# Exercise 1

Show the corrected sample variance is an unbiased estimator of population variance.

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

# Exercise

$$
\begin{aligned}
S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2}{n-1} = \\
&= \frac{\sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2}{n-1} \\
&= \frac{\sum_{i=1}^n \left((X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)\right)}{n-1} = \\
&= \frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1} + \frac{\sum_i (\bar{X} - \mu)^2}{n-1} - \frac{\sum_i \left(2(X_i - \mu)(\bar{X} - \mu)\right)}{n-1}
\end{aligned}
$$

# Exercise

$$\begin{aligned}
E((X_i-\mu)^2) &= \sigma^2 \\
E(\bar{X}-\mu)^2 &= \frac{\sigma^2}{n} \\
\left(\sum_i (X_i-\mu)(\bar{X}-\mu)\right) &= (\bar{X}-\mu)\left(\sum_i (X_i-\mu)\right) = n(\bar{X}-\mu)^2 \\
E(S^2) &= \frac{n\sigma^2}{n-1} + \frac{\sigma^2}{n-1} - 2\frac{n\sigma^2}{n(n-1)} \\
E(S^2) &= \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} \\
E(S^2) &= \sigma^2
\end{aligned}$$

# Exercise 2: Uniform Distribution

**Example:** Let $X_1, \ldots, X_n$ be iid r.v. distributed as continuous uniform distribution on $[0, \theta]$. The probability distribution function of $X_i$ for each $i$ is:

$$f(x|\theta) = \begin{cases} \theta^{-1}, & 0 \leq x \leq \theta \\ 0, & otherwise \end{cases}$$

Find $\hat{\theta}_{MOM}$ and discuss properties

# Uniform Distribution: $\hat{\theta}_{MOM}$

**Example:**

$$E(X) = \int xf(x|\theta)dx = \int_0^\theta \frac{x}{\theta}dx = |_0^\theta \frac{x^2}{2\theta} = \frac{\theta}{2}$$

$$\bar{X} = \frac{\theta}{2}$$

$$\hat{\theta}_{MOM} = 2\bar{X}$$

# Uniform Distribution: $\hat{\theta}_{MOM}$

**Example:**

$$
\begin{aligned}
E\left(\hat{\theta}_{MOM}\right) &= \theta \\
Var\left(\hat{\theta}_{MOM}\right) &= 4Var\left(\bar{X}\right) = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n}
\end{aligned}
$$

$$
\begin{aligned}
MSE\left(\hat{\theta}_{MLE}\right) &\leq Var\left(\hat{\theta}_{MOM}\right) \\
\left(\frac{1}{(n+1)(n+2)}\theta^2\right) &\leq \frac{\theta^2}{3n} \\
\left(\frac{2}{(n+1)(n+2)}\right) &\leq \frac{1}{3n}
\end{aligned}
$$

# Exercise 3

Let $T_1$ and $T_2$ be two independent and unbiased estimators of the parameter $\theta$, with $Var(T_1) = \sigma_1^2$ and $Var(T_2) = \sigma_2^2$. Find the UMVUE for $\theta$ among all linear combinations of $T_1$ and $T_2$. What is its variance?

# Solution Exercise 3

$$
\begin{aligned}
T &= a_1 T_1 + a_2 T_2 \\
E(T) &= a_1 E(T_1) + a_2 E(T_2) \\
E(T) &= (a_1 + a_2)\theta
\end{aligned}
$$

To be unbiased: $a_1 + a_2 = 1$, $a_2 = 1 - a_1$

$$
\begin{aligned}
Var(T) &= a^2 Var(T_1) + (1-a)^2 Var(T_2) \\
Var(T) &= a^2 \sigma_1^2 + (1-a)^2 \sigma_2^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{dVar(T)}{da} &= 2a\sigma_1^2 - 2(1-a)\sigma_2^2 \\
2a\sigma_1^2 + 2(1-a)\sigma_2^2 &= 0 \\
a\star &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}
\end{aligned}
$$

The UMVUE estimator for $\theta$ among all linear combinations of $T_1$ and $T_2$ is

$$T = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} T_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} T_2$$

# Exercise 4: Pareto

Let $(X_1, \dots, X_n)$ be a random sample of i.i.d. random variables distributed as a Pareto distribution with unknown parameters $\alpha$ and $x_0$ known

$$f(x; \alpha, x_0) = \alpha\, x_0^{\alpha} x^{-(\alpha+1)} \qquad \text{for } x \geq x_0$$

.

