ENM 531: Data-driven Modeling and Probabilistic Scientific Computing

Lecture #1: Primer on Probability and Statistics

Paris Perdikaris
January 17, 2023

Penn
University *of* Pennsylvania

Probability space

Probability space

A probability space W is a unique triple $W = \{\Omega, \mathcal{F}, P\}$:

- Ω is its sample space
- $\mathcal{F}$ its σ-algebra of events
- P its probability measure

Remarks: (1) The sample space Ω is the set of all possible samples or elementary events ω: $\Omega = \{\omega \mid \omega \in \Omega\}$.
(2)The σ-algebra $\mathcal{F}$ is the set of all of the considered events A, i.e., subsets of Ω: $\mathcal{F} = \{A \mid A \subseteq \Omega, A \in \mathcal{F}\}$.
(3) The probability measure P assigns a probability $P(A)$ to every event $A \in \mathcal{F}$: $P : \mathcal{F} \to [0, 1]$.

Sample space

The sample space Ω is sometimes called the $universe$ of all samples or possible outcomes ω.

Example 1. Sample space

- *Toss of a coin (with head and tail):* $\Omega = \{H, T\}$.
- *Two tosses of a coin:* $\Omega = \{HH, HT, TH, TT\}$.
- *A cubic die:* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.
- *The positive integers:* $\Omega = \{1, 2, 3, \dots\}$.
- *The reals:* $\Omega = \{\omega \mid \omega \in \mathbb{R}\}$.

Note that the ωs are a mathematical construct and have per se no real or scientific meaning. The ωs in the die example refer to the numbers of dots observed when the die is thrown.

Events

An event A is a subset of Ω. If the outcome ω of the experiment is in the subset A, then the event A is said to have occurred. The set of all subsets of the sample space are denoted by 2^{Ω}.

Example 2. Events

- *Head in the coin toss:* $A = \{H\}$.
- *Odd number in the roll of a die:* $A = \{\omega_1, \omega_3, \omega_5\}$.
- *An integer smaller than 5:* $A = \{1, 2, 3, 4\}$, *where* $\Omega = \{1, 2, 3, \dots\}$.
- *A real number between 0 and 1:* $A = [0, 1]$, *where* $\Omega = \{\omega \mid \omega \in \mathbb{R}\}$.

We denote the complementary event of A by $A^c = \Omega \backslash A$. When it is possible to determine whether an event A has occurred or not, we must also be able to determine whether A^c has occurred or not.

σ-algebras

Not every subset of $[0, 1]$ has a determinable length $\Rightarrow$ collect the ones with a determinable length in $\mathcal{F}$. Such a mathematical construct, which has additional, desirable properties, is called σ-algebra.

Definition 2. σ-algebra

A collection $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if

- $\Omega \in \mathcal{F}$ *and* $\emptyset \in \mathcal{F}$ *($\emptyset$ denotes the empty set)*
- *If* $A \in \mathcal{F}$ *then* $\Omega \backslash A = A^c \in \mathcal{F}$*: The complementary subset of A is also in* Ω
- *For all* $A_i \in \mathcal{F}$*:* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The intuition behind it: collect all events in the σ-algebra $\mathcal{F}$, make sure that by performing countably many elementary set operation $(\cup, \cap, ^c)$ on elements of $\mathcal{F}$ yields again an element in $\mathcal{F}$ (closeness).

The pair $\{\Omega, \mathcal{F}\}$ is called $measure\ space$.

Probability measures

Definition 7. Probability measure
A $\mathit{probability\ measure}$ P on the sample space Ω with σ-algebra $\mathcal{F}$ is a set function

$$P : \mathcal{F} \to [0,1],$$

satisfying the following conditions

- $P(\Omega) = 1$.
- *If* $A \in \mathcal{F}$ *then* $P(A) \geq 0$.
- *If* $A_1, A_2, A_3, \ldots \in \mathcal{F}$ *are mutually disjoint, then*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called a $\mathit{probability\ space}$.

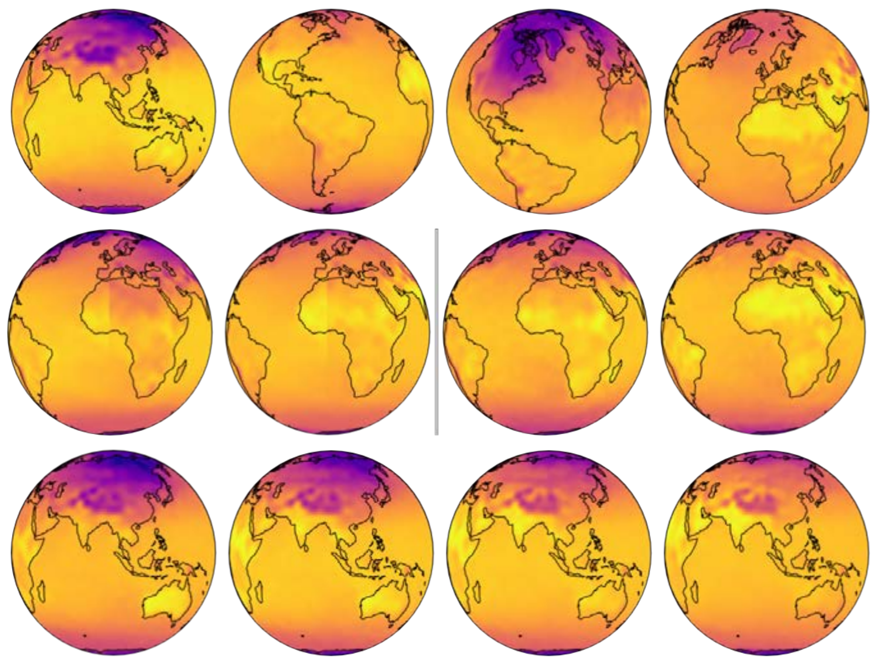
Random variables

Definition 11. Random variable
A real-valued random variable X *is a* $\mathcal{F}$*-measurable function defined on a probability space* $(\Omega, \mathcal{F}, P)$ *mapping its sample space* Ω *into the real line* $\mathbb{R}$*:*

