STA3020 Assignment 1.

[1]*School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)*

1. Change of Measure

•*Exercise* 1.1. Assume $Z_1, Z_2, \cdots, Z_{n+1}$ are i.i.d random variables with standard exponential distribution. Let

$$Y = (Y_1, \cdots, Y_n) = \left(\frac{Z_1}{\sum_{i=1}^{n+1} Z_i}, \frac{Z_1 + Z_2}{\sum_{i=1}^{n+1} Z_i}, \cdots, \frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n+1} Z_i} \right),$$

please give the joint distribution of $(Y_1, \cdots, Y_n)$.

Hint. Define the simplex

$$\bar{\mathcal{S}}_n = \Big\{ (y_1, \cdots, y_n) : 0 \le y_1 \le \cdots \le y_n \le 1 \Big\},$$

1. First we look at the joint distribution of $Y' = (Y'_1, \cdots, Y'_{n+1}) = (Z_1, Z_1 + Z_2, \cdots, \sum_{i=1}^{n+1} Z_i)$.
2. Second, we look at the joint distribution of $Y'' = (Y_1, \cdots, Y_n, Y_{n+1}) = (Y_1, \cdots, Y_n, Y'_{n+1})$.
3. Third, we look at the joint distribution of Y.

$\square$

2. Delta Method

•*Exercise* 2.1. Let $X_1, \cdots, X_n \overset{i.i.d}{\sim} N(\theta, \theta)$ with some $\theta > 0$. Apparently, $\bar{X} = \sum_{i=1}^{n} X_i / n$ is a "good" estimator for θ, where based on the central limit theorem, we have

$$\frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\theta}} \xrightarrow{d} N(0, 1).$$

Please give a $1 - \alpha$ confidence interval for θ using Delta methods.

3. Exponential Family

•*Exercise* 3.1. In a Bernoulli sequence of trials with success probability p, let $X+m$ be the number of trials required to achieve m successes.

1. Show that the distribution of X.
2. Verify that the above probabilities mass (negative binomial distribution) add up to 1 by expanding
$$1=\left(\frac{1}{p}-\frac{q}{p}\right)^{-m}.$$
3. Show that the derived distributions above constitute a one-parameter exponential family.
4. Please give the moment generating function of X.
5. Please give the mean and the variance of X.
6. let X_i+1 be the number of trials required after the $(i-1)$-th success has been obtained until the next success occurs. Use the fact that $X=\sum_{i=1}^m X_i$ to find an alternative derivation of the mean and variance of X.

•*Exercise* 3.2. Efron (1975) generalized the definitions of curvature in expoential distribution. For X follows a k-dimensional expoential family

$$f(x|\theta)=h(x)c(\theta)\exp\left(\sum_{i=1}^k\omega_i(\theta)T_i(x)\right)$$

with the covariance matrix of $T=(T_1(X),\cdots,T_k(X))$ being denoted as Σ_θ. if θ is a scalar,define the statistical curvature to be $\gamma_\theta=(|M_\theta|/m_{11}^3)^{1/2}$ where

$$M_\theta=\begin{pmatrix}m_{11} & m_{12}\\ m_{21} & m_{22}\end{pmatrix}=\begin{pmatrix}\dot\omega_\theta^T\Sigma_\theta\dot\omega_\theta & \dot\omega_\theta^T\Sigma_\theta\ddot\omega_\theta\\ \ddot\omega_\theta^T\Sigma_\theta\dot\omega_\theta & \ddot\omega_\theta^T\Sigma_\theta\ddot\omega_\theta\end{pmatrix}$$

with $\omega_\theta=(\omega_1(\theta),\cdots,\omega_k(\theta))^T$, $\dot\omega_\theta=(\frac{\partial\omega_1(\theta)}{\partial\theta},\cdots,\frac{\partial\omega_k(\theta)}{\partial\theta})^T$ and

$$\ddot\omega_\theta=(\frac{\partial^2\omega_1(\theta)}{(\partial\theta)^2},\cdots,\frac{\partial^2\omega_k(\theta)}{(\partial\theta)^2})^T.$$

Please calculate the statistical curvature of the family

$$f(x|\theta)=C\exp\big[-\sum_{i=1}^n(x_i-\theta)^m\big]$$

for $m=2,3,4$.

•*Exercise* 3.3. For a independent sample $\{(X_i,Y_i)\}_{1\le i\le n}$ where $\{X_i\}_{1\le i\le n}$ are i.i.d follow $N(0,1)$ and the conditional distribution of $Y_i|X_i$ has density

$$f(y_i|x_i,\alpha,\beta)=\exp\Big(y_i\eta_i-\phi(\eta_i)\Big)h(y_i),\ \text{ with }\ \eta_i=\alpha+x_i\beta.$$

1. Please write out the log-likelihood $\ell(\alpha,\beta)$ and give the MLE of (α,β).
2. Please give the procedure of Newton's method to obtain the MLE of (α,β).
3. Please use Stein's identity to simplify the procedure of Newton's method.

4. Location-Scale Family

•*Exercise* 4.1. If X is distributed according to uniform$(0,\theta)$, show that $-\log X$ form a location family.