Centro Brasileiro de Pesquisas Físicas

Métodos para Análise de grande volume de dados e Astroinformática

Clécio Roque De Bom – debom@cbpf.br

clearnightsrthebest.com

CBPF

Error Theory

O *mensurando*[1] é a grandeza a ser determinada num processo de medição. Como regra geral, *valor verdadeiro do mensurando é uma quantidade sempre desconhecida*, Isto é, mesmo após a medição, o valor verdadeiro do mensurando só pode ser conhecido aproximadamente, devido a *erros de medição*.

Se y_v é o valor verdadeiro de um mensurando e y é o resultado de uma medição, o *erro* em y é definido por

$$\eta = y - y_v \,. \tag{3.1}$$

Error Theory

Geralmente, o erro η em um valor experimental y tem diversas causas[4]. Isto significa que o erro total η pode ser escrito como uma soma de q *erros elementares* η_1, η_2, $\cdots$, η_q:

$$\eta = \eta_1 + \eta_2 + \cdots + \eta_q . \tag{3.2}$$

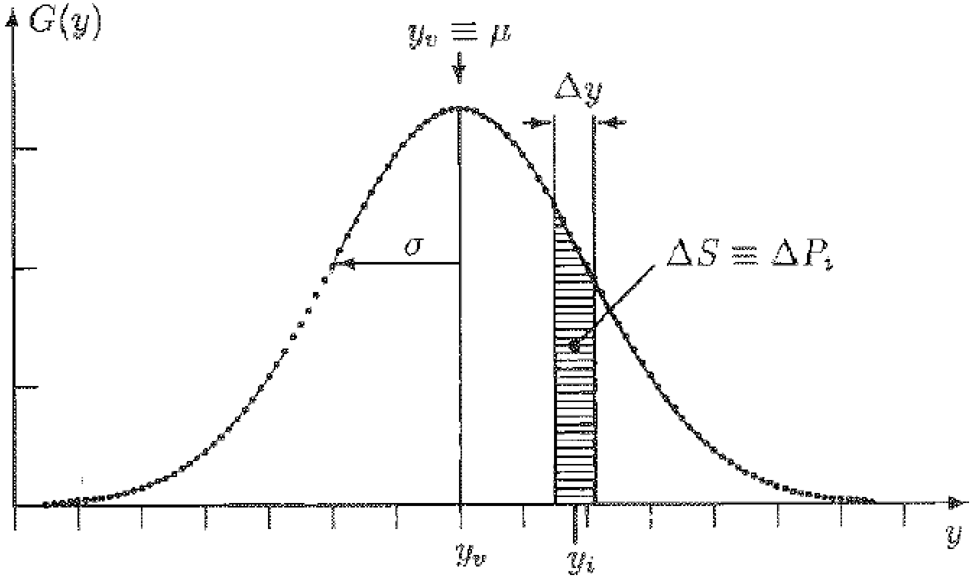


Figura 3.1. *Função gaussiana de probabilidades. A probabilidade ΔP_i de obter uma medida y no intervalo Δy é a área ΔS. A área total sob a curva é 1, devido à condição de normalização 2.11.*

Error Theory

A distribuição de Laplace-Gauss ou distribuição gaussiana de erros é definida pela função de densidade de probabilidade[5] dada por

$$G(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} , \tag{3.3}$$

Teorema do Limite Central

Erros aleatórios independentes η_1, η_2, $\cdots$ *e* η_q *são admitidos como tendo distribuições de probabilidade quaisquer com variâncias finitas e tais que nenhum* η_i *particular é muito maior que os demais. Nestas condições, se o erro total é* $\eta = \eta_1 + \eta_2 + \cdots + \eta_q$, *então, a distribuição de erros para* η *converge para uma distribuição gaussiana, no limite* $q \to \infty$.