$$X : \Omega \to \mathbb{R}.$$

vectors

matrices



functions

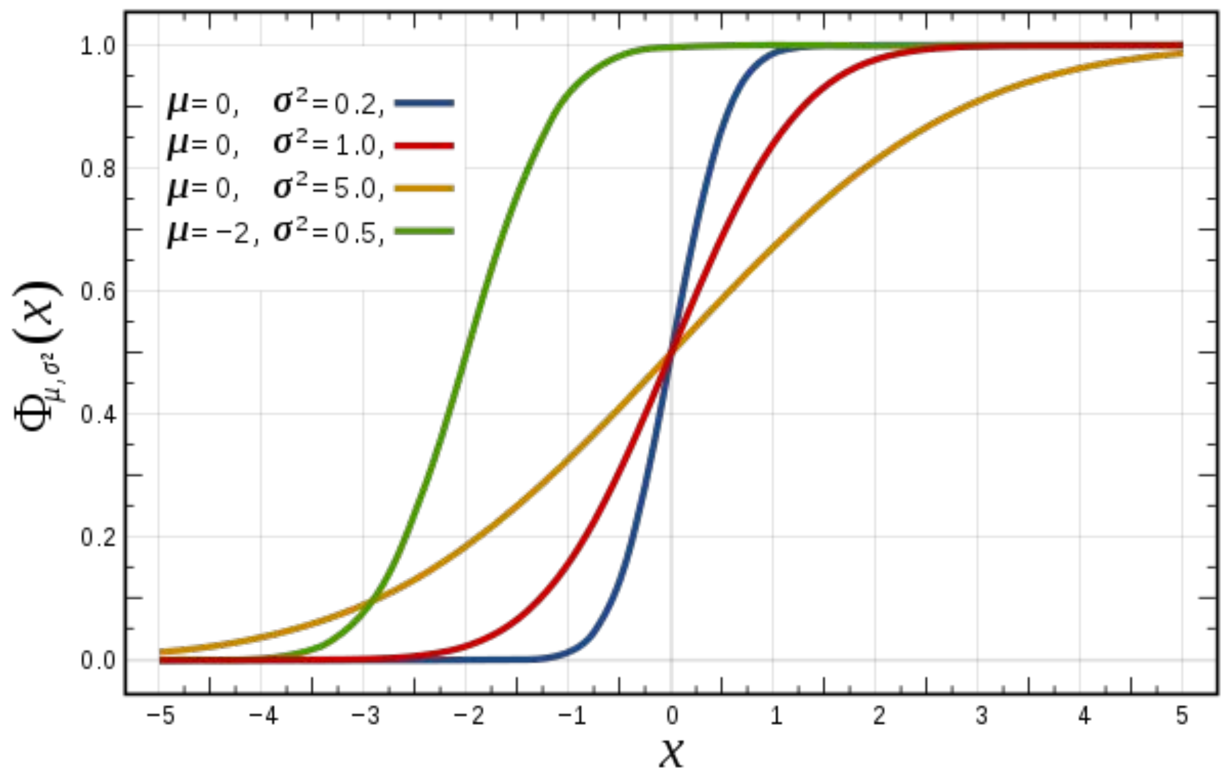
Distribution function

Definition 12. Distribution function
The distribution function of a random variable X, defined on a probability space $(\Omega, \mathcal{F}, P)$*, is defined by:*

$$F(x) = P(X(\omega) \leq x) = P(\{\omega : X(\omega) \leq x\}).$$

From this the probability measure of the half-open sets in $\mathbb{R}$ is

$$P(a < X \leq b) = P(\{\omega : a < X(\omega) \leq b\}) = F(b) - F(a).$$



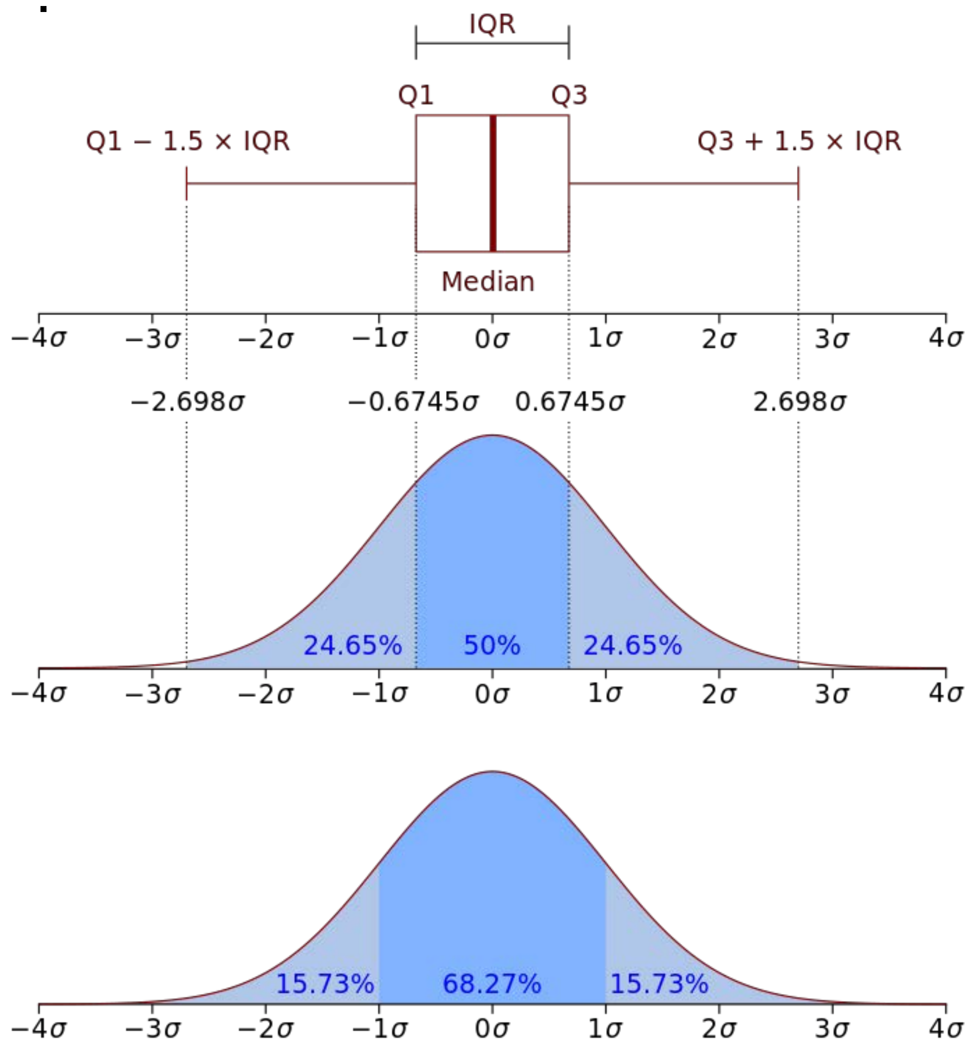
cdf of univariate normal random variables

