ENM 360: Introduction to Data-driven Modeling

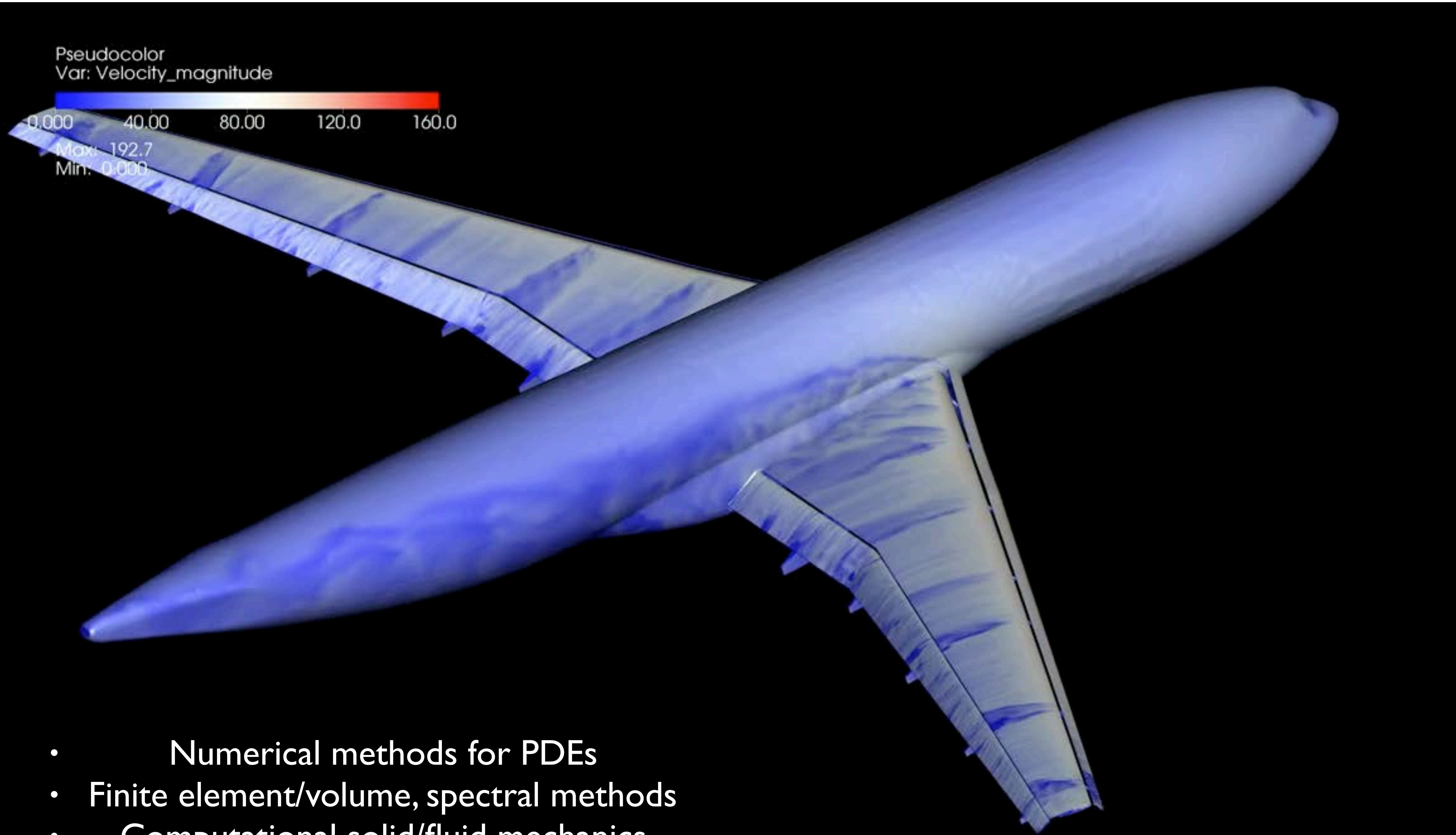
Lecture #4: Primer on Probability and Statistics

