This is a take home midterm. You can use your notes, books or any other source you need. You are supposed to work on your own text without external help. I'll be available to answer question in person or via email.

Name (print): ______________________________

Question 1 .... *130 point*

Vilfredo Pareto (15 July 1848 – 19 August 1923) argued that in any country the fraction n of the population with income larger than x follows the law:

$$\log n = A(\log(x_m) - \log(x))$$

for $x > x_m$, Here A is a parameter. This law is quite well verified experimentally (data available on wikipedia). Pareto justification of this law is not accepted today but research is active in the field.

In modern term we say that a r.v. X has a Pareto distribution with minimum x_m and shape A if:

$$f(x|x_m, A) = \frac{Ax_m^A}{x^{A+1}}$$

where we assume $A > 1$.

This law is at the basis of many discussions on economic inequality due to the following property:

(a) (10 points) Show that the wealthiest quintile of the population own a fraction $(0.2)^{\frac{A-1}{A}}$ of the total wealth.

Solution: The fourth quintile $q(0.8)$ is defined by

$$\int_{q(0.8)}^{\infty} f(x|x_m, A)dx = 0.2 \qquad \Longrightarrow \qquad \frac{x_m}{q(0.8)} = (0.2)^{\frac{1}{A}}.$$

Assume the total population size is N. Then the number of individuals with wealth in $(x, x + dx)$ is $f(x|x_m, A)dx$. Thus the total wealth W is

$$W = N \int_{x_m}^{\infty} xf(x|x_m, A)dx = N\frac{A}{A-1}x_m$$

while the wealth $W(0.8)$ of the upper quintile is

$$W(0.8) = N \int_{q(0.8)}^{\infty} xf(x|x_m, A)dx = N\frac{A}{A-1}x_m \left(\frac{x_m}{q(0.8)}\right)^{A-1}$$

Thus we get

$$\frac{W(0.8)}{W} = (0.2)^{\frac{A-1}{A}}$$

If $A = 1.16$, pretty close to the experimental value, you get that the above fraction is 0.8. This is the so called “20-80 rule”, that is the welthiest 20% of the population owns 80% of the total wealth.

Let's do some statistics with the Pareto distribution assuming that x_m is known.

(b) (10 points) Show that the family of Pareto distributions with given minimum x_m and unknown shape A is conjugated to the $\Gamma(\alpha,\beta)$ family of prior distributions. Find the updating rule for α and β.

Solution: Assume that the prior distribution is $\Gamma(\alpha,\beta)$ that is

$$g_0(A) \propto A^{\alpha-1}e^{-\beta A}.$$

Given one observation x we have that the posterior is

$$g_1(A|x) \propto A^{\alpha}e^{-\beta A}\left(\frac{x_m}{x}\right)^A \frac{1}{x} \propto A^{\alpha}e^{-\beta A-\log\left(\frac{x}{x_m}\right)A}$$

since $\frac{1}{x}$ is a constant as a function of A. Thus the updating is

$$\alpha \to \alpha+1 \qquad \beta \to \beta + \log\left(\frac{x}{x_m}\right).$$

Clearly for a sample of size N we get

$$\alpha \to \alpha+n \qquad \beta \to \beta + \sum_{i=1}^{N}\log\left(\frac{x_i}{x_m}\right).$$

(c) (10 points) You decide to use a quadratic loss function and you run a sample of size N. If $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ is your sample, find the Bayes estimator $\hat{A}_B(\mathbf{X})$ for A.

Solution: We know that for quadratic loss function the Bayes estimator is the expected value of the posterior distribution. Thus we get:

$$\hat{A}_B(\mathbf{X}) = \frac{\alpha+N}{\beta+\sum_{i=1}^{N}\log\left(\frac{X_i}{x_m}\right)}$$

(d) (10 points) Under the condition of part (c), find the MLE estimator $\hat{A}_{MLE}(\mathbf{X})$ for A.

