1. A. $Z_1, \cdots, Z_N$ are i.i.d. Pareto(θ) with density $p_\theta(x) = \theta x^{-(\theta+1)} 1_{[1,\infty)}(x)$, $\theta > 0$, and consider the joint density $p_\theta(\underline{z})$. Is $\{p_\theta : \theta \in (0, \infty)\}$ an exponential family? If so, identify the various components in the exponential family form explicitly.

Solution: Now the density of one z is

$$\begin{aligned} p_\theta(z) &= \theta z^{-(\theta+1)} 1_{[1,\infty)}(z) \\ &= \theta \exp(-(\theta+1)\log(z)) 1_{[1,\infty)}(z) \\ &= \theta \exp((\theta+1)\log(1/z)) 1_{[1,\infty)}(z), \end{aligned}$$

and hence the joint density of $\underline{Z} = (Z_1, \cdots, Z_N)$ is given by

$$\begin{aligned} p_\theta(\underline{z}) &= \prod_{i=1}^N p_\theta(z_i) = \theta^N \exp((\theta+1)\sum_{i=1}^N \log(1/z_i)) 1_{[1,\infty)}(\min_{1\le i\le N} z_i) \\ &= c(\theta)\exp(Q(\theta)T(\underline{z}))h(\underline{z}) \end{aligned}$$

where $c(\theta) = \theta^N$, $Q(\theta) = \theta + 1$, and $T(\underline{z}) = \sum_{i=1}^N \log(1/z_i)$, and $h(\underline{z}) = 1_{[1,\infty)}(\min z_i)$. This is an exponential family.

B. Now suppose that we condition on $(Z_{(1)}, \cdots, Z_{(N)}) = \underline{z}$, the order statistics of the Z_i's in A, and we sample $n < N$ of the z_i's without replacement; call the resulting sample $Y_1, \cdots, Y_n$.

(i) What are the (conditional) mean μ_N and variance σ_N^2 of $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$?

(ii) Under what further conditions does $(\overline{Y}_n - \overline{z}_N)/\sigma_N \to_d N(0,1)$ in probability under the model in A? Explain what this means and, in particular, specify for what values of θ the asymptotic normality holds and for what values of θ the asymptotic normality fails.

Solution: (i) $\mu_N \equiv E\,\overline{Y}_n = \overline{z}_N = N^{-1}\sum_{i=1}^N z_i$, and

$$\sigma_N^2 \equiv Var(\overline{Y}_n) = (1 - \frac{n-1}{N-1})\frac{\sigma_z^2}{n}$$

where $\sigma_z^2 \equiv N^{-1}\sum_1^N (z_i - \overline{z})^2$.

(ii) If $0 < \liminf(n/N) \le \limsup(n/N) < 1$ and

$$\eta_N \equiv \frac{\max_{1\le i\le N}|Z_i - \overline{Z}|^2}{\sum_1^N |Z_i - \overline{Z}|^2} \to_p 0,$$

then the WWNH finite - sampling CLT is in force (in probability), and the Kolmogorov - distance between the conditional distribution function of $(\overline{Y}_n - \overline{z}_N)/\sigma_N$ given Z and the standard normal d.f. Φ converges to zero in probability:

$$(1) \qquad \sup_{-\infty < x < \infty} \left| P\left(\frac{\overline{Y}_n - \overline{z}_N}{\sigma_N} \le x \mid \underline{Z} \right) - \Phi(z) \right| \to_p 0 .$$

But in order for $\eta_N \to_{a.s.} 0$ we need $E|Z_1|^2 < \infty$, and

$$EZ_1^2 = \int_1^\infty \theta z^2 z^{-\theta - 1}\, dz = \theta \int_1^\infty z^{-\theta + 1}\, dz = \frac{\theta}{\theta - 2} < \infty$$

only if $\theta > 2$. Thus for $\theta > 2$ the convergence in (1) holds almost surely. If $\theta \le 2$, then $E|Z_1|^2 = \infty$, and we suspect that that the convergence in (1) fails, even in probability.