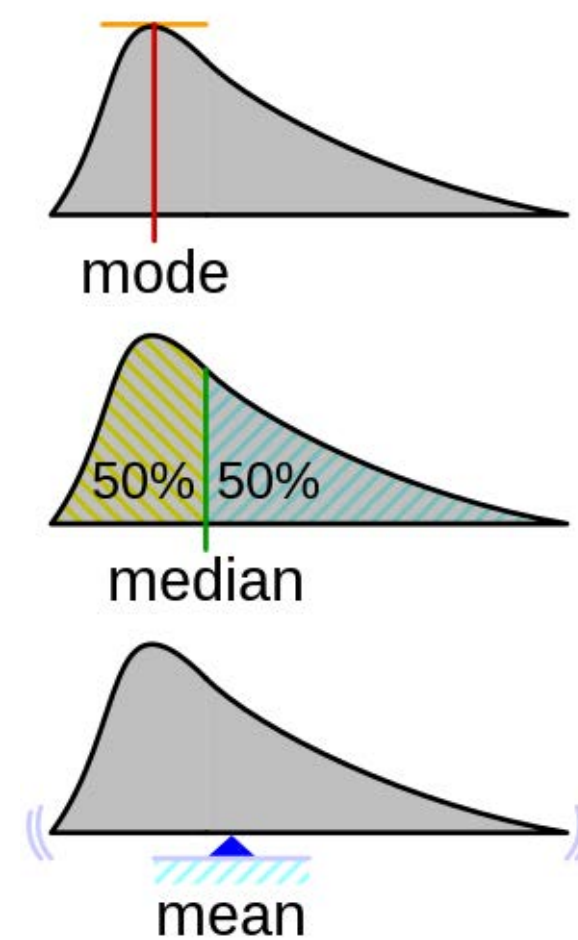
Density function

Closely related to the distribution function is the density function. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a nonnegative function, satisfying $\int_{\mathbb{R}} f d\lambda = 1$. The function f is called a density function (with respect to the Lebesgue measure) and the associated probability measure for a random variable X, defined on $(\Omega, \mathcal{F}, P)$, is

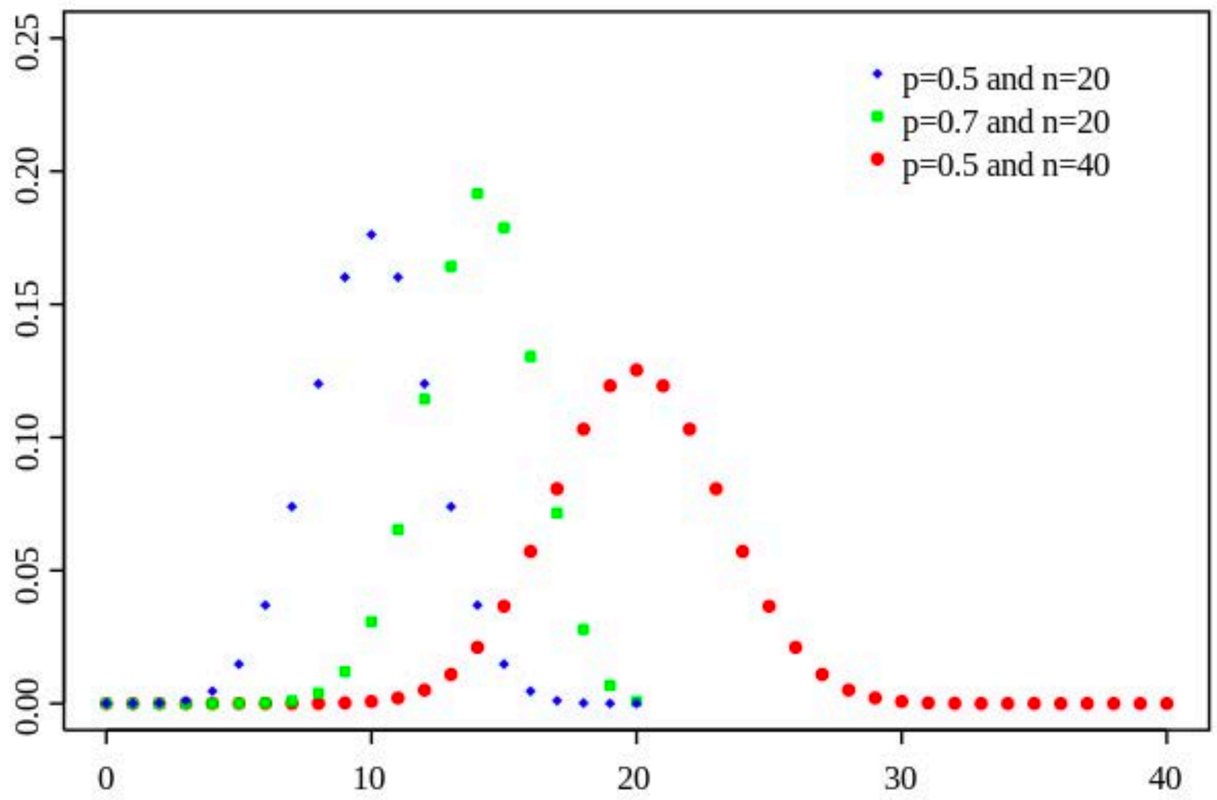
$$P(\{\omega : \omega \in A\}) = \int_A f d\lambda.$$

