Frequently Asked Questions

1. Name the person after whom the Pareto distribution is named.

Answer:

The **Pareto distribution** is named after the Italian economist Vilfredo Pareto.

2. Define Pareto distribution.

Answer:

A continuous random variable X is said to have Pareto distribution with parameters α and b if its pdf is given by, $f(x) = \frac{\alpha b^{\alpha}}{x^{\alpha+1}}; x \geq b, \alpha > 0$. Here α is called shape parameter and b is called scale parameter.

3. Define incomplete gamma integral.

Answer:

$$\int_{\beta}^{\infty} e^{-x} x^{\alpha-1} dx = \Gamma(\alpha, \beta) = (\alpha-1)! e^{-\beta} \sum_{k=0}^{\alpha-1} \frac{\beta^k}{k!}$$

4. Obtain distribution function of Pareto distribution.

Answer:

Distribution function is given by,

$$F(x) = \int_b^x f(x)dx = \int_b^x \frac{\alpha b^{\alpha}}{x^{\alpha+1}} dx$$

$$= \alpha b^{\alpha} \int_b^x x^{-\alpha-1} dx = \alpha b^{\alpha} \left| \frac{x^{-\alpha}}{-\alpha} \right|_b^x = -b^{\alpha}(x^{-\alpha} - b^{-\alpha}) = 1 - \frac{b^{\alpha}}{x^{\alpha}}, \ x \geq b$$

5. Obtain mean of the Pareto distribution.

Answer:

Mean of the distribution is given by,

μ1'= E(X)

$$= \int_b^{\infty} x f(x) dx = \alpha b^{\alpha} \int_b^{\infty} x x^{-\alpha-1} dx$$

$$= \alpha b^{\alpha} \int_b^{\infty} x^{-(\alpha-1)-1} dx = \alpha b^{\alpha} \frac{1}{(\alpha-1) b^{(\alpha-1)}} = \frac{\alpha b}{(\alpha-1)}$$

6. Obtain variance of Pareto distribution.

Answer:

Variance of the Pareto distribution is given by $\mu_2 = \mu_2' - \mu_1'^2$

Hence, first let us find

$\mu_2' = E(X^2)$

$$= \int_b^{\infty} x^2 f(x) dx = \alpha b^{\alpha} \int_b^{\infty} x^2 x^{-\alpha-1} dx$$

$$= \alpha b^{\alpha} \int_b^{\infty} x^{-(\alpha-2)-1} dx = \alpha b^{\alpha} \frac{1}{(\alpha-2) b^{(\alpha-2)}} = \frac{\alpha b^2}{(\alpha-2)}$$

Also, μ1'= E(X)

$$= \int_b^\infty x f(x)dx = \alpha b^\alpha \int_b^\infty x x^{-\alpha-1} dx$$

$$= \alpha b^\alpha \int_b^\infty x^{-(\alpha-1)-1} dx = \alpha b^\alpha \frac{1}{(\alpha-1)b^{(\alpha-1)}} = \frac{\alpha b}{(\alpha-1)}$$

Therefore, variance

$\mu_2 = \mu_2' - \mu_1'^2$

$$= \frac{\alpha b^2}{\alpha-2} - \left(\frac{\alpha b}{\alpha-1}\right)^2 = \alpha b^2 \left[\frac{1}{\alpha-2} - \frac{\alpha}{(\alpha-1)^2}\right]$$

$$= \frac{\alpha b^2 [(\alpha-1)^2 - \alpha(\alpha-2)}{(\alpha-1)^2(\alpha-2)} = \frac{\alpha b^2}{(\alpha-1)^2(\alpha-2)}$$

For $\alpha>2$, If α is 1 or 2, variance is infinity.

7. Obtain an expression for r^{th} raw moment.

Answer:

r^{th} raw moment $\mu r'$ of Pareto distribution is given by

$\mu_r' = E(X^r)$

$$= \int_b^\infty x^r f(x)dx = \alpha b^\alpha \int_b^\infty x^r x^{-\alpha-1} dx$$

$$= \alpha b^\alpha \int_b^\infty x^{-(\alpha-r)-1} dx = \alpha b^\alpha \frac{1}{(\alpha-r)b^{(\alpha-r)}} = \frac{\alpha b^r}{(\alpha-r)}$$

8. Obtain expressions for first 4 raw moments using rth raw moment.

Answer:

By putting r=1, we get,

Mean $\mu_1' = \alpha\, b/(\alpha -1)$

Putting r=2, we get, $\mu_2' = \alpha\, b^2/(\alpha -2)$.

