

5.40: 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.

Basic Theory

The Zeta Function

The *Riemann zeta function* ζ , named after Bernhard Riemann, is defined as follows:

$$\zeta(a) = \sum_{n=1}^{\infty} \frac{1}{n^a}, \quad a \in (1, \infty) \quad (5.40.1)$$

You might recall from calculus that the series in the zeta function converges for $a > 1$ and diverges for $a \leq 1$.

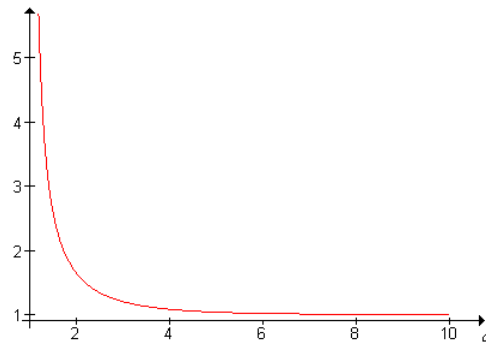


Figure 5.40.1: Graph of ζ on the interval $(1, 10]$

The zeta function satisfies the following properties:

1. ζ is decreasing.
2. ζ is concave upward.
3. $\zeta(a) \downarrow 1$ as $a \uparrow \infty$
4. $\zeta(a) \uparrow \infty$ 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

The *zeta distribution* with *shape parameter* $a \in (1, \infty)$ is a discrete distribution on \mathbb{N}_+ with probability density function f given by.

$$f(n) = \frac{1}{\zeta(a)n^a}, \quad n \in \mathbb{N}_+ \quad (5.40.2)$$

1. f is decreasing with mode $n = 1$.
2. When smoothed, f is concave upward.

Proof

Clearly f is a valid PDF, since by definition, $\zeta(a)$ is the normalizing constant for the function $n \mapsto \frac{1}{n^a}$ on \mathbb{N}_+ . Part (a) is clear. For part (b), note that the function $x \mapsto x^{-a}$ on $[1, \infty)$ has a positive second derivative.

Open the special distribution simulator and select the zeta distribution. Vary the shape parameter and note the shape of the probability density function. For selected values of the parameter, run the simulation 1000 times and compare the empirical density function to the probability density function.

The distribution function and quantile function do not have simple closed forms, except in terms of other special functions.

Open the special distribution calculator and select the zeta distribution. Vary the parameter and note the shape of the distribution and probability density functions. For selected values of the parameter, compute the median and the first and third quartiles.

Moments

Suppose that N has the zeta distribution with shape parameter $a \in (1, \infty)$. The moments of X can be expressed easily in terms of the zeta function.

If $k \geq a - 1$, $\mathbb{E}(X) = \infty$. If $k < a - 1$,

$$\mathbb{E}(N^k) = \frac{\zeta(a-k)}{\zeta(a)} \quad (5.40.3)$$

Proof

Note that

$$\mathbb{E}(N^k) = \sum_{n=1}^{\infty} n^k \frac{1}{\zeta(a)n^a} = \frac{1}{\zeta(a)} \sum_{n=1}^{\infty} \frac{1}{n^{a-k}} \quad (5.40.4)$$

If $a - k \leq 1$, the last sum diverges to ∞ . If $a - k > 1$, the sum converges to $\zeta(a - k)$

The mean and variance of N are as follows:

1. If $a > 2$,

$$\mathbb{E}(N) = \frac{\zeta(a-1)}{\zeta(a)} \quad (5.40.5)$$

2. If $a > 3$,

$$\text{var}(N) = \frac{\zeta(a-2)}{\zeta(a)} - \left(\frac{\zeta(a-1)}{\zeta(a)} \right)^2 \quad (5.40.6)$$

Open the special distribution simulator and select the zeta distribution. Vary the parameter and note the shape and location of the mean \pm standard deviation bar. For selected values of the parameter, run the simulation 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

The skewness and kurtosis of N are as follows:

1. If $a > 4$,

$$\text{skew}(N) = \frac{\zeta(a-3)\zeta^2(a) - 3\zeta(a-1)\zeta(a-2)\zeta(a) + 2\zeta^3(a-1)}{[\zeta(a-2)\zeta(a) - \zeta^2(a-1)]^{3/2}} \quad (5.40.7)$$

2. If $a > 5$,

$$\text{kurt}(N) = \frac{\zeta(a-4)\zeta^3(a) - 4\zeta(a-1)\zeta(a-3)\zeta^2(a) + 6\zeta^2(a-1)\zeta(a-2)\zeta(a) - 3\zeta^4(a-1)}{[\zeta(a-2)\zeta(a) - \zeta^2(a-1)]^2} \quad (5.40.8)$$

Proof

These results follow from the [general moment result](#) above and standard computational formulas for skewness and kurtosis.

The probability generating function of N can be expressed in terms of the *polylogarithm function* Li that was introduced in the section on the exponential-logarithmic distribution. Recall that the polylogarithm of order $s \in \mathbb{R}$ is defined by

$$\text{Li}_s(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^s}, \quad x \in (-1, 1) \quad (5.40.9)$$

N has probability generating function P given by

$$P(t) = \mathbb{E}(t^N) = \frac{\text{Li}_a(t)}{\zeta(a)}, \quad t \in (-1, 1) \quad (5.40.10)$$

Proof

Note that

$$\mathbb{E}(t^N) = \sum_{n=1}^{\infty} t^n \frac{1}{n^a \zeta(a)} = \frac{1}{\zeta(a)} \sum_{n=1}^{\infty} \frac{t^n}{n^a} \quad (5.40.11)$$

The last sum is $\text{Li}_a(t)$.

Related Distributions

In an algebraic sense, the zeta distribution is a discrete version of the Pareto distribution. Recall that if $a > 1$, the Pareto distribution with shape parameter $a - 1$ is a continuous distribution on $[1, \infty)$ with probability density function

$$f(x) = \frac{a-1}{x^a}, \quad x \in [1, \infty) \quad (5.40.12)$$

Naturally, the limits of the zeta distribution with respect to the shape parameter a are of interest.

The zeta distribution with shape parameter $a \in (1, \infty)$ converges to point mass at 1 as $a \rightarrow \infty$.

Proof

For the PDF f above, note that $f(1) = \zeta(a) \rightarrow 1$ as $a \rightarrow \infty$ and for $n \in \{2, 3, \dots\}$, $f(n) = 1/n^a \zeta(a) \rightarrow 0$ as $a \rightarrow \infty$

Finally, the zeta distribution is a member of the family of general exponential distributions.

Suppose that N has the zeta distribution with parameter a . Then the distribution is a one-parameter exponential family with natural parameter a and natural statistic $-\ln N$.

Proof

This follows from the definition of the general exponential distribution, since the zeta PDF can be written in the form

$$f(n) = \frac{1}{\zeta(a)} \exp(-a \ln n), \quad n \in \mathbb{N}_+ \quad (5.40.13)$$

Computational Exercises

Let N 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 $\mathbb{P}(N > 4)$.

Answer

$$\mathbb{P}(N > 4) = 1 - \frac{49}{6\pi^2} \approx 0.1725$$

Suppose that N has the zeta distribution with parameter $a = 6$. Approximate each of the following:

1. $\mathbb{E}(N)$
2. $\text{var}(N)$
3. $\text{skew}(N)$
4. $\text{kurt}(N)$

Answer

1. $\mathbb{E}(N) \approx 1.109$
2. $\text{var}(N) \approx 0.025$
3. $\text{skew}(N) \approx 11.700$
4. $\text{kurt}(N) \approx 309.19$

This page titled [5.40: The Zeta Distribution](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.