Exercises l4

Example 4.1

(a) is there a sufficient statistic for θ?
I use the normalization property of the pdf:
$1 = \sum\limits_{x=1}^{\infty} c_\theta \theta^x = c_\theta(\frac{1}{1-\theta} - 1)$
and solving the equation we get $f(x;\theta) = (1-\theta)\theta^{x-1}$ a geometric rv with parameter $1-\theta$.

$f(\vec{x};\theta) = \prod\limits_{i=1}^{n} f(x_i;\theta) = (1-\theta)^n \theta^{\sum\limits_{i=1}^{n} x_i} \theta^{-n}$
if we choose $h(x) = 1$ and $g(\overline{x};\theta) = \frac{(1-\theta)^n}{\theta^n}\theta^{n\overline{x}}$
we see how $f(\vec{x};\theta) = h(x)g(\overline{x};\theta)$ which means that $\overline{x}$ is a sufficient statistic because of the factorization criterion. □

(c) method of moments and asymptotic bias:
Since X_i is a geometric rv with parameter $1-\theta$ we have that $\mathbb{E}[X] = \frac{1}{1-\theta}$, solving the equation for the 1st moment we find:
$\hat{\theta}_{MM} = \frac{\overline{X}-1}{\overline{X}}$

The asymptotic distribution of $\overline{X}$ can be computed using CLT $\frac{\overline{X}-\mathbb{E}[X_i]}{\sqrt{\frac{\text{var}(X_i)}{n}}} \dot{\sim} \mathcal{N}(0,1)$ we get that $\overline{X} \dot{\sim} \mathcal{N}(\frac{1}{1-\theta}, \frac{\theta}{n(1-\theta)^2})$
And then from this we can use the delta method to compute the asymptotic distribution of $\hat{\theta}_{MM}$, we have $g(x) = \frac{x-1}{x}$ which has $g'(\frac{1}{1-\theta}) \neq 0$ (since $0 < \theta < 1$).

we get that $\hat{\theta}_{MM} \dot{\sim} \mathcal{N}(g(\mathbb{E}[\overline{X}]), (g'(\mathbb{E}[\overline{X}]))^2 \frac{\text{var}(\overline{X})}{n})$

so $\hat{\theta}_{MM} \dot{\sim} \mathcal{N}(\theta, \frac{\theta}{n^2})$ (using the formula for variance of geometric rv, and the formula for the variance of the sample average)

The asymptotic bias is 0 because for $n \to \infty$ we have that $\mathbb{E}_\theta[\hat{\theta}_{MM}] \to \theta$

(d) MLE
$L(\theta) = \frac{(1-\theta)^n}{\theta^n}\theta^{\sum\limits_{i=1}^{n} x_i} = (1-\theta)^n \theta^{n(\overline{x}-1)}$
$L'(\theta) = -n(1-\theta)^{n-1}\theta^{n(\overline{x}-1)} + (1-\theta)^n n(\overline{x}-1)\theta^{n(\overline{x}-1)}$
we solve $L'(\hat{\theta}) = 0$ and get $\hat{\theta} = \frac{\overline{x}-1}{\overline{x}}$, the same estimator obtained by the method of moments.

The limiting distribution is the same as in the previous point.

For the sample $(8,2,3,1)$ $\overline{x} = \frac{8+2+3+1}{4} = 3.5$, $\hat{\theta} = 0.71428$

(e) The answer is the same as in (d) because the sample is iid.

Example 4.2

$f(x,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ pdf of X_i
$L(\sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{\sum\limits_{i=1}^{n} x_i^2}{2\sigma^2}}$
$\ell(\sigma) = \ln(\frac{1}{\sigma^n(2\pi)^{\frac{n}{2}}}) - \frac{\sum\limits_{i=1}^{n} x_i^2}{2\sigma^2} =$

$= -n\ln(\sigma(2\pi)^{\frac{1}{2}}) - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} =$
$= -n\ln(\sigma) - (2\pi)^{\frac{1}{2}} - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2}$

$\ell'(\sigma) = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n} x_i^2}{\sigma^3}$

solve $\ell'(\hat{\sigma}) = 0$ we get
$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$

and since MLE is `equivariant` then $\hat{\psi} = \ln(\hat{\sigma})$.

