Exercise 1. Find MLE (Maximum Likelihood Estimator) for the following parameters:

1. Probability of success p in Bernoulli(p) model.

Solution. Let X be a Bernoulli random variable with parameter p. Let $X_1, \cdots, X_n$ be the independent random samples of X. Recall that the probability density function for the Bernoulli distribution with parameter p is $f(x) = p^x(1-p)^{1-x}$ where $x = 0, 1$. Then the likelihood function of the sample is

$$\begin{aligned} l(x_1, \cdots, x_n, p) &= f(x_1, \cdots, x_n, p) \\ &= \prod_{i=1}^{n} f(x_i, p) \qquad \{\text{since } X_i \text{ are independent}\} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}. \end{aligned}$$

Taking the natural logarithm on both sides gives,

$$\begin{aligned} L(x_1, \cdots, x_n, p) &= \ln l(x_1, \cdots, x_n, p) \\ &= \ln\left(p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right) \\ &= \left(\sum_{i=1}^{n} x_i\right) p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p). \end{aligned}$$

Since $L(x_1, \cdots, x_n, p)$ is a continuous function of p, it has a maximum value. This value can be found by taking the derivative of $L(x_1, \cdots, x_n, p)$ with respect to p, and setting it equal to 0. So,

$$\frac{\partial L(x_1, \cdots, x_n, p)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \left(\frac{n - \sum_{i=1}^{n} x_i}{1-p}\right) = 0$$

gives that $p = \frac{\sum_{i=1}^{n} x_i}{n}$. Hence, $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}_n$. □

2. Probability of success p in Binomial(n, p) model.

Proof. Let X be a Binomial random variable with parameter p. Let $X_1, \cdots, X_m$ be the independent random samples of X. Recall that the probability density function for the Binomial distribution with parameter p is $f(x) = \binom{n}{x} p^x(1-p)^{n-x}$ where $x = 0, \cdots, n$. Then the likelihood function of the sample is

$$\begin{aligned} l(x_1, \cdots, x_m, p) &= f(x_1, \cdots, x_m, p) \\ &= \prod_{i=1}^{m} f(x_i, p) \qquad \{\text{since } X_i \text{ are independent}\} \\ &= \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \end{aligned}$$

Taking the natural logarithm on both sides gives,

$$\begin{aligned}
L(x_1, \cdots, x_m, p) &= \ln l(x_1, \cdots, x_m, p) \\
&= \ln \left(\prod_{i=1}^{m} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} \right) \\
&= \sum_{i=1}^{m} \left(\ln \binom{n}{x_i} + x_i \ln p + (n - x_i) \ln(1-p) \right) \\
&= \sum_{i=1}^{m} \ln \binom{n}{x_i} + \left(\sum_{i=1}^{m} x_i \right) \ln p + \left(mn - \sum_{i=1}^{m} x_i \right) \ln(1-p)
\end{aligned}$$

Since $L(x_1, \cdots, x_n, p)$ is a continuous function of p, it has a maximum value. This value can be found by taking the derivative of $L(x_1, \cdots, x_n, p)$ with respect to p, and setting it equal to 0. So,

$$\frac{\partial L(x_1, \cdots, x_n, p)}{\partial p} = 0 + \frac{\sum\limits_{i=1}^{m} x_i}{p} - \left(\frac{mn - \sum\limits_{i=1}^{m} x_i}{1-p} \right) = 0$$

gives that $p = \frac{\sum_{i=1}^{m} x_i}{mn}$. Hence, $\hat{p} = \frac{\sum_{i=1}^{m} x_i}{mn}$. □

3. Probability of success p in Geometric(p) model.

Proof. Let X be a Geometric random variable with parameter p. Let $X_1, \cdots, X_n$ be the independent random samples of X. Recall that the probability density function for the Geometric distribution with parameter p is $f(x) = p(1-p)^x$ where $x = 0, 1, 2, \cdots$. Then the likelihood function of the sample is

$$\begin{aligned}
l(x_1, \cdots, x_n, p) &= f(x_1, \cdots, x_n, p) \\
&= \prod_{i=1}^{n} f(x_i, p) \qquad \{\text{since } X_i \text{ are independent}\} \\
&= \prod_{i=1}^{n} p(1-p)^{x_i} \\
&= p^n (1-p)^{\sum_{i=1}^{n} x_i}.
\end{aligned}$$