Fundamental Concepts

Average

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 ,$$

Standard Deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Uncertainty

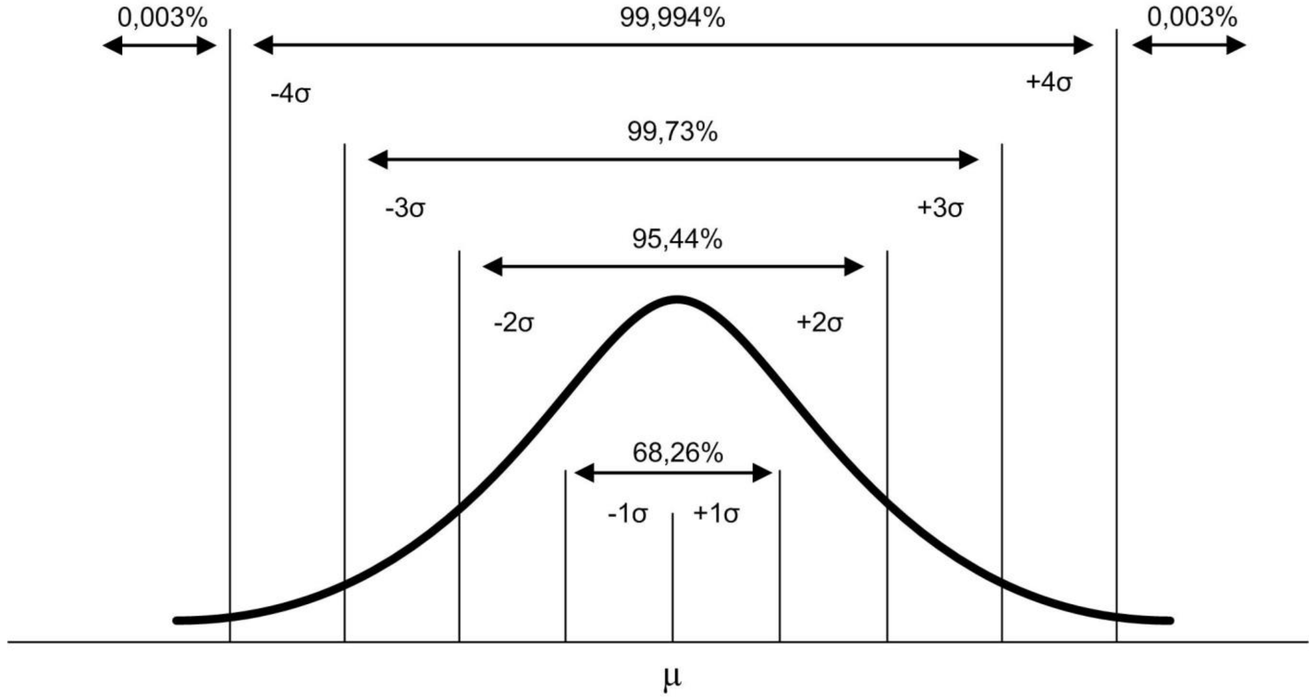
Se y_v é o valor verdadeiro de um mensurando e y é o resultado de um processo de medição, a probabilidade $P(\delta)$ de se obter um resultado y no intervalo

$$y_v - \delta \ < \ y \ < \ y_v + \delta \tag{4.2}$$

pode ser obtida integrando-se a função densidade de probabilidade (Equação 2.10). Admitindo uma distribuição gaussiana para os erros,

$$P(\delta) \ = \ \frac{1}{\sigma_v \sqrt{2\,\pi}} \int_{y_v-\delta}^{y_v+\delta} e^{-\frac{1}{2}\left(\frac{y-y_v}{\sigma_v}\right)^2} \, dy\,, \tag{4.3}$$