C. Under K_1 the joint density of $\underline{X}, \underline{Y}$ is given by

$$\begin{aligned} h(\underline{x}, \underline{y}) &= \theta_1^m \theta_2^n \prod_{i=1}^m x_i \prod_{j=1}^n y_j \exp(- \theta_1 \sum_{i=1}^m \log x_i - \theta_2 \sum_{j=1}^n \log y_j) \\ &= \theta_1^m \theta_2^n \prod_{i=1}^m x_i \prod_{j=1}^n y_j \exp(- (\theta_2 - \theta_1) \sum_{j=1}^n \log y_j - \theta_1 (\sum_{i=1}^n \log x_i + \sum_{j=1}^n \log y_j) \end{aligned}$$

Hence rejecting for those permutations with large valued of $h(\underline{z}')$ is equivalent to rejecting for small valued of $(\theta_2 - \theta_2) \sum_{j=1}^n \log Y_j$, or, for large valued of $\sum_{j=1}^n \log Y_j$, or, for large values of $\overline{\log Y} - \overline{\log z} = (m/N)(\overline{\log Y} - \overline{\log X})$. Thus a most powerful similar test of H_c versus K_1 is given by

$$\phi(\underline{X}, \underline{Y}) = \begin{cases} 1 & \text{if } \overline{\log Y} - \overline{\log X} > c_\alpha(\underline{Z}) \\ 0 & \text{if } \overline{\log Y} - \overline{\log X} \le c_\alpha(\underline{Z}) \end{cases}$$

where $c_\alpha(\underline{Z})$ is determined so that exactly $\alpha N!/m!n!$ of the $N!/m!n!$ assignments of the $\underline{Z}$'s to be $\underline{X}$'s and $\underline{Y}$'s lead to rejection.

2. Suppose that $X_1, \cdots, X_n$ are i.i.d.

$$p_\theta(x) = p(x; \theta) = f(x - \theta)$$

where $f(x)$ is the logistic density

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2} .$$

A. Use the generalized NP lemma to find the locally most powerful test of $H : \theta = 0$ versus $K : \theta > 0$ and show how to approximate the appropriate critical points to carry out your test.

B. If instead of the null hypothesis $H : \theta = 0$ in A, we consider testing $H_s : X_1, \cdots, X_n$ are i.i.d. $F \in \mathbf{F}_s$, the collection of all distribution functions symmetric about 0 and seek a locally most powerful test against the logistic alternatives, how would you proceed to construct a LMP level α similar test?

Solution: A. We want to choose ϕ to maximize

$$\frac{d}{d\theta}\,\beta_\phi(\theta)|_{\theta=\theta_0} \;=\; \frac{d}{d\theta}\,E_\theta\phi(\underline{X})|_{\theta=\theta_0}$$

subject to $E_{\theta_0}\phi(\underline{X}) = \alpha$. If the interchange of differentiation and expectation can be justified (and in the present case it can be justified quite easily), this is equivalent to: maximize

$$\int\cdots\int \phi(\underline{x})\,\frac{\partial}{\partial\theta}\,p_\theta(\underline{x})|_{\theta=\theta_0}\,d\underline{x}$$

subject to $E_{\theta_0}\phi(\underline{X}) = \alpha$. But the generalized NP lemma tells us that the following test ϕ solves this problem:

$$\phi(\underline{x}) \;=\; \begin{cases} 1 & \text{if } \frac{\partial}{\partial\theta}\,p_\theta(\underline{x})|_{\theta=\theta_0} \;>\; kp_{\theta_0}(\underline{x}) \\ \gamma & \text{if } \frac{\partial}{\partial\theta}\,p_\theta(\underline{x})|_{\theta=\theta_0} \;=\; kp_{\theta_0}(\underline{x}) \\ 0 & \text{if } \frac{\partial}{\partial\theta}\,p_\theta(\underline{x})|_{\theta=\theta_0} \;<\; kp_{\theta_0}(\underline{x}) \end{cases}.$$