Taking the natural logarithm on both sides gives,

$$\begin{aligned}
L(x_1, \cdots, x_n, p) &= \ln l(x_1, \cdots, x_n, p) \\
&= \ln \left(p^n (1-p)^{\sum_{i=1}^{n} x_i} \right) \\
&= n \ln p + \left(\sum_{i=1}^{n} x_i \right) \ln(1-p).
\end{aligned}$$

Since $L(x_1, \cdots, x_n, p)$ is a continuous function of p, it has a maximum value. This value can be found by taking the derivative of $L(x_1, \cdots, x_n, p)$ with respect to p, and setting it equal to 0. So,

$$\frac{\partial L(x_1, \cdots, x_n, p)}{\partial p} = \frac{n}{p} - \left(\frac{\sum\limits_{i=1}^{n} x_i}{1-p} \right) = 0$$

gives that $p = \dfrac{n}{n + \sum\limits_{i=1}^{n} x_i} = \dfrac{1}{1+\bar{X}_n}$. Hence, $\hat{p} = \dfrac{1}{1+\bar{X}_n}$. □

4. Intensity λ in Poisson(λ) model.

Solution. Let X be a Poisson random variable with parameter λ. Let $X_1, \cdots, X_n$ be the independent random samples of X. Recall that the probability density function for the Poisson distribution with parameter λ is $f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ where $x = 0, 1, \cdots$. Then the likelihood function of the sample is

$$\begin{aligned}
l(x_1, \cdots, x_n, \lambda) &= f(x_1, \cdots, x_n, \lambda) \\
&= \prod_{i=1}^{n} f(x_i, \lambda) \qquad \{\text{since } X_i \text{ are independent}\} \\
&= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\
&= e^{-\lambda n} \cdot \lambda^{\sum\limits_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} \frac{1}{x_i!}.
\end{aligned}$$

Taking the natural logarithm on both sides gives,

$$\begin{aligned}
L(x_1, \cdots, x_n, \lambda) &= \ln l(x_1, \cdots, x_n, \lambda) \\
&= \ln\left(e^{-\lambda n} \cdot \lambda^{\sum\limits_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} \frac{1}{x_i!} \right) \\
&= -\lambda n + \left(\sum_{i=1}^{n} x_i \right) \ln \lambda + \ln \prod_{i=1}^{n} \frac{1}{x_i!}.
\end{aligned}$$

Since $L(x_1, \cdots, x_n, \lambda)$ is a continuous function of λ, it has a maximum value. This value can be found by taking the derivative of $L(x_1, \cdots, x_n, \lambda)$ with respect to λ, and setting it equal to 0. So,

$$\frac{\partial L(x_1, \cdots, x_n, \lambda)}{\partial \lambda} = -n + \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} + 0 = 0$$

gives that $\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$. Hence, $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}_n$. □

Exercise 2. Find the Fisher information for the models from Exercise 1.

1. Probability of success p in Bernoulli(p) model.

   Solution. The Fisher information for the parameter $\hat{p}$ is given by $I(\hat{p}) = E\left[\left(\dfrac{\partial L(x,p)}{\partial p}\right)^2\right]$ or $I(\hat{p}) = -E\left(\dfrac{\partial^2 L(x,p)}{\partial^2 p}\right)$. From Exercise 1 part 1, we have

$$\begin{aligned}
\frac{\partial L(x,p)}{\partial p} &= \frac{\partial\left(xp + (1-x)\ln(1-p)\right)}{\partial p} \\
&= \frac{x}{p} - \left(\frac{1-x}{1-p}\right).
\end{aligned}$$