Example 4.3

(a) MLE of $\psi = \frac{\sigma}{\mu}$ and its distribution:

$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pdf of X_i
$L(\mu, \sigma) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n} (x_i-\mu)^2}{2\sigma^2}}$
$\ell(\mu, \sigma) = -n\ln(\sigma\sqrt{2\pi}) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$

$$\nabla\ell(\mu, \sigma) = [\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}, \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}]^\top$$

Solving $\nabla\ell(\mu, \sigma) = \vec{0}$ we get:
$\begin{cases} \hat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \overline{x} \\ \hat{\sigma} = \overline{x^2} - \overline{x}^2 \end{cases}$
and for the `equivariant` property $\hat{\psi} = \frac{\hat{\sigma}}{\hat{\mu}} = \frac{\overline{x^2} - \overline{x}^2}{\overline{x}}$

We know that the asymptotic distribution of the MLE is normal.
$\hat{\psi} \tilde{\ } \mathcal{N}(\frac{\sigma}{\mu}, \text{var}(\hat{\psi}))$

(b) sample size 10, $\overline{x} = 1.5$ and $\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = 3$
provide an approximation for $P(\hat{\psi} > 1 + \psi)$
$\hat{\mu} = 1.5, \hat{\sigma} = \frac{3}{n} = 0.3 \rightarrow \hat{\psi} = 0.2$
$P(\hat{\psi} > 1 + \psi) = P(\psi < -0.8)$
...... need the distribution.............

Example 4.4

$X_1, \ldots, X_n$ iid sample with $X_i \sim \mathcal{N}(\theta, \theta^2)$ with $\theta > 0$
(a) sufficient statistic for θ is $T = [\overline{x^2}, \overline{x}]$
$f(\vec{x}, \theta) = \frac{1}{\theta^n (2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\theta^2}} =$

$= \frac{1}{\theta^n (2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i^2+\theta^2-2x_i\theta)}{2\theta^2}} =$
$= \frac{1}{\theta^n (2\pi)^{n/2}} e^{-\frac{n\overline{x^2}+n\theta^2-2n\theta\overline{x}}{2\theta^2}}$
and for the likelihood factorization criterion T is sufficient

(b) MLE of θ:

$f(x,\theta) = \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\theta^2}}$ pdf of X_i

$L(\theta) = \frac{1}{\theta^n (2\pi)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\theta^2}}$

$\ell(\theta) = -n\ln(\theta\sqrt{2\pi}) - \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\theta^2}$
$= -n\ln(\theta\sqrt{2\pi}) - \frac{\sum_{i=1}^{n}(x_i^2+\theta^2-2x_i\theta)}{2\theta^2} =$
$= -n\ln(\theta\sqrt{2\pi}) - \frac{n\overline{x^2}}{\theta^2} - \frac{-n}{2} + \frac{n\overline{x}}{\theta}$

$\frac{\partial\ell(\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{n\overline{x^2}}{\theta^3} - \frac{n\overline{x}}{\theta^2}$
We solve $\frac{\partial\ell(\theta)}{\partial\theta} = 0$ and get
$\hat{\theta}^2 + \theta\overline{x} - \overline{x^2} = 0$, we discard the negative solution and get:
$\hat{\theta} = -\frac{\overline{x}+\sqrt{\overline{x}^2+4\overline{x^2}}}{2}$
we can try an example with the following R script:

```
x = rnorm(10000,3,3) sample_average = mean(x) stat_power = mean(x^2) MLE = (-sample_average+sqrt(sample_average^2+4*stat_power))[/2](tg://bot_command?command=2) print(MLE)
> print(MLE)
[1] 2.999047
```

Example 4.5

$X_1, \dots, X_n$ iid random sample with $X_i \sim \mathrm{U}nif(\theta, 0)$
(i) Distribution of MLE of θ:
$f(x,\theta) = -\frac{1}{\theta}\mathbb{1}_{[\theta,0]}(x)$ pdf of X_i (θ is negative)

$$L(\theta) = \begin{cases} (-1)^n \frac{1}{\theta^n} & \text{if } x_{(1)} \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

which reaches the maximum for $\hat{\theta} = x_{(1)}$

this means that the distribution of the MLE is the distribution of the sample minimum:
$f_{\hat{\theta}}(t) = n(1-F(t))^{n-1} f(t) = n(1-\frac{t-\theta}{-\theta})^{n-1}(\frac{-1}{\theta}\mathbb{1}_{[\theta,0]}(t))$
$= -nt^{n-1}\frac{1}{\theta}\mathbb{1}_{[\theta,0]}(t)$

(ii) We can use the property that tell us that MLE is asymptotically normal:
$\hat{\theta} \to^d \mathcal{N}(\theta, \mathrm{var}(\hat{\theta}))$

Example 4.6

X_1, X_2, X_3 iid sample with $X_i \sim \text{Ber}(\theta)$

(a) $S = \sum_{i=1}^{3} X_i$ is sufficient, prove it using the definition:

We need to check that the following $(\star)$ conditional probability does not depend on θ

$(\star) = P_\theta(\vec{X} = \vec{x} | T(\vec{X}) = t(\vec{x})) = \frac{f(\vec{x};\theta)}{q(t(\vec{x});\theta)}$