The log-likelihood function is

$$l(\alpha, x_0) = n log \alpha + n\alpha log(x_0) - (\alpha + 1) \sum_{i=1}^{n} log x_i$$

# Pareto

Thus

$$\frac{\delta l(\alpha, x_0)}{\delta\alpha} = \frac{n}{\alpha} + nlog(x_0) - \sum_{i=1}^{n} log x_i$$

Solving for $\frac{\delta l(\alpha, x_0)}{\delta\alpha} = 0$, the mle of $\alpha$ is given by

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log x_i - nlog(x_0)}$$

# Pareto: sufficiency

Observe that the joint pdf of $X = (X_1, \ldots, X_n)$

$$
\begin{aligned}
f(x; \alpha, x_{\mathrm{m}}) &= \prod_{i=1}^{n} \frac{\alpha\, x_{\mathrm{m}}^{\alpha}}{x_i^{\alpha+1}} \\
&= \frac{\alpha^n\, x_{\mathrm{m}}^{n\alpha}}{\prod_{i=1}^{n} x_i^{\alpha+1}} \\
&= g(t, \alpha) h(x)
\end{aligned}
$$

where $t = \prod_{i=1}^{n} x_i$ $g(t, \alpha) = c\alpha^n\, x_{\mathrm{m}}^{n\alpha} t^{-(\alpha+1)}$ and $h(x) = 1$. By the factorization theorem, $T(X) = \prod_{i=1}^{n} X_i$ is sufficient for $\alpha$.

# Pareto: Fisher Information

Thus

$$\frac{\delta^2 l(\alpha, x_0)}{\delta \alpha^2} = -\frac{n}{\alpha^2}$$

$$I_n(\theta) = \frac{n}{\alpha^2}$$

# Exercise 5

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables distributed as follows:

$$f(x;\theta) = \frac{\theta 2^\theta}{x^{\theta+1}} \qquad x > 2$$

1. Show that $\sum_i log(X_i)$ is a sufficient statistics for $\theta$
2. Find $\hat{\theta}_{MLE}$ maximum likelihood estimator (MLE) for $\theta$ and discuss properties of this estimator.
3. Find $\hat{\theta}_{MOM}$ method of moment estimator (MOM) for $\theta$.

# Exercise 5: solution

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n; \theta) &= \prod_i \frac{\theta 2^{\theta}}{x_i^{\theta+1}} \\
f(x_1, x_2, \ldots, x_n; \theta) &= \theta^n 2^{n\theta} \left( \prod_i \frac{1}{x_i} \right)^{\theta+1}
\end{aligned}
$$

Sufficient statistics $\prod_i \frac{1}{X_i}$, or any trasformation, as for example, $\sum_i log(X_i)$

# Exercise 5: solution

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n; \theta) &= \theta^n 2^{n\theta} \left( \prod_i \frac{1}{x_i} \right)^{\theta+1} \\
L(\theta | x_1, x_2, \ldots, x_n) &= \theta^n 2^{n\theta} \left( \prod_i \frac{1}{x_i} \right)^{\theta+1} \\
l(\theta | x_1, x_2, \ldots, x_n) &= n log(\theta) + n\theta log(2) - (\theta + 1) \left( \sum_i log(x_i) \right) \\
\frac{\partial l(\theta)}{\partial \theta} &= \frac{n}{\theta} + n log 2 - \left( \sum_i log(x_i) \right) \\
\frac{\partial^2 l(\theta)^2}{\partial \theta} &= -\frac{n}{\theta^2}
\end{aligned}
$$

# Exercise 5: solution

$$
\begin{aligned}
\frac{\partial l(\theta)}{\partial \theta} &= 0 \\
\hat{\theta}_{MLE} &= \frac{n}{\sum_i log(x_i) - nlog(2)} = \frac{n}{\left(\sum_i log\left(\frac{x_i}{2}\right)\right)}
\end{aligned}
$$

# Exercise 5: solution

$$
\begin{aligned}
\int xf(x;\theta)dx &= \int_2^{\infty} x\frac{\theta 2^{\theta}}{x^{\theta+1}}dx \\
\int xf(x;\theta)dx &= \int_2^{\infty} \frac{\theta 2^{\theta}}{x^{\theta}}dx \\
E(X) &= |_2^{\infty}\theta 2^{\theta}\frac{x^{(-\theta+1)}}{-\theta+1}dx \\
E(X) &= 2\frac{\theta}{\theta-1} \\
\bar{x} &= 2\frac{\theta_{\hat{MOM}}}{\theta_{\hat{MOM}}-1} \\
\hat{\theta}_{MOM} &= \frac{\bar{x}}{\bar{x}-2}
\end{aligned}
$$

