# University of Toronto
# STA261 (Summer 2021) - Quiz 4
# August 3, 2021

- Quiz 4 is open-notes and consists of three questions. There are 15 points available. Take a quick scan through the questions first and prioritize your time accordingly.
- Show all of your work for full marks, and ensure your notation is legible, correct, and consistent with that used in the course. Be sure to clearly distinguish between random variables and constants, and between vectors and scalars (you can write $\mathbf{X}$ as $\vec{X}$ and $\mathbf{x}$ as $\vec{x}$).
- Upload clear, legible photos/screenshots of your handwritten answers to the questions within the time window, one question at a time.
- You may refer to the lecture slides and assignments. If you need to use a result from them, either refer to it by its number or its name (if it is numbered/named), or describe the result.

Good luck!

---

1. (5 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with pdf

$$f_\theta(x) = \theta x^{\theta-1} e^{1-x^\theta}, \quad x > 1, \quad \theta > 0.$$

(a) (2 points) Show that $X_i^\theta$ is a pivotal quantity, for any $i$.

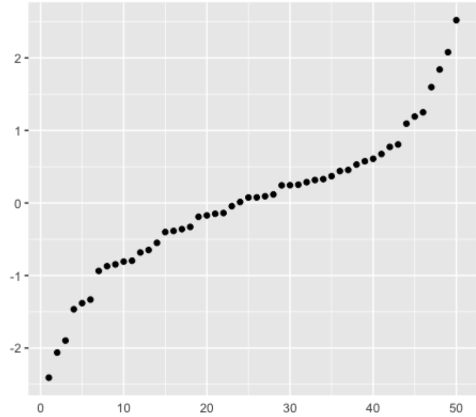
(b) (1 point) Explain why this implies that $Q(\mathbf{X}, \theta) := \left(\prod_{i=1}^n X_i\right)^\theta$ is also a pivotal quantity.

(c) (1 point) Invert the statement "$a < Q(\mathbf{X}, \theta) < b$" into one of the form "____ $< \theta <$ ____".

(d) (1 point) Let $F_Q(\cdot)$ be the cdf of $Q(\mathbf{X}, \theta)$, and let $F_Q^{-1}(\cdot)$ be its inverse. Given that $F_Q(1) = 0$ and $F_Q^{-1}(1-\alpha) = 1 - \log\left(1 - (1-\alpha)^{1/n}\right)$, find a $(1-\alpha)$-confidence interval for $\theta$.

2. (5 points) Answer each of the following questions by writing YES or NO (1 point), and justify your answer in *at most* three sentences (1.5 points).

   (a) (2.5 points) I have 50 observations $x_1, x_2, \ldots, x_{50}$ that I suspect arise from some $\mathcal{N}(\mu, \sigma^2)$ distribution. To visually assess this, I compute $r_i^* = (x_i - \bar{x})/\sqrt{\frac{49}{50}s^2}$ and plot the $r_i^*$'s in increasing order:




   Since this looks nothing like a straight line, I conclude that the data probably didn't come from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Have I done anything wrong?

   (b) (2.5 points) A scientific study reports a 95%-confidence interval of $(-2.32, -1.40)$ for some parameter $\theta$. "95%? What nonsense!" your friend exclaims. "Since $\theta$ is a fixed quantity, it either *is* or *isn't* between $-2.32$ and $-1.40$. Since we'll never know which is true, publishing these numbers is simply useless." Is your friend right?

3. (5 points) It turns out that we can easily extend the chi-squared goodness of fit test to test whether our data arises from a parametric model $\{f_\theta : \theta \in \Theta\}$ on a finite sample space with $k$ elements, by taking $p_i = \mathbb{P}_{\hat{\theta}}(\{i\})$, where $\hat{\theta}$ is the MLE of $\theta$. The only difference is that the asymptotic distribution is now $\chi^2_{(k-1-\dim(\Theta))}$, where $\dim(\Theta) = 1$ when $\Theta \subseteq \mathbb{R}$.

   Determine the chi-squared statistic for testing whether a random sample $W_1, W_2, \ldots, W_n$ arises from a $\text{Bin}(2, \theta)$ distribution with pmf $f_\theta(i) = \binom{2}{i}\theta^i(1-\theta)^{2-i}$ for $i \in \{0, 1, 2\}$, and write down its asymptotic distribution under $H_0$. You can write $\bar{W}_n$ for the sample mean, but otherwise, express your test statistic in terms of the $W_i$'s alone. (*Hint*: it should be a sum of three things).