10. The Zeta Distribution

The zeta distribution is used to model the size or ranks of certain types of objects randomly chosen from certain types of populations. Typical examples include the frequency of occurrence of a word randomly chosen from a text, or the population rank of a city randomly chosen from a country. The zeta distribution is also known as the Zipf distribution, in honor of the American linguist George Zipf.

The Zeta Function

The Riemann zeta function $\zeta$, named after Bernhard Riemann, is defined as follows:

$$\zeta(a) = \sum_{n=1}^{\infty} \frac{1}{n^a}, \quad a > 1$$

(You might recall from calculus that the series in the zeta function converges for $a > 1$ and diverges for $a \leq 1$. A graph of the zeta function on the interval $[1, 10]$ is given below:

---

1. Try to verify the main properties of the graph analytically. In particular, show that

   a. $\zeta$ is decreasing.
   b. $\zeta$ is concave upward.
   c. $\zeta(a) \downarrow 1$ as $a \uparrow \infty$
   d. $\zeta(a) \uparrow \infty$ as $a \downarrow 1$

The zeta function is transcendental, and most of its values must be approximated. However, $\zeta(a)$ can be given explicitly for even integer values of $a$; in particular, $\zeta(2) = \frac{\pi^2}{6}$ and $\zeta(4) = \frac{\pi^4}{90}$.
The Probability Density Function

2. Show that the function \( f \) given below is probability density function for any \( a > 1 \).

\[
f(n) = \frac{1}{\zeta(a)n^a}, \quad n \in \mathbb{N}_+
\]

The discrete distribution defined by the density function in Exercise 2 is called the zeta distribution with parameter \( a \). In an algebraic sense, the zeta distribution is a discrete version of the Pareto distribution.

3. Let \( X \) denote the frequency of occurrence of a word chosen at random from a certain text, and suppose that \( X \) has the zeta distribution with parameter \( a = 2 \). Find \( \mathbb{P}(X > 4) \).

4. Suppose that \( X \) has the zeta distribution with parameter \( a \). Show that the distribution is a one-parameter exponential family with natural parameter \( a \) and natural statistic \(-\ln(X)\).

Moments

The moments of the zeta distribution can be expressed easily in terms of the zeta function.

5. Suppose that \( X \) has the zeta distribution with parameter \( a \) and that \( k \geq 0 \). Show that

\[
\mathbb{E}(X^k) = \begin{cases} 
\infty, & a \leq k + 1 \\
\frac{\zeta(a-k)}{\zeta(a)}, & a > k + 1 
\end{cases}
\]

6. In particular, show that
   a. \( \mathbb{E}(X) = \frac{\zeta(a-1)}{\zeta(a)} \) if \( a > 2 \).
   b. \( \text{var}(X) = \frac{\zeta(a-2)}{\zeta(a)} \left( \frac{\zeta(a-1)}{\zeta(a)} \right)^2 \) if \( a > 3 \).

7. Let \( X \) denote the frequency of occurrence of a word chosen at random from a certain text, and suppose that \( X \) has the zeta distribution with parameter \( a = 4 \). Approximate each of the following:
   a. \( \mathbb{E}(X) \).
   b. \( \text{var}(X) \).