DISCRETE PROBABILITY DISTRIBUTIONS

https://phitter.io

2024-03-02

Abstract

This document provides an overview of the continuous probability distributions utilized in Phitter. It includes a detailed description for each distribution, covering aspects such as the definition, domain, parameter definitions and domains, probability density function, cumulative distribution function, percentile point function, raw moments, mean, variance, skewness, kurtosis, median, and mode in a concise and clear manner.

Contents

1 Bernoulli Distribution

1.1 Distribution definition

$$X \sim \text{Bernoulli}\,(p)$$

1.2 Distribution domain

$$x \in \{0, 1\}$$

1.3 Parameters domain and parameters constraints

$$p \in (0, 1) \subseteq \mathbb{R}$$

1.4 Cumulative distribution function

$$F_X\,(x) = \begin{cases} 1-p & \text{if} \;\; x = 0 \\ 1 & \text{if} \;\; x = 1 \end{cases}$$

1.5 Probability density function

$$f_X\,(x) = p^x (1-p)^{1-x}$$

1.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \begin{cases} 1 & \text{if} \;\; u \leq p \\ 0 & \text{if} \;\; u > p \end{cases}$$

1.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{1} x^k f_X\,(x) = p$$

1.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = p$$

1.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = p(1-p)$$

1.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \frac{1-2p}{\sqrt{p(1-p)}}$$

1.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 + \frac{1-6p(1-p)}{p(1-p)}$$

1.12 Parametric median

$$\text{Median}(X) = \begin{cases} 0 & \text{if } p < 1/2 \\ [0, 1] & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$$

1.13 Parametric mode

$$\text{Mode}(X) = \begin{cases} 0 & \text{if} \;\; p < 1/2 \\ 0, 1 & \text{if} \;\; p = 1/2 \\ 1 & \text{if} \;\; p > 1/2 \end{cases}$$

1.14 Additional information and definitions

- u : Uniform[0,1] random varible

1.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

2 Binomial Distribution

2.1 Distribution definition

$$X \sim \text{Binomial}\,(n, p)$$

2.2 Distribution domain

$$x \in \mathbb{N} \equiv \{0, 1, 2, \dots\}$$

2.3 Parameters domain and parameters constraints

$$n \in \mathbb{N}, p \in (0, 1) \subseteq \mathbb{R}$$

2.4 Cumulative distribution function

$$F_X\,(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} = I(1-p, n-x, 1+x)$$

2.5 Probability density function

$$f_X\,(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

2.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \arg\min_x |F_X\,(x) - u|$$

2.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{\infty} x^k f_X\,(x) = \sum_{i=0}^{k} \tfrac{n!}{(n-i)!} S(k, i) p^i$$

2.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = np$$

2.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = np(1-p)$$

2.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \frac{1-2p}{\sqrt{np(1-p)}}$$

2.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 + \frac{1 - 6p(1-p)}{np(1-p)}$$

2.12 Parametric median

$$\text{Median}(X) = \lfloor np \rfloor \vee \lceil np \rceil$$

2.13 Parametric mode

$$\text{Mode}(X) = \lfloor (n+1)p \rfloor \vee \lceil (n+1)p \rceil - 1$$

2.14 Additional information and definitions

- Computing an analytic expression for the inverse of the cumulative distribution function is not feasible. However, it is possible to calculate the Percentile Point Function by approximating it to the nearest integer.
- u : Uniform[0,1] random varible
- $\lfloor x \rfloor$: Floor function
- $\lceil x \rceil$: Ceiling Function
- $I\left(x, a, b\right)$: Regularized incomplete beta function
- $S(a, b)$: Stirling numbers of the second kind $= \frac{1}{b!}\sum_{j=0}^{b}(-1)^{b-j}\binom{b}{j}j^a$

2.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

3 Geometric Distribution

3.1 Distribution definition

$$X \sim \text{Geometric}\,(p)$$

3.2 Distribution domain

$$x \in \mathbb{N} \equiv \{0, 1, 2, \dots\}$$

3.3 Parameters domain and parameters constraints

$$p \in (0, 1) \subseteq \mathbb{R}$$

3.4 Cumulative distribution function

$$F_X\,(x) = 1 - (1-p)^{\lfloor x \rfloor}$$

3.5 Probability density function

$$f_X\,(x) = (1-p)^{x-1} p$$

3.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \left\lceil \frac{\ln\,(1-u)}{\ln\,(1-p)} \right\rceil$$

3.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{\infty} x^k f_X\,(x) = \sum_{x=0}^{\infty} (1-p)^x p \cdot x^k$$

