{\theta_2 - \theta_1} - \frac{50}{\frac{1}{2} - \theta_2}$$

Setting this equal to 0, we solve for $\hat{\theta_2}$:

$$\frac{100}{\hat{\theta_2} - \hat{\theta_1}} - \frac{50}{\frac{1}{2} - \hat{\theta_2}} = 0$$

$$100\left(\frac{1}{2} - \hat{\theta_2}\right) - 50\,(\hat{\theta_2} - \hat{\theta_1}) = 0$$

$$50 - 150\,\hat{\theta_2} + 50\,\hat{\theta_1} = 0$$

$$\hat{\theta_2} = \frac{\hat{\theta_1} + 1}{3}$$

We can now solve the simultaneous equations we have for θ_1 and θ_2 to obtain the maximum likelihood estimators for each parameter.

$$\hat{\theta_2} = \frac{\hat{\theta_1} + 1}{3}$$

Plugging in the equation for θ_1, we find

$$\hat{\theta_2} = \frac{\frac{11}{31}\,\hat{\theta_2} + 1}{3}$$

$$3\,\hat{\theta_2} = \frac{11}{31}\,\hat{\theta_2} + 1$$

$$93\,\hat{\theta_2} = 11\,\hat{\theta_2} + 31$$

$$\hat{\theta_2} = \frac{31}{82}$$

Plugging in the value for θ_2 into the equation for θ_1,

$$\hat{\theta_1} = \frac{11}{31} \cdot \frac{31}{82} = \frac{11}{82}$$

To confirm that this is in fact a maximum, we could do a second derivative test. We won't ask you do this for this multivariate case, but it would still be good to check!

Task 6 – Elections

Individuals in a certain country are voting in an election between 3 candidates: A, B and C. Suppose that each person makes their choice independent of others and votes for candidate A with probability θ_1, for candidate B with probability θ_2 and for candidate C with probability $1 - \theta_1 - \theta_2$. (Thus, $0 \leqslant \theta_1 + \theta_2 \leqslant 1$.) The parameters θ_1, θ_2 are unknown.

Let n_A, n_B, and n_C be the number of votes for candidate A, B, and C, respectively. What are the maximum likelihood estimates for θ_1 and θ_2 in terms of n_A, n_B, and n_C?

(You don't need to check second order conditions.)

The likelihood is

$$\mathcal{L}(n_A, n_B, n_C\,; \theta_1, \theta_2) = (\theta_1)^{n_A}(\theta_2)^{n_B}(1-\theta_1-\theta_2)^{n_C}\ .$$

Therefore, the log-likelihood is

$$\ln \mathcal{L}(n_A, n_B, n_C\,; \theta_1, \theta_2) = n_A \ln(\theta_1) + n_B \ln(\theta_2) + n_C \ln(1-\theta_1-\theta_2)\ .$$

We take the partial derivative of the log-likelihood with respect to θ_1 and θ_2, separately

$$\frac{\partial}{\partial \theta_1} \ln \mathcal{L}(n_A, n_B, n_C\,; \theta_1, \theta_2) = \frac{n_A}{\theta_1} - \frac{n_C}{1-\theta_1-\theta_2}$$
$$\frac{\partial}{\partial \theta_2} \ln \mathcal{L}(n_A, n_B, n_C\,; \theta_1, \theta_2) = \frac{n_B}{\theta_2} - \frac{n_C}{1-\theta_1-\theta_2}$$

Now, we set both partial derivatives to 0 and solve (here we replace θ_1 and θ_2 with $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively)

We now want to find the solutions $\hat{\theta}_1$ and $\hat{\theta}_2$ for the system of equations

$$\frac{n_A}{\hat{\theta}_1} - \frac{n_C}{1-\hat{\theta}_1-\hat{\theta}_2} = 0 \tag{1}$$
$$\frac{n_B}{\hat{\theta}_2} - \frac{n_C}{1-\hat{\theta}_1-\hat{\theta}_2} = 0 \tag{2}$$