Recall that $E(X) = p$ where X is the Bernoulli random variable. Then the Fisher information is

$$\begin{aligned} I(\hat{p}) &= -E\left(\frac{\partial^2 L(x,p)}{\partial^2 p}\right) \\ &= -E\left(-\frac{X}{p^2} - \left(\frac{1-X}{(1-p)^2}\right)\right) \\ &= \frac{E(X)}{p^2} + \frac{E(1-X)}{(1-p)^2} \\ &= \frac{p}{p^2} + \left(\frac{1-p}{(1-p)^2}\right) \\ &= \frac{1}{p} + \frac{1-p}{(1-p)^2} \\ &= \frac{(1-p)^2 + p - p^2}{p(1-p)^2} \\ &= \frac{1-p}{p(1-p)^2} \\ &= \frac{1}{p(1-p)}. \end{aligned}$$

□

2. Probability of success p in Binomial(n, p) model.

Solution. The Fisher information for the parameter $\hat{p}$ is given by $I(\hat{p}) = E\left[\left(\frac{\partial L(x,p)}{\partial p}\right)^2\right]$ or $I(\hat{p}) = -E\left(\frac{\partial^2 L(x,p)}{\partial^2 p}\right)$. From Exercise 1 part 2, we have

$$\begin{aligned} \frac{\partial L(x,p)}{\partial p} &= \frac{\partial\left(\ln\binom{n}{x} + x\ln p + (n-x)\ln(1-p)\right)}{\partial p} \\ &= \frac{x}{p} - \left(\frac{n-x}{1-p}\right). \end{aligned}$$

Recall that $E(X) = np$ where X is the Binomial random variable. Then the Fisher information is

$$\begin{aligned} I(\hat{p}) &= -E\left(\frac{\partial^2 L(x,p)}{\partial^2 p}\right) \\ &= -E\left(-\frac{X}{p^2} - \left(\frac{n-X}{(1-p)^2}\right)\right) \\ &= \frac{np}{p^2} - \left(\frac{n-np}{(1-p)^2}\right) \\ &= n\left(\frac{1}{p} - \frac{1}{1-p}\right) \\ &= \frac{n}{p(1-p)}. \end{aligned}$$

□

3. Probability of success p in Geometric(p) model.

Solution. The Fisher information for the parameter $\hat{p}$ is given by $I(\hat{p}) = E\left[\left(\frac{\partial L(x,p)}{\partial p}\right)^2\right]$ or $I(\hat{p}) = -E\left(\frac{\partial^2 L(x,p)}{\partial^2 p}\right)$. From Exercise 1 part 3, we have

$$\begin{aligned}\frac{\partial L(x,p)}{\partial p} &= \frac{\partial\left(\ln p + (x-1)\ln(1-p)\right)}{\partial p}\\ &= \frac{1}{p} - \left(\frac{x}{1-p}\right).\end{aligned}$$

Recall that $E(X) = \frac{1}{p}$ where X is the Geometric random variable. Then the Fisher information is

$$\begin{aligned}I(\hat{p}) &= -E\left(\frac{\partial^2 L(x,p)}{\partial^2 p}\right)\\ &= -E\left(-\frac{1}{p^2} - \left(\frac{X}{(1-p)^2}\right)\right)\\ &= \frac{1}{p^2} + \frac{E(X)}{(1-p)^2}\\ &= \frac{1}{p^2} + \frac{1}{p(1-p)^2}\\ &= \frac{(1-p)^2 + p}{p^2(1-p)^2}\\ &= \frac{1-p+p^2}{p^2(1-p)^2}\\ &= \frac{1-p(1-p)}{p^2(1-p)^2}\\ &= \frac{1}{p^2(1-p)^2} - \frac{1}{p(1-p)}\\ &= \frac{1-p}{p^2(1-p)^2}\\ &= \frac{1}{p^2(1-p)}.\end{aligned}$$

□

4. Intensity λ in Poisson(λ) model.

Solution. The Fisher information for the parameter $\hat{\lambda}$ is given by $I(\hat{\lambda}) = E\left[\left(\frac{\partial L(x,\lambda)}{\partial \lambda}\right)^2\right]$ or $I(\hat{\lambda}) = -E\left(\frac{\partial^2 L(x,\lambda)}{\partial^2 \lambda}\right)$. From Exercise 1 part 4, we have $\frac{\partial L(x,\lambda)}{\partial p} = \frac{\partial\left(-\lambda + x\ln\lambda + \ln\frac{1}{x!}\right)}{\partial \lambda} = \frac{x}{\lambda} - 1$, and so $\frac{\partial^2 L(x,\lambda)}{\partial^2 \lambda} = -\frac{x}{\lambda^2}$. Recall that $E(X) = \lambda$ where X is a Poisson random variable with parameter λ.