Paris Perdikaris
September 10, 2019

Penn
University *of* Pennsylvania

Computational science and engineering

Data ⟷ *Models*

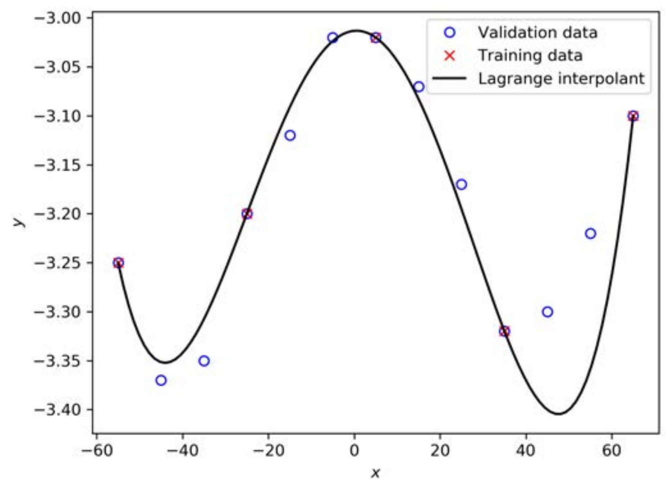


- Numerical methods for PDEs
- Finite element/volume, spectral methods
- Computational solid/fluid mechanics

Deterministic vs probabilistic modeling

$\boldsymbol{x}$

Inputs
(deterministic)



$$\boldsymbol{y} = f_\theta(\boldsymbol{x})$$

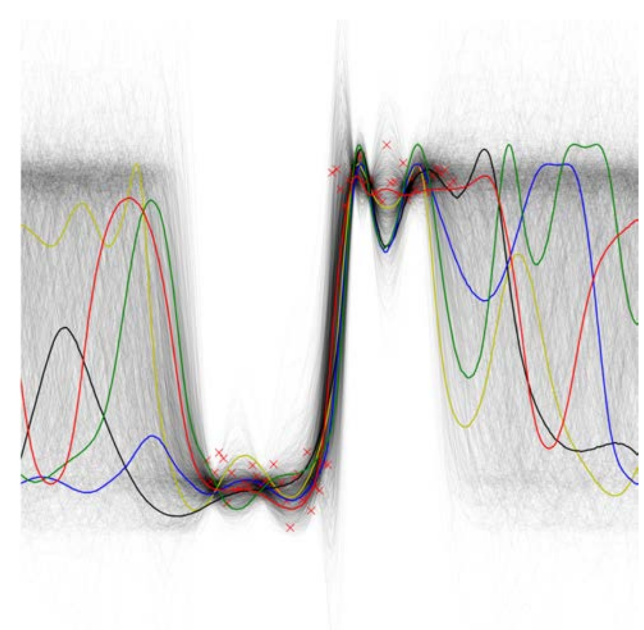
Outputs

Inputs
(random)

$\boldsymbol{z} \sim p(\boldsymbol{z})$

$\boldsymbol{x}$

Inputs
(deterministic)