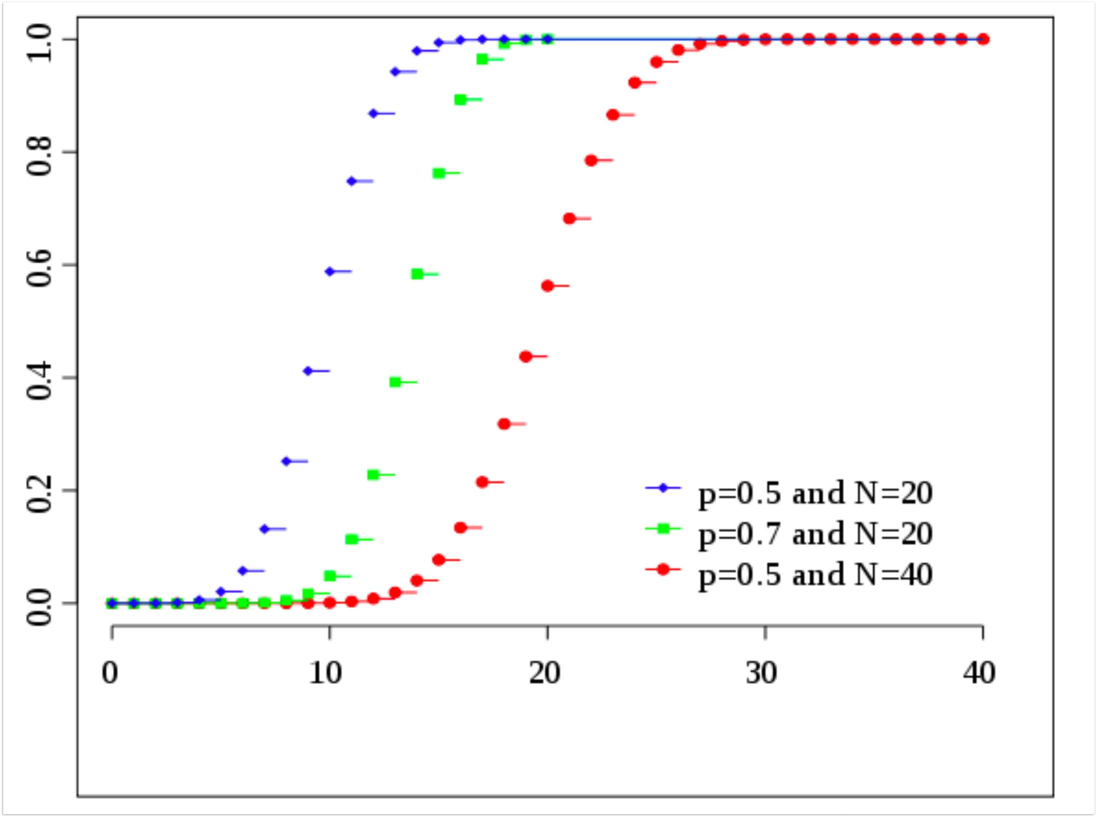
for all $A \in \mathcal{F}$.