3.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \frac{1}{p}$$

3.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = \frac{1-p}{p^2}$$

3.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \frac{2-p}{\sqrt{1-p}}$$

3.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 9 + \frac{p^2}{1-p}$$

3.12 Parametric median

$$\text{Median}(X) = \left\lceil \frac{-1}{\log_2(1-p)} \right\rceil$$

3.13 Parametric mode

$$\text{Mode}(X) = 1$$

3.14 Additional information and definitions

- u : Uniform[0,1] random varible
- $\lfloor x \rfloor$: Floor function
- $\lceil x \rceil$: Ceiling Function

3.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

4 Hypergeometric Distribution

4.1 Distribution definition

$$X \sim \text{Hypergeometric}\,(N, K, n)$$

4.2 Distribution domain

$$x \in \{\max\,(0, n + K - N), \min\,(n, K)\}$$

4.3 Parameters domain and parameters constraints

$$N \in \mathbb{N}, K \in \{0 \ldots N\}\,, n \in \{0 \ldots N\}$$

4.4 Cumulative distribution function

$$F_X\,(x) = \sum_{i=0}^{x} \binom{K}{i}\binom{N-K}{n-i}\Big/\binom{N}{n}$$

4.5 Probability density function

$$f_X\,(x) = \binom{K}{x}\binom{N-K}{n-x}\Big/\binom{N}{n}$$

4.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \arg\min_x |F_X\,(x) - u|$$

4.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=\max\,(0,n+K-N)}^{\min\,(n,K)} x^k f_X\,(x)$$

4.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \frac{nK}{N}$$

4.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$$

4.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \frac{(N-2K)(N-1)^{\frac{1}{2}}(N-2n)}{[nK(N-K)(N-n)]^{\frac{1}{2}}(N-2)}$$

4.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 + \frac{1}{nK(N-K)(N-n)(N-2)(N-3)}$$

4.12 Parametric median

$$\text{Median}(X) = F_X^{-1}\,(0.5)$$

4.13 Parametric mode

$$\text{Mode}(X) = \left\lfloor \frac{(n+1)(K+1)}{N+2} \right\rfloor$$

4.14 Additional information and definitions

- Computing an analytic expression for the inverse of the cumulative distribution function is not feasible. However, it is possible to calculate the Percentile Point Function by approximating it to the nearest integer.
- u : Uniform[0,1] random varible
- $\lfloor x \rfloor$: Floor function
- $\lceil x \rceil$: Ceiling Function

4.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

5 Logarithmic Distribution

5.1 Distribution definition

$$X \sim \text{Logarithmic}\,(p)$$

5.2 Distribution domain

$$x \in \mathbb{N}_{\geqslant 1} \equiv \{1, 2, \dots\}$$

5.3 Parameters domain and parameters constraints

$$p \in (0, 1) \subseteq \mathbb{R}$$

5.4 Cumulative distribution function

$$F_X\,(x) = \sum_{i=0}^{x} \frac{1}{-\ln(1-p)} \frac{p^i}{i}$$

5.5 Probability density function

$$f_X\,(x) = \frac{1}{-\ln(1-p)} \frac{p^x}{x}$$

5.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \arg\min_x |F_X\,(x) - u|$$

5.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{\infty} x^k f_X\,(x) = \frac{(k-1)!}{-\ln(1-p)} \left(\frac{p}{1-p}\right)^k$$

5.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \frac{1}{-\ln(1-p)} \frac{p}{1-p}$$

5.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = -\frac{p^2 + p\ln(1-p)}{(1-p)^2(\ln(1-p))^2}$$

5.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}}$$

5.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2}$$

5.12 Parametric median

$$\text{Median}(X) = F_X^{-1}\,(0.5)$$

5.13 Parametric mode

$$\text{Mode}(X) = 1$$

5.14 Additional information and definitions

- Computing an analytic expression for the inverse of the cumulative distribution function is not feasible. However, it is possible to calculate the Percentile Point Function by approximating it to the nearest integer.
- u : Uniform[0,1] random varible

5.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

6 Negative Binomial Distribution

6.1 Distribution definition

$$X \sim \text{NegativeBinomial}\,(r, p)$$

6.2 Distribution domain

$$x \in \mathbb{N} \equiv \{0, 1, 2, \dots\}$$

6.3 Parameters domain and parameters constraints

$$r \in \mathbb{N}_{\geqslant 1}, p \in (0, 1) \subseteq \mathbb{R}$$

6.4 Cumulative distribution function

$$F_X\,(x) = I(p, r, x + 1)$$

6.5 Probability density function

$$f_X\,(x) = \binom{r + x - 1}{x} p^r (1 - p)^x$$

6.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \arg\min_x |F_X\,(x) - u|$$