$$p_\theta(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z})$$

Outputs

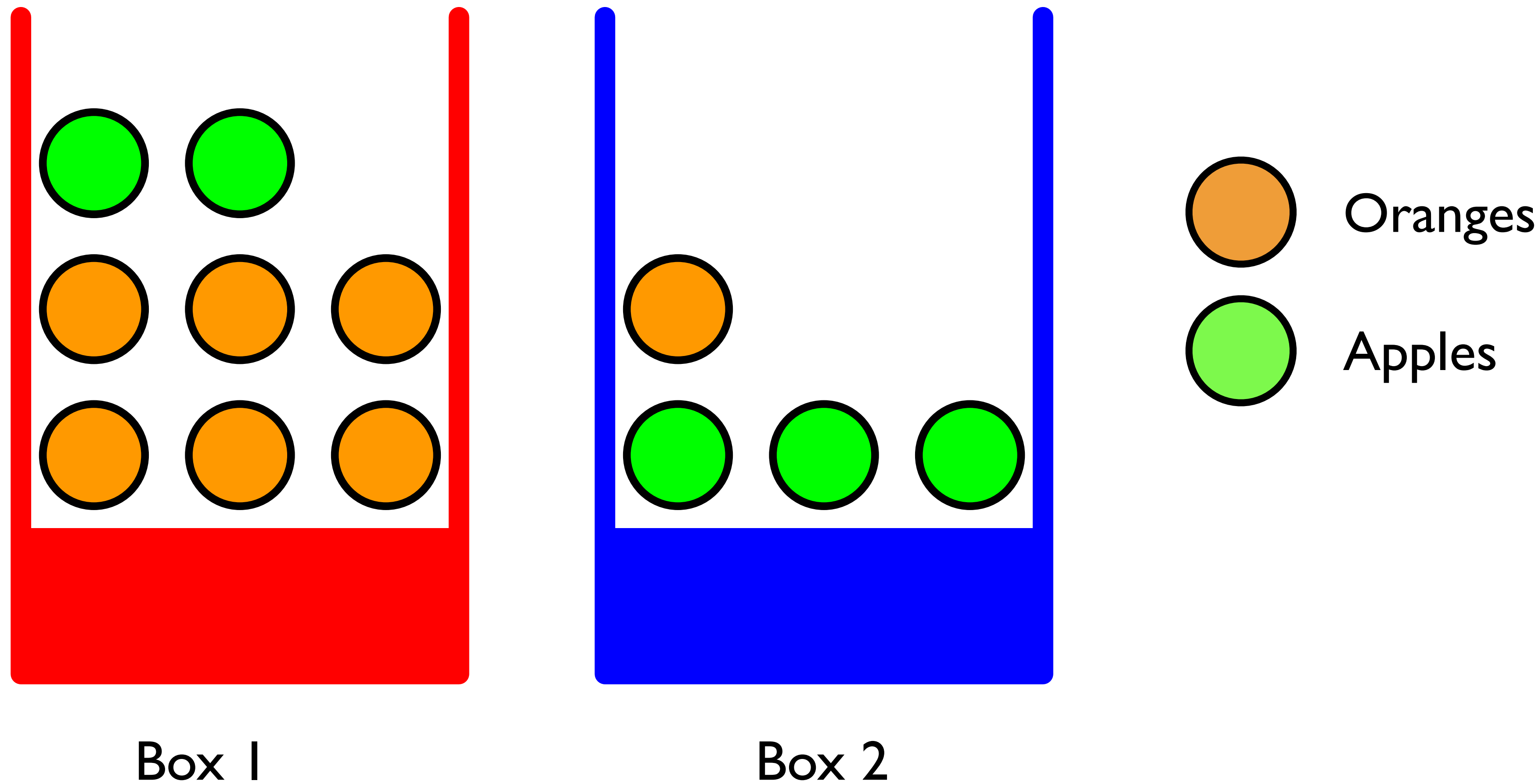
Discrete random variables

Random variables:

- Box ID
- Fruit

$$B = \{r, b\}$$

$$F = \{a, o\}$$



Suppose we randomly pick one of the boxes and from that box we randomly select an item of fruit, and having observed which sort of fruit it is we replace it in the box from which it came. We could imagine repeating this process many times. Let us suppose that in so doing we pick the red box 40% of the time and we pick the blue box 60% of the time, and that when we remove an item of fruit from a box we are equally likely to select any of the pieces of fruit in the box.

Probability of an event: fraction of times that event occurs out of the total number of trials, in the limit that the total number of trials goes to infinity .

We can now ask questions such as: "what is the overall probability that the selection procedure will pick an apple?", or "given that we have chosen an orange, what is the probability that the box we chose was the blue one?"

Discrete random variables

- A ***discrete random variable*** is one which may take on only a countable number of distinct values such as 0,1,2,3,4,........ Discrete random variables are usually (but not necessarily) counts. If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a b
- The ***probability distribution*** of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.