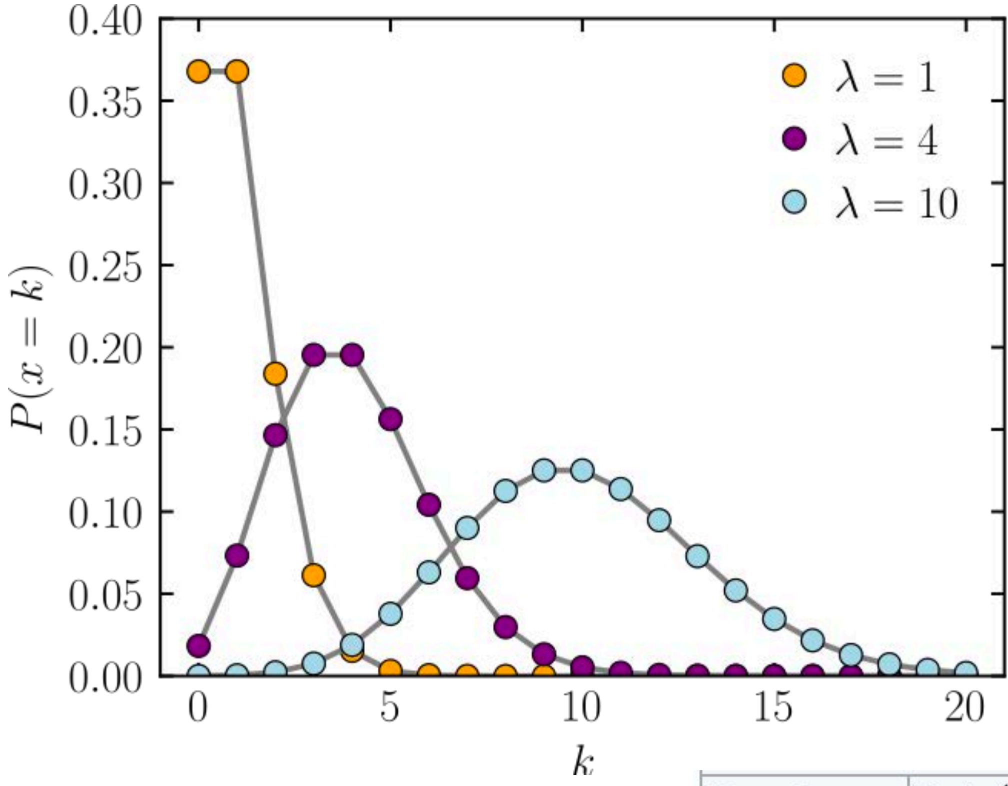
The binomial distribution

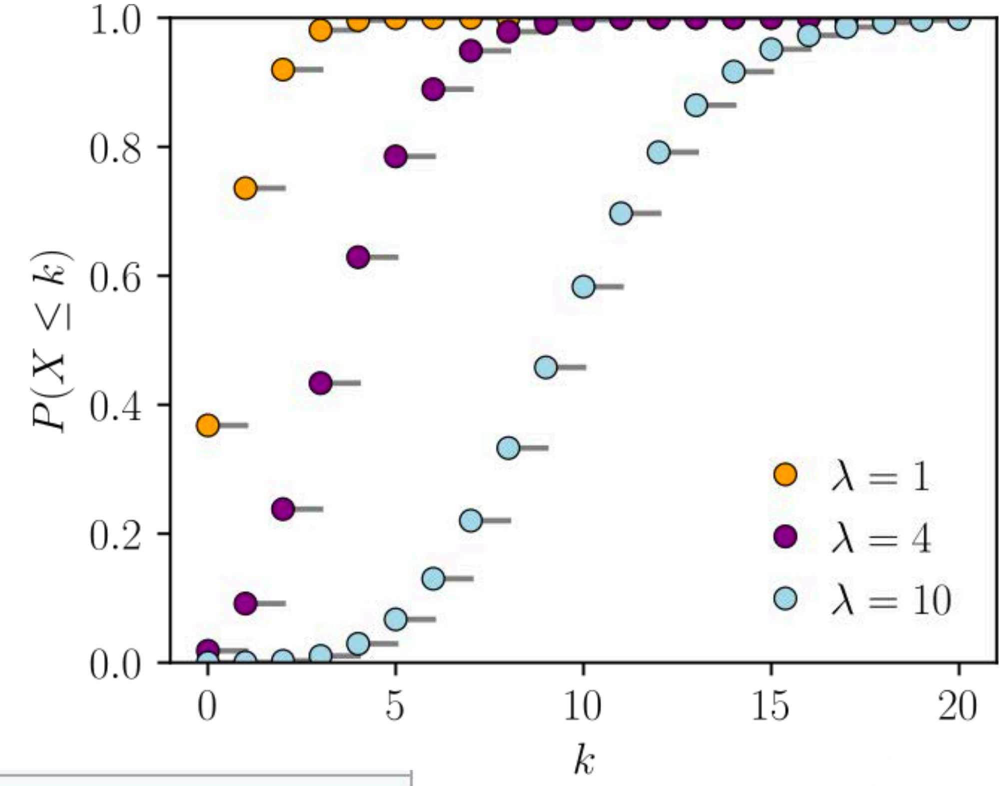




Notation	$B(n, p)$
Parameters	$n \in \{0, 1, 2, \ldots\}$ – number of trials $p \in [0, 1]$ – success probability for each trial $q = 1 - p$
Support	$k \in \{0, 1, \ldots, n\}$ – number of successes
PMF	$\binom{n}{k} p^k q^{n-k}$
CDF	$I_q(n - k, 1 + k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n + 1)p \rfloor$ or $\lceil (n + 1)p \rceil - 1$
Variance	npq

