Exercise 1 (25 p.)
Suppose that \( X = (X_1, \ldots, X_n) \) is a random sample such that each \( X_i \) can take values in \( \mathbb{N} \cup \{0\} \) and has the probability mass function

\[
p_\theta(x_i) = ce^{-\theta x_i}
\]

with \( \theta = \mathbb{R}_+ \) and \( x_i \in \mathbb{N} \cup \{0\} \).

(i) (3 p.) Show that \( c = 1 - e^{-\theta} \).

(ii) (5 p.) Find the MoM estimator \( \hat{\theta}_n \) of \( \theta \). (Hint: You can use that \( \sum_{k=0}^{\infty} k z^k = \frac{z}{(1-z)^2} \) for \( |z| < 1 \).)

(iii) (6 p.) Show that \( \text{Var}_\theta(X_1) = e^{-\theta}(1 - e^{-\theta})^2 \) and that \( \hat{\theta}_n \) is asymptotically normal. Determine the variance of the asymptotic distribution. (Hint: You can use that \( \sum_{k=0}^{\infty} k^2 z^k = \frac{z(z+1)}{(1-z)^3} \) for \( |z| < 1 \).)

(iv) (6 p.) Show that the MLE \( \theta^*_n \) is equal to the MoM estimator.

(v) (5 p.) Assuming that \( \theta^*_n \) is consistent and asymptotically normal, verify that the asymptotic variance is the same as the one found in (iii).

Solution:

(i) Follows from solving \( \sum_{k=0}^{\infty} p_\theta(k) = 1 \).

(ii) The expected value is equal to \( \mathbb{E}_\theta(X_1) = \frac{e^{-\theta}}{(1-e^{-\theta})} \). We identify \( q(x) = \frac{e^{-x}}{1-e^{-\theta}} \), so \( q^{-1}(y) = \ln(\frac{1+y}{y}) \). The estimator is equal to

\[
\hat{\theta}_n = \ln \left( 1 + \frac{n}{\sum_{i=1}^{n} X_i} \right).
\]

(iii) We have that \( \text{Var}_\theta(X_1) = \frac{e^{-\theta}}{(1-e^{-\theta})^2} \) which is finite, \( q \in C^2 \) and \( q'(x) = -\frac{e^{-x}}{(1-e^{-\theta})^2} \neq 0 \) for all \( \theta > 0 \). By Theorem 2.7 the estimator is asymptotically normal with variance \( e^\theta(1-e^{-\theta})^2 \).

(iv) Call \( z = e^{-\theta} \) and find maxima of the map

\[
z \mapsto (1 - z)^n z^{\sum_{i=1}^{n} x_i},
\]

we obtain that the maximum is attained at \( z = \frac{\sum_{i=1}^{n} x_i}{n+\sum_{i=1}^{n} x_i} \), solving for \( \theta \) yields the claim.
(v) We compute the Fisher information. First note that
\[ \ln(p_{\theta}(x_1)) = \ln(1 - e^{-\theta}) - \theta x_1 \]
which has the second derivative equal to
\[ \frac{\partial^2}{\partial \theta^2} \ln(p_{\theta}(x_1)) = -\frac{e^{-\theta}}{(1 - e^{-\theta})} - \frac{e^{-2\theta}}{(1 - e^{-\theta})^2} = -\frac{e^{-\theta}}{(1 - e^{-\theta})^2}. \]
We have that \( I^{-1}(\theta) = e^{\theta}(1 - e^{-\theta})^2 \).

Exercise 2 (20 p.)
Two surveys were independently conducted to estimate a population mean \( \mu \) in a population of size \( N \) using simple random sampling. Denote their unbiased estimators by \( \overline{X}_{n,1} \) and \( \overline{X}_{n,2} \) respectively and standard errors by \( \sigma_{\overline{X}_{n,1}} \) and \( \sigma_{\overline{X}_{n,2}} \). We want to understand if combining both estimators will give a better result. Let \( a, b \in \mathbb{R} \) and define:
\[ X'_n := a \overline{X}_{n,1} + b \overline{X}_{n,2}. \]

(i) (2 p.) Under which condition on \( a, b \) is the new estimator \( X'_n \) unbiased?