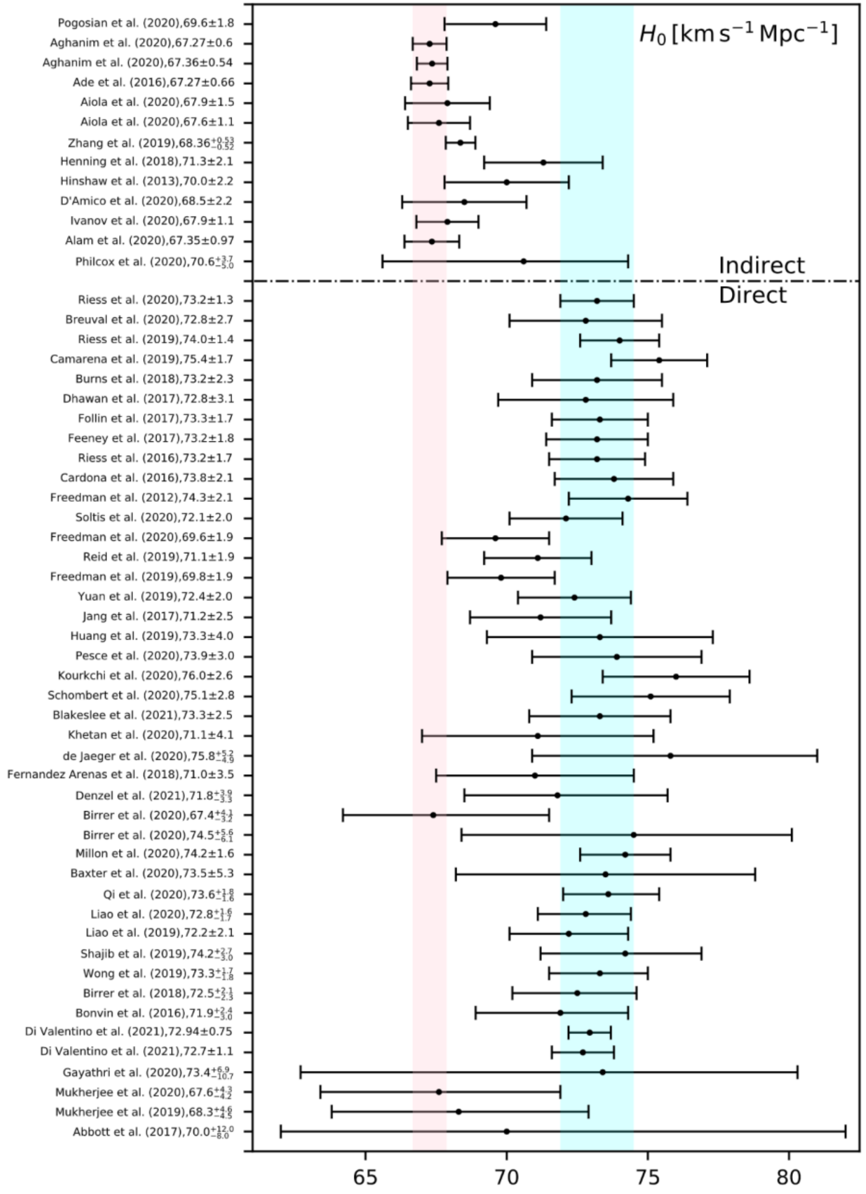
Confidence Levels

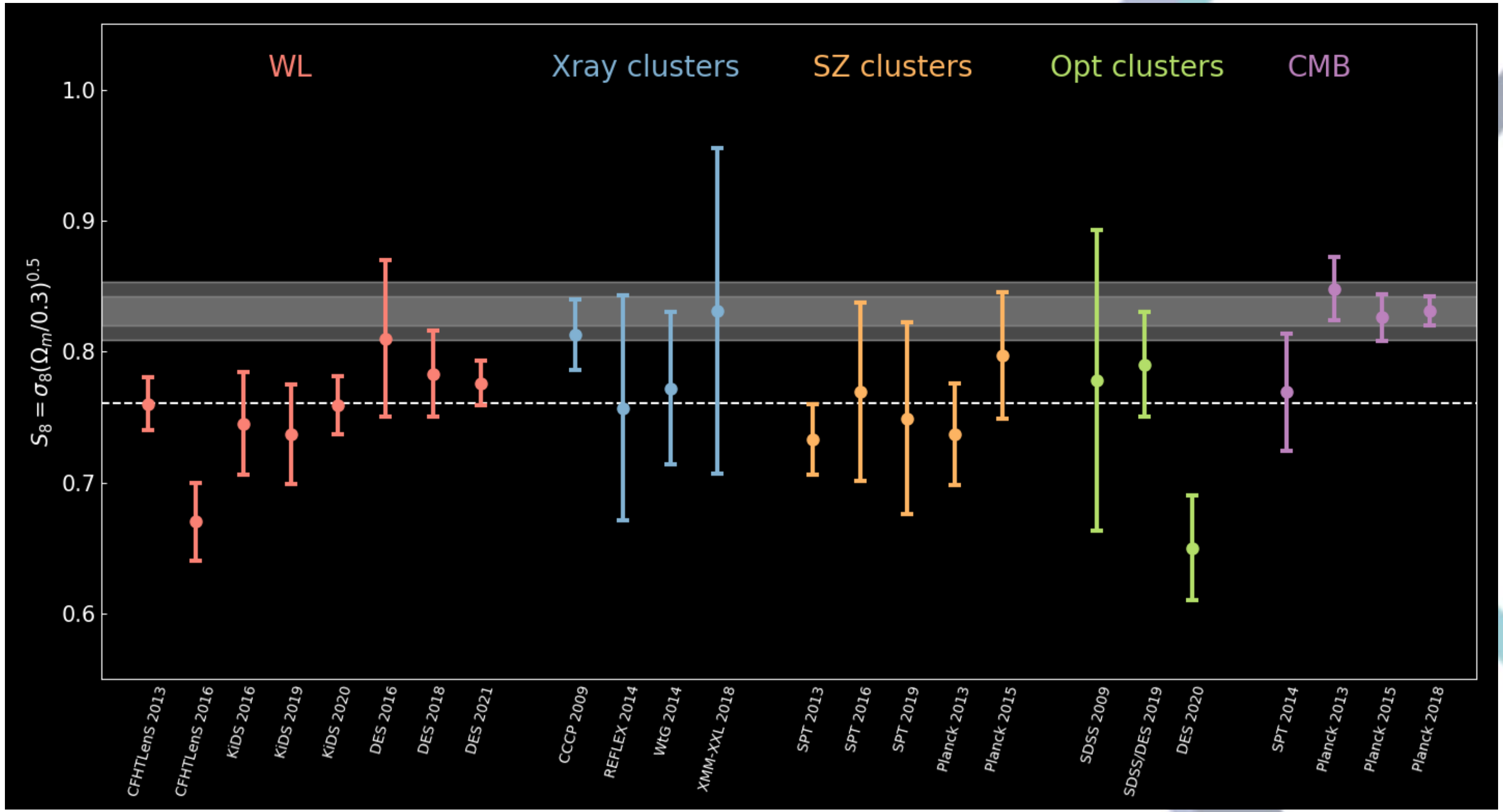


Confidence Levels

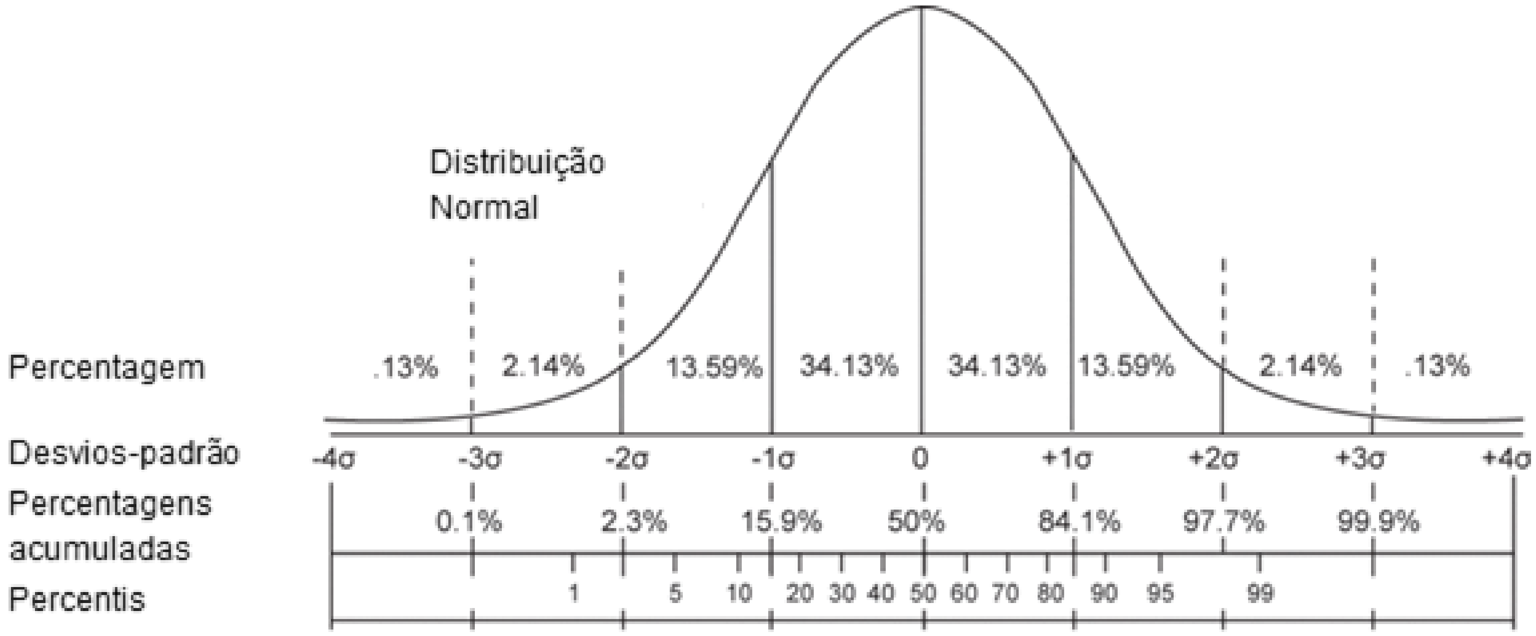
Incerteza		*Intervalo de confiança*	*Confiança*
Incerteza padrão	σ_v	$(y - \sigma_v) < y_v < (y + \sigma_v)$	68,27 %
δ	$2\sigma_v$	$(y - 2\sigma_v) < y_v < (y + 2\sigma_v)$	95,45 %
δ	$3\sigma_v$	$(y - 3\sigma_v) < y_v < (y + 3\sigma_v)$	99,73 %
δ	$1,645\sigma_v$	$(y - \delta) < y_v < (y + \delta)$	90 %
δ	$2,576\sigma_v$	$(y - \delta) < y_v < (y + \delta)$	99 %
Erro provável	Δ	$(y - \Delta) < y_v < (y + \Delta)$	50 %

Tensions?