# Exercise 6: Maximum Likelihood Estimator for the Geometric distribution

Considering for $n$ observations from a Geometric distribution:

$$p(x|\pi) = \pi(1-\pi)^x$$

Find maximum likelihood estimator for $E(X) = \frac{1-\pi}{\pi}$

$$\frac{dlogL(\pi|x)}{d\pi} = \frac{n}{\pi} - \frac{\sum_i x_i}{1-\pi}$$

The second derivative:

$$\frac{d^2 logL(\pi|x)}{d\pi^2} = -\frac{n}{\pi^2} - \frac{\sum_i x_i}{(1-\pi)^2}$$

# Exercise 6: Maximum Likelihood Estimator for the Geometric distribution

$$\begin{aligned}\frac{n}{\pi} - \frac{\sum_i x_i}{1-\pi} &= 0 \\ \hat{\pi} = \frac{n}{n + \sum_i x_i}\end{aligned}$$

# Exercise 7: Exam January 2016

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables distributed as a Pareto distribution with parameters $\alpha$ and $x_m$ both un known

$$f(x; \alpha, x_{\mathrm{m}}) = \alpha\, x_{\mathrm{m}}^{\alpha} x^{-(\alpha+1)} \quad \text{for } x \geq x_m$$

.

Calculate the Fisher information matrix for the parameter vector $\theta = (x_m, \alpha)$. How do you interpret the off-diagonal terms?

# Exercise 7: Exam January 2016

The log-likelihood function is

$$
\begin{aligned}
l(\alpha, x_m) &= n log \alpha + n\alpha log(x_m) - (\alpha+1)\sum_{i=1}^{n} log x_i \\
\frac{\partial l(\alpha, x_m)}{\partial \alpha} &= \frac{n}{\alpha} + n log(x_m) - \sum_{i=1}^{n} log x_i \\
\frac{\partial l(\alpha, x_m)}{\partial x_m} &= \frac{n\alpha}{x_m} \\
\frac{\partial^2 l(\alpha, x_m)}{\partial \alpha^2} &= -\frac{n}{\alpha^2} \\
\frac{\partial^2 l(\alpha, x_m)}{\partial x_m^2} &= -\frac{n\alpha}{x_m^2} \\
\frac{\partial^2 l(\alpha, x_m)}{\partial \alpha \partial x_m} &= \frac{n}{x_m}
\end{aligned}
$$

# Exercise 8

We consider two continuous independent random variables $U$ and $W$ normally distributed with $N(0, \sigma^2)$. The variable X defined by

$$X = \sqrt{U^2 + V^2}$$

has a Rayleigh distribution with a parameter $\sigma^2$

$$f(x; \theta) = \frac{x}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad x \geq 0$$

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables distributed as $X$

1. Apply the method of the moments to find the estimator $\hat{\sigma}_{MOM}$ of the parameter $\sigma$.
2. Find $\hat{\sigma^2}_{MLE}$ maximum likelihood estimator (MLE) for $\sigma^2$ and discuss properties of this 1estimator.
3. Compute the score function and the Fisher information.
4. Specify asymptotic distribution of $\hat{\theta}_{MLE}$.

# Exercise 8

$$
\begin{aligned}
\int_0^\infty x f(x;\theta) dx &= \int_0^\infty \frac{x^2}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sigma}\int_0^\infty \frac{x^2}{\sigma} exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= \frac{\sqrt{2\pi}}{\sigma}\frac{1}{2}\int_{-\infty}^\infty \frac{x^2}{\sqrt{2\pi}\sigma} exp\left(-\frac{x^2}{2\sigma^2}\right) dx
\end{aligned}
$$

$$
\begin{aligned}
Y &\sim N(0,\sigma^2) \\
\int_{-\infty}^\infty \frac{y^2}{\sqrt{2\pi}\sigma} exp\left(-\frac{y^2}{2\sigma^2}\right) dy &= E(Y^2) = Var(Y) + E(Y)^2 = \sigma^2
\end{aligned}
$$

$$\int_0^\infty x f(x;\theta) dx = \frac{\sqrt{2\pi}}{\sigma} \frac{1}{2} \times \sigma^2 = \frac{\sigma\sqrt{\pi}}{\sqrt{2}}$$

$$E(X) = \sigma\sqrt{\frac{\pi}{2}}$$

$$\hat{\sigma}_{MOM} = \bar{x}\sqrt{\frac{2}{\pi}}$$

# Exercise 8

Find $\hat{\sigma}^2_{MLE}$ maximum likelihood estimator (MLE) for $\sigma^2$ and discuss properties of this 1estimator.