By putting r=3, we get, $\mu_3' = \alpha\, b^3/(\alpha -3)$ and r=4, we get $\mu_4' = \alpha\, b^4/(\alpha-4)$

9. Find coefficient of skewness of Pareto distribution.

Answer:

Coefficient of skewness is given by,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\left[\frac{2\alpha(\alpha+1)b^3}{(\alpha-1)^3(\alpha-2)(\alpha-3)}\right]^2}{\left[\frac{\alpha b^2}{(\alpha-1)^2(\alpha-2)}\right]^3} = \frac{4((\alpha+1)^2(\alpha-2)}{\alpha(\alpha-3)^2}$$

For $\alpha>3$, Pareto distribution is positively skewed.

10. Find coefficient of Kurtosis of Pareto distribution.

Answer:

Coefficient of kurtosis is given by,

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\frac{3\alpha(3\alpha^3+\alpha+3)b^4}{(\alpha-1)^4(\alpha-2)(\alpha-3)(\alpha-4)}}{\left[\frac{\alpha b^2}{(\alpha-1)^2(\alpha-2)}\right]^2} = \frac{3(3\alpha^3+\alpha+3)(\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}$$

For $\alpha>4$ the Pareto distribution has leptokurtic curve.

11. Derive moment generating function of Pareto distribution.

Answer:
Mgf of Pareto distribution is given by,
$M_X(t)=E(e^{tX})$

$$= \int_b^\infty e^{tx} f(x)dx = \alpha b^\alpha \int_b^\infty e^{tx} x^{-\alpha-1} dx$$

$$= \alpha b^\alpha \int_b^\infty e^{-x(-t)} (x)^{-\alpha-1} dx, = \alpha b^\alpha \frac{\Gamma(-\alpha,-bt)}{(-t)^{-\alpha}} = \alpha(-tb)^\alpha \Gamma(-\alpha,-bt)$$

12. Find median of Pareto distribution.
Answer:
If M is the median of the Pareto distribution, then $\int_b^M f(x)dx = \frac{1}{2} = \int_M^\infty f(x)dx$

Consider the first integral and solve for M

$$\int_b^M f(x)dx = \frac{1}{2} \Rightarrow \alpha b^\alpha \int_b^M x^{-\alpha-1} dx = \frac{1}{2}$$

$$\Rightarrow \alpha b^\alpha \left| \frac{x^{-\alpha}}{-\alpha} \right|_b^M = \frac{1}{2} or - b^\alpha [M^{-\alpha} - b^{-\alpha}] = \frac{1}{2}$$

$$i.e., 1 - b^\alpha M^{-\alpha} = \frac{1}{2} \Rightarrow b^\alpha M^{-\alpha} = \frac{1}{2}$$

$$\therefore M = b\sqrt[\alpha]{2}$$

13. What is the mode of Pareto distribution?
Answer:
Mode is that value of x for which f(x) is maximum. Observe that Pareto distribution is strictly a decreasing function. Hence, it attains its maximum at the very first value taken by the variable, i.e. If X has Pareto distribution with parameters α and b, then f(x) attains its maximum at x=b. Therefore, b is mode of the distribution.

14. Write an application of Pareto distribution.
Answer:
Pareto originally used this distribution to describe the allocation of wealth among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income. This idea is sometimes expressed more simply as the Pareto principle or the "80-20 rule" which says that 20% of the population controls 80% of the wealth. However, the 80-20 rule corresponds to a particular value of α, and in fact, Pareto's data on British income taxes in his *Cours d'économie politique* indicates that about 30% of the population had about 70% of the income.

15. Mention some of the examples of Pareto distribution.
Answer:
- The sizes of human settlements (few cities, many hamlets/villages)
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)
- Hard disk drive error rates
- Clusters of Bose-Elistein condensate near absolute zero.
- The values of oil reserves in oil fields (a few large fields, many small fields)
- The standardized price returns on individual stocks
- Sizes of sand particles

- Sizes of meteorites
- Numbers of species per genus (There is subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it)
- Areas burnt in forest fires
- Severity of large casualty losses for certain lines of business such as general liability, commercial auto, and workers compensation
- The length distribution in jobs assigned supercomputers (a few large ones, many small ones)