Expected Information and its Applications

Uwe Saint-Mont, Hochschule Nordhausen

Berlin (online), 20 April 2026

Motivation

$\ln(1/p) = -\ln p$ gives the information in the **probability** p, $0 \leq p \leq 1$

about the origin

$-\ln(x)$ gives the information in the **observation** x, $x \in (0, \infty)$

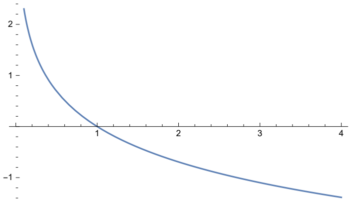


Figure: The function $-ln\ x$ on the non-negative reals

Motivation

Given a pdf $f(x)$ on $(0, \infty)$, the function $(-\ln x) \cdot f(x)$ indicates how much information about the origin is contained in the observation x

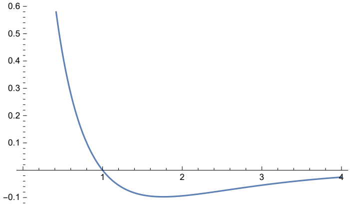
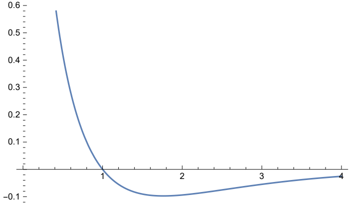


Figure: The function $(-ln\, x)e^{-x}$ for the Standard Exponential

Motivation



The overall information in a Standard Exponential is positive, since

$$\int_0^\infty (-\ln x) e^{-x} \, \mathrm{dx} = \gamma$$

where $\gamma \approx 0.5772$ is Euler's constant

Definition: Expected Information

Given a pdf $f(x)$, the corresponding expected information is

$$
\begin{aligned}
E_I(X) &= \int_0^\infty (-\ln x) \cdot f(x)\,\mathrm{dx} \\
E_I(X) &= \int_{-\infty}^\infty (-\ln |x|) \cdot f(x)\,\mathrm{dx} \\
E_I(X) &= \int_{x\in\mathbb{R}^n} (-\ln ||x||) \cdot f(x)\,\mathrm{dx}
\end{aligned}
$$

Only the distance from the origin matters, not the direction

Center vs. Periphery: positive vs. negative information

Out[ ]=

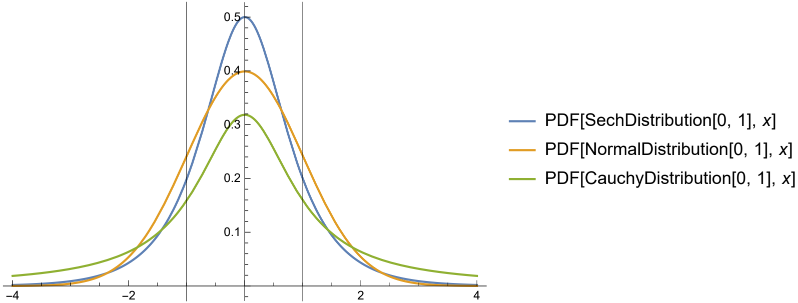
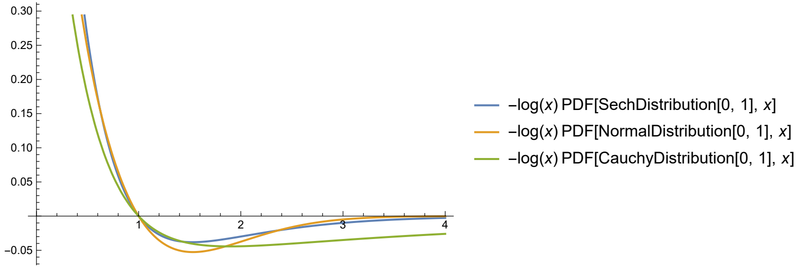


Figure: Center ($|x| < 1$) and Periphery ($|x| > 1$) both contribute to Expected Information

