

Problem

The Pareto distribution with $\alpha > 0$ and $c > 0$ has p.d.f.

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

This is a power-law distribution in which the parameter α is the “scaling constant”. Suppose x_1, \dots, x_n are the sizes of the populations of the n largest cities in North Carolina. Consider the following model for this data:

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Pareto}(\alpha, c).$$

1. What is the distribution of $Y \mid Y > c$, when $Y \sim \text{Pareto}(\alpha, c_0)$ and $0 < c_0 < c$?
2. There is a potential issue with using an i.i.d. model for this data, due to the fact that we are only seeing the largest n cities — explain the issue. However, supposing that we are interested only in α (not c), how does the answer to the first part alleviate this issue?
3. Derive the Jeffreys prior for α , that is, $\pi(\alpha) \propto \sqrt{I(\alpha)}$ for $\alpha > 0$, where $I(\alpha) = -\mathbb{E}\left(\frac{\partial^2}{\partial \alpha^2} \log p(X_i|\alpha, c)\right)$. Is this an improper prior?
4. Using the Jeffreys prior for α and a flat prior on c , derive the full conditionals for Gibbs sampling from the posterior on α and c . (If you were unable to derive the Jeffreys prior for α , use a semi-conjugate prior.) Identify the full conditional for α as a well-known distribution.
5. The full conditional for c is not a very well-known distribution. Use the inverse CDF method to derive a formula for transforming $\text{Uniform}(0, 1)$ samples into samples from this distribution.
6. Now, suppose we have data on the sizes of the largest cities from all 50 states, say, the largest n_i cities from state i . Using the Pareto distribution as above, write down a hierarchical model in which each state has its own α_i and c_i , using semi-conjugate (i.e., conditionally-conjugate) priors on the α_i ’s and c_i ’s. (Make sure your model enables information sharing among states.) Draw the directed graphical model.

Solution

1. The density of $Y \mid Y > c$ is

$$\begin{aligned} p(y \mid Y > c) &= \frac{p(y)}{\mathbb{P}(Y > c)} \mathbb{1}(y > c) \propto \frac{\mathbb{1}(y > c_0)}{y^{\alpha+1}} \mathbb{1}(y > c) = \frac{\mathbb{1}(y > c)}{y^{\alpha+1}} \\ &\propto \text{Pareto}(y \mid \alpha, c), \end{aligned}$$

hence, $Y \mid Y > c \sim \text{Pareto}(\alpha, c)$.

2. Using only the largest cities is a form of selection bias. This could make the results invalid, since the distribution of the data will depend on how many of the largest cities we decide to look at, and consequently, the i.i.d. model is questionable. Using only the n largest cities can be thought of as conditioning on the set of cities with size above a certain cutoff point c between the n th and $(n+1)$ th largest sizes. If we are only interested in α , then by part 1, if the distribution of all cities (large and small) is Pareto, then the distribution of cities with size greater than c will be Pareto with the same α — thus, inferences for α will still be valid.
3. As a function of α ,

$$\begin{aligned} \log p(x \mid \alpha, c) &= \text{const} + \log \alpha + \alpha \log c - (\alpha + 1) \log x \\ \frac{\partial}{\partial \alpha} \log p(x \mid \alpha, c) &= 1/\alpha + \log c - \log x \\ \frac{\partial^2}{\partial \alpha^2} \log p(x \mid \alpha, c) &= -1/\alpha^2 \end{aligned}$$

and thus,

$$\begin{aligned} I(\alpha) &= -\mathbb{E}\left(\frac{\partial^2}{\partial \alpha^2} \log p(X_i \mid \alpha, c)\right) = 1/\alpha^2 \\ \pi(\alpha) &\propto \sqrt{I(\alpha)} \mathbb{1}(\alpha > 0) = (1/\alpha) \mathbb{1}(\alpha > 0). \end{aligned}$$

This is an improper prior, since $\int_0^\infty (1/\alpha) d\alpha = \infty$.

4. The likelihood is

$$\frac{\alpha^n c^{n\alpha}}{\prod x_i^{\alpha+1}} \mathbb{1}(c < x_*)$$

where $x_* = \min\{x_1, \dots, x_n\}$, so the posterior is

$$p(\alpha, c \mid x_{1:n}) \propto \frac{\alpha^n c^{n\alpha}}{\prod x_i^{\alpha+1}} \mathbb{1}(c < x_*) (1/\alpha) \mathbb{1}(\alpha, c > 0).$$

Hence,

$$p(\alpha|x_{1:n}, c) \propto \alpha^{n-1} \left(\frac{c^n}{\prod x_i} \right)^\alpha \mathbf{1}(\alpha > 0) \propto \text{Gamma}(\alpha \mid n, \sum_i \log(x_i/c))$$

and

$$p(c|x_{1:n}, \alpha) \propto c^{n\alpha} \mathbf{1}(0 < c < x_*).$$

5. Consider a density of the form $p(x) \propto x^{a-1} \mathbf{1}(0 < x < b)$ where $a, b > 0$. (The full conditional for c is of this form.) Since $\int_0^t x^{a-1} dx = t^a/a$, we have

$$p(x) = \frac{ax^{a-1}}{b^a} \mathbf{1}(0 < x < b)$$

and the CDF is

$$F(x) = \int_0^x p(\xi) d\xi = \frac{x^a}{b^a}$$

for $x \in [0, b]$, $F(x) = 0$ for $x < 0$, and $F(x) = 1$ for $x > b$. For $u \in (0, 1)$, we have $u = F(x)$ if and only if $x = bu^{1/a}$. Hence, to sample $X \sim p(x)$, we can draw $U \sim \text{Uniform}(0, 1)$ and set $X = bU^{1/a}$.

6. There is more than one correct answer. Here is one approach: for $i = 1, \dots, 50$, let

$$\begin{aligned} X_{i1}, \dots, X_{in_i} | \alpha_i, c_i &\stackrel{\text{iid}}{\sim} \text{Pareto}(\alpha_i, c_i) \\ \alpha_i | a, b &\sim \text{Gamma}(a, b) \\ c_i | r, s &\sim \text{Gamma}(r, s) \end{aligned}$$

and place independent Gamma priors on a, b, r , and s . Gamma distributions (including truncated Gammas) are semi-conjugate for both the α_i 's and the c_i 's since

$$p(\alpha_i | \dots) \propto \text{Gamma}(\alpha_i \mid a + n_i, b + \sum_j \log(x_{ij}/c_i))$$

and, letting $x_{i*} = \min\{x_{i1}, \dots, x_{in_i}\}$,

$$p(c_i | \dots) \propto \text{Gamma}(c_i \mid r + n_i \alpha_i, s) \mathbf{1}(0 < c_i < x_{i*})$$

which is proportional to a truncated Gamma. Note that it is necessary to place a prior on at least one of a, b, r, s in order to have information shared among states.