$$
\begin{aligned}
L(\sigma^2|\underline{x}) &= \frac{\prod_i x_i}{\sigma^{2n}} exp\left(-\frac{\sum_i x_i^2}{2\sigma^2}\right) \\
logL(\sigma^2|\underline{x}) &= \sum_i log(x_i) - n \times log\sigma^2 - \frac{\sum_i x_i^2}{2\sigma^2} \\
\frac{\partial logL(\sigma^2|\underline{x})}{\partial\sigma^2} &= -\frac{n}{\sigma^2} + \frac{\sum_i x_i^2}{2\sigma^4}
\end{aligned}
$$

# Exercise 8

$$
\begin{aligned}
\frac{\partial logL(\sigma^2|\underline{x})}{\partial\sigma^2} = 0 \qquad & if \qquad \sigma^2 = \frac{\sum_i x_i^2}{2n} \\
\frac{\partial^2 logL(\sigma^2|\underline{x})}{\partial^2\sigma^2} \qquad & = \qquad +\frac{n}{\sigma^4} - \frac{\sum_i x_i^2}{\sigma^6} \\
\frac{\partial^2 logL(\sigma^2|\underline{x})}{\partial^2\sigma^2}\Big|_{\hat{\sigma^2}} \qquad & = \qquad +\frac{n}{\hat{\sigma}^4} - 2\frac{n}{\hat{\sigma}^4} < 0 \\
\hat{\sigma}^2_{MLE} \qquad & = \qquad \frac{\sum_i x_i^2}{2n}
\end{aligned}
$$

# Exercise 8

Compute the score function and the Fisher information.

$$
\begin{aligned}
Score(\sigma) = \frac{\partial logL(\sigma^2|\underline{x})}{\partial\sigma^2} &= -\frac{n}{\sigma^2} + \frac{\sum_i x_i^2}{2\sigma^4} \\
\frac{\partial^2 logL(\sigma^2|\underline{x})}{\partial^2\sigma^2} &= \frac{n}{\sigma^4} - \frac{\sum_i x_i^2}{\sigma^6}
\end{aligned}
$$

$$
\int_0^\infty x^2 f(x;\theta)dx = \int_0^\infty \frac{x^3}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right) dx
$$

# Exercise 8

Integration by parts

$$
\begin{aligned}
\int_0^\infty f \times g' &= f \times g - \int f' \times g \\
\int_0^\infty x^2 \frac{x}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right) dx &= |_0^\infty x^2 \left(-exp\left(-\frac{x^2}{2\sigma^2}\right)\right) \\
&\quad - \int_0^\infty 2x \left(-exp\left(-\frac{x^2}{2\sigma^2}\right)\right) dx \\
&= 2\sigma^2 \times \int_0^\infty \frac{x}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
E(X^2) &= 2\sigma^2
\end{aligned}
$$

# Exercise 8

$$
\begin{aligned}
I(\theta) = -E\left(\frac{\partial^2 logL(\sigma^2|\underline{x})}{\partial^2\sigma^2}\right) &= E\left(\frac{n}{\sigma^4} - \frac{\sum_i x_i^2}{\sigma^6}\right) \\
I(\theta) &= -\frac{n}{\sigma^4} - \frac{2n\sigma^2}{\sigma^6} \\
I(\theta) &= \frac{n}{\sigma^4}
\end{aligned}
$$

# Exercise 8

Specify asymptotic distribution of $\hat{\theta}_{MLE}$.

$$\hat{\sigma}^2_{MLE} \quad \sim \quad N\left(\sigma^2, \frac{\sigma^4}{n}\right)$$

# Exercise 9: Exam November 2016

Let $(X_1, \ldots, X_n)$ be independent identically distributed geometric random variables with

$$p(x) = \theta(1-\theta)^x, \qquad x = 0, 1, \ldots$$

Show that the $\bar{X}$ is a consistent estimator for $\frac{1-\theta}{\theta}$, that represent the expected number of trials needed to get one success.

# Exam Statistics $29^{th}$ October 2021 (C)

The literacy rate of a nation measures the proportion of people age 15 and over who can read and write. A statistician is interested in the estimation of parameter $\theta$, the literacy rate for women in Afghanistan and she has to choose between random samples from a Bernoulli or a geometric distributions (with the same $\theta$).
i.e. Given $n$ she can choose between the two following experiments:

1. She randomly select $n$ Afghani women and count the number of women who are literate.
2. She run $n$ different experiments, for each experiment, she keeps selecting women untill she finds a literate one; $x_i$ is the number of Afghani women she asks until one says that she is literate.

