# Assignment 6

Benjamin Jakubowski

November 2, 2015

## 1. Uniform distribution

### a. Possible values of $u$

Let $X_1, X_2, ..., X_n \sim U(0, u)$. Then, given a realization $x_1, x_2, ..., x_n$, the possible values of the parameter u are

$$u \in [x_{max}, \infty)$$

where $x_{max} = \max\{x_1, x_2, ..., x_n\}$.

### b. Maximum-likelihood estimator $\hat{U}_{ML}$

We now find the maximum-likelihood estimator $\hat{U}_{ML}$. First, the likelihood function is defined as follows: for $u \in [x_{max}, \infty)$,

$$\begin{aligned}\mathcal{L}_{\vec{x}}(u) &= \prod_{i=1}^{n} P_{X_i}(x_i, u) \\ &= \prod_{i=1}^{n} \left(\frac{1}{u-0}\right) = \prod_{i=1}^{n} \left(\frac{1}{u}\right) \\ &= \left(\frac{1}{u}\right)^n\end{aligned}$$

And, for $u \in (-\infty, x_{max}), \mathcal{L}_{\vec{x}}(u) = 0$.
Now note

$$\mathcal{L}'_{\vec{x}}(u) = (-n)\frac{1}{u}^{n+1} < 0$$

for all $u \in [x_{max}, \infty)$, so $\mathcal{L}_{\vec{x}}$ is monotonically decreasing on this interval. Therefore $\mathcal{L}_{\vec{x}}(u)$ is maximized by $\hat{U}_{ML} = x_{max}$.

### c. PDF of $\hat{U}_{ML}$

We want to find the PDF of $\hat{U}_{ML} = x_{max}$. We'll first find the CDF, then differentiate to find the PDF.

$$\begin{aligned}
F_{\hat{U}_{ML}}(x) &= P(\hat{U}_{ML} < x) \\
&= P(x_{max} < x) \\
&= P\left(\cap_{i=1}^{n} (x_i < x)\right) \\
&= \prod_{i=1}^{n} P(x_i < x) = \prod_{i=1}^{n} \left(\frac{x-0}{u-0}\right) \qquad \text{(since the } X_i \text{ are IID)} \\
&= \left(\frac{x}{u}\right)^n \\
&= \left(\frac{1}{u}\right)^n x^n
\end{aligned}$$

So

$$f_{\hat{U}_{ML}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{\hat{U}_{ML}}(x) = n\left(\frac{1}{u}\right)^n x^{n-1}$$

### d. $\hat{U}_{ML}$ is biased

We now show $\hat{U}_{ML}$ is biased:

$$\begin{aligned}
E(\hat{U}_{ML}) &= \int_0^u x \cdot n \left(\frac{1}{u}\right)^n x^{n-1} \mathrm{d}x \\
&= n\left(\frac{1}{u}\right)^n \int_0^u x^n \mathrm{d}x \\
&= n\left(\frac{1}{u}\right)^n \left[\frac{1}{n+1} x^{n+1}\right]_{x=0}^{x=u} \\
&= n\left(\frac{1}{u}\right)^n \frac{1}{n+1} u^{n+1} \\
&= \frac{n}{n+1} \cdot u
\end{aligned}$$

Since $E(\hat{U}_{ML}) = \frac{n}{n+1} \cdot u \neq u$, $\hat{U}_{ML}$ is biased.

### e. $\hat{U}_{ML}$ converges to u in probability

We now show $\hat{U}_{ML}$ converges in probability to $u$. Let $\epsilon > 0$ be given. Then

$$\begin{aligned}
P\left(\left|\hat{U}_{ML} - u\right| > \epsilon\right) &= 1 - P(-\epsilon \leq \hat{U}_{ML} - u \leq \epsilon) \\
&= 1 - P(u - \epsilon \leq \hat{U}_{ML} \leq u + \epsilon) \\
&= 1 - \left(P(u - \epsilon \leq \hat{U}_{ML} \leq u) + P(u < \hat{U}_{ML} \leq u + \epsilon)\right)
\end{aligned}$$