We can use algebra to solve for $\hat{\theta}_1$ and $\hat{\theta}_2$: We observe that the second term is the same in both equations (1) and (2) is the same so we obtain

$$\frac{n_A}{\hat{\theta}_1} = \frac{n_B}{\hat{\theta}_2}$$

or equivalently $n_A \cdot \hat{\theta}_2 = n_B \cdot \hat{\theta}_1$. We also can multiply out the denominators in equation (1) to get

$$n_C \cdot \hat{\theta}_1 = n_A \cdot (1-\hat{\theta}_1-\hat{\theta}_2) = n_A(1-\hat{\theta}_1) - n_A \cdot \hat{\theta}_2 = n_A \cdot (1-\hat{\theta}_1) - n_B \cdot \hat{\theta}_1.$$

and rearrange the equation to get that

$$(n_A + n_B + n_C) \cdot \hat{\theta}_1 = n_A.$$
$$\hat{\theta}_1 = \frac{n_A}{n_A + n_B + n_C}$$

Furthermore $\hat{\theta}_2 = \frac{n_B}{n_A} \cdot \hat{\theta}_1$ so

$$(n_A + n_B + n_C) \cdot \hat{\theta}_2 = (n_A + n_B + n_C)\hat{\theta}_1 \frac{n_B}{n_A} = n_A \cdot \frac{n_B}{n_A} = n_B.$$

Therefore

$$\boxed{\hat{\theta}_1 = \frac{n_A}{n_A + n_B + n_C}}$$
$$\boxed{\hat{\theta}_2 = \frac{n_B}{n_A + n_B + n_C}}$$

Note, the likelihood expression used has no binomial/multinomial term since the samples are in a particular order. Even if the samples didn't have a specified order and a binomial term was included, it would disappear when taking the derivative of the log likelihood.

Task 7 – Continuous Law of Total Probability

Suppose that the time until server 1 crashes is $X \sim \text{Exp}(\lambda)$ and the time until server 2 crashes is independent, with $Y \sim \text{Exp}(\mu)$.
What is the probability that server 1 crashes before server 2?

We have $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$, and want to find $\text{P}(X < Y)$. Let A be the event that $X < Y$. By the Law of Total Probability for Continuous Variables, using A as the event and Y as the continuous random variable, we have

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|Y = y) f_Y(y)\ dy$$

As $Y \sim \text{Exp}(\mu)$, $f_Y(y) = 0$ for all $y < 0$, so the integral is only nonzero over the range from 0 to ∞, so we have

$$\mathbb{P}(A) = \int_{0}^{\infty} \mathbb{P}(A|Y = y) f_Y(y)\ dy$$

As A is the event that $X < Y$, we know

$$\mathbb{P}(A|Y = y) = \mathbb{P}(X < Y|Y = y) = \mathbb{P}(X < y|Y = y) = \mathbb{P}(X < y)$$

Note the last equality above is true since X,Y are independent so $P(X < y|Y = y) = P(X < y)$ for any y. Also note $\mathbb{P}(X < y) = \mathbb{P}(X \leqslant y)$ as X is continuous. So,

$$\mathbb{P}(A) = \int_{0}^{\infty} \mathbb{P}(A|Y = y) f_Y(y)\ dy = \int_{0}^{\infty} \mathbb{P}(X \leqslant y) f_Y(y)\ dy$$

Here, as $X \sim \text{Exp}(\lambda)$, we can plug in the CDF of $\text{Exp}(\lambda)$ to get $\mathbb{P}(X \leqslant y)$, and as $Y \sim \text{Exp}(\mu)$, we can plug in the PDF of $\text{Exp}(\mu)$ to get $f_Y(y)$ as follows:

$$\mathbb{P}(A) = \int_{0}^{\infty} \mathbb{P}(X \leqslant y) f_Y(y)\ dy = \int_{0}^{\infty} (1 - e^{-\lambda y}) f_Y(y)\ dy = \int_{0}^{\infty} (1 - e^{-\lambda y}) \mu e^{-\mu y}\ dy$$

We can simply evaluate the integral to finish the problem as follows:

$$\mathbb{P}(A) = \int_{0}^{\infty} (1 - e^{-\lambda y}) \mu e^{-\mu y}\ dy = \int_{0}^{\infty} \mu e^{-\mu y}\ dy + \int_{0}^{\infty} (-e^{-\lambda y}) \mu e^{-\mu y}\ dy$$

Let's now evaluate each of the above integrals:

$$\int_{0}^{\infty} \mu e^{-\mu y}\ dy = [-e^{-\mu y}]_0^\infty = 0 - (-1) = 1$$

$$\int_{0}^{\infty} (-e^{-\lambda y}) \mu e^{-\mu y}\ dy = -\mu \int_{0}^{\infty} (e^{-(\lambda+\mu)y})\ dy = \frac{-\mu}{\lambda+\mu}[-e^{-(\lambda+\mu)y}]_0^\infty = \frac{-\mu}{\lambda+\mu}(0 - (-1)) = \frac{-\mu}{\lambda+\mu}$$

So, we can combine these results to get:

$$\mathbb{P}(A) = \int_{0}^{\infty} \mu e^{-\mu y}\ dy + \int_{0}^{\infty} (-e^{-\lambda y}) \mu e^{-\mu y}\ dy = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}$$

So, as A is the event that $X < Y$, our final answer is that the probability that server 1 crashes before server 2 is $\frac{\lambda}{\lambda+\mu}$.