Basic rules of probability

Sum rule

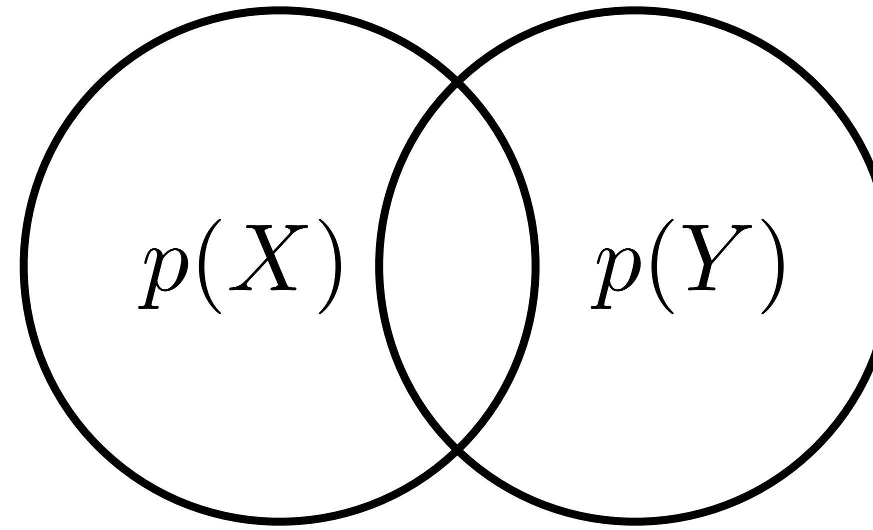
$$p(X) = \sum_{Y} p(X, Y)$$

Product rule

$$p(X, Y) = p(Y|X)p(X)$$

Bayes rule

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$



Venn diagrams

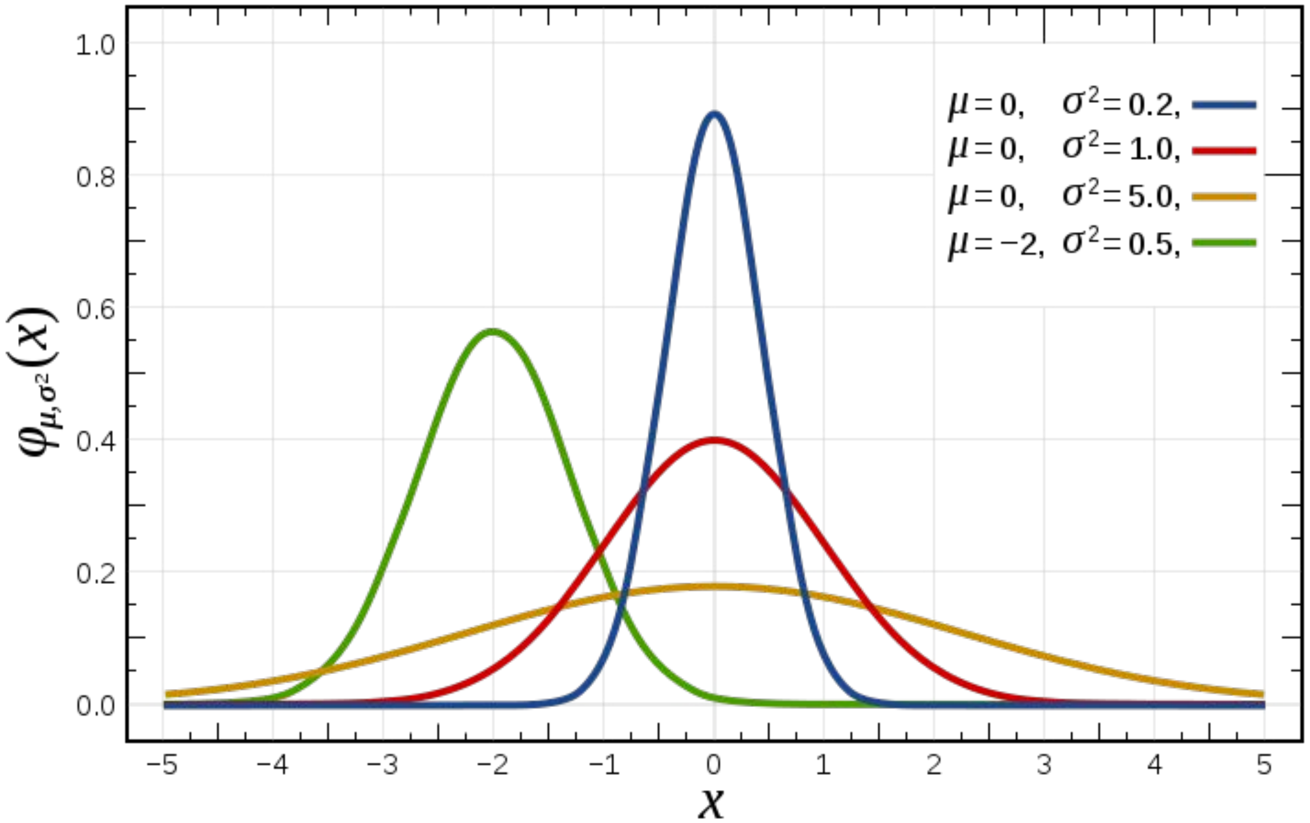
Reading: Chapters 1,2 from Bishop “Pattern Recognition and Machine Learning