Thus,

$$\begin{aligned} I(\hat{\lambda}) &= -E\left(\frac{\partial^2 L(x,\lambda)}{\partial^2 \lambda}\right) \\ &= -E\left(-\frac{X}{\lambda^2}\right) \\ &= \frac{1}{\lambda^2}E(X) \\ &= \frac{\lambda}{\lambda^2} \\ &= \frac{1}{\lambda}. \end{aligned}$$

□

Exercise 3. Assess the efficiency of the MLE estimators for the following models.

1. Probability of success p in Bernoulli(p) model.

   Solution. The Cramer-Rao inequality,

   $$Var(\hat{p}) \geq \frac{1}{nI(\hat{p})},$$

   states that if there exists an estimator that achieves this lower bound, then it is the most efficient estimator for that parameter.

   As shown in Excercise 1 part 1, $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$. So, the variance of this parameter is

   $$\begin{aligned} var(\hat{p}) &= var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \\ &= \frac{\sum_{i=1}^{n} var(x_i)}{n^2} \qquad \{\text{since } X_i \text{ are independent}\} \\ &= \frac{\sum_{i=1}^{n} p(1-p)}{n^2} \\ &= \frac{np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n}. \end{aligned}$$

   Also, as shown in Exercise 2 part 1, $I(\hat{p}) = \frac{1}{p(1-p)}$. Thus, we see that the Cramer-Rao inequality holds, and so $\hat{p}$ is the most efficient estimator. □

2. Probability of success p in Binomial(n, p) model.

Solution. As shown in Excercise 1 part 2, $\hat{p} = \frac{\sum_{i=1}^{m} x_i}{mn}$. The variance of this parameter is

$$\begin{aligned}
var(\hat{p}) &= var\left(\frac{\sum_{i=1}^{m} x_i}{mn}\right) \\
&= \frac{\sum_{i=1}^{m} var(x_i)}{m^2n^2} \qquad \{\text{since } X_i \text{ are independent}\} \\
&= \frac{\sum_{i=1}^{m} np(1-p)}{m^2n^2} \\
&= \frac{mnp(1-p)}{m^2n^2} \\
&= \frac{p(1-p)}{mn}.
\end{aligned}$$

Also, from Exercise 2 part 2, we have $I(\hat{p}) = \dfrac{n}{p(1-p)}$. Hence, the Cramer-Rao inequality $Var(\hat{p}) \geq \frac{1}{m \cdot I(\hat{p})}$ holds, and $\hat{p}$ is the most efficient estimator. □

3. Intensity λ in Poisson(λ) model

Solution. As shown in Excercise 1 part 4, $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$. The variance of this parameter is

$$\begin{aligned}
var(\hat{\lambda}) &= \frac{\sum_{i=1}^{n} var(x_i)}{n^2} \qquad \{\text{since } X_i \text{ are independent}\} \\
&= \frac{\sum_{i=1}^{n} \lambda}{n^2} \\
&= \frac{n\lambda}{n^2} \\
&= \frac{\lambda}{n}.
\end{aligned}$$

Also, from Exercise 2 part 4, we have $I(\hat{\lambda}) = \dfrac{1}{\lambda}$. Hence, the Cramer-Rao inequality $Var(\hat{p}) \geq \frac{1}{n \cdot I(\hat{p})}$ holds, and $\hat{p}$ is the most efficient estimator. □

Exercise 4. Consider the 2×2 contingency table given below that describes the Belief in After Life (Y) by Gender (X). Compute the following:

1. point estimations for π_1, π_2, Δ.
2. distribution for $\hat{\Delta}$.
3. 95% confidence interval for $\hat{\Delta}$.
4. hypothesis test testing whether the gender has an effect on belief in after life.

