

Exercises for seminars week 44

Exercise 1

a) Let $\hat{\theta}_n$ be an estimator for an unknown parameter θ . Suppose we know from theory that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{D} Z \sim N(0, b^2)$ where $b > 0$ is a constant. Explain why this is the same as saying that $\sqrt{n} \frac{\hat{\theta}_n - \theta}{b} \xrightarrow[n \rightarrow \infty]{D} \frac{Z}{b} \sim N(0, 1)$. [**Hint:** Use the definition of convergence in distribution directly, or, alternatively, use Slutsky's lemma combined with property (3) in "Lecture Notes to Rice chap. 5".]

b) Let θ be an unknown parameter in a model and $\hat{\theta}_n$ an estimator based on n observations. Suppose we have proved that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{D} X \sim N(0, b^2)$ where $b > 0$ is some constant. Show that this implies that $\hat{\theta}_n$ must be a consistent estimator for θ .

[**Hint:** Set $X_n = \sqrt{n}(\hat{\theta}_n - \theta)$, solve with respect to $\hat{\theta}_n$, and use Slutsky. Also properties (3) and (5) in "Lecture Notes to Rice chap. 5" may be relevant.]

Exercise 2

a. Suppose that X is Pareto distributed with *pdf*

$$(1) \quad f(x) = \begin{cases} \theta b^\theta \frac{1}{x^{\theta+1}} & \text{for } x > b \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ and $b > 0$ are parameters. Show that¹

¹ Note that the result implies that the moment generating function (*mgf*) for the pareto distribution does not exist in an open interval containing 0, which would have implied that moments of all orders exist.

$$E(X^r) = \begin{cases} b^r \frac{\theta}{\theta - r} & \text{for } 0 < r < \theta \\ \text{does not exist} & \text{for } r \geq \theta \end{cases}$$

b. Let X_1, X_2, \dots, X_n be *iid* and pareto distributed as in (1) with parameters (b, θ) , where b is known. The preferred estimator for θ based on X_1, X_2, \dots, X_n , is the so called maximum likelihood estimator (mle) that was derived in the lectures:

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^n \ln(X_i/b)} = \frac{1}{\bar{Y}} \quad \text{where } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \text{ and } Y_i = \ln(X_i/b)$$

Show that $\hat{\theta}_n$ is a consistent estimator for θ by using the law of large numbers and the continuity properties of plim. [**Hint:** You may also use **supplementary exercise 2d** (on the net)]

c. Show that the exact distribution of $\hat{\theta}_n$, for any n , has the following pdf:

$$(2) \quad f_{\hat{\theta}}(t) = \begin{cases} \frac{(n\theta)^n}{\Gamma(n)} t^{-n-1} e^{-\frac{n\theta}{t}} & \text{for } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

[**Hint:** Use exercise 1 from no-seminar week 41, and remember the supplementary exercise 2d to find the distribution for \bar{Y} first. Then, finally, transform this distribution.]

d. It is rather seldom that we can find (like here) the exact distribution of a mle estimator since such estimators are often complicated. Luckily the general theory for mle's provides approximate distributions (usually normal) for mle's that are valid in large samples. For example, in the present case, the general theory (details on derivation are given later in the course) states that $\hat{\theta}_n$ is approximately $N(\theta, \theta^2/n)$ distributed for large n . We are now in the fortunate position that we can study the quality of the approximation result by comparing the exact with the approximate distribution for various n .

Make graphs, using e.g. Stata, for $n = 5, 10, 25, 80$ respectively. For each given n , plot both the exact pdf in (2) and the approximate normal density for $\hat{\theta}_n$ in the same graph. Use the value $\theta = 3.8$ (which corresponds to the estimate for Norwegian female incomes above $b = 250\,000$ in Norway 1998 given in supplementary exercise 2, 3). Comment on the results.

[Hint: Choose an interval, from 0 to 8 for example, that covers most of the variation. Make a column of arguments, for example 0,1, 0,2, 0,3, ..., 8,0 (use the fill command and the expression 0.1(0.1)8). (Remember to specify first the length of the column by the command: set obs 80). Then calculate the values of the two densities for each arguments and the four n values, giving altogether 8 columns. Then plot, using the *line* option, the two corresponding densities in the same graph for each n . The easiest is to use “Overlaid twoway graphs” from the Graphics menu.

When calculating the density in (2), it is convenient to calculate $\ln(f_{\hat{\theta}}(t))$ first and then $\exp[\ln(f_{\hat{\theta}}(t))]$. Note that Stata has a function `lngamma(x)` that calculates the log of the gamma function. The normal density you can calculate with the function *normalden* (use the command “help normalden” for info on this, or “help functions” for more on functions.]

Exercise 3

Refer to supplementary exercise 1-3 (on the net) that compare incomes (1998) for Norwegian women and men. In particular, for incomes larger than $b = 250\,000$ kr, the mle’s for θ are given in table 1, assuming the Pareto model in exercise 2b above:

Table 1: Incomes $> b = 250\,000$ kr. Data samples from Norway 1998

	Sample size n	$\bar{Y} = (1/n) \sum \ln(X_i/b)$	MLE $\hat{\theta}_n$
Women	2361	0.262	3.813
Men	7280	0.438	2.283

- a. Calculate approximate 95% confidence intervals (CI) for θ both for women and for men using results from exercise 2.

[**Hints:** From ex. 2d above we have that $\hat{\theta}_n$ is approximately $N(\theta, \theta^2/n)$ distributed, which means, more precisely, that

$$V_n = \sqrt{n} \frac{\hat{\theta}_n - \theta}{\theta} \xrightarrow[n \rightarrow \infty]{D} Z \sim N(0,1)$$

Use Slutsky's lemma (explain how) to conclude that also

$$Z_n = \sqrt{n} \frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \xrightarrow[n \rightarrow \infty]{D} Z \sim N(0,1)$$

Use the last result to derive an approximate 95% CI for θ .]

[**Note.** As is common in the literature whenever the estimator has a normal limit distribution, the expression θ/\sqrt{n} (i.e., the square root of the second parameter in the limit distribution) is called the *asymptotic standard error* of the estimator $\hat{\theta}_n$. The term *standard error* used in statistical computer packages usually refers to an estimated version of this, i.e., $\hat{\theta}_n/\sqrt{n}$ in this case where θ is replaced by a consistent estimate.]

b. The Gini-coefficient (see supplementary exercise 3) in the Pareto model (well defined for $\theta > 1$) becomes

$$\gamma = \frac{1}{2\theta - 1}$$

Use the results from **a.** to calculate approximate 95% CI's for the Gini-coefficients for women and men incomes above 250 000 kr.

[**Hint:** Let $[L, U]$ be an approximate $1 - \alpha$ CI for a parameter θ , satisfying

$$P(L \leq \theta \leq U) \approx 1 - \alpha$$

Let $\gamma = g(\theta)$ be a transformed parameter where $y = g(x)$ is a strictly decreasing function. Show by drawing a suitable graph that

$$P(L \leq \theta \leq U) = P(g(U) \leq \gamma \leq g(L)) \quad]$$