But, since $\hat{U}_{ML} = x_{max} \leq u$ we know $P(u < \hat{U}_{ML} \leq u + \epsilon) = 0$. Thus,

$$1 - \Big(P(u - \epsilon \leq \hat{U}_{ML} \leq u) + P(u < \hat{U}_{ML} \leq u + \epsilon)\Big) = 1 - P(u - \epsilon \leq \hat{U}_{ML} \leq u)$$

Then,

$$\begin{aligned}
1 - P(u - \epsilon \leq \hat{U}_{ML} \leq u) &= 1 - P(u - \epsilon \leq x_{max} \leq u) \\
&= 1 - (1 - P(x_{max} < u - \epsilon)) \\
&= P(x_{max} < u - \epsilon) \\
&= P\left(\cap_{i=1}^{n} (x_i < u - \epsilon)\right) \\
&= \prod_{i=1}^{n} P(x_i < u - \epsilon) \\
&= \prod_{i=1}^{n} \frac{u - \epsilon}{u} \\
&= \left(\frac{u - \epsilon}{u}\right)^n
\end{aligned}$$

So, to recap: we have shown (given $\epsilon > 0$),

$$P\left(\left|\hat{U}_{ML} - u\right| > \epsilon\right) = \left(\frac{u - \epsilon}{u}\right)^n$$

Therefore,

$$\lim_{n \to \infty} P\left(\left|\hat{U}_{ML} - u\right| > \epsilon\right) = \lim_{n \to \infty} \left(\frac{u - \epsilon}{u}\right)^n = 0$$

So $\hat{U}_{ML}$ converges in probability to $u$.

## 2. Half life

### a. Parameter of exponential distribution modeling nuclear decay

Let $h_l$ be the half-life of an element with decay wait time modeled by $T \sim \text{Exp}(\lambda)$. Then

$$\begin{aligned}
P(T < h_l) &= 1/2 \\
\int_0^{h_l} \lambda e^{-\lambda t} \mathrm{d}t &= 1/2 \\
1 - e^{-\lambda h_l} &= 1/2 \\
e^{-\lambda h_l} &= 1/2 \\
-\lambda h_l &= \ln(1/2) = -\ln(2) \\
\lambda &= \frac{\ln(2)}{h_l}
\end{aligned}$$

Using this equation, the parameter of the distribution for each isotope is calculated below:

| **Element (isotope)** | **Parameter($\lambda$)** |
|---|---|
| Carbon-10 ($C^{10}$) | $\lambda_{C^{10}} = \frac{\ln(2)}{19.29} \approx 0.0359$ |
| Carbon-15 ($C^{15}$) | $\lambda_{C^{15}} = \frac{\ln(2)}{2.45} \approx 0.283$ |
| Seaborgium-266 ($Sg^{266}$) | $\lambda_{Sg^{266}} = \frac{\ln(2)}{30} \approx 0.0231$ |

### B. MLE for exponential $\lambda$

First we find the general MLE for an exponential $\lambda$. Then we consider the question of restricting the set of possible values, or considering $\lambda \in [0, \infty)$.

Let $t_1, t_2, ..., t_n$ be the realization of the IID vector $T_1, T_2, ..., T_n \sim \text{Exp}(\lambda)$. Then

$$\begin{aligned}\mathcal{L}_{\vec{t}}(\lambda) &= \prod_{i=1}^{n} f_{T_i}(t_i, \lambda) \\ &= \prod_{i=1}^{n} \left(\lambda e^{-\lambda t_i}\right) \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}\end{aligned}$$

Now we consider

$$\begin{aligned}\log \mathcal{L}_{\vec{t}}(\lambda) &= \log \left[\lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}\right] \\ &= n \cdot \log(\lambda) - \lambda \sum_{i=1}^{n} t_i\end{aligned}$$

Then $\frac{\mathrm{d}}{\mathrm{d}\lambda} \log \mathcal{L}_{\vec{t}}(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i$. Setting this to zero, we find

$$\begin{aligned}\frac{n}{\lambda} - \sum_{i=1}^{n} t_i &= 0 \\ \lambda &= \frac{n}{\sum_{i=1}^{n} t_i}\end{aligned}$$

Now, let's confirm this is a maximum:

$$\frac{\mathrm{d}^2}{\mathrm{d}^2\lambda} \log \mathcal{L}_{\vec{t}}(\lambda) = -n \cdot \lambda^{-2}$$

Since the second derivative is negative for all $n \geq 1, \lambda \in (0, \infty)$, this is indeed a maximum.