(ii) (8 p.) Determine which \( a, b \) minimizes the variance, subjected to unbiasedness.

(iii) (5 p.) Assume now that we have a standard error for the first survey of \( \sqrt{0.1} \) and for the second of \( \sqrt{0.3} \). What is the standard error of the combined estimator? What do you observe?

(iv) (5 p.) Assume that \( \overline{X}_{n,1} \) and \( \overline{X}_{n,2} \) are sample mean estimators. Compute the standard error \( \sigma_{X'_n} \) when \( \sigma^2 = 0.2 \), \( N = 200 \) and \( n = 50 \) and compare with the standard error for the estimator \( X'_n \) from random sampling. Which one is smaller?

Solution

(i) We see that
\[ \mathbb{E}(X'_n) := \mathbb{E}(a \overline{X}_{n,1} + b \overline{X}_{n,2}) = (a + b)\mu \]
Hence \( X'_n \) is unbiased if \( a + b = 1 \).

(ii) Since the surveys were conducted independently, we get
\[ \sigma^2_{X'_n} = a^2 \sigma^2_{\overline{X}_{n,1}} + b^2 \sigma^2_{\overline{X}_{n,2}} \]
Using the condition we found in part (i) we get
\[ \sigma^2_{X'_n} = a^2 \sigma^2_{\overline{X}_{n,1}} + (1 - a)^2 \sigma^2_{\overline{X}_{n,2}} \]
\[ = a^2 \left( \sigma^2_{\overline{X}_{n,1}} + \sigma^2_{\overline{X}_{n,2}} \right) - 2a \sigma_{\overline{X}_{n,1}} \sigma_{\overline{X}_{n,2}} + \sigma^2_{\overline{X}_{n,2}} \]
This function reaches its minimum at
\[ a_{\text{min}} = \frac{\sigma^2_{\overline{X}_{n,2}}}{\sigma^2_{\overline{X}_{n,1}} + \sigma^2_{\overline{X}_{n,2}}} \]
Hence
\[ b_{\text{min}} = \frac{\sigma^2_{\overline{X}_{n,1}}}{\sigma^2_{\overline{X}_{n,1}} + \sigma^2_{\overline{X}_{n,2}}}. \]
(iii) Plugging in, we get that, \(a = \frac{3}{4}, b = \frac{1}{4}\) and \(\sigma_{X_n'} = \sqrt{0.075} \approx 0.27\) which much smaller than any of the others.

(iv) For the simple random sampling we get

\[\sigma_{X_n'} = \sqrt{\frac{0.2}{50}} \left( \frac{200 - 50}{200 - 1} \right) \approx 0.055\]

and for the random sampling

\[\sigma_{X_n'} = \sqrt{\frac{0.2}{50}} \approx 0.063\]

The first one is smaller.

**Exercise 3** (25 p.)

Consider a random sample \(X = (X_1, ..., X_n)\) such that \(X_i \sim N(\theta, \theta)\) were \(\theta \in \mathbb{R}_+\).

(i) (5 p.) Prove that the parametric family is of exponential type.

(ii) (5 p.) Design a u.m.p. test for testing \((H_0) : \theta \geq \theta_0\) against \((H_1) : \theta < \theta_0\) at significance \(\alpha\). Describe the rejection region implicitly.

(iii) (1 p.) Does the rejection region change when we change the null hypothesis to \((H_0) : \theta = \theta_0\)?