Comparison

Out[ ]=



$$\begin{aligned} E_I(\text{Sech}(0,1)) &= 2\ln\varpi - \ln\pi \approx 0.783 \\ E_I(\text{Normal}(0,1)) &= (\gamma + \ln 2)/2 \approx 0.635 \\ E_I(\text{Cauchy}(0,1)) &= 0 \end{aligned}$$

with the lemniscate constant $\varpi \approx 2.622$

Some Examples

Continuous Distribution	Expected Inform.	Remarks
Laplace(1)	γ	positive
$N \sim$ Normal$(0,1)$	$(\gamma + \ln 2)/2$	$> \gamma$
Lévy$(0,1)$	$-\gamma - \ln 2$	$1/N^2$
'Lognormal'$(0,1)$	0	exp(Normal)
$U \sim$ Uniform$(0,1)$	1	elementary
Pareto$(1,\infty)$	-1	$1/U$
$C \sim$ Cauchy$(0,1)$	0	symmetries!
Weibull$(a,1)$	$-\gamma/a$	
Lindley(b)	$\gamma - \frac{1}{b+1} + \ln b$	
Logistic$(0,1)$	$\gamma + \ln 2 - \ln \pi$	
Sech$(0,1)$	$2 \ln \varpi - \ln \pi$	pdf $= \frac{1}{e^{x\pi/2} + e^{-x\pi/2}}$

$E_I(\cdot)$ highlights how classical distributions are related.

Basic Properties

Whenever expected information exists,

$$E_I(X \cdot Y) = E_I(X) + E_I(Y)$$

In particular

$$E_I(r \cdot X) = E_I(X) + E_I(r) = E_I(X) - \ln|r|$$

Moreover,

$$E_I(X^s) = s \cdot E_I(X),$$

and $E_I(1/X) = -E_I(X)$ in particular.

Exponential mapping: If $X = \exp(-Y)$ then

$$E_I(X) = EY$$

Constructing the Normal: exponential component

Distribution	E_I	**Domain**	**Remark**
$U(0,1)$	1	unit interval	- symmetric -
$\downarrow$ inverse			
$P \sim$ Pareto	-1	$(1, \infty)$	
$\downarrow$ square			polarization
P^2	-2	$(1, \infty)$	
$\downarrow \ln(\cdot)$			
Exponential(1/2)	$\gamma - \ln 2$	$\mathbb{R}^+$	nontrivial info.

By the way: the median of an Exponential(1/2) is $2 \ln 2$.

Constructing the Normal: circular component

Distribution	E_I	**Domain**	**Remark**
$U(0,1)$	1	unit interval	- symmetric -
$\downarrow \cdot\pi/2$			
$U(0,\pi/2)$	$1-\ln\pi+\ln 2$	segment	
$\downarrow \sin(\cdot)$			
pdf $= \frac{2}{\pi\sqrt{1-x^2}}$	$\ln 2$	unit interval	
$\downarrow$ square			polarization
pdf $= \frac{1}{\pi\sqrt{x(1-x)}}$			
$\text{Arcsin}(0,1)$	$2\ln 2$	unit interval	- symmetric -

Constructing the Normal: putting all together

Distribution	E_I	**Domain**	**Remark**
$X \sim$ Exponential$(1/2)$	$\gamma - \ln 2$	$\mathbb{R}^+$	
$Y \sim$ Arcsin$(0,1)$	$2\ln 2$	$(0,1)$	
$X \cdot Y$	$\gamma + \ln 2$	$\mathbb{R}^+$	
$\sqrt{XY}$	$(\gamma + \ln 2)/2$	$\mathbb{R}^+$	depolarization
$S \cdot \sqrt{XY}$	$(\gamma + \ln 2)/2$	$\mathbb{R}$	symmetrization

