Problem Set 3

Due: May 24, 2021 (in class; subject to change if COVID restrictions apply)

Problem 1 (Gibbs Sampling)

Distributions of sizes and frequencies often tend to follow a Pareto distributions. Examples include:

- wealth of individuals
- size of oil reserves
- size of cities
- word frequency
- return on stocks

The Pareto distribution with shape $\alpha > 0$ and scale $c > 0$ has pdf

$$\text{Pareto}(x|\alpha, c) = \frac{\alpha c^\alpha}{x^{\alpha+1}} 1(x > c)$$

- This is referred to as a power law distribution, because the pdf is proportional to x raised to a power
- α tells us the scaling relationship between the size of cities and their probability of occurring.

  - Let $\alpha = 1$
  - Density looks like $1/x^{\alpha+1} = 1/x^2$
  - Cities with 10,000-20,000 inhabitants occur roughly $10^{\alpha+1} = 100$ times as frequently as cities with 100,000-110,000 inhabitants.
- c is a lower bound on the observed values; c represents the cut off point

- We will use Gibbs sampling to perform inference for α and c.
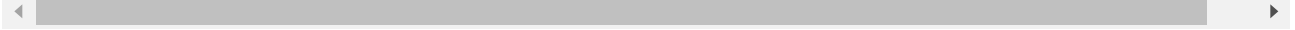
Let us use an improper **prior**:

$$p(\alpha, c) \propto 1(\alpha, c > 0)$$

Note: An improper prior is a nonnegative function of the parameters which integrates to infinity, so it can't really be considered to define a prior distribution. But, we can still plug it into Bayes' formula, and often (but not always!) the resulting "posterior" will be proper—in other words, the likelihood times the prior integrates to a finite value, and so this "posterior" is a well-defined a probability distribution. It is important that the "posterior" be proper, since otherwise the whole Bayesian framework breaks down. Improper priors are often used in an attempt to make a prior as non-informative as possible, in other words, to represent as little prior knowledge as possible. They are sometimes also mathematically convenient.

Plugging into the Bayes' theorem, we define the **posterior** to be proportional to the likelihood times the prior:

$$p(\alpha, c|x_{1:n}) \propto p(x_{1:n}|\alpha, c)p(\alpha, c) \propto 1(\alpha, c > 0) \prod_{i=1}^{n} \frac{\alpha c^{\alpha}}{x_i^{\alpha+1}} 1(x_1 > c) = \frac{\alpha c^{n\alpha}}{(\prod x_i)^{\alpha+1}} 1(c < x_*)1(\alpha, c$$

where $x_* = \min x_1, \ldots, x_n$.

To use Gibbs, we need to be able to sample $\alpha|x_{1:n}$ and $c|x_{1:n}$:

$$\alpha|x_{1:n} \sim \text{Gamma}(n+1, \sum \log x_i - n \log c)$$

$$c|x_{1:n} \sim \text{Mono}(n\alpha + 1, x_*)$$

To see the derivation, please reference: https://jwmi.github.io/BMS/chapter6-gibbs-sampling.pdf.

(1.a) Download the data that is of your interest. It could be any data from the population size of the cities in Spain to income distribution of the U.S. working population, and etc.

(1.b) Estimate the posterior distribuiton of α and c. Show the plot of the distributions.

(1.c) Calculate the mean of the α and c as you increase the iteration size. In other words plot $k \in 1, \ldots,$ number of simulations on x-axis and $\frac{1}{k} \sum_i^k \alpha_i$ on y-axis.

(1.d) Given the estimated α and c, generate data using $Pareto(x|\alpha, c)$. Plot the distribution. Compare the distribution with the data you have downloaded.