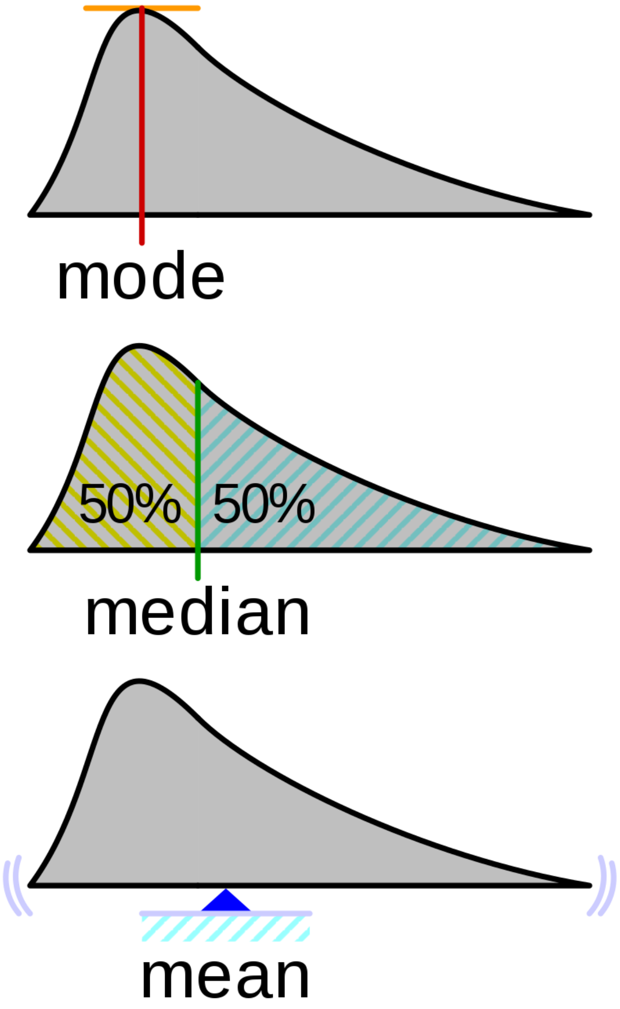

$H_0\,[\mathrm{km\,s^{-1}\,Mpc^{-1}}]$
Pogosian et al. (2020),69.6±1.8
Aghanim et al. (2020),67.27±0.6
Aghanim et al. (2020),67.36±0.54
Ade et al. (2016),67.27±0.66
Aiola et al. (2020),67.9±1.5
Aiola et al. (2020),67.6±1.1
Zhang et al. (2019),$68.36^{+0.53}_{-0.52}$
Henning et al. (2018),71.3±2.1
Hinshaw et al. (2013),70.0±2.2
D'Amico et al. (2020),68.5±2.2
Ivanov et al. (2020),67.9±1.1
Alam et al. (2020),67.35±0.97
Philcox et al. (2020),$70.6^{+3.7}_{-5.0}$
Indirect
Direct
Riess et al. (2020),73.2±1.3
Breuval et al. (2020),72.8±2.7
Riess et al. (2019),74.0±1.4
Camarena et al. (2019),75.4±1.7
Burns et al. (2018),73.2±2.3
Dhawan et al. (2017),72.8±3.1
Follin et al. (2017),73.3±1.7
Feeney et al. (2017),73.2±1.8
Riess et al. (2016),73.2±1.7
Cardona et al. (2016),73.8±2.1
Freedman et al. (2012),74.3±2.1
Soltis et al. (2020),72.1±2.0
Freedman et al. (2020),69.6±1.9
Reid et al. (2019),71.1±1.9
Freedman et al. (2019),69.8±1.9
Yuan et al. (2019),72.4±2.0
Jang et al. (2017),71.2±2.5
Huang et al. (2019),73.3±4.0
Pesce et al. (2020),73.9±3.0
Kourkchi et al. (2020),76.0±2.6
Schombert et al. (2020),75.1±2.8
Blakeslee et al. (2021),73.3±2.5
Khetan et al. (2020),71.1±4.1
de Jaeger et al. (2020),$75.8^{+5.2}_{-4.9}$
Fernandez Arenas et al. (2018),71.0±3.5
Denzel et al. (2021),$71.8^{+3.9}_{-3.3}$
Birrer et al. (2020),$67.4^{+4.1}_{-3.2}$
Birrer et al. (2020),$74.5^{+5.6}_{-6.1}$
Millon et al. (2020),74.2±1.6
Baxter et al. (2020),73.5±5.3
Qi et al. (2020),$73.6^{+1.8}_{-1.6}$
Liao et al. (2020),$72.8^{+1.6}_{-1.7}$
Liao et al. (2019),72.2±2.1
Shajib et al. (2019),$74.2^{+2.7}_{-3.0}$
Wong et al. (2019),$73.3^{+1.7}_{-1.8}$
Birrer et al. (2018),$72.5^{+2.1}_{-2.3}$
Bonvin et al. (2016),$71.9^{+2.4}_{-3.0}$
Di Valentino et al. (2021),72.94±0.75
Di Valentino et al. (2021),72.7±1.1
Gayathri et al. (2020),$73.4^{+6.9}_{-10.7}$
Mukherjee et al. (2020),$67.6^{+4.3}_{-4.2}$
Mukherjee et al. (2019),$68.3^{+4.6}_{-4.5}$
Abbott et al. (2017),$70.0^{+12.0}_{-8.0}$
65
70
75
80