Actually,

$$Z = \pm\sqrt{XY} = \pm\sqrt{\ln(U_1^{-2}) \cdot \sin^2\left(\frac{\pi}{2} U_2\right)} \sim N(0,1)$$

where U_i are independent, Standard Uniform

Entropy and Information Functions

Expected Information:

$$E_I(X) = \int_0^\infty (-\ln x) \cdot f(x)\,\mathrm{dx}$$

Entropy is a close non-parametric cousin:

$$H(X) = \int_0^\infty (-\ln f(x)) \cdot f(x)\,\mathrm{dx}$$

In general - Integral with respect to an information function $g(x)$:

$$G(X) = \int_0^\infty (-\ln g(x)) \cdot f(x)\,\mathrm{dx}$$

Integral Transforms

Laplace:

$$L_f(s) = \int_0^\infty e^{-sx} \cdot f(x)\,\mathrm{dx}$$

Mellin:

$$M_f(s) = \int_0^\infty x^{s-1} \cdot f(x)\,\mathrm{dx}$$

Logarithmic:

$$T_f(s) = \int_0^\infty (-\ln x)^s \cdot f(x)\,\mathrm{dx}$$

Logarithmic Moments and Information Variance

Logarithmic Moments:

$$M_I^s(X) = \int_0^\infty (-\ln x)^s f(x)\,\mathrm{dx}$$

Information Variance:

$$\begin{aligned}\sigma_I^2(X) &= M_I^2(X) - (E_I(X))^2 \\ &= \int ((-\ln|x|) - E_I(X))^2 f(x)\,\mathrm{dx}\end{aligned}$$

Exponential mapping: If $X = \exp(-Y)$ then

$$\sigma_I^2(X) = \sigma^2(Y)$$

Properties of Information Variance

Given independence,

$$\sigma_I^2(X \cdot Y) = \sigma_I^2(X) + \sigma^2(Y)$$

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$$

Further Analogies:

$$\sigma_I^2(a \cdot X) = \sigma_I^2(X) \qquad \sigma_I^2(X^b) = b^2 \cdot \sigma_I^2(X)$$

$$\sigma^2(a + X) = \sigma^2(X) \qquad \sigma^2(b \cdot X) = b^2 \cdot \sigma^2(X)$$

In particular: $\sigma_I^2(1/X) = \sigma_I^2(X)$

Some Examples

Distribution	E_l	σ_l^2
Normal$(0,1)$	$(\gamma + \ln 2)/2$	$\pi^2/8$
Laplace(1)	γ	$\pi^2/6 = 0.1\bar{6}\pi^2$
Cauchy$(0,1)$	0	$\pi^2/4 = 0.25\pi^2$
Lévy$(0,1)$	$-\gamma - \ln 2$	$\pi^2/2$
Holtsmark	$\gamma/3$	$17\pi^2/108 \approx 0.157\pi^2$
Stable$(\alpha, 0)$	$\gamma(1 - 1/\alpha)$	$(1 + 2/\alpha^2)\pi^2/12$
Student t_n	$\frac{\gamma+\ln 2}{2} + \frac{\psi(n/2)-\ln(n/2)}{2}$	$\pi^2/8 + (\psi^{(1)}(n/2))/4$
Gamma(a,b,c)	$-\psi(a)/c - \ln b$	$\psi^{(1)}(a)/c^2$

where the pdf of a Gamma(a, b, c) is
$f(x) = \frac{c}{b\,\Gamma(a)}(x/b)^{ac-1} \exp\left(-(x/b)^c\right)$,
with the gamma function $\Gamma(x)$, the digamma function
$\psi(x) = (\ln \Gamma(x))'$, and the trigamma function $\psi^{(1)}(x) = \psi'(x)$

Convergence of Student's t towards the Normal

Out[ ]=

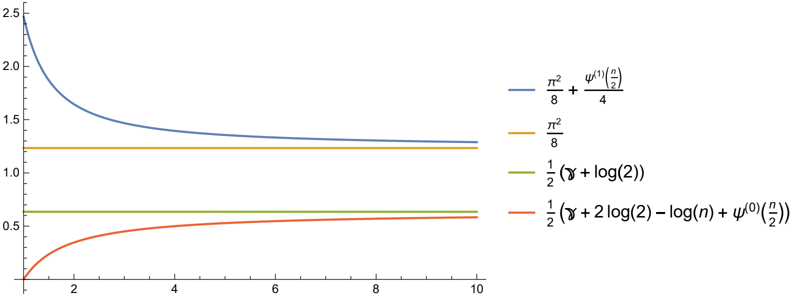


Figure: Expected information and information variance of t_n.