Which of the two experiment will give the more precise inference on $\theta$

# Exercise 10

*The Economist* collects data each year on the price of a Big Mac in various countries around the world. A sample of McDonald's restaurants in Europe in July 2018 resulted in the following Big Mac prices (after conversion to U.S. dollars).

$$4.44, 3.94, 2.40, 3.97, 4.36, 4.49, 4.19, 3.71, 4.61, 3.89$$

Assuming that the price of a Big Mac, X, is well modeled by a normal distribution

- Compute an estimate of $P(X < 4.2)$.

# Exercise 11

You studied the lifetime of electronic components, but since your study can not last forever, some of the components were still in working order when you had to stop your data collection. Among the $n$ components you studied, $m$ failed at time $X_i, \quad i = 1, ..., m$ (in hours) and the $n - m$ others did last for the 10000 hours that the study lasted, so $X_{m+1}, \ldots, X_n$ are all greater than 10000, but could not be observed. The exponential distribution is a good model for the lifetime of these components. Provide the MLE estimator for the parameter $\lambda$ of that exponential.

# Exercise 11: Continue

1. Find the survival function of of $X$,

$$S(x) = P(X > x),$$

   where $X$ follows an exponential with mean $1/\lambda$.

2. Find $\hat{\lambda}_{MLE}$ the maximum likelihood estimator of $\lambda$ from the sample above. Because you did not observe the exact value of each variable, the likelihood is

$$L(\lambda) = \left\{ \prod_{i=1}^{m} f(Y_i) \right\} \left\{ \prod_{i=m+1}^{n} S(Y_i) \right\}$$

   where $Y_i = min(X_i, 10000)$ are the values you actually observed.

# Exercise 12

A gas station estimates that it takes at least $\alpha$ minutes for a change of oil. The actual time varies from costumer to costumer. However, one can assume that this time will be well represented by an exponential random variable. The random variable $X$, therefore, possess the following density function

$$f(t) = exp(\alpha - t) I_{[t \geq \alpha]}$$

The following values were recorded from 10 clients randomly selected (the time is in minutes):

5.2, 4.1, 3.4, 4.4, 5.6, 7.3, 2.4, 3.3, 2.8, 2.3.

Estimate the parameter $\alpha$ using the estimator of maximum likelihood.

# Exercise 13

A factory producing some electronic devices puts a series number on each of them. The numbers start at 1 and end at $N$, where $N$ is the total number of produced devices. Four devices are chosen at random and their series numbers are 24, 20, 16, 20 respectively.

a) What is the estimation of $N$ by Maximum Likelihood Method?
b) What is the estimation of N by the Method of Moments ?
c) Which on, among the previous estimates, do you consider more accurate?

# Exercise 3- Exam 20230906

In a casino, you observe a table where the game involves tossing a die. You suspect that the die of the dealer may be tricked. You record the result of each game. For game $i$, $n_i$ is the number of times the die was tossed and $X_i$ the number of times a 6 appeared. You recorded N games in all.

1. What is the distribution of $X_i$?
2. Find $\hat{p}_{MLE}$ maximum likelihood estimator (MLE) for $p = P(get\, a\, 6)$
3. Find the Likelihood Ratio Test statistics of level $\alpha = 0.05$ for $H_0 : p = 1/6$ against the alternative $H_1 : p \neq 1/6$
4. You repeat the experiment 100 times and observe $\sum_i x_i = 1100$ and $\sum_i n_i = 2000$, decide whether accept or reject the null hypothesis.

# Exercise 4- Exam 20230906

Total precipitation (in mm) has been recorded on a daily basis at the Vancouver airport for many years. The largest of these daily totals for a given years is called the annual maximum; it describes the amount of precipitation on the "wettest" day of the year. You have a data file listing the annual maxima for the last $n$ years. These $n$ annual maxima do not show any trend over time, so a statistical model in which these are modelled as independent and identically distributed random variables seems reasonable. Assume that the exponential distribution provides a reasonable model. So, if $X_i$ denotes the annual maxima for year $i$, our statistical models is $X_1, X_2, \ldots, X_n$ is a simple random sample from the population with density function $f(x)$ given by:

$$f(x) = \lambda exp(-\lambda x) \qquad for\ x > 0$$

# Exercise 4- Exam 20230906 b)

Our primary interest is in $\theta$, the probability that next year's annual maximum will exceed 100mm.

1. Provide $\hat{\lambda}_{MLE}$ maximum likelihood estimator (MLE) for $\lambda$
2. Provide and discuss an approximate confidence interval for $\lambda$
3. Express $\theta$ as a function of $\lambda$
4. Find $\hat{\theta}_{MLE}$ maximum likelihood estimator (MLE) for $\theta$ and its asymptotic distribution