Maximum Entropy Principle

Prof. Americo Cunha Jr

Rio de Janeiro State University – UERJ

americo.cunha@uerj.br

www.americocunha.org

@AmericoCunhaJr @AmericoCunhaJr

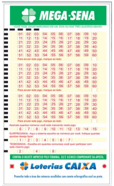
@AmericoCunhaJr @AmericoCunhaJr

In which scenario is the outcome is more uncertain ?

A: **Tossing a fair coin**

$$\mathcal{P}(\text{Head}) = 1/2$$
$$\mathcal{P}(\text{Tail}) = 1/2$$

B: **Play the lottery**



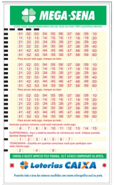
$$\mathcal{P}(\text{Win}) \approx 0.00000002$$
$$\mathcal{P}(\text{Lose}) \approx 0.99999998$$

In which scenario is the outcome is more uncertain ?

A: **Tossing a fair coin**

$$\mathcal{P}(\text{Head}) = 1/2$$
$$\mathcal{P}(\text{Tail}\ \ ) = 1/2$$

B: **Play the lottery**



$$\mathcal{P}(\text{Win}\ \ ) \approx 0.00000002$$
$$\mathcal{P}(\text{Lose}) \approx 0.99999998$$

The result of A is much less certain (more uncertain) !

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Not a lot! There is a lot of "uncertainty" about this distribution!

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Not a lot! There is a lot of "uncertainty" about this distribution!

1. $p_1 \geq 0, \; p_2 \geq 0, \; \cdots \; , p_n \geq 0$ for arbitrary n

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Not a lot! There is a lot of "uncertainty" about this distribution!

1. $p_1 \geq 0, \; p_2 \geq 0, \; \cdots \; , p_n \geq 0$ for arbitrary n
2. $p_1 + p_2 + \cdots + p_n = 1$

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Not a lot! There is a lot of "uncertainty" about this distribution!

1. $p_1 \geq 0, \; p_2 \geq 0, \; \cdots \; , p_n \geq 0$ for arbitrary n
2. $p_1 + p_2 + \cdots + p_n = 1$

A plethora of distributions are possible!

Probabilistic uncertainty

Imagine you have a dice with an unknown number of faces

What is known about the dice probability distribution?

Not a lot! There is a lot of "uncertainty" about this distribution!

1. $p_1 \geq 0, \; p_2 \geq 0, \; \cdots \;, p_n \geq 0$ for arbitrary n
2. $p_1 + p_2 + \cdots + p_n = 1$

A plethora of distributions are possible!

This probabilistic uncertainty is called Entropy

How to reduce part of this uncertainty (entropy) ?

Known information:

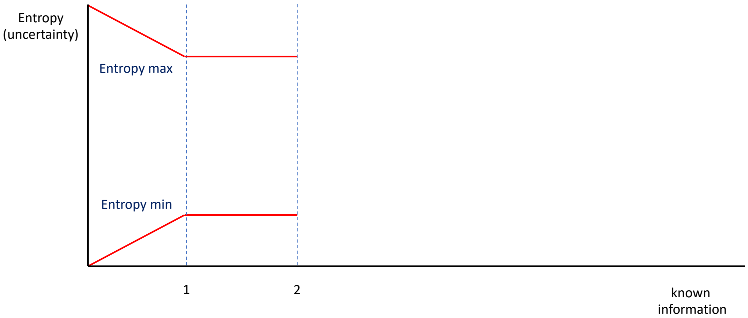
- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)

How to reduce part of this uncertainty (entropy) ?

Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)

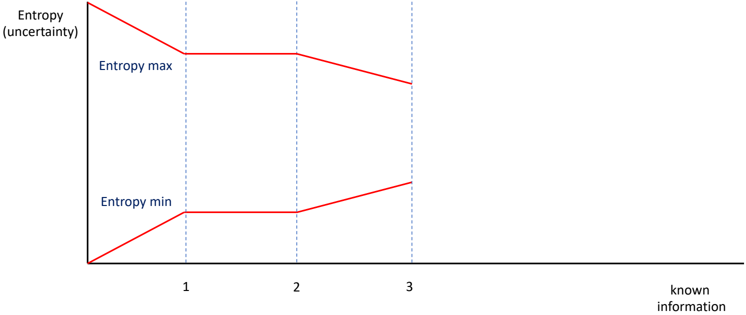
How to reduce part of this uncertainty (entropy) ?



Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)
- $1^2 p_1 + 2^2\,p_2 + 3^2\,p_3 + 4^2\,p_4 + 5^2\,p_5 + 6^2\,p_6 = 15$ (variance)

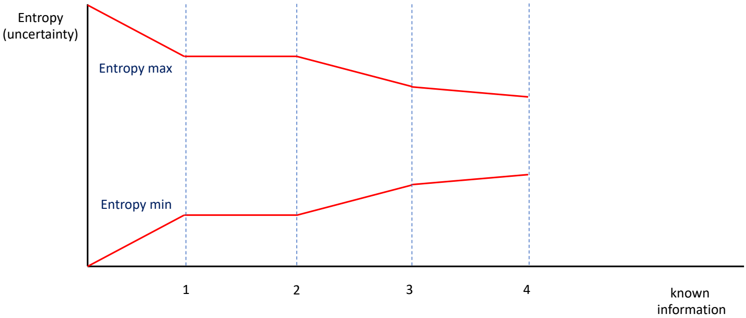
How to reduce part of this uncertainty (entropy) ?



Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)
- $1^2 p_1 + 2^2\,p_2 + 3^2\,p_3 + 4^2\,p_4 + 5^2\,p_5 + 6^2\,p_6 = 15$ (variance)
- ...

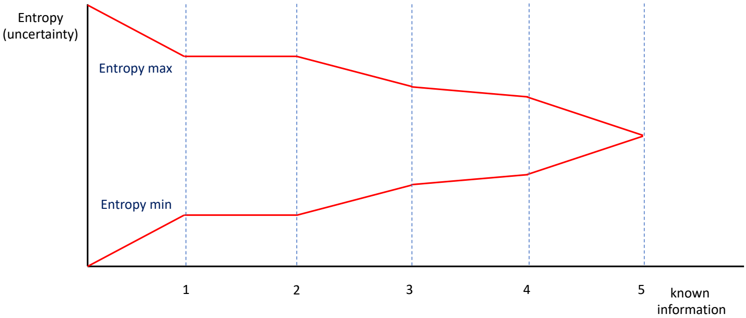
How to reduce part of this uncertainty (entropy) ?



Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)
- $1^2 p_1 + 2^2\,p_2 + 3^2\,p_3 + 4^2\,p_4 + 5^2\,p_5 + 6^2\,p_6 = 15$ (variance)
- ...

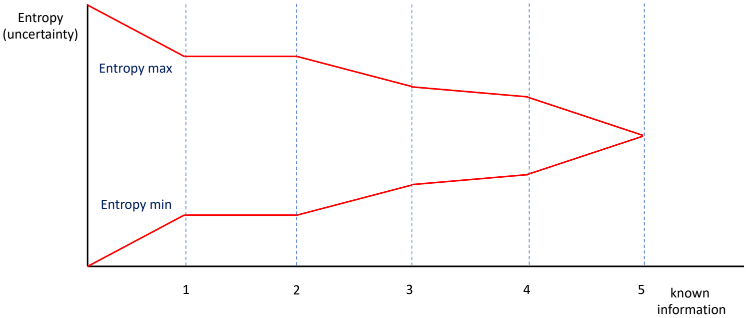
How to reduce part of this uncertainty (entropy) ?



Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)
- $1^2 p_1 + 2^2\,p_2 + 3^2\,p_3 + 4^2\,p_4 + 5^2\,p_5 + 6^2\,p_6 = 15$ (variance)
- ...

How to reduce part of this uncertainty (entropy) ?

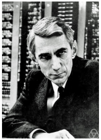


Known information:

- $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$ (6 faces)
- $1\,p_1 + 2\,p_2 + 3\,p_3 + 4\,p_4 + 5\,p_5 + 6\,p_6 = 4.5$ (mean)
- $1^2 p_1 + 2^2\,p_2 + 3^2\,p_3 + 4^2\,p_4 + 5^2\,p_5 + 6^2\,p_6 = 15$ (variance)
- ...

New information reduce the distribution entropy !