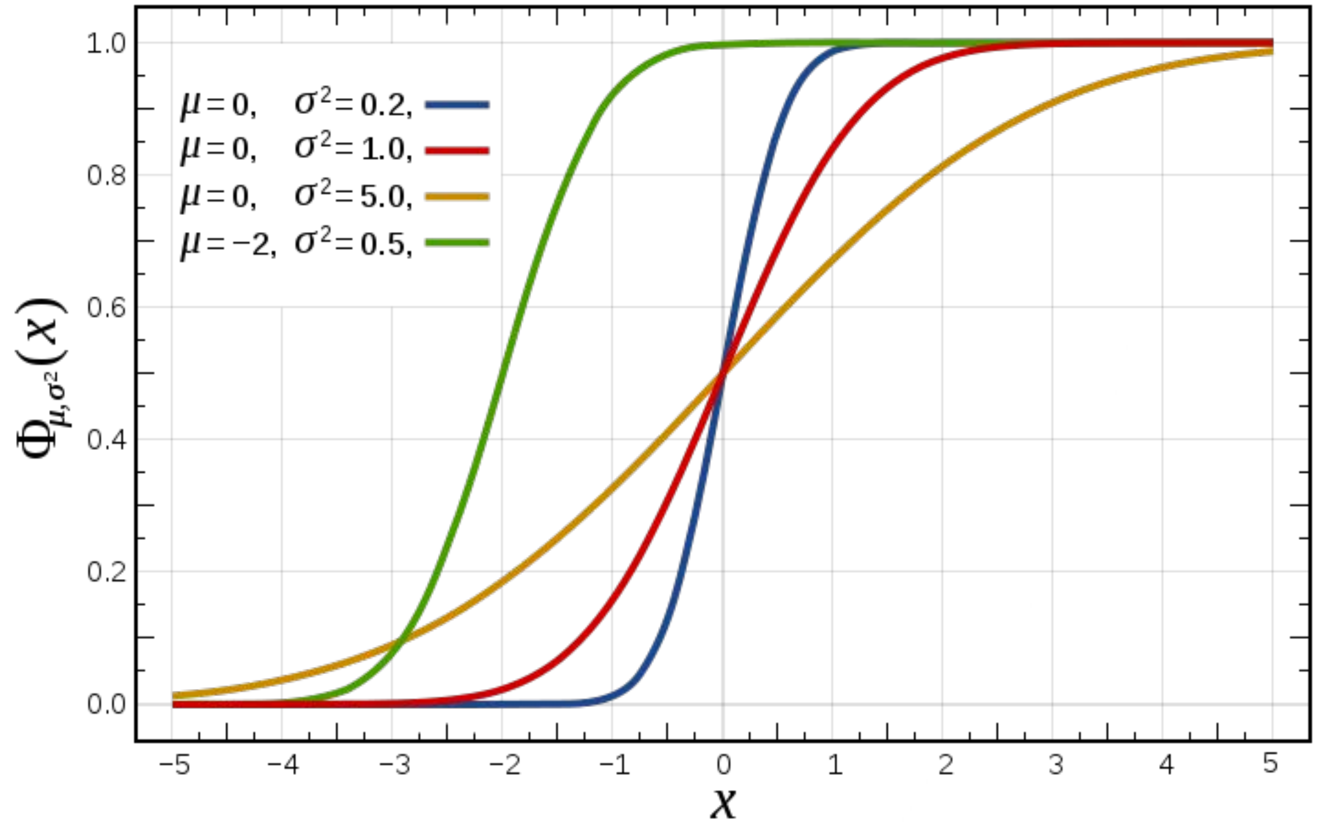
Continuous random variables

- A ***continuous random variable*** is one which takes an infinite number of possible values. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.
- A continuous random variable is not defined at specific values. Instead, it is defined over an *interval* of values, and is represented by the ***area under a curve*** (in advanced mathematics, this is known as an *integral*). The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.

- Suppose a random variable X may take all values over an interval of real numbers. Then the probability that X is in the set of outcomes A, $P(A)$, is defined to be the area above A and under a curve. The curve, which represents a function $p(x)$, must satisfy the following:
- ***1:*** *The curve has no negative values* ($p(x) \geq 0$ *for all* x)
- ***2:*** *The total area under the curve is equal to 1.*
- A curve meeting these requirements is known as a ***density curve***.