Since

$$\dot{\mathbf{l}}_{n\theta}(\underline{x};\theta_0) \;=\; \frac{\partial}{\partial\theta}\log p_\theta(\underline{x};\theta)|_{\theta=\theta_0} \;=\; \frac{1}{p(\underline{x};\theta)}\,\frac{\partial}{\partial\theta}\,p(\underline{x};\theta)|_{\theta=\theta_0} \;=\; \sum_{i=1}^{n}\dot{\mathbf{l}}_\theta(x_i;\theta_0)\,,$$

this test ϕ is equivalent to

$$\phi(\underline{x}) \;=\; \begin{cases} 1 & \text{if } S_n(0) \;>\; k' \\ \gamma & \text{if } S_n(0) \;=\; k' \\ 0 & \text{if } S_n(0) \;<\; k' \end{cases}$$

where $S_n(0) \equiv n^{-1/2}\sum_{i=1}^{n}\dot{\mathbf{l}}_\theta(x_i;\theta_0)$. Since $E_{\theta_0}\dot{\mathbf{l}}_\theta(X_i;\theta_0) = 0$ and $E_{\theta_0}\dot{\mathbf{l}}_\theta^2(X_i;\theta_0) = I(\theta_0)$, we can approximate the constant k' for large sample sizes by $z_\alpha\sqrt{I(\theta_0)}$. In the particular present case, $p_\theta(x) = f(x-\theta)$, so $\dot{\mathbf{l}}_\theta(x) = -(f'/f)(x-\theta)$ where $\log f(x) = -x - 2\log(1+e^{-x})$ and hence

$$\begin{aligned} -(f'/f)(x) \;&=\; 1 - 2e^{-x}/(1+e^{-x}) \;=\; (1-e^{-x})/(1+e^{-x}) \\ &=\; F(x) - (1-F(x)) \;=\; 2(F(x)-1/2) \end{aligned}$$

where F is the distribution function corresponding to f. Hence

$$\begin{aligned} I(\theta) \;&=\; \int\left(\frac{f'}{f}\right)^2 f\,dx \;=\; 4\int (F(x)-1/2)^2\,dF(x) \\ &=\; 4Var[F(X)] \;=\; \frac{4}{12} \;=\; 1/3\,. \end{aligned}$$

Thus the natural large sample approximation to the locally MP test is "reject $\theta = 0$ if $n^{-1/2}\sum_{i=1}^{n}\dot{\mathbf{l}}_\theta(X_i;0) > \sqrt{1/3}z_\alpha$ " where

$\dot{\mathbf{l}}_\theta(x;0) \;=\; (1-e^{-x})/(1+e^{-x})\,.$

B. First consider the situation under the null hypothesis $F \in \mathbf{F}_s$. Let $sign(\underline{X}) \;=\; (sign(X_1)\,,\cdots,sign(X_n))$ where

$$sign(x) \;=\; \begin{cases} 1 & \text{if} \quad x \;>\; 0 \\ 0 & \text{if} \quad x \;=\; 0 \\ -1 & \text{if} \quad x \;<\; 0 \end{cases},$$

$\underline{R}^+ \;=\; (R_1^+\,,\cdots,R_n^+)$ where $R_i^+ \;\equiv\; \#\{j \le n:\; |X_j| \;\le\; |X_i|\,\}$, and $|\underline{X}|_{(\cdot)} \;=\; (|X|_{(1)}\,,\cdots,|X|_{(n)})$, the order statistics of absolute values. When the null hypothesis $F \in \mathbf{F}_s$ is true, $|\underline{X}|_{(\cdot)}$ is a sufficient statistic for F. Note that

$$\begin{aligned} P_F(|X| \le x) \;&=\; P(\,-x \le X \le x\,) \\ &=\; F(x) \;-\; F(-x) \;=\; F(x) \;-\; (1 - F(x)) \;=\; 2F(x) - 1 \\ &\equiv\; F^+(x) \end{aligned}$$