(iv) (5 p.) Show that the MLE \(\hat{\theta}_n\) for \(\theta\) is equal to

\[\hat{\theta}_n = \frac{1}{2} \left( \sqrt{\frac{4}{n} \sum_{i=1}^{n} X_i^2} + 1 - 1 \right)\]

(v) (5 p.) Compute the Fisher information \(I(\theta)\) and determine the Wald random confidence interval for \(\theta\) at significance \(\alpha\).

(vi) (4 p.) For \(\alpha = 0.05\), \(z(0.025) = 1.96\) and \(\sum_{i=1}^{100} x_i^2 = 80\) determine the confidence interval.

**Solution:**

(i) We write the pdf as

\[f_\theta(x_1) = \frac{1}{\sqrt{2\pi\theta}} e^{-(x_1 - \theta)^2} = \frac{e^{x_1}}{\sqrt{2\pi}} e^{-\frac{1}{2\theta} x_1^2 - \frac{\theta}{2} - \ln(\theta)}\]

so we have an exponential family by identifying \(h(x_1) = \frac{e^{x_1}}{\sqrt{2\pi}}, V(\theta) = -\frac{\theta}{2} - \frac{\ln(\theta)}{2}, a(\theta) = -\frac{1}{2\theta}\) and \(U(x_1) = x_1^2\). \(a(\theta)\) is decreasing in \(\theta\).

(ii) By Theorem 2.19 we have that

\[\phi(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i^2 > c \\ 0 & \text{else} \end{cases}\]

(since \(a(\theta)\) is decreasing and we reversed the null hypothesis) and

\[\alpha = \mathbb{P}_{\theta_0} \left( \sum_{i=1}^{n} X_i^2 > c \right)\]

We reject if \(\sum_{i=1}^{n} x_i^2 > c\).
(iii) No.
(iv) The log-likelihood function is equal to

\[
l(x; \theta) = -\frac{n}{2} \ln(2\pi \theta) - \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{2\theta}.
\]

Taking the derivative

\[
\frac{\partial}{\partial \theta} l(x; \theta) = -\frac{n}{2\theta} + \sum_{i=1}^{n} \frac{(x_i - \theta)}{\theta} + \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{2\theta^2},
\]
equating to 0 and checking the second derivative we obtain the claim. One solution of the quadratic equation can be neglected since \(\theta > 0\).

(v) \[I(\theta) = -\mathbb{E}_\theta \left( \frac{\partial^2}{\partial \theta^2} \ln(f_\theta(X)) \right) = \frac{1}{2\theta^2} + \frac{1}{\theta^3} \mathbb{E}_\theta(X_1^2) = \frac{1 + 2\theta}{2\theta^2}.\]

The Wald random confidence interval is equal to \([L(X), R(X)]\) where

\[
L(X) = \frac{1}{2} \left( \sqrt{\frac{4 \sum_{i=1}^{n} X_i^2}{n} + 1} - 1 \right) - \frac{z(\alpha/2)(\sqrt{4/n \sum_{i=1}^{n} X_i^2} + 1 - 1)}{\sqrt{2n}(4/n \sum_{i=1}^{n} X_i^2 + 1)^{1/4}}
\]
and

\[
R(X) = \frac{1}{2} \left( \sqrt{\frac{4 \sum_{i=1}^{n} X_i^2}{n} + 1} + 1 \right) + \frac{z(\alpha/2)(\sqrt{4/n \sum_{i=1}^{n} X_i^2} + 1 - 1)}{\sqrt{2n}(4/n \sum_{i=1}^{n} X_i^2 + 1)^{1/4}}
\]

(vi) [0.38, 0.66]

Exercise 4 (30 p.) Let \((X_n)_{n \geq 1}\) be a sequence of i.i.d random variables with common density \(f\). Define for \(s > 0\)

\[\tau(s) = \inf\{n \geq 1; X_n > s\}\]

the index of the first \(X_n\)-random variable which is exceeding a certain level \(s\).

(i) (3 p.) Let \(k \in \mathbb{N}\), express the event \(\{\tau(s) = k\}\) in terms of \(X_1, \ldots, X_k\).