Solution: The likelihood function is:

$$\ell(\mathbf{x}|A) = \prod_{i=1}^{N} f(x_i|A) = A^N \prod_{i=1}^{N} \left(\frac{x_m}{x_i}\right)^{A+1} x_m^{-N}.$$

To find the maximum in A we first neglect the constant x_m^{-N}, then take the logarithm and differentiate. This give us

$$\frac{N}{A} + \sum_{i=1}^{N} \log\left(\frac{x_m}{x_i}\right) = 0.$$

Thus we get

$$\hat{A}_{MLE}(\mathbf{X}) = \frac{N}{\sum_{i=1}^{N} \log\left(\frac{X_i}{x_m}\right)}$$

(e) (10 points) Is there an improper prior you can use? Discuss the updating rule and the relation between the Bayes and the MLE estimator in this case.

Solution: If we choose $\alpha = \beta = 0$ as a prior in point (c) we see that $\hat{A}_B = \hat{A}_{MLE}$, that is the updating rule when new data are collected consists in computing the MLE estimator of the new data together any previous data available. Clearly this is not a real prior since for $\alpha = \beta = 0$ that Gamma distribution is not a distribution.

(f) (10 points) Under the condition of part (c), find the moment estimator $\hat{A}_M(\mathbf{X})$ for A.

Solution: In point one we found out that

$$\mathbb{E}(X) = \frac{A}{A-1} x_m$$

thus $\hat{A}_M$ is defined by the equation

$$\frac{\hat{A}_M}{\hat{A}_M - 1} x_m = \bar{X}_N$$

that is

$$\hat{A}_M = \frac{\bar{X}_N}{\bar{X}_N - x_m}.$$

(g) (10 points) Do $\hat{A}_M$ form a consistent sequence of estimators when the sample size $N \to \infty$? Observe that, if $A \leq 2$, then $\sigma_X = +\infty$ and you cannot use Theorem 6.3.4.

Solution:

If $A > 2$, we cab use Theorem 6.3.4 to show that $\bar{X}_N$ converges in probability to $\mathbb{E}(X)$. Moreover, since $g(x) = \frac{x}{x - x_m}$ is a continuous function if $x > x_m$, we have that

$$\hat{A}_M = g(\bar{X}_N) \to_p g(\mathbb{E}(X)) = A$$

so that $\hat{A}_M$ form a consistent sequence of estimators.

If $A < 2$, it is still true that $\bar{X}_N$ converges in probability to $\mathbb{E}(X)$. Indeed the law of large number is true even if the variance is infinite. Thus $\hat{A}_M$ still form a consistent sequence of estimators.

The major difference is that if σ_X is finite, the proof of the law of large number via Chebyshev inequality gives an explicit bound on the probability that $\bar{X}$ may turn out to be away from $\mathbb{E}(X)$. The proof used in the case $\sigma_X = \infty$, based on Fourier transform techniques, does not readily provide such a bound.

(h) (10 points) Discuss the asymptotic distribution of $\hat{A}_{MLE}(\mathbf{X})$. Use the extra material posted on the class webpage taken from R.V. Hoggs and E.A. Tanis "Probability ans Statistical Inference".

Solution:
We need to compute

$$\sqrt{-\mathbb{E}\left(\partial_A^2 \log(f(X|A, x_m))\right.}.$$

We find

$$\log(f(x|A, x_m) = \log(A) + A\log(x_m) - (A+1)\log(x)$$

so that

$$-\partial_A^2 \log(f(x|A, x_m) = \frac{1}{A^2}$$

So that the asymptotic value of the standard deviation of $\hat{A}_{MLE}(\mathbf{X})$ can be estimated as $\frac{A}{\sqrt{N}}$. This predict that the asymptotic distribution of $\hat{A}_{MLE}(\mathbf{X})$ is normal with mean A and variance $\frac{A^2}{N}$.

(i) Use the delta method (see page 364 of the textbook) to discuss the asymptotic distribution of $\hat{A}_M(\mathbf{X})$. Keep in mind the comment in point (g).

Solution: If $A > 2$ we have that $\sigma_X < \infty$ so that

$$\sqrt{N}(\bar{X}_N - \mathbb{E}(X)) \to_d \mathcal{N}(0, \sigma_X^2).$$

Consider the function $g(x) = \frac{x}{x - x_m}$. Using the delta method and the fact that $\mathbb{E}(X) = x_m \frac{A}{A-1}$, we get that

$$\frac{\sqrt{N}}{g'\left(x_m \frac{A}{A-1}\right)}(\hat{A}_M - A) \to_d \mathcal{N}(0, \sigma_X^2).$$

Finally, using that $\sigma_X = \frac{A}{A-2} x_m^2$, gives

$$\sqrt{N}(\hat{A}_M - A) \to_d \mathcal{N}\left(0, \frac{A(A-1)^2}{A-2}\right).$$

or that $\hat{A}_M$ asymptotic distribution is normal with mean A and variance $\frac{A(A-1)^2}{N(A-2)}$. This clearly makes sense only if $A > 2$.