Application: Statistics

Since expected information and information variances add up,

$\Rightarrow$ **Analysis of Information Variance (ANOIVA)**

based on $E_I(XY) = E_I(X) + E_I(Y)$ and $\sigma_I^2(XY) = \sigma_I^2(X) + \sigma_I^2(Y)$

Example: Products of (independent) Normals and Cauchys

	$N(0,1)$	$N_1 N_2$	N^{-2}	$C(0,1)$	$C_1 C_2$	C^2
$E(\cdot)$	0	0	–	–	–	–
$\sigma^2(\cdot)$	1	1	–	–	–	–
$E_I(\cdot)$	$(\gamma + \ln 2)/2$	$\gamma + \ln 2$	$-\gamma - \ln 2$	0	0	0
$\sigma_I^2(\cdot)$	$\pi^2/8$	$\pi^2/4$	$\pi^2/2$	$\pi^2/4$	$\pi^2/2$	π^2

The Normal and some of its Cousins

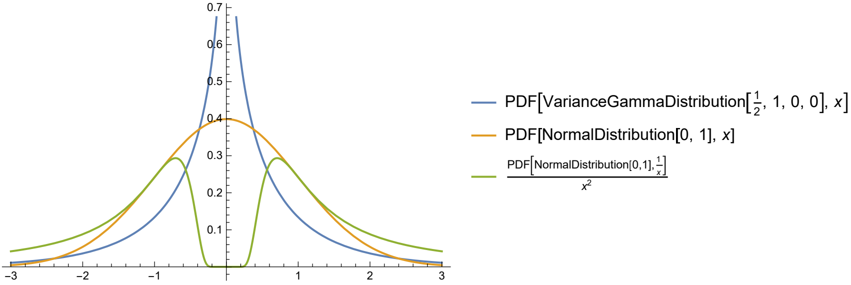


Figure: Product of two Normals, the Standard Normal and its Inverse

More general idea: consider sums, products and powers of random variables, i.e., the **Algebra of Distributions**

Major difference: There are many distributions $\mathcal{D}$ with $E_I(\mathcal{D}) = 0$

A Map of Distributions: the Stochastic Askey Scheme

Notation:

- OPS = Orthogonal Polynomial System
- *OPS* = Sheffer OPS = Appell and Binomial OPS
- (generalized) hypergeometric function

$${}_rF_s\left(\begin{array}{c} a_1,\ldots,a_r \\ b_1,\ldots,b_s \end{array};z\right)=\sum_{k=0}^{\infty}\frac{(a_1,\ldots,a_r)_k}{(b_1,\ldots,b_s)_k}\frac{z^k}{k!}$$

where $(a_1,\ldots,a_r)_k=(a_1)_k\cdots(a_r)_k$,

and $(a)_k=\prod_{i=1}^{k}(a+i-1)$
is the Pochhammer symbol (rising factorial)

OPS and their corresponding Distributions

Tier	Hypergeom. Function	Orthogonal Poly. System	Distribution (weights)	Convergence
(III)	${}_2F_1(\cdot)$	*M.-P.*	**Sech**	$\rightarrow$ (I, IIa)
	${}_2F_1(\cdot)$	Jacobi	Beta	$\rightarrow$ (I, IIa, IIb)
	${}_2F_1(\cdot)$	Pseudo Jacobi	Student t	
	${}_2F_1(\cdot)$	*Meixner*	**Neg. Binom.**	$\rightarrow$ (IIa, IIc)
	${}_2F_1(\cdot)$	*Krawtchouk*	**Binomial**	$\rightarrow$ (I, IIc)
(IIa)	${}_1F_1(\cdot)$	*Laguerre*	**Gamma**	$\rightarrow$ (I)
(IIb)	${}_1F_1(\cdot), {}_2F_0(\cdot)$	Bessel	Inv. Gamma	
(IIc)	${}_2F_0(\cdot)$	*Charlier*	**Poisson**	$\rightarrow$ (I)
(I)	${}_2F_0(\cdot)$	*Hermite*	**Normal**	