Now that we've found the general MLE $\hat{\lambda}$, we consider the question of maximizing over a restricted set versus all non-negative reals. In the context of this problem, it would make more sense to consider only $\lambda \in \{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\}$. Our goal is to identify which of the three isotopes is most likely given the observed decay times. In a Bayesian interpretation, this means our prior $P_\Lambda(\lambda) = 0$ for all $\lambda \notin \{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\}$.

## C. Posterior distribution

First, note our prior distribution is

$$P_\Lambda(\lambda) = \begin{cases} P(\lambda_{C^{10}}) \\ P(\lambda_{C^{15}}) \\ P(\lambda_{Sg^{266}}) \end{cases}$$

The likelihood function is

$$P_{\vec{T}|\Lambda}(\vec{t}|\lambda) = \mathcal{L}_{\vec{t}}(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

So, for $\lambda_I$ (the parameter for a given isotope)

$$P_{\Lambda|\vec{T}}(\lambda_I|\vec{t}) = \frac{P_\Lambda(\lambda_I) \cdot {\lambda_I}^n e^{-\lambda_i \sum_{i=1}^n t_i}}{\sum_{\lambda \in \{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\}} \left[P_\Lambda(\lambda) \cdot \lambda^n e^{-\lambda \sum_{i=1}^n t_i}\right]}$$

## D. Point estimate for the parameter

It does not make sense to use the posterior mean as a point estimate of the parameter. The reason is that we are trying to identify the most likely isotope given the data (and our prior beliefs). The posterior mean would likely be an intermediate value- generally

$$E(\Lambda|\vec{T}) \notin \{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\}$$

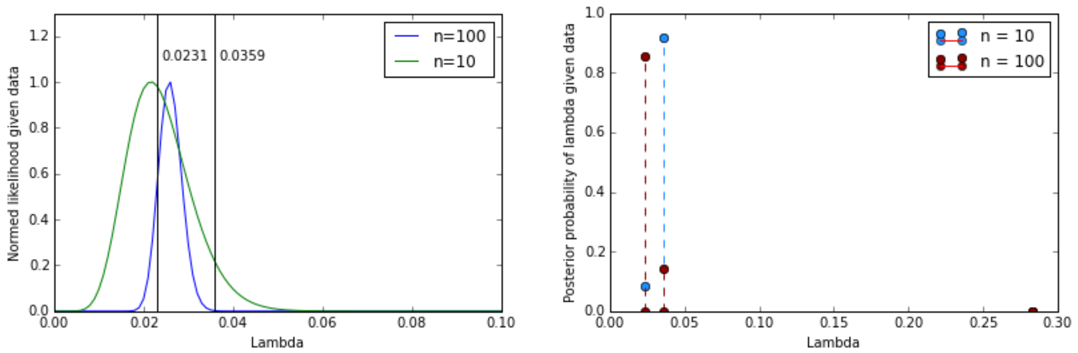
Instead, our point estimate should just be

$$\hat{\lambda} = \max \left\{ P_{\Lambda|\vec{T}}(\lambda|\vec{t}) \middle| \lambda \in \{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\} \right\}$$

Note this estimate is called the maximum a posteriori (MAP) estimate (a pedagogical digression- I didn't know what MAP estimates were until this problem, and as a former teacher I found this a really neat way to learn about them).

## E. Plotting the posterior distribution

To inform discussion, two plots are submitted below (while contrary to instructions, I find it necessary to submit these plots to fully explain the results):



First, consider the left plot (showing the normed likelihood function). The likelihood function for $n = 10$ is less concentrated compared to the likelihood function for $n = 100$ (which makes sense, given the MLE converges to $\lambda$ as $n \to \infty$). As such, the likelihood of $\lambda_{C^{10}} = 0.0359$ is much greater when $n = 10$ than when $n = 100$ (even though both functions have similar modes).