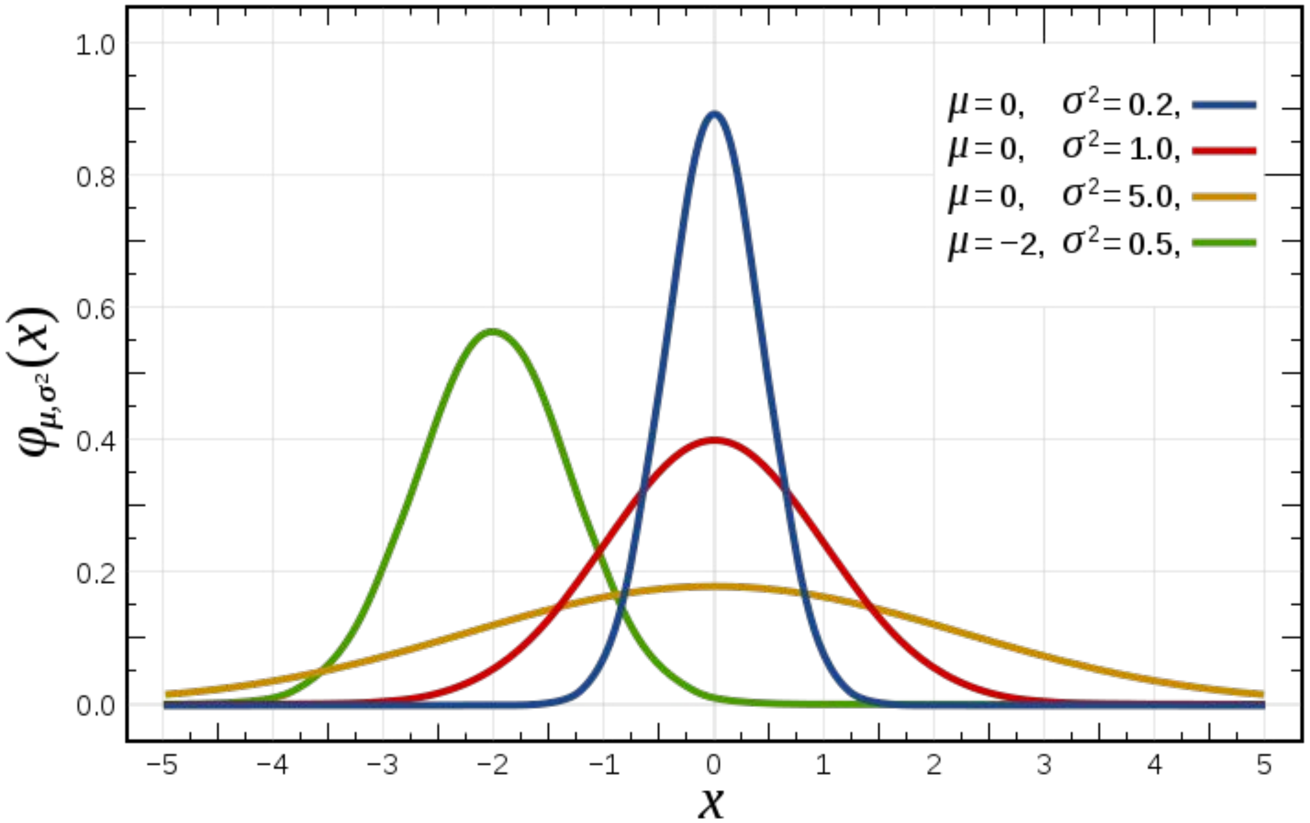
The Poisson distribution

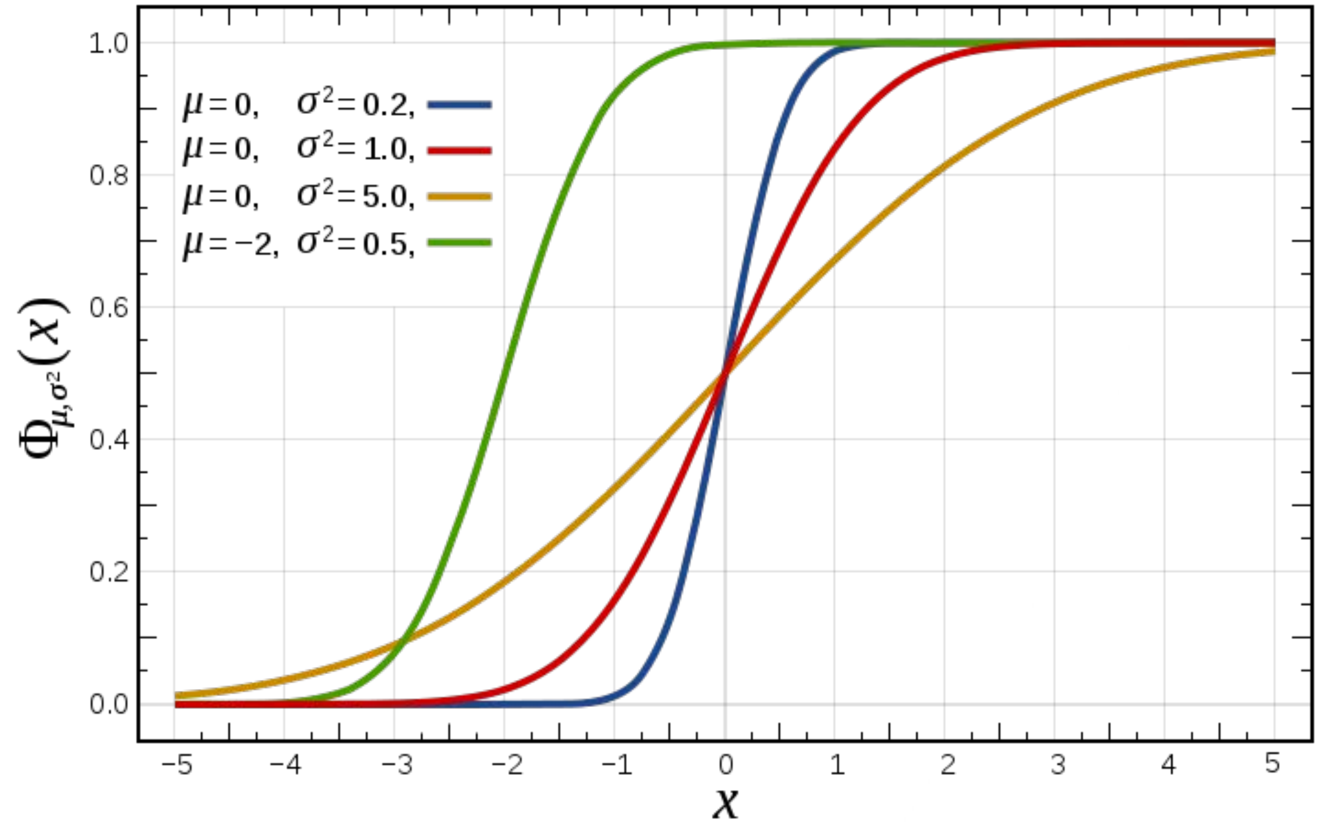




Notation	$\mathrm{Pois}(\lambda)$
Parameters	$\lambda \in (0, \infty)$ (rate)
Support	$k \in \mathbb{N}_0$ (Natural numbers starting from 0)
PMF	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$, or $e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$, or $Q(\lfloor k+1 \rfloor, \lambda)$ (for $k \geq 0$, where $\Gamma(x, y)$ is the upper incomplete gamma function, $\lfloor k \rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Mode	$\lceil \lambda \rceil - 1, \lfloor \lambda \rfloor$
Variance	λ