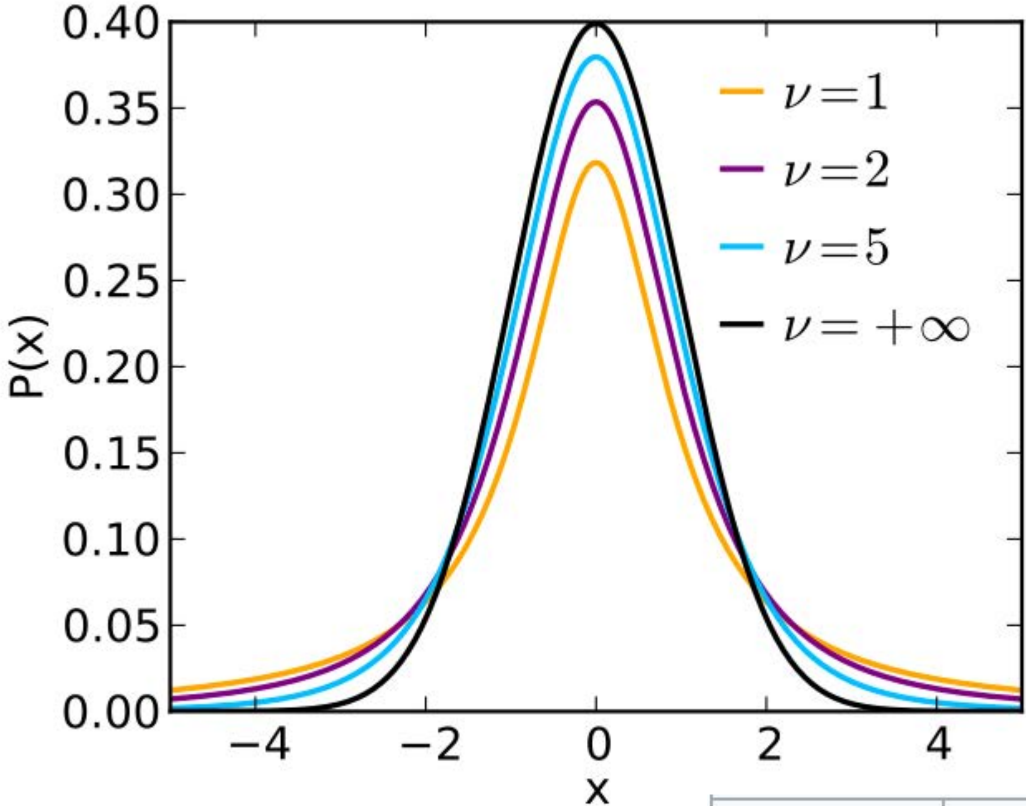
The Gaussian distribution

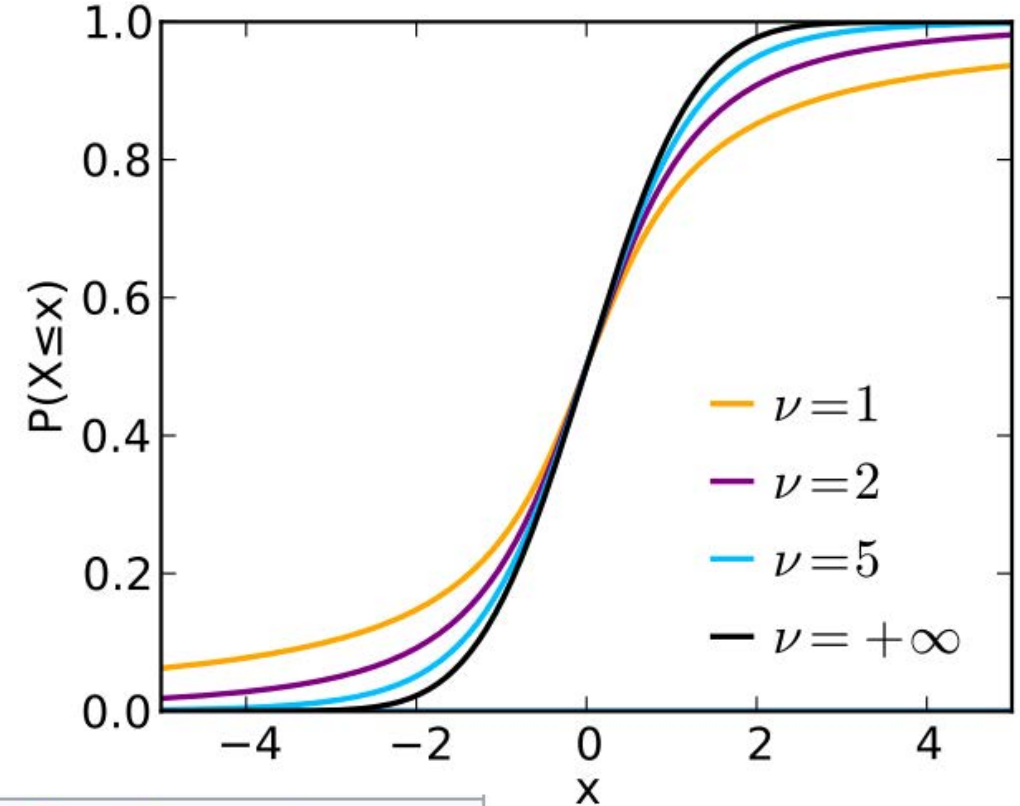




Notation	$\mathcal{N}(\mu,\, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2