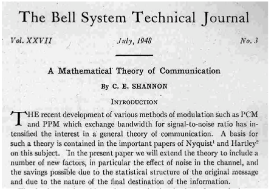
The Entropy functional

 ©

Claude Shannon
(1916-2001)

 ©

John von Neumann
(1903–1957)



The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

Entropy of the random variable X is defined as

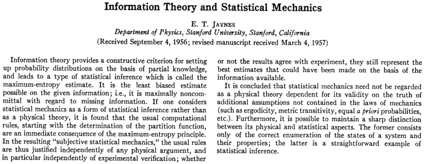
$$\mathcal{S}\left(p_X\right) = - \int_{\mathbb{R}} p_X(x) \ln p_X(x) \, dx$$

"measure for the level of uncertainty about p_X"

Maximum Entropy Principle (MaxEnt)

Edwin T. Jaynes

(1922–1998)



Information Theory and Statistical Mechanics

E. T. Jaynes

Department of Physics, Stanford University, Stanford, California

(Received September 4, 1956; revised manuscript received March 4, 1957)

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum-entropy estimate. It is the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information. If one considers statistical mechanics as a form of statistical inference rather than as a physical theory, it is found that the usual computational rules, starting with the determination of the partition function, are an immediate consequence of the maximum-entropy principle. In the resulting "subjective statistical mechanics," the usual rules are thus justified independently of any physical argument, and in particular independently of experimental verification; whether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.

It is concluded that statistical mechanics need not be regarded as a physical theory dependent for its validity on the truth of additional assumptions not contained in the laws of mechanics (such as ergodicity, metric transitivity, equal *a priori* probabilities, etc.). Furthermore, it is possible to maintain a sharp distinction between its physical and statistical aspects. The former consists only of the correct enumeration of the states of a system and their properties; the latter is a straightforward example of statistical inference.

"Among all the probability distributions, consistent with the known information about a random parameter, choose the one which corresponds to the maximum of entropy (MaxEnt)."

MaxEnt distribution = most unbiased distribution

- use all information available
- avoid to use information not available
- scientific objectivity and honesty

MaxEnt optimization problem

Maximize

$$\mathcal{S}\left(p_X\right) = -\int_{\mathbb{R}} p_X(x) \ln p_X(x)\, dx,$$

respecting $N+1$ constraints (known information) given by

$$\int_{\mathbb{R}} g_k\left(X\right) p_X(x)\, dx = m_k, \qquad k = 0, \cdots, N,$$

where the g_k are known real functions, with $g_0(x) = 1$.

MaxEnt optimization problem

Maximize

$$\mathcal{S}\left(p_X\right) = -\int_{\mathbb{R}} p_X(x)\, \ln p_X(x)\, dx,$$

respecting $N+1$ constraints (known information) given by

$$\int_{\mathbb{R}} g_k\left(X\right) p_X(x)\, dx = m_k, \qquad k = 0, \cdots, N,$$

where the g_k are known real functions, with $g_0(x) = 1$.

MaxEnt general solution:

$$p_X(x) = \mathbb{1}_{\mathcal{K}}(x)\, \exp\left(-\lambda_0\right)\, \exp\left(-\sum_{k=1}^{N} \lambda_k\, g_k(x)\right)$$

- λ_k – hyperparameters (usually obtained numerically)

MaxEnt example 1: theoretical information only

Known information about X: $\texttt{Supp}\, p_X = [a, b]$

MaxEnt example 1: theoretical information only

Known information about X: $\texttt{Supp}\, p_X = [a, b]$

To obtain a consistent (unbiased) probability distribution for X, it is necessary to solve the optimization problem:

MaxEnt example 1: theoretical information only

Known information about X: $\mathrm{Supp}\, p_X = [a, b]$

To obtain a consistent (unbiased) probability distribution for X, it is necessary to solve the optimization problem:

Maximize

$$S\,(p_X) = -\int_{\mathbb{R}} p_X(x) \ln p_X(x)\, dx$$

subjected to

$$\int_a^b p_X(x)\, dx = 1$$

MaxEnt example

Define the Lagrangian:

$$\mathcal{L}\left(p_X, \lambda_0\right) = -\int_a^b p_X(x)\, \ln\left(p_X(x)\right) dx - (\lambda_0 - 1)\left(\int_a^b p_X(x)\, dx - 1\right)$$

MaxEnt example

Define the Lagrangian:

$$\mathcal{L}\left(p_X, \lambda_0\right) = -\int_a^b p_X(x) \ln\left(p_X(x)\right) dx - (\lambda_0 - 1)\left(\int_a^b p_X(x)\, dx - 1\right)$$

Apply the necessary conditions for an extreme:

$$\frac{\partial \mathcal{L}}{\partial p_X}\left(p_X, \lambda_0\right) = 0$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda_0}\left(p_X, \lambda_0\right) = 0$$