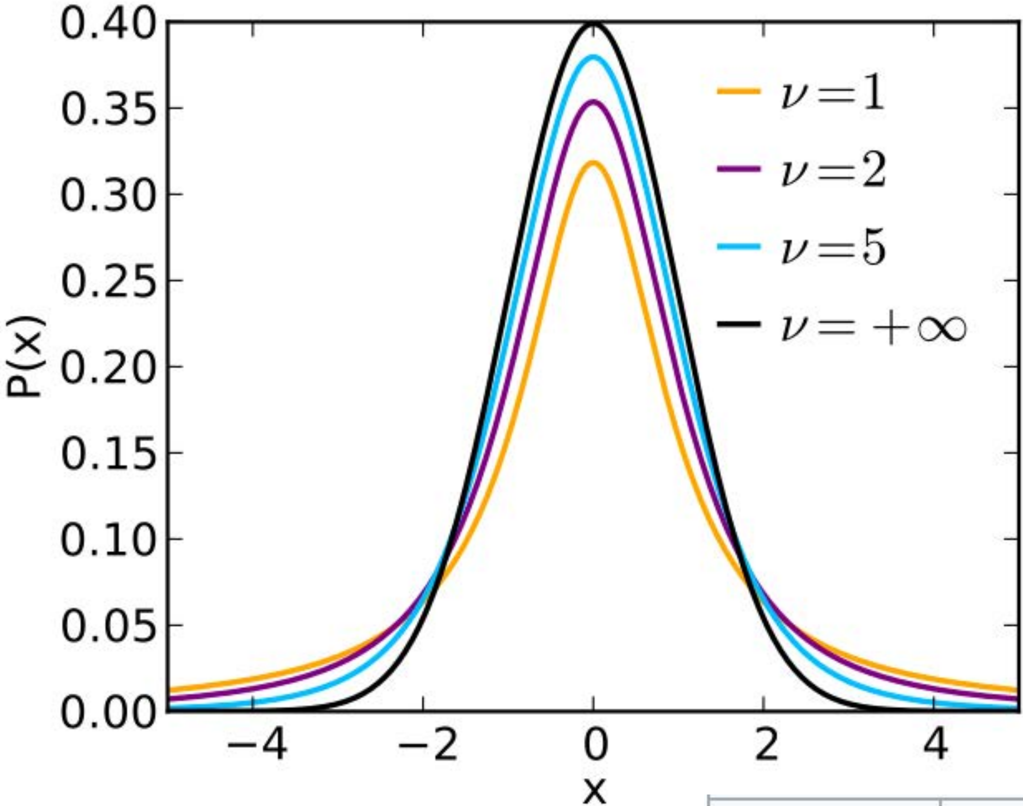
The Gaussian distribution

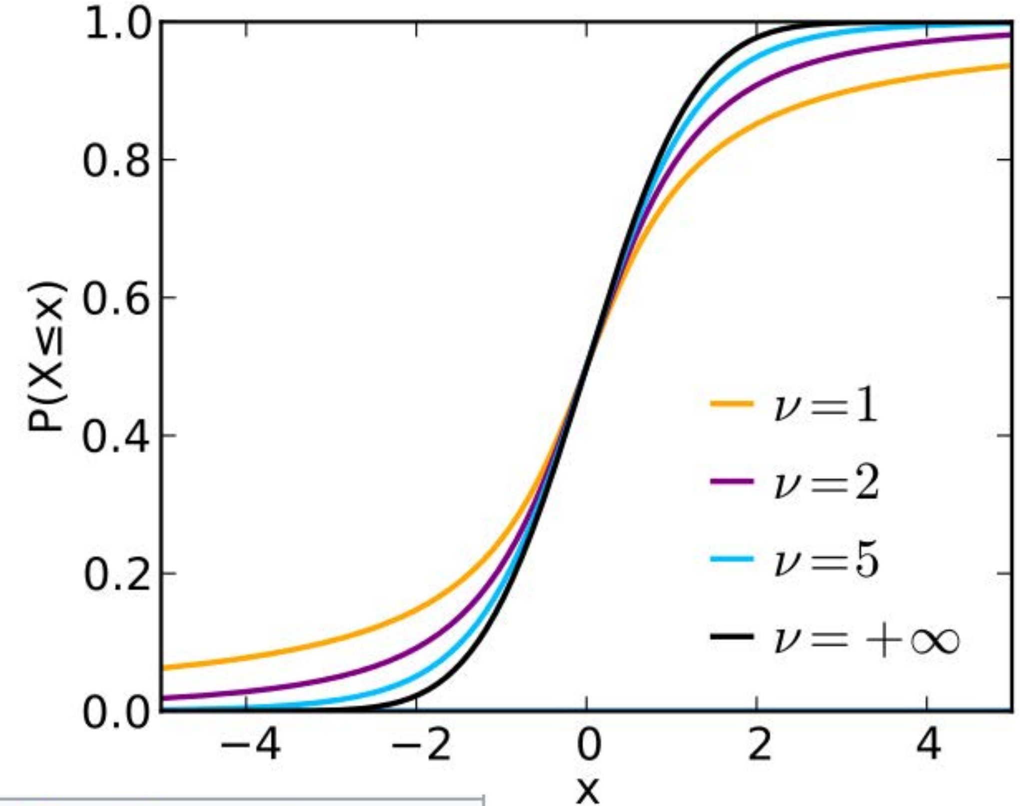




Notation	$\mathcal{N}(\mu,\, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$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