(j) (10 points) Compute $\mathbb{E}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})}\right)$ and $\mathrm{var}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})}\right)$. Using the CLT find the approximate asymptotic distribution of $\frac{1}{\hat{A}_{MLE}(\mathbf{X})}$. (**Hint** You can use integration by parts or the substitution $y = \log(x)$ to compute $\int_{x_m}^{\infty} \log(x) x^{-(A+1)} dx$ and $\int_{x_m}^{\infty} \log(x)^2 x^{-(A+1)} dx$.)

Solution: Since the variables in the sample are independent and

$$\frac{1}{\hat{A}_{MLE}(\mathbf{X})} = \frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{X_i}{x_m}\right),$$

we have

$$\mathbb{E}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})}\right) = \mathbb{E}\left(\log\left(\frac{X}{x_m}\right)\right) = \int_{x_m}^{\infty} \log\left(\frac{x}{x_m}\right) \frac{A x_m^A}{x^{A+1}} dx = \frac{1}{A}.$$

In the same way we get

$$\mathrm{var}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})}\right) = \frac{1}{N} \mathrm{var}\left(\log\left(\frac{X}{x_m}\right)\right).$$

After observing that

$$\int_{x_m}^{\infty} \log\left(\frac{x}{x_m}\right)^2 \frac{A x_m^A}{x^{A+1}} dx = \frac{2}{A^2}$$

we get

$$\mathrm{var}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})}\right) = \frac{1}{A^2}.$$

Using the CLT we can conclude that $\frac{1}{\hat{A}_{MLE}(\mathbf{X})}$ has an approximate normal distribution with mean $\frac{1}{A}$ and standard deviation $\frac{1}{A\sqrt{N}}$. More precisely we should say that

$$\sqrt{N}\left(\frac{1}{\hat{A}_{MLE}(\mathbf{X})} - \frac{1}{A}\right) \to_d \mathcal{N}\left(0, \frac{1}{A^2}\right)$$

(k) (10 points) Use the delta method (see page 364 of the textbook) to find the approximate distribution of $\hat{A}_{MLE}$. What can you say about the asymptotic efficiency of $\hat{A}_{MLE}$? Use the Rao-Cramèr Inequality from R.V. Hoggs and E.A. Tanis "Probability ans Statistical Inference".

Solution: As for point (i), using $g(x) = \frac{1}{x}$, we get that

$$\sqrt{N}\left(\hat{A}_{MLE} - A\right) \to_d \mathcal{N}\left(0, A^2\right).$$

Thus, in this case, we can show directly that the MLE estimator is asymptotically the MVUE and that Rao-Cramèr efficiency is 1.

(l) (10 points) Show that if Y is uniform in $[0, x_m^{-A}]$ then $X = Y^{-\frac{1}{A}}$ has a Pareto distribution with minimum x_m and shape A. Use this fact to create a random number generator for a Pareto distribution (assume only that you have a random number generator for a uniform distribution in $[0, 1]$.)

Solution: Observe that, if Y is uniform in $[0, x_m^{-A}]$ then

$$\mathbb{P}(X \geq x) = \mathbb{P}(Y < x^{-A}) = \frac{x_m^A}{x^A}$$

so that X has a Pareto distribution with minimum x_m and shape A.

(m) (10 points) Using Matlab or your favourite programming language, write a code that generates a random sample of size N from a Pareto distribution with minimum x_m and shape A and computes $\hat{A}_{MLE}$ and $\hat{A}_M$ from the sample.

(n) (10 points) Use your code to numerically compute $\mathbb{E}(\hat{A}_M)$, $\mathrm{var}(\hat{A}_M)$, $\mathbb{E}(\hat{A}_{MLE})$, and $\mathrm{var}(\hat{A}_{MLE})$ when $A = 1.2, 2.2, 3.2$ and $N = 10, 20, 30, \ldots, 100$. Compare with the theoretical results.