Moreover (now considering $n = 10$), even though the likelihood of $\lambda_{C^{10}} = 0.0359$ is still lower than the likelihood $\lambda_{Sg^{266}} = 0.0231$, the dramatic difference in priors (recall $P_\Lambda(\lambda_{C^{10}}) = 0.49$, while $P_\Lambda(\lambda_{Sg^{266}}) = 0.01$) means the posterior probability of $\lambda_{C^{10}}$ is much greater.

In conclusion, even though the MLE (maximizing over the set $\{\lambda_{C^{10}}, \lambda_{C^{15}}, \lambda_{Sg^{266}}\}$) $\hat{\lambda}_{MLE} = \lambda_{Sg^{266}}$ is the same for $n = 10, n = 100$, the MAP estimate $\hat{\lambda}_{MP}$ changes with $n$.

## 3. Empirical probability mass function

### a. Nonparametric estimator for the PMF of A (age)

Given we are trying to estimate the distribution of ages in years, a reasonable nonparametric estimator is

$$\hat{P}_A(a) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i = a}$$

where $a$ is an age in years.

### B. $\hat{P}_A(a)$ IS UNBIASED

To show $\hat{P}_A(a)$ is unbiased, we must show $E(\hat{P}_A(a)) = P_A(a)$ for all $a$. Well,

$$\begin{aligned} E(\hat{P}_A(a)) &= E\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{x_i=a}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left(\mathbb{1}_{x_i=a}\right) \end{aligned}$$

But, for all $i \in \{1, 2, ..., n\}, E(\mathbb{1}_{x_i=a}) = P_{X_i}(a)$ Now, assume the $X_i$ are IID, and follow the underlying distribution of the population (i.e. $P_{X_i}(a) = P_A(a)$). Then

$$\begin{aligned} E(\hat{P}_A(a)) &= \frac{1}{n}\sum_{i=1}^{n} E\left(\mathbb{1}_{x_i=a}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} P_A(a) \\ &= \frac{1}{n} n P_A(a) \\ &= P_A(a) \end{aligned}$$

So $\hat{P}_A(a)$ is unbiased (under the previously articulated assumptions).

### C. $\hat{P}_A(a)$ IS CONSISTENT

First, recall $\hat{P}_A(a)$ is consistent if $\lim_{n\to\infty} E\left((\hat{P}_A(a) - P_A(a))^2\right) = 0$. Now note

$$\begin{aligned} E\left((\hat{P}_A(a) - P_A(a))^2\right) &= E\left(\hat{P}_A(a)^2 - 2P_A(a)\hat{P}_A(a) + P_A(a)^2\right) \\ &= E\left(\hat{P}_A(a)^2\right) - 2P_A(a)E\left(\hat{P}_A(a)\right) + P_A(a)^2 \\ &= E\left(\hat{P}_A(a)^2\right) - 2P_A(a)^2 + P_A(a)^2 \qquad \text{(since } \hat{P}_A(a) \text{ is unbiased)} \\ &= E\left(\hat{P}_A(a)^2\right) - P_A(a)^2 \qquad (3.1) \end{aligned}$$

Thus, to determine if $\hat{P}_A(a)$ is consistent, we need to evaluate $E\left(\hat{P}_A(a)^2\right)$.

$$
\begin{aligned}
E\left(\hat{P}_A(a)^2\right) &= E\left(\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{x_i=a}\right)^2\right) \\
&= E\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{1}_{x_i=a}\mathbb{1}_{x_j=a}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}P_A(a) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(P_A(a))^2 \\
&= \frac{1}{n}P_A(a) + \frac{n-1}{n}P_A(a)^2
\end{aligned}
$$