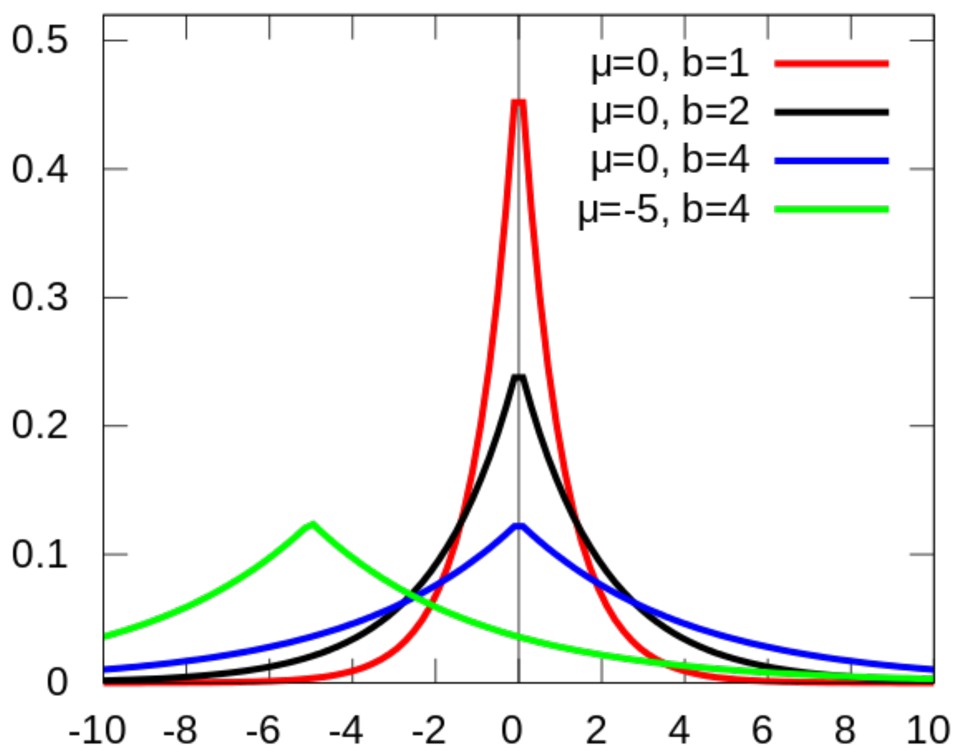
The Student-t distribution

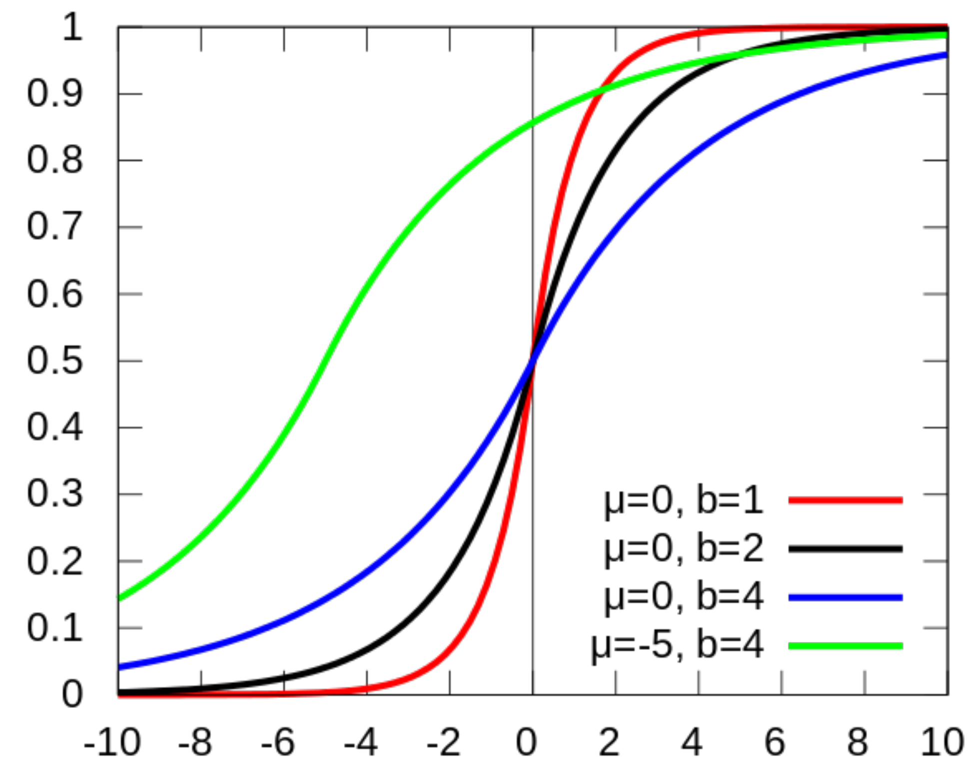




Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
CDF	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \times \frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$ 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	$\frac{1}{2\,b} \exp\left(-\frac{\lvert x-\mu \rvert}{b}\right)$
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

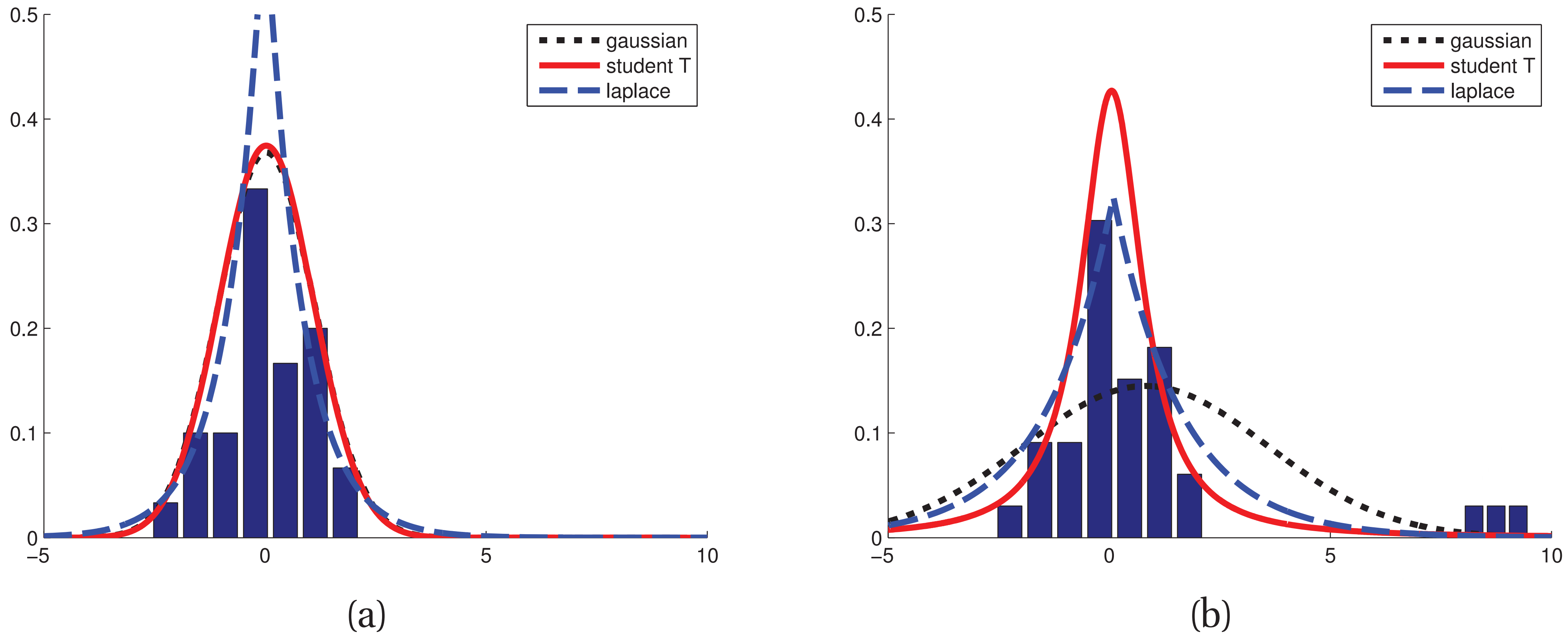


Figure 2.8 Illustration of the effect of outliers on fitting Gaussian, Student and Laplace distributions. (a) No outliers (the Gaussian and Student curves are on top of each other). (b) With outliers. We see that the Gaussian is more affected by outliers than the Student and Laplace distributions. Based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