	Belief in After Life			
Gender		Y	N	Total
	M	398 (n_{11})	104 (n_{12})	502
	F	509 (n_{21})	116 (n_{22})	625
	Total	907	220	$n = 1127$

1. Point estimations for π_1, π_2 and Δ are as follows:

a. $\hat{\pi}_1 = p_1 = \dfrac{n_{11}}{n_{11}+n_{12}} = \dfrac{n_{11}}{n_{1.}} = \dfrac{398}{502} = 0.7928$

b. $\hat{\pi}_2 = p_2 = \dfrac{n_{21}}{n_{121}+n_{22}} = \dfrac{n_{21}}{n_{2.}} = \dfrac{509}{625} = 0.8144$

c. $\hat{\Delta} = \hat{\pi}_1 - \hat{\pi}_2 = -0.0216$

2. Each subject in the given data can be described in terms of Bernoulli trials. Let

$$M_i = \begin{cases} 1 & \text{males believe in after life} \\ 0 & \text{otherwise} \end{cases}$$

and

$$F_i = \begin{cases} 1 & \text{females believe in after life} \\ 0 & \text{otherwise} \end{cases}$$

Then $n_{11} = \sum_{i=1}^{n_{1.}} M_i \sim \text{Binomial}(n_{1.}, \pi_1)$ and $n_{21} = \sum_{i=1}^{n_{2.}} F_i \sim \text{Binomial}(n_{2.}, \pi_2)$. Since $n_{1.}$, $n_{2.}$ are large, by Central Limit Theorem, the binomial random variable can be approximated by a normal random variable. Thus, $n_{11} = \sum_{i=1}^{n_{1.}} M_i \sim N\left(n_{1.}\pi_1, n_{1.}\pi_1(1-\pi_1)\right)$ and $n_{21} = \sum_{i=1}^{n_{2.}} F_i \sim N\left(n_{2.}\pi_2, n_{2.}\pi_2(1-\pi_2)\right)$. Thus,

$$\begin{aligned} \hat{\Delta} &= \hat{\pi}_1 - \hat{\pi}_2 \\ &= p_1 - p_2 \\ &= \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}} \\ &\sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_{1.}} + \frac{\pi_2(1-\pi_2)}{n_{2.}}\right). \end{aligned}$$

3. 95% confidence interval / interval estimation for $\hat{\Delta}$ is given by $\hat{\Delta} \pm Z_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\Delta}}$. The standard deviation for $\hat{\Delta}$ is given by

$$\begin{aligned} \sigma_{\hat{\Delta}} &= \sqrt{\frac{\pi_1(1-\pi_1)}{n_{1.}} + \frac{n_2(1-\pi_2)}{n_{2.}}} \\ &\approx \sqrt{\frac{p_1(1-p_1)}{n_{1.}} + \frac{p_2(1-p_2)}{n_{2.}}} \\ &= \sqrt{\frac{0.7928(1-0.7928)}{502} + \frac{0.8144(1-0.8144)}{625}} \\ &= 0.02386. \end{aligned}$$

Thus, the 95% confidence interval is

$$\begin{aligned} \hat{\Delta} \pm Z_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\Delta}} &= -0.0216 \pm 1.96 \times 0.02386 \\ &= (-0.06837, 0.02517). \end{aligned}$$

We are 95% confident that the true value of Δ lies within this interval.

4. Our null hypothesis is that gender has no effect on belief in after life, and the alternate hypothesis is gender has an effect on belief in after life.

$$\begin{aligned} H_0 &: \Delta = 0 \\ H_a &: \Delta \neq 0 \end{aligned}$$

The distribution of Δ under the null hypothesis is $N(0, 0.02386^2)$. The p-value which is the probability that the true value Δ is greater than the estimated value $\hat{\Delta}$ under the null hypothesis is calculated below.

$$\begin{aligned} P(\Delta > \hat{\Delta} \mid H_0) &= P(\Delta > -0.0216) \\ &= 1 - P(\Delta < -0.0216) \\ &> 0.05. \end{aligned}$$

Since the p-value is greater than the significance level $\alpha = 5\%$, we do not reject the null hypothesis, and conclude that the belief in after life is gender independent.