Now, substituting into expression (3.1) yields

$$
\begin{aligned}
E(\hat{P}_A(a)^2) - P_A(a)^2 &= \frac{1}{n}P_A(a) + \frac{n-1}{n}P_A(a)^2 - P_A(a)^2 \\
&= \frac{1}{n}P_A(a) + \frac{1}{n}P_A(a)^2 \\
&= \frac{1}{n}\left(P_A(a) + P_A(a)^2\right)
\end{aligned}
$$

So

$$
\lim_{n\to\infty} E\left(\left(\hat{P}_A(a) - P_A(a)\right)^2\right) = \lim_{n\to\infty}\left(\frac{1}{n}\left(P_A(a) + P_A(a)^2\right)\right) = 0
$$

Thus, $\hat{P}_A(a) \to P_A(a)$ in mean square (and, also, by implication, in probability).

### D. Estimator performance at higher granularity

If we record the age at a higher granularity, we will likely end up with many empty bins (i.e. values of $a \in R_A$ such that $x_i \neq a$ for all $i \in \{1, 2, ..., n\}$. For these $a$, our estimate $\hat{P}_A(a) = 0$ (even though we would want non-zero probability estimates for all reasonable ages). More generally, our estimator is not very robust (in the sense that it won't perform well for small $n$ or at higher granularity).

## 4. Call Center

### A. MLE estimator for Poisson $\lambda$

Let $x_1, x_2, ..., x_n$ be realizations of IID $X_1, X_2, ..., X_n \sim$ Poisson($\lambda$). Recall, first, that

$$
f_{X_i}(k) = \frac{\lambda^k}{k!}e^{-\lambda}
$$

Then

$$\begin{aligned}\mathcal{L}_{\vec{x}}(\lambda) &= \prod_{i=1}^{n} f_{X_i}(x_i, \lambda)\\ &= \prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right)\\ &= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-n\lambda}\end{aligned}$$

So

$$\log\mathcal{L}_{\vec{x}}(\lambda) = \log(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n}\log(x_i!)$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\log\mathcal{L}_{\vec{x}}(\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

Setting this equal to zero, we find

$$\begin{aligned} 0 &= \frac{\sum_{i=1}^{n} x_i}{\lambda} - n\\ \implies \qquad \lambda &= \frac{1}{n}\sum_{i=1}^{n} x_i\end{aligned}$$

Again, let's check this is a maximum:

$$\frac{\mathrm{d}^2}{\mathrm{d}^2\lambda}\log\mathcal{L}_{\vec{x}}(\lambda) = -\sum_{i=1}^{n} x_i\lambda^{-2} < 0$$

(since $\lambda \geq 0, x_i \geq 0$ for all $i$). Thus this is indeed a maximum.

### B. Comparing parametric and non-parametric estimation

First, recall our goal was to estimate the distribution of the calls in October using (i) two days of data from the beginning of October, and (ii) data from the entire month of September. Parametric (MLE) and non-parametric estimators were constructed and estimation errors were determined using the $l_1$ norm of the difference in distributions. These errors are presented below:

| **Data** | **MLE fitting errors** | **Nonparametric fitting errors** |
|---|---|---|
| 2 days | 0.4334 | 0.5185 |
| September | 0.5072 | 0.4028 |

### c. Comparing parametric and non-parametric estimation-continued

Based on these errors, it appears parametric (MLE) estimation performs best for the 2-day data. In contrast, nonparametric estimation performs better for the September data. This suggests that parametric estimation is better when less data is available, while non-parametric estimation is better when more data is available. This assertion is not only supported by the results from this particular modeling problem- it also makes theoretical sense through the lens of bias-variance tradeoff.

First, looking at the plotted non-parametric estimated distribution, the 2-day estimate is very noisy. This illustrates that with less data, model variance is higher and dominates the total estimation error. On the other hand, using parametric estimation constrains the model, reducing model variance (though potentially inflating bias). With less data, this reduction in variance results in overall lower error (i.e. it is more significant than any potential increase in bias).

Now, consider the estimates constructed using September data. When more data is available, non-parametric estimation outperforms parametric estimation. This is because the parametric estimate may be biased if the underlying process does not truly follow the Poisson distribution. With more data, the reduction in variance achieved through parametric estimation (i.e by constraining our model space) does not outweigh the error cost incurred due to inflated bias.