WL
Xray clusters
SZ clusters
Opt clusters
CMB
$S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$
1.0
0.9
0.8
0.7
0.6
CFHTLenS 2013
CFHTLenS 2016
KiDS 2016
KiDS 2019
KiDS 2020
DES 2016
DES 2018
DES 2021
CCCP 2009
REFLEX 2014
WtG 2014
XMM-XXL 2018
SPT 2013
SPT 2016
SPT 2019
Planck 2013
Planck 2015
SDSS 2009
SDSS/DES 2019
DES 2020
SPT 2014
Planck 2013
Planck 2015
Planck 2018

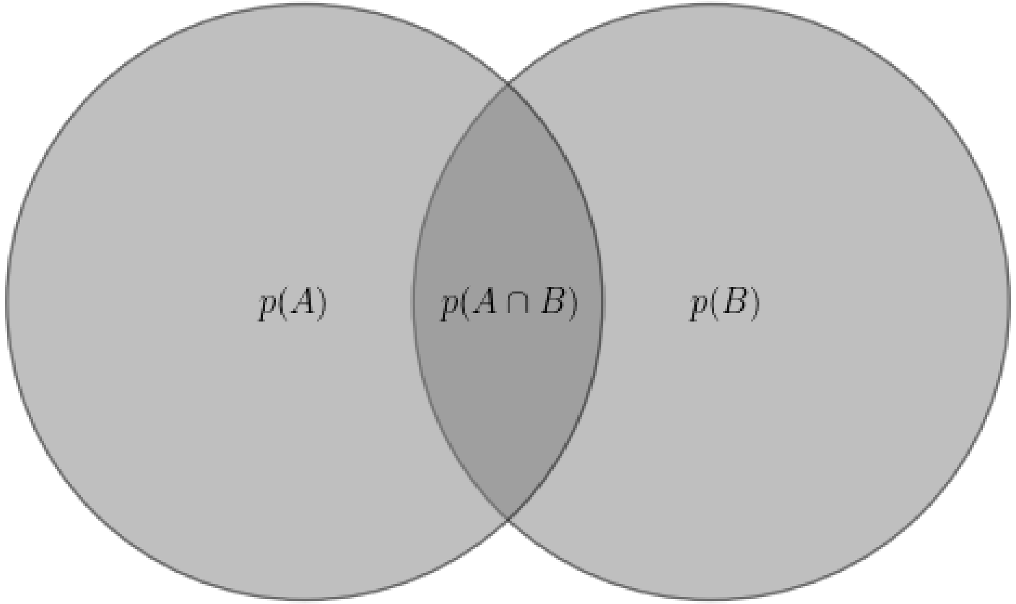
Confidence Levels - Percentil



Confidence Levels - Percentil



Probabilities



$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A)$$

Probabilities

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A)$$

A = hit head on door frame, B = { is tall, is average, is short }
P(A) = P(A|is tall) + P(B|is average) + P(C|is short)

if A and B are independent, then

$$p(A, B) = p(A)p(B)$$

Types of Errors

ML / Physics	Aleatoric	Epistemic
Statistical	noise in data, stdev of measurements (noise in period T)	—
Systematic	noise in data, not stdev of measurements (noise in length L)	model fidelity, not stdev of measurements (far from training set)

Expected Values

$$\mathbf{E}[X] = \int_{\mathbb{R}} x f(x)\, dx,$$

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x)\, dx.$$

Centro Brasileiro de Pesquisas Físicas

Métodos para Análise de grande volume de dados e Astroinformática

Clécio Roque De Bom – debom@cbpf.br

clearnightsrthebest.com

CBPF