(ii) (5 p.) Compute \(P(\tau(s) = k)\) for \(k \in \mathbb{N}\) in terms of \(p_s := P(X_1 > s)\). Show that it is geometric. Identify the parameter.

(iii) (5 p.) Show that the MGF \(M_s(t)\) of \(p_s \tau(s)\) is equal to

\[
M_s(t) = \frac{p_s e^{tp_s}}{1 - e^{tp_s}(1 - p_s)}.
\]
(iv) (5 p.) Show that \( \lim_{s \to \infty} p_s = 0 \) and compute \( \lim_{s \to \infty} M_s(t) \). Show the limiting distribution is \( Exp(1) \).

(v) (7 p.) Application: every day the water level of the Maas river is measured. \( X_n \) measures the water level on day \( n \) in standard units. If the level exceeds 8 then there is a high chance of floods to occur and the Rijkswaterstaat needs to be called. Assume now that the common density of \( (X_n)_{n \geq 1} \) is given by

\[
f(x) = \frac{3}{(1 + x)^4} \mathbb{1}_{x \geq 0}.
\]

What is the probability that there is danger of flooding for the first time after exactly 5 days?

(vi) (5 p.) Find the maximal \( k \geq 1 \) such that \( \mathbb{P}(\tau(8) > k) \geq 0.9 \). What does the result mean in combination with (v)?

**Solution:**

(i) For \( k > 1 \),

\[
\{ \tau(s) = k \} = \{ X_1 \leq s, X_2 \leq s, ..., X_k-1 \leq s, X_k > s \}
\]

for \( k = 1 \), \( \{ \tau(s) = 1 \} = \{ X_1 > s \} \).

(ii) By independence, \( k \in \mathbb{N} \),

\[
\mathbb{P}(\tau(s) = k) = (1 - p_s)^{k-1} p_s.
\]

The distribution is geometric with parameter \( p_s \).

(iii) For \( t \in (-\infty, -\frac{\log(1-p_s)}{p_s}) \), assuming \( p_s < 1 \),

\[
M_s(t) = \mathbb{E}(e^{tp_s \tau(s)}) = \sum_{k=1}^{\infty} e^{tp_k} (1 - p_s)^{k-1} p_s
\]

\[
= \frac{p_s}{1 - p_s} \sum_{k=1}^{\infty} e^{tp_k} (1 - p_s)^k
\]

\[
= \frac{p_se^{tp_s}}{1 - e^{tp_s}(1 - p_s)}
\]

(iv)

\[
\lim_{s \to \infty} p_s = \lim_{s \to \infty} (1 - F_{X_1}(s)) = 0
\]

by property of the CDF. The map \( s \mapsto p_s \) is differentiable, \( \frac{d}{ds} p_s = -f(s) \), using L'Hôpital we get

\[
\lim_{s \to \infty} \frac{p_se^{tp_s}}{1 - e^{tp_s}(1 - p_s)} = \lim_{s \to \infty} \frac{-f(s)e^{tp_s} - f(s)p_se^{tp_s}}{-f(s)e^{tp_s} + (1 - p_s)t(-f(s)e^{tp_s})} = \frac{1}{1-t}.
\]

(v)

\[
p_s = \mathbb{P}(X_1 > s) = \int_s^{\infty} \frac{3}{(1 + x)^4} dx = \frac{1}{(1 + s)^3}
\]

and

\[
\mathbb{P}(\tau(8) = 5) = \frac{1}{9^3} \left( 1 - \frac{1}{9^3} \right)^4 \approx 0.0013.
\]
(vi) Look for $k$ such that
\[ \mathbb{P}(\tau(8) > k) \geq 0.9 \]
and
\[ \mathbb{P}(\tau(8) > k) = \left( \frac{9^3 - 1}{9^3} \right)^k \]
which implies $k \leq 76$ with at least 90% probability the water level will stay below 8 in the first 76 days.