MaxEnt example

$$\frac{\partial \mathcal{L}}{\partial p_X}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad p_X(x) = \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_0}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad \int_a^b p_X(x)\, dx = 1$$

MaxEnt example

$$\frac{\partial \mathcal{L}}{\partial p_X}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad p_X(x) = \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_0}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad \int_a^b p_X(x)\, dx = 1$$

Hence,

$$\int_a^b \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}\, dx = 1 \qquad \Longrightarrow \qquad e^{-\lambda_0} = \frac{1}{b-a},$$

MaxEnt example

$$\frac{\partial \mathcal{L}}{\partial p_X}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad p_X(x) = \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_0}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad \int_a^b p_X(x)\, dx = 1$$

Hence,

$$\int_a^b \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}\, dx = 1 \qquad \Longrightarrow \qquad e^{-\lambda_0} = \frac{1}{b-a},$$

so that

$$p_X(x) = \mathbb{1}_{[a,b]}(x)\, \frac{1}{b-a}.$$

MaxEnt example

$$\frac{\partial \mathcal{L}}{\partial p_X}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad p_X(x) = \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_0}(p_X, \lambda_0) = 0 \qquad \Longrightarrow \qquad \int_a^b p_X(x)\, dx = 1$$

Hence,

$$\int_a^b \mathbb{1}_{[a,b]}(x)\, e^{-\lambda_0}\, dx = 1 \qquad \Longrightarrow \qquad e^{-\lambda_0} = \frac{1}{b-a},$$

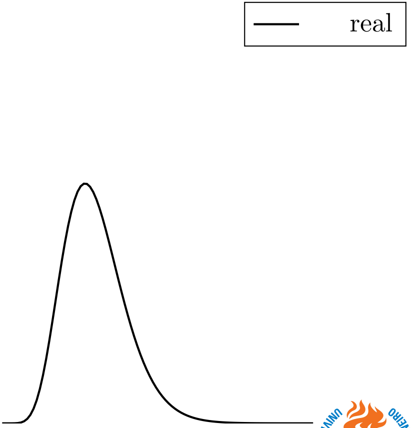
so that

$$p_X(x) = \mathbb{1}_{[a,b]}(x)\, \frac{1}{b-a}.$$

X has uniform distribution on $[a, b]$

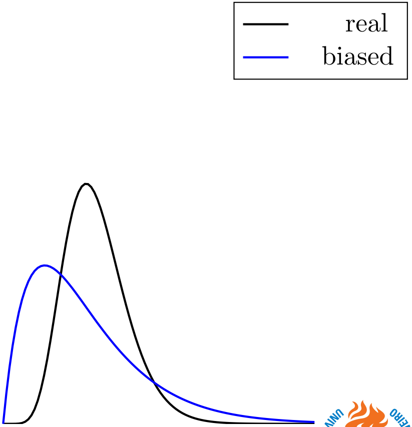
Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution



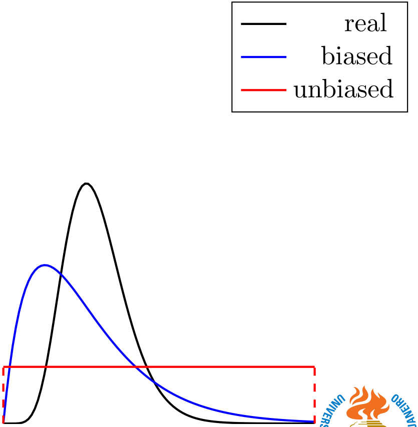
Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution
- Distributions arbitrarily chosen can be coarse and biased



Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution
- Distributions arbitrarily chosen can be coarse and biased
- A conservative strategy is to use the most unbiased (MaxEnt) distribution



Examples of MaxEnt distributions

Available information:

- $\operatorname{Supp} X = [a, b]$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{[a,b]}(x) \, \frac{1}{b - a}$$

Uniform in $[a, b]$ — $\mathcal{U}(a, b)$

Examples of MaxEnt distributions

Available information:

- $\operatorname{Supp} X = [a, b]$
- $\mathbb{E}\{X\} = \mu \in [a, b]$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{[a,b]}(x) \, \exp\left(-\lambda_0 - x\,\lambda_1\right)$$

- $\lambda_0 = \lambda_0(a, b, \mu)$
- $\lambda_1 = \lambda_1(a, b, \mu)$

Truncated Exponential in $[a, b]$ with mean μ

Examples of MaxEnt distributions

Available information:

- $\operatorname{Supp} X = [a, b]$
- $\mathbb{E}\{X\} = \mu \in [a, b]$
- $\mathbb{E}\left\{X^2\right\} = \mu^2 + \sigma^2$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{[a,b]}(x)\, \exp\left(-\lambda_0 - x\,\lambda_1 - x^2\,\lambda_2\right)$$

- $\lambda_0 = \lambda_0(a, b, \mu, \sigma)$
- $\lambda_1 = \lambda_1(a, b, \mu, \sigma)$
- $\lambda_2 = \lambda_2(a, b, \mu, \sigma)$

Truncated Exponential in $[a, b]$ with mean μ and variance σ^2

Examples of MaxEnt distributions

Available information:

- $\mathrm{Supp}\, X = [0, 1]$
- $\mathbb{E}\left\{\ln(X)\right\} = p, \; |p| < +\infty$
- $\mathbb{E}\left\{\ln(1 - X)\right\} = q, \; |q| < +\infty$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{[0,1]}(x) \, \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

- $a = \left(\mu/\delta^2\right) \left(1/\mu - \delta^2 - 1\right)$
- $b = \left(\mu/\delta^2\right) \left(1/\mu - \delta^2 - 1\right) \left(1/\mu - 1\right)$

Beta with mean μ and coefficient of dispersion $\delta = \mu/\sigma$

Examples of MaxEnt distributions

Available information:

- $\operatorname{Supp} X = (0, +\infty)$
- $\mathbb{E}\{X\} = \mu > 0$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{(0,+\infty)}(x)\,\frac{1}{\mu}\,\exp\left(-\frac{x}{\mu}\right)$$

Exponential with mean μ – $Exp(1/\mu)$

Examples of MaxEnt distributions

Available information:

- $\operatorname{Supp} X = (0, +\infty)$
- $\mathbb{E}\{X\} = \mu > 0$
- $\mathbb{E}\left\{\ln(X)\right\} = q, \; |q| < +\infty$

MaxEnt distribution:

$$p_X(x) = \mathbb{1}_{(0,+\infty)}(x) \frac{1}{\Gamma(k)\,\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)$$

- $k = 1/\delta^2$
- $\theta = \mu\,\delta^2$

Gamma with mean μ and coefficient of dispersion $\delta = \mu/\sigma$

Examples of MaxEnt distributions

Available information:

- $\text{Supp}\, X = (-\infty, +\infty)$
- $\mathbb{E}\left\{X\right\} = \mu \in \mathbb{R}$
- $\mathbb{E}\left\{X^2\right\} = \mu^2 + \sigma^2$

MaxEnt distribution:

$$p_X(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\,\sigma^2}\right\}$$

Normal with mean μ and variance σ^2 — $\mathcal{N}(\mu, \sigma^2)$

Examples of MaxEnt distributions

Available information:

- $\text{Supp}\, X = (0, +\infty)$
- $\mathbb{E}\left\{\ln(X)\right\} = \mu \in \mathbb{R}$
- $\mathbb{E}\left\{\left(\ln(X) - \mu\right)^2\right\} = \sigma^2, \;\; \sigma > 0$

MaxEnt distribution:

$$p_X(x) = \frac{1}{x\sqrt{2\pi\,\sigma^2}} \exp\left\{-\frac{(\ln(x) - \mu)^2}{2\,\sigma^2}\right\}$$

- $\mu = \ln\left(\mu / \sqrt{1 + \delta^2}\right)$
- $\sigma = \sqrt{\ln\left(1 + \delta^2\right)}$

Log-normal with parameters μ and σ — $\ln \mathcal{N}(\mu, \sigma^2)$

Examples of MaxEnt distributions

Available information:

- $\text{Supp}\, X = [a, b]$
- $\mathbb{E}\left\{X\right\} = \mu = (a+b)/2$
- $\mathbb{E}\left\{X^2\right\} = \mu^2 + \sigma^2 < (a+b)^2/4 + (b-a)^2/12$

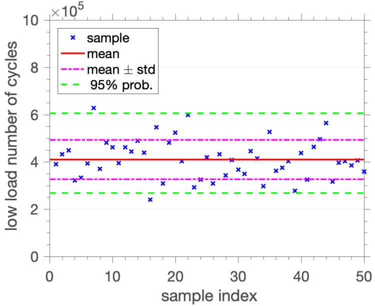
MaxEnt distribution:

$$p_X(x) = \frac{\phi(\frac{x-\mu}{\sigma})}{\sigma\,\left(\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})\right)}$$