where M.-P. stands for Meixner-Pollaczek

Sheffer OPS ⇔ Stochastic Askey Scheme

		(Bell) **Poisson** *Charlier*		
	↗		↖	
	(Stirling, 2nd) **Binomial** *Krawtchouk*		(Stirling, 2nd) **Neg. Binomial** *Meixner*	
↙				↘
(double factorial) **Normal** *Hermite*	←	(Euler) **Sech** *M.-P.*	→	(rising factorial) **Gamma** *Laguerre*

Table: *Sheffer OPS*, corresp. **Distributions** and (Moments / Numbers)

Extended Stochastic Askey Scheme: Cauchy

Discrete Distributions

		Poisson		
	$\swarrow$		$\searrow$	
Binomial	⇠	**Bernoulli**	⇢	**Geometric**
Normal	$\leftarrow$	**Sech**	$\rightarrow$	**Exponential**
$\diagdown$				$\nearrow \ln(\cdot)$
$\downarrow \exp(\cdot)$	**Student t** Pseudo Jacobi	$\uparrow \ln\lvert\cdot\rvert$	**Uniform** Legendre	⇢
	$\diagdown$		$\diagup$	
'Lognormal'		**Cauchy**		

Continuous Distributions and corresponding OPS

Extended Stochastic Askey Scheme: Beta

Discrete Distributions

Poisson				
	↘			
Bernoulli	⇢	**Geometric**	⇢	**Negative Binomial**
Sech	→	**Exponential**	⇢	**Gamma**
		↗ ln(·)	↙	ratios
↑ ln|·|	**Uniform**	⇢	**Beta1**	↓
	Legendre		Jacobi	
	↗			
Cauchy		*Zero-Information Tier*		**Beta2**(p, p)

Continuous Distributions and corresponding OPS

Extended Stochastic Askey Scheme: Below the Line

E_I	Distribution	Ratio Distributions
positive	**Gamma** *Laguerre*	**Beta1**(p, q) Jacobi
zero		**Beta2**(p, p) Jacobi
negative	**Inverse Gamma** Bessel	**Beta1**(q, p) Jacobi

E_I	Product and Ratio Distributions		
positive	**Uniform**$(0, 1)$ Legendre	**Normal** N *Hermite*	**VarianceGammaDistrib.** $N_1 \cdot N_2$ pdf$= K_0(\|x\|)/\pi$
zero **ratios:**	**'NoName'** U_1/U_2	**Cauchy** $C = N_1/N_2$	**'Product-Ratio Cauchy'** $\frac{N_1 N_2}{N_3 N_4}$ pdf$= \frac{ln(x^2)}{\pi^2(x^2-1)}$
negative	**Pareto** OPS = ?	**InvNormal** $1/N$	**Inverse VarianceGamma** pdf$= K_0(\|1/x\|)/(\pi x^2)$

Logarithmic Transform of Ratio Distributions

Distribution	Ratio	Difference
Laplace	$\ln(U_1/U_2)$	$\text{Exponential}_1 - \text{Exponential}_2$
Logistic	$\ln(\text{Exp}_1/\text{Exp}_2)$	$\text{Gumbel}_1 - \text{Gumbel}_2$
Sech	$\ln\|C\|$	$\ln\|N_1\| - \ln\|N_2\|$
Csch	$\ln\|C_1 C_2\|$	$\ln\|N_1 N_3\| - \ln\|N_2 N_4\|$
'Morris Family'	$\ln\|C_1 \cdots C_n\|$	...