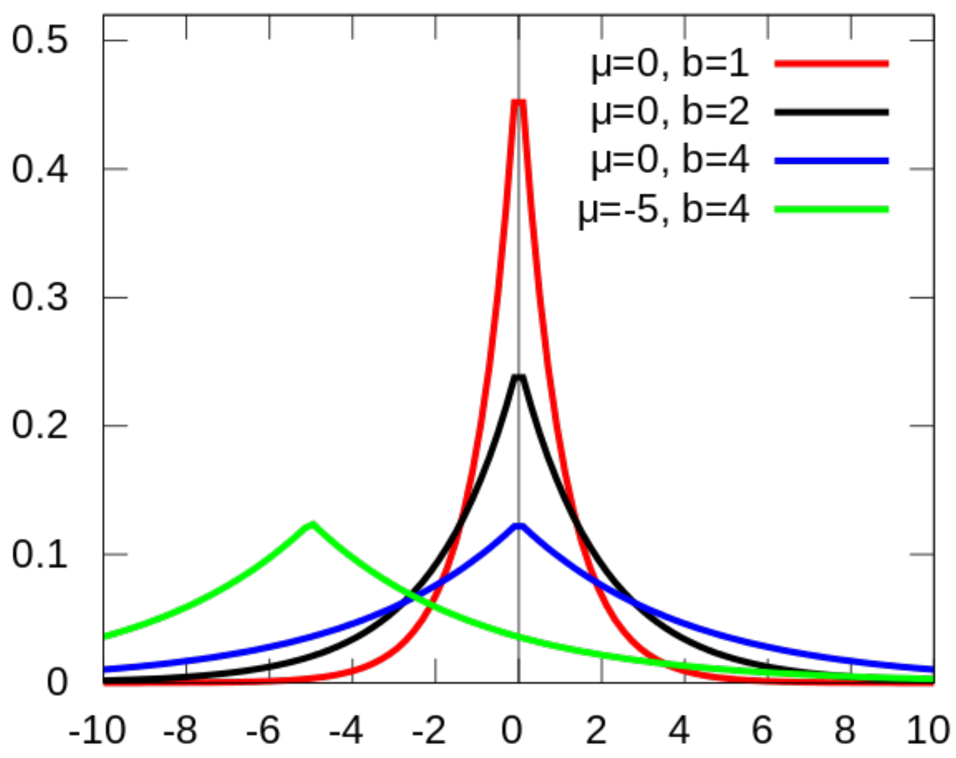
The Student-t distribution

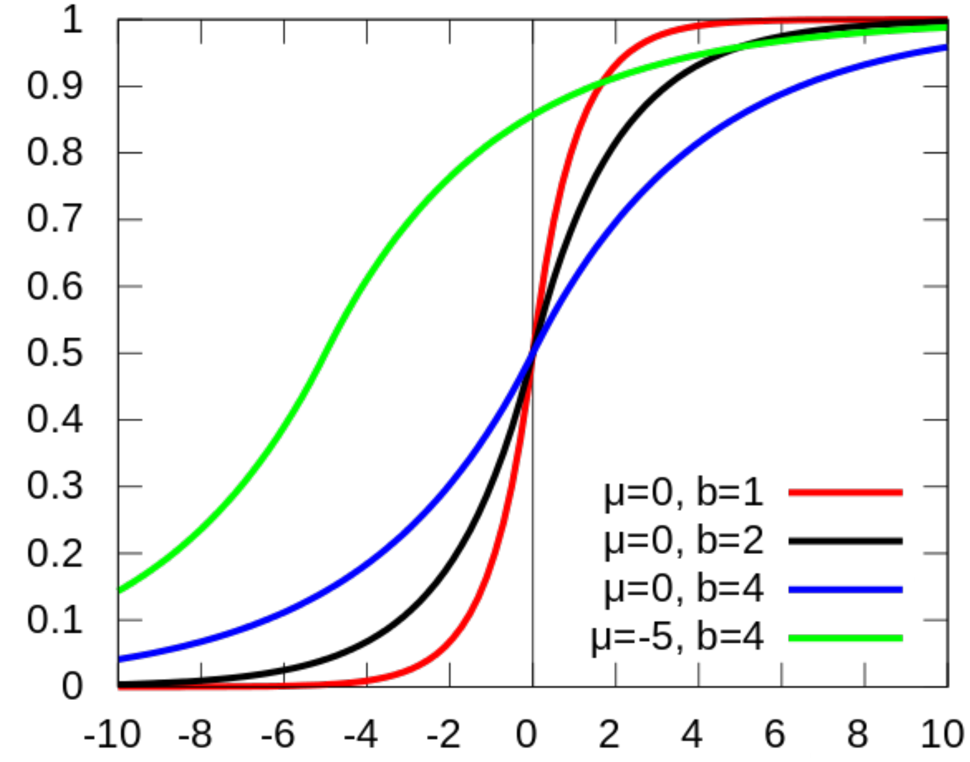




Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
CDF	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \times$ $\frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$ where ${}_2F_1$ is the hypergeometric function
Mean	0 for $\nu > 1$, otherwise undefined
Median	0
Mode	0
Variance	$\frac{\nu}{\nu-2}$ for $\nu > 2$, ∞ for $1 < \nu \le 2$, otherwise undefined

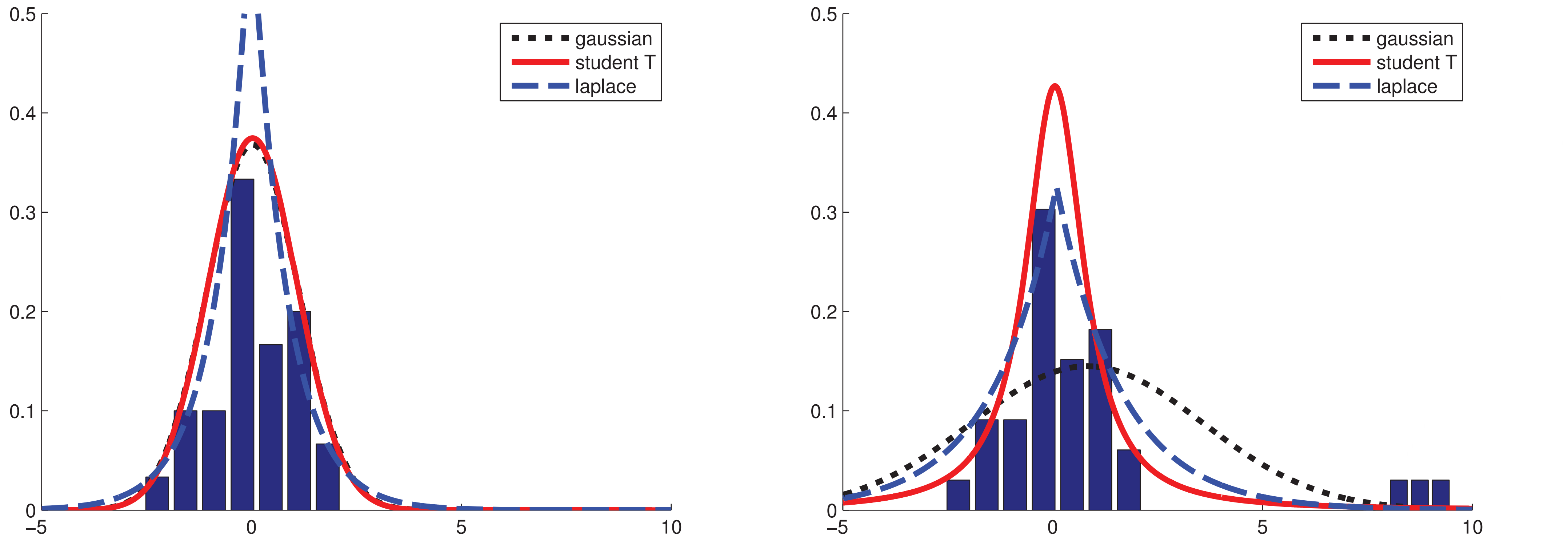
The Laplace distribution





Parameters	μ location (real) $b > 0$ scale (real)
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{1}{2b}\exp\left(-\frac{\|x-\mu\|}{b}\right)$
CDF	$\begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x < \mu \\ 1-\frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$
Mean	μ
Median	μ
Mode	μ
Variance	$2b^2$

Gaussian vs Student-t vs Laplace



(a) (b)

Figure 2.8 Illustration of the effect of outliers on fitting Gaussian, Student and Laplace distributions. (a) No outliers (the Gaussian and Student curves are on top of each other). (b) With outliers. We see that the Gaussian is more affected by outliers than the Student and Laplace distributions. Based on Figure 2.16 of (Bishop 2006a). Figure generated by `robustDemo`.