Task 8 – Y Me?

Let $y_1, y_2, ...y_n$ be i.i.d. samples of a random variable with density function

$$f_Y(y\,;\,\theta) = \frac{1}{2\theta}\exp\left(-\frac{|y|}{\theta}\right)\,.$$

Find the MLE for θ in terms of $|y_i|$ and n.

Since the samples are i.i.d., the likelihood of seeing n samples of them is just their PDFs multiplied together. From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for for $\hat{\theta}$.

$$\begin{aligned}
L(y_1,\ldots,y_n\,;\,\theta) &= \prod_{i=1}^{n}\frac{1}{2\theta}\exp\left(-\frac{|y_i|}{\theta}\right)\\
\ln L\left(y_1,\ldots,y_n \mid \theta\right) &= \sum_{i=1}^{n}\left[-\ln 2-\ln\theta-\frac{|y_i|}{\theta}\right]\\
\frac{\partial}{\partial\theta}\ln L(y_1,\ldots,y_n\,;\,\theta) &= \sum_{i=1}^{n}\left[-\frac{1}{\theta}+\frac{|y_i|}{\theta^2}\right]\\
\sum_{i=1}^{n}\left[-\frac{1}{\hat{\theta}}+\frac{|y_i|}{\hat{\theta}^2}\right] &= 0\\
-\frac{n}{\hat{\theta}}+\frac{\Sigma_{i=1}^{n}|y_i|}{\hat{\theta}^2} &= 0\\
\hat{\theta} = \frac{\Sigma_{i=1}^{n}|y_i|}{n}
\end{aligned}$$

Task 9 – Elevator rides

[This is the problem we did in class.] The number X of people who enter an elevator on the ground floor is a Poisson random variable with mean 10. If there are N floors above the ground floor, and if each person is equally likely to get off at any one of the N floors, independently of where others get off, compute the expected number of stops the elevator will make before discharging all the passengers.

Let S be the number of stops the elevator makes, and $X \sim Poi(10)$. We shall calculate $E[S]$.
By the law of total expectation, partitioning on the value of X, we have

$$E[S] = \sum_{i=0}^{\infty} E[S|X=i]P(X=i)$$

By the definition of Poison distribution, we know

$$P(X=i) = e^{-10}\frac{10^i}{i!}$$

To calculate $E[S|X=i]$, let $S = Y_1 + Y_2 + ... + Y_N$, where

$$Y_j = \begin{cases} 1 & \text{if someone gets off at the } j^{th} \text{ floor} \\ 0 & \text{otherwise} \end{cases}$$

Then, by the linearity of conditional expectation, we have

$$E[S|X=i] = E[Y_1+Y_2+...+Y_N|X=i] = \sum_{j=1}^{N} E[Y_j|X=i] = \sum_{j=1}^{N}\mathbb{P}\left(Y_j=1|X=i\right)$$

To figure out $\mathbb{P}\left(Y_j = 1|X = i\right)$, it would be more convenient to find its complement, $\mathbb{P}\left(Y_j = 0|X = i\right)$, which represents the probability that nobody gets off at j^{th} floor. Since each person is equally likely to get off at any one of N floor, we know $\mathbb{P}\left(Y_j = 0|X = i\right) = (\frac{N-1}{N})^i$. Thus, we have

$$E[S|X = i] = \sum_{j=1}^{N} \mathbb{P}\left(Y_j = 1|X = i\right) = \sum_{j=1}^{N} 1 - \mathbb{P}\left(Y_j = 0|X = i\right) = \sum_{j=1}^{N} 1 - (\frac{N-1}{N})^i$$

Finally, we find

$$E[S] = \sum_{i=0}^{\infty} E[S|X = i]P(X = i) = \sum_{i=0}^{\infty} \left(\sum_{j=1}^{N} 1 - (\frac{N-1}{N})^i \right) e^{-10} \frac{10^i}{i!}$$