where symmetry of F has been used in the third equality. Thus

$$I\!F_n^+(x) \;\equiv\; \frac{1}{n}\sum_{i=1}^{n} 1_{[|X_i| \,\le\, x]}$$

estimates F^+ and $(I\!F_n^+ + 1)/2$ estimates F consistently. Also note that $sign(X)$ and $|X|$ are independent:

$$\begin{aligned} P_F(sign(X) = 1\,,\, |X| \le x) \;&=\; P(0 < X \le x) \;=\; F(x) - 1/2 \\ &=\; \frac{1}{2}\,(2F(x) - 1) \;=\; P(sign(X) = 1)P(|X| \le x) \end{aligned}$$

and similarly for -1. Thus $sign(\underline{X})$ and $|\underline{X}|$ are independent. Furthermore $\underline{R}^+$ and $|\underline{X}|_{(\cdot)}$ are independent from problem 1.4. Thus under $F \in \mathbf{F}_s$,

$$P(sign(\underline{X}) \;=\; \underline{v}) \;=\; \frac{1}{2^n}\,,$$

for all $\underline{v} \in \mathbf{V} \;\equiv\; \{v = (v_1\,,\cdots,v_n):\; v_i \in \{\,-1\,,1\}\}\,;$

$$P(\underline{R}^+ \;=\; r) \;=\; \frac{1}{n!}$$

for all $\underline{r} \in \Pi$, and if the X_i's have density f, then $|\underline{X}|_{(\cdot)}$ has density

$$2^n n! \prod_{i=1}^{n} f(x_i) 1_{\{0 \,<\, x_1 \,<\, \cdots \,<\, x_n\}}\,.$$

Note that $X_i \;=\; signX_i|X|_{(R_i^+)}$. Thus there is a one-to-one correspondence between $\underline{X}$ and $(sign(\underline{X})\,,\underline{R}^+\,,|\underline{X}|_{(\cdot)})$.

Now we turn to the alternatives. Conditioning on the sufficient statistic $|\underline{X}|_{(\cdot)})$, the random part left in the data is $(sign(\underline{X})\,,\underline{R}^+)$; and the conditional distribution of $\underline{X}$ given $|\underline{X}|_{(\cdot)}) = \underline{z}$ under a particular alternative $h(\underline{x}) \;=\; \prod_1^n f(x_i)$ is given by

$$\frac{h(\underline{z}')}{\sum_{\underline{z}''} h(\underline{z}'')}$$

where the sum is over all $2^n n!$ possible permutations and sign changes of $\underline{z}$. When $h = p_\theta$ is given by $p_\theta(\underline{x}) = \prod_{i=1}^n f(x_i - \theta)$ and we seek a locally most powerful similar test, we are naturally lead by use of the generalized Neyman - Pearson lemma, to rejection of H_s for large values of

$$\frac{\partial}{\partial\theta} \log \left\{ \frac{p_\theta(\underline{z}')}{\sum_{\underline{z}''} p_\theta(\underline{z}'')} \right\} |_{\theta=0}$$

$$= \dot{\mathbf{l}}_{n\theta}(\underline{z}', 0) - E\dot{\mathbf{l}}_{n\theta}(\underline{Z}', 0)$$

where $\underline{Z}' \equiv (V_1 z_{R_1^+}, \cdots, V_n z_{R_n^+})$ with $P(\underline{V} = \underline{v}) = 2^{-n}$, $\underline{v} \in \mathbf{V}$, $P(\underline{R}^+ = \underline{r}) = 1/n!$, $\underline{r} \in \Pi$. Thus we see that the locally most powerful similar test of H_s versus logistic shift alternatives is given by the "permutation version" of the score test derived in part A:

$$\phi(\underline{X}) = 1\{\dot{\mathbf{l}}_{n\theta}(\underline{X}; 0) > c_\alpha(|\underline{X}|_{(\cdot)})\}$$

where $c_\alpha(|\underline{X}|_{(\cdot)})$ is chosen so that we reject for exactly $\alpha n! 2^n$ of the $n! 2^n$ permutation and sign changes of $|\underline{X}|_{(\cdot)}$.