6.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{\infty} x^k f_X\,(x)$$

6.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \frac{r(1 - p)}{p}$$

6.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = \frac{r(1 - p)}{p^2}$$

6.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \frac{2 - p}{\sqrt{r\,(1 - p)}}$$

6.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 + \frac{6}{r} + \frac{p^2}{r\,(1 - p)}$$

6.12 Parametric median

$$\text{Median}(X) = F_X^{-1}\,(0.5)$$

6.13 Parametric mode

$$\text{Mode}(X) = \lfloor (r - 1)\,(1 - p)/p \rfloor$$

6.14 Additional information and definitions

- Computing an analytic expression for the inverse of the cumulative distribution function is not feasible. However, it is possible to calculate the Percentile Point Function by approximating it to the nearest integer.
- u : Uniform[0,1] random varible
- $I\left(x, a, b\right)$: Regularized incomplete beta function
- $\lfloor x \rfloor$: Floor function

6.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

7 Poisson Distribution

7.1 Distribution definition

$$X \sim \text{Poisson}\,(\lambda)$$

7.2 Distribution domain

$$x \in \mathbb{N} \equiv \{0, 1, 2, \dots\}$$

7.3 Parameters domain and parameters constraints

$$\lambda \in \mathbb{R}^{+}$$

7.4 Cumulative distribution function

$$F_X\,(x) = e^{-\lambda} \sum_{i=0}^{x} \frac{\lambda^i}{i!} = 1 - \frac{\gamma(x+1, \lambda)}{x!} = 1 - P(x-1, \lambda)$$

7.5 Probability density function

$$f_X\,(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

7.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \arg\min_x |F_X\,(x) - u|$$

7.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=0}^{\infty} x^k f_X\,(x)$$

7.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \lambda$$

7.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = \lambda$$

7.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = \lambda^{-1/2}$$

7.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 + \lambda^{-1}$$

7.12 Parametric median

$$\text{Median}(X) = \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$$

7.13 Parametric mode

$$\text{Mode}(X) = \lfloor \lambda \rfloor$$

7.14 Additional information and definitions

- Computing an analytic expression for the inverse of the cumulative distribution function is not feasible. However, it is possible to calculate the Percentile Point Function by approximating it to the nearest integer.
- u : Uniform[0,1] random varible
- $\lfloor x \rfloor$: Floor function
- $P\left(a,x\right)=\frac{\gamma(a,x)}{\Gamma(a)}$: Regularized lower incomplete gamma function
- $\gamma\left(a,x\right)$: Lower incomplete Gamma function

7.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

8 Uniform Distribution

8.1 Distribution definition

$$X \sim \text{Uniform}\,(a, b)$$

8.2 Distribution domain

$$x \in \{a, a+1, \ldots, b-1, b\}$$

8.3 Parameters domain and parameters constraints

$$a \in \mathbb{N}, b \in \mathbb{N}, a < b$$

8.4 Cumulative distribution function

$$F_X\,(x) = \frac{x - a + 1}{b - a + 1}$$

8.5 Probability density function

$$f_X\,(x) = \frac{1}{b - a + 1}$$

8.6 Percent point function/Sample

$$F_X^{-1}\,(u) = \lceil u(b - a + 1) + a - 1 \rceil$$

8.7 Parametric centered moments

$$E[X^k] = \mu'_k = \sum_{x=a}^{b} x^k f_X\,(x) = \frac{1}{b - a + 1} \sum_{x=a}^{b} x^k$$

8.8 Parametric mean

$$\text{Mean}(X) = \mu'_1 = \frac{a + b}{2}$$

8.9 Parametric variance

$$\text{Variance}(X) = (\mu'_2 - \mu'^2_1) = \frac{(b - a + 1)^2 - 1}{12}$$

8.10 Parametric skewness

$$\text{Skewness}(X) = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{1.5}} = 0$$

8.11 Parametric kurtosis

$$\text{Kurtosis}(X) = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2} = 3 - \frac{6((b - a + 1)^2 + 1)}{5((b - a + 1)^2 - 1)}$$

8.12 Parametric median

$$\text{Median}(X) = \frac{a + b}{2}$$

8.13 Parametric mode

$$\text{Mode}(X) \in [a, b]$$

8.14 Additional information and definitions

- u : Uniform[0,1] random varible
- $\lceil x \rceil$: Ceiling Function

8.15 Spreadsheet documents

- Excel file from GitHub repository
- Google spreadsheet document

References

[1] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[2] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.