## 5. Method of Moments (MM) estimators

Note that in all cases, let $x_1, x_2, ..., x_n$ be realizations of $X_1, X_2, ...X_n \sim D(c)$, where the $D$ is the distribution specified in the subproblem, and $c$ is the target parameter.

### a. Bernoulli distribution

First, recall (from the notes) that $\hat{p}_{MLE} = \frac{n_1}{n_1+n_0}$. Now, $E(X_i) = p$, and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$. Thus, setting $\bar{x}_n = E(X_i)$ yields:

$$\begin{aligned} \hat{p}_{MM} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{n_1}{n_1 + n_0} = \hat{p}_{MLE} \end{aligned}$$

### b. Geometric distribution

First, we derive the MLE for a geometric distribution. Recall $f_{X_i}(k) = (1-p)^{k-1}p$. Thus

$$\begin{aligned}
\mathcal{L}_{\vec{x}}(p) &= \prod_{i=1}^{n} f_{X_i}(x_i, p) \\
&= \prod_{i=1}^{n} \big((1-p)^{x_i-1}p\big) \\
&= p^n(1-p)^{\sum_{i=1}^{n}(x_i-1)} \\
&= p^n(1-p)^{\sum_{i=1}^{n}(x_i)-n}
\end{aligned}$$

Now we consider

$$\begin{aligned}
\log\, \mathcal{L}_{\vec{x}}(p) &= \log\left(p^n(1-p)^{\sum_{i=1}^{n}(x_i)-n}\right) \\
&= n \cdot \log(p) + \log(1-p)\left(\sum_{i=1}^{n}(x_i) - n\right)
\end{aligned}$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}p}\log\, \mathcal{L}_{\vec{x}}(p) = n\frac{1}{p} - \left(\sum_{i=1}^{n}(x_i) - n\right)\frac{1}{1-p}$$

Setting this to zero, we find

$$\begin{aligned}
n\frac{1}{p} - \left(\sum_{i=1}^{n}(x_i) - n\right)\frac{1}{1-p} &= 0 \\
\implies \qquad \left(\sum_{i=1}^{n}(x_i) - n\right)\frac{1}{1-p} &= n\frac{1}{p} \\
\implies \qquad p\left(\sum_{i=1}^{n}(x_i) - n\right) &= n(1-p) \\
\implies \qquad p\sum_{i=1}^{n} x_i &= n \\
\implies \qquad p &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}$$

Now, let's confirm this is a maximum:

$$\frac{\mathrm{d}^2}{\mathrm{d}^2p}\log\, \mathcal{L}_{\vec{x}}(p) = -n \cdot p^{-2} - \left(\sum_{i=1}^{n}(x_i) - n\right)(1-p)^{-2}$$

Since the second derivative is negative for all relevant values of $x_i, n$, and $p$, this is indeed a maximum.

Now, $E(X_i) = 1/p$. Thus, setting $\bar{x}_n = E(X_i)$ yields:

$$\begin{aligned} \frac{1}{\hat{p}_{MM}} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \implies \qquad \hat{p}_{MM} &= \frac{n}{\sum_{i=1}^{n} x_i} = \hat{p}_{MLE} \end{aligned}$$

### c. Poisson distribution

First, recall (from problem 4(a)) that $\hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
Now, $E(X_i) = \lambda$. Thus, setting $\bar{x}_n = E(X_i)$ yields:

$$\hat{\lambda}_{MM} = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\lambda}_{MLE}$$

### d. Exponential distribution

First, recall (from problem 2(b)) that $\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i}$.
Now, $E(X_i) = \frac{1}{\lambda}$. Thus, setting $\bar{x}_n = E(X_i)$ yields:

$$\begin{aligned} \frac{1}{\hat{\lambda}_{MM}} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \implies \qquad \hat{\lambda}_{MM} &= \frac{n}{\sum_{i=1}^{n} x_i} = \hat{\lambda}_{MLE} \end{aligned}$$