Out[ ]=

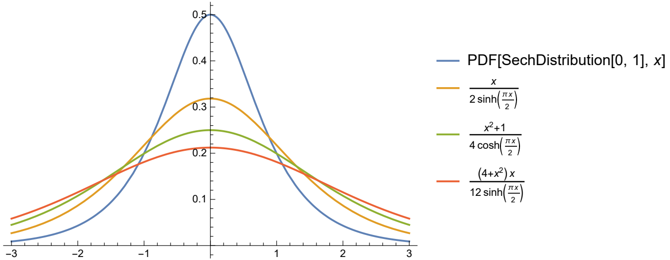


Figure: Generalized Hyperbolic Secant Distributions ('Morris Family')

Interesting Facts / Questions

Cauchy and Normal in several Dimensions p

p	1	2	3	4
$E_I(\mathbf{C})$	0	$-\ln 2 \approx -0.69$	-1	$-\ln 2 - \frac{1}{2} \approx -1.19$
$\sigma_I^2(\mathbf{C})$	$\frac{\pi^2}{4}$ ≈ 2.47	$\frac{\pi^2}{6}$ ≈ 1.64	$\frac{\pi^2}{4} - 1$ ≈ 1.47	$\frac{\pi^2}{6} - \frac{1}{4}$ ≈ 1.39
$E_I(\mathbf{N})$	$\frac{\gamma+\ln 2}{2}$ ≈ 0.64	$\frac{\gamma-\ln 2}{2}$ ≈ -0.06	$\frac{\gamma+\ln 2-1}{2}$ ≈ -0.36	$\frac{\gamma-\ln 2-1}{2}$ ≈ -0.56
$\sigma_I^2(\mathbf{N})$	$\pi^2/8$ ≈ 1.23	$\pi^2/24$ ≈ 0.41	$\pi^2/8 - 1$ ≈ 0.23	$\pi^2/24 - 1/4$ ≈ 0.16

Logarithmic Moments of other Multidimensional Distributions (Exponential, Uniform, Pareto,...) ?

Interesting Facts / Questions

'Simultaneous' Moments

For instance,

$$\int_0^1 \frac{(-\ln x)^{s-1} x^n}{1-x}\,\mathrm{dx} = \Gamma(s)\,\zeta(s, n+1)$$

$$\int_0^1 \int_0^1 \frac{(-\ln xy)^{s-2}(xy)^n}{1-xy}\,\mathrm{dx}\,\mathrm{dy} = \Gamma(s)\,(\zeta(s,n) - n^{-s})$$

$$\int_0^1 \frac{(-\ln x)^{s-1} x^n}{1+x^2}\,\mathrm{dx} = \frac{\Gamma(s)}{4^s}\left(\zeta\left(s, \frac{n+1}{4}\right) - \zeta\left(s, \frac{n+3}{4}\right)\right)$$

where $\zeta(s,m) = \sum_{k=0}^{\infty} \frac{1}{(k+m)^s}$ is the Hurwitz zeta function.

Note the shift '$+m$' in the denominator!

Applications / Outlook: Statistics

Traditional statistical modelling:
Data = Structure + Error
where the error is normally distributed, since $E(N) = 0$

Extension:
Data = Structure · Error
Here, a Cauchy error is a natural choice, since $E_I(C) = 0$

General Models
should employ both kinds of error

Applications / Outlook: Signal $\leftrightarrow$ Noise

Signal vs. Noise Paradigm

- Convergence (towards a point),
  e.g. of opinions, sedimentation
- Divergence (towards infinity),
  e.g. polarization and confusion, erosion
- Middle Ground: Fractals and Distributions

Stochastic Distributions are a nice **compromise:**

- there are many distributions, and they are closely related
- centripetal 'force' $\approx$ centrifugal 'force' (center vs. periphery)
- precise analysis: positive vs. negative expected information...
- they model Mandelbrot's States of Randomness:
  Mild - Slow - Wild (e.g. the Normal vs. Power Laws)

Thank you!

Uwe Saint-Mont (2026): Information is Key.

Parametric Information Theory, The Signal-to-Noise Paradigm, Stochastic Distributions, and Logarithmic Functions.

Published on Amazon:

ISBN Hardcover (colour): 979-8-249695-18-7
ISBN Softcover (b/w): 979-8-249687-57-1