We have that the pdf of $\vec{X}$ is $f(\vec{x};\theta) = \theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3}$

and since S is the sum of 3 independent Bernoulli rvs we have that the pdf of S is $q(x_1+x_2+x_3;\theta) = \binom{3}{x_1+x_2+x_3}\theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3}$

which means that $(\star) = \frac{1}{\binom{3}{x_1+x_2+x_3}}$ that doesn't depend on θ. $\square$

(b) Show that $X_1 + 2X_2 + 3X_3$ is not sufficient:

The pdf of $\vec{X}$ is the same as in point (a)

$A \triangleq X_1$ is a Bernoulli rv, $B \triangleq 2X_2$ has a pmf $f_B(b) = \theta^b(1-\theta)^{2-b}$ and $C \triangleq 3X_3$ has a pmf $f_C(c) = \theta^c(1-\theta)^{3-c}$

functions of independent rvs are independent, which means A, B, C are independent rvs.

The rv $T = A + B + C$ takes values in $0, 1, 2, 3, 4, 5, 6$ and (since they are independent) we have that:

A+B+C	A	B	C	Pr
0	0	0	0	$(1-\theta)^3$
1	1	0	0	$(1-\theta)^2\theta$
2	0	2	0	$(1-\theta)^2\theta$
3	1	2	0	$(1-\theta)\theta^2$
3	0	0	3	$(1-\theta)^2\theta$
4	1	0	3	$(1-\theta)\theta^2$
5	0	2	3	$(1-\theta)\theta^2$
6	1	2	3	θ^3

and for example for $x_1 = 1, x_2 = 1, x_3 = 0$

$f(1,1,0;\theta) = \theta^{1+1+0}(1-\theta)^{3-1-1-0} = \theta^2(1-\theta)$

and $Pr(A+B+C=3) = (1-\theta)\theta^2 + (1-\theta)^2\theta = \theta(1-\theta)$

which means that the conditional probability

$(\star) = \theta$: it depends on θ when the outcome is $A+B+C=3$ which means that $A+B+C = X_1 + 2X_2 + 3X_3$ is not a sufficient statistic.

(this is because there isn't a function that maps $A+B+C$ in (A, B, C))

Example 4.7

Compute the method of moments estimator for

$X_1, \ldots, X_n$ iid random sample with $X_i \sim \text{Unif}(0, \theta)$

Method of moments, we solve: $\mathbb{E}[Y] = \overline{Y}$

$\mathbb{E}[Y] = \frac{\theta}{2} = \overline{Y}$

which means that $\hat{\theta}_{MM} = 2\overline{Y}$

the bias of the method of moments estimator is

$b(\theta; \hat{\theta}_{MM}) = \mathbb{E}_\theta(\hat{\theta}_{MM}) - \theta = 2\frac{\theta}{2} - \theta = 0$, it's unbiased.

To compute the MSE we also need

$$\text{var}(\hat{\theta}_{MM}) = \text{var}(2\overline{Y}) \underbrace{=}_{\text{prop. of } \overline{Y}} \frac{4\sigma^2}{n}$$

so $\text{mse}(\theta; \hat{\theta}_{MM}) = \frac{4\sigma^2}{n}$.

For the MLE we have that $\hat{\theta} = Y_{(n)}$ is only asymptotically unbiased in fact:

$f_{Y_{(n)}}(t) = n(F(t))^{n-1}f(t) = n(\frac{t}{\theta})^{n-1}\frac{1}{\theta}\mathbb{1}_{[0,\theta]}$

$E_\theta(\hat{\theta}) = \int\limits_0^\theta t\, f_{Y_{(n)}}(t)\ \mathrm{dt} =$
$= \int\limits_0^\theta n \frac{t^n}{\theta^n}\ \mathrm{dt} =$
$= \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1}\theta$

$b(\theta; \hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta = \frac{n}{n+1}\theta - \theta = \frac{-1}{n-1}$,

$\mathrm{var}(\hat{\theta}) = \mathrm{var}(Y_{(n)}) \underbrace{=}_{\text{prop. of var}} \mathbb{E}[Y_{(n)}^2] - \mathbb{E}[Y_{(n)}]^2 =$

$(\mathbb{E}[Y_{(n)}^2] = \cdots = \frac{n}{n+2}\theta^2) = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 =$

$= \theta^2 (\frac{n(n+1)^2 - n^3 - 2n^2}{(n+2)(n+1)^2}) =$
$= \theta^2 (\frac{n^3 + 2n^2 + n - n^3 - 2n^2}{(n+2)(n+1)^2}) =$
$= \theta^2 (\frac{n}{(n+2)(n+1)^2})$

$\mathrm{mse}(\theta, \hat{\theta}) = \theta^2 (\frac{n}{(n+2)(n+1)^2}) + \frac{1}{(n-1)^2}$

We see that both estimators are consistent because the mse of both goes to 0 for $n \to \infty$

But the MM estimator is unbiased while the MLE is only asymptotically unbiased, also mse of MLE depends on θ while the mse of MM does not.