- $\phi(\xi)$ – standard normal PDF
- $\Phi(\xi)$ – standard normal CDF

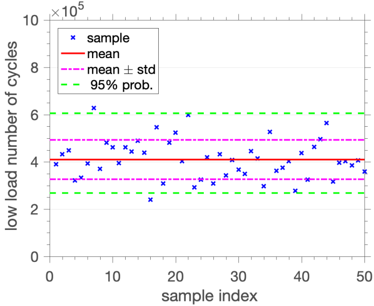
Truncated Normal in $[a, b]$ with mean μ and variance σ^2

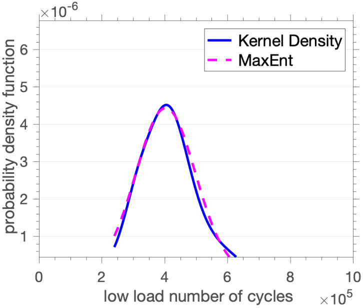
MaxEnt example 2: fatigue data from laboratory



J. P. Dias et al., *Parametric probabilistic approach for cumulative fatigue damage using double linear damage rule considering limited data*, **International Journal of Fatigue**, 127: 246-258, 2019.

MaxEnt example 2: fatigue data from laboratory





J. P. Dias et al., *Parametric probabilistic approach for cumulative fatigue damage using double linear damage rule considering limited data*, **International Journal of Fatigue**, 127: 246-258, 2019.

MaxEnt - Maximum Entropy Code

MaxEnt

https://github.com/americocunhajr/MaxEnt

References

A. Cunha Jr, *Modeling and quantification of physical systems uncertainties in a probabilistic framework*, In: S. Ekwaro-Osire; A. C. Gonçalves; F. M. Alemayehu (Org.), **Probabilistic Prognostics and Health Management of Energy Systems**, Springer, 2017. http://dx.doi.org/10.1007/978-3-319-55852-3_8

S. Guiasu and A. Shenitzer, *The principle of maximum entropy*, **The Mathematical Intelligencer**, vol. 7, pp. 42-48, 1985. https://doi.org/10.1007%2Fbf03023004

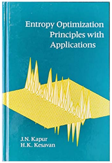
J. N. Kapur and H. K. Kesavan, **Entropy Optimization Principles with Applications**, Academic Press, 1992.

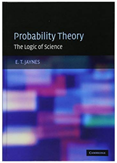
E. T. Jaynes, **Probability Theory: The Logic of Science**, Cambridge University Press, 2003.

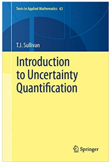
T. J. Sulivan, **Introduction to Uncertainty Quantification**, Springer, 2015.

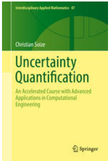
C. Soize **Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering**, Springer, 2017.

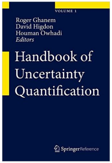
R. Ghanem, D. Higdon and H. Owhadi (Editors) **Handbook of Uncertainty Quantification**, Springer, 2017.











How to cite this material?

A. Cunha Jr, *Maximum Entropy Principle*, Rio de Janeiro State University – UERJ, 2021.

@AmericoCunhaJr @AmericoCunhaJr

@AmericoCunhaJr @AmericoCunhaJr

These class notes may be shared under the terms of
Creative Commons BY-NC-ND 4.0 license,
for educational purposes only.

© copyrighted material

Content excluded from our Creative Commons license

- Portinari Coin:
  https://www.bcb.gov.br/cedulasemoedas/moedascomemorativas
- Volante Mega Sena:
  Wikimedia Commons, File:Volante da Megasena.jpg — Wikimedia Commons, the free media repository
  https://commons.wikimedia.org/w/index.php?title=File:Volante_da_Megasena.jpg
- Several dice figure:
  Wikimedia Commons, File:6dice(cropped).jpg — Wikimedia Commons, the free media repository
  https://commons.wikimedia.org/wiki/File:6dice(cropped).jpg
- Claude Shannon picture:
  Wikimedia Commons, File:ClaudeShannon MFO3807.jpg — Wikimedia Commons, the free media repository https://commons.wikimedia.org/wiki/File:ClaudeShannon_MFO3807.jpg
- John von Neumann picture:
  copyright Los Alamos National Laboratory
  https://en.wikipedia.org/wiki/File:JohnvonNeumann-LosAlamos.gif