Probability and Random Processes

ECS 315

Asst. Prof. Dr. Prapun Suksompong

prapun@siit.tu.ac.th

10 Continuous Random Variables

Office Hours:

BKD, 6th floor of Sirindhralai building

Wednesday	14:00-15:30
Friday	14:00-15:30

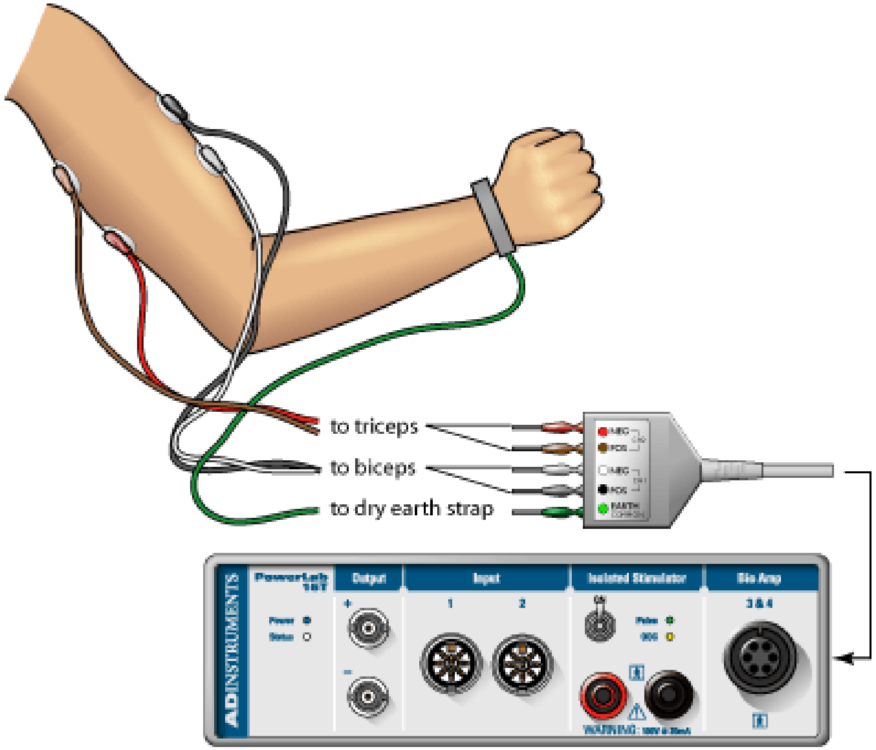
Ex. `rand` function

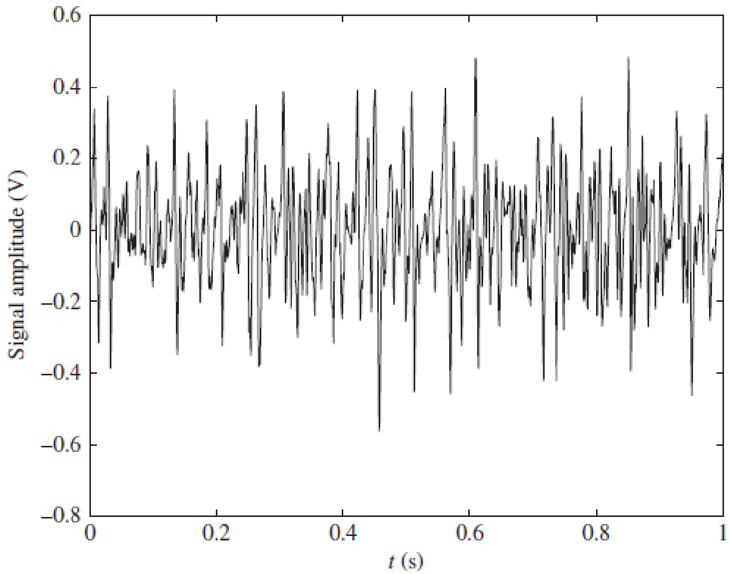
- Generate an array of uniformly distributed pseudorandom numbers.
  - The pseudorandom values are drawn from the **standard uniform distribution** on the open **interval (0,1)**.
- `rand` returns a scalar.
- `rand(m,n)` or `rand([m,n])` returns an *m*-by-*n* matrix.
  - `rand(n)` returns an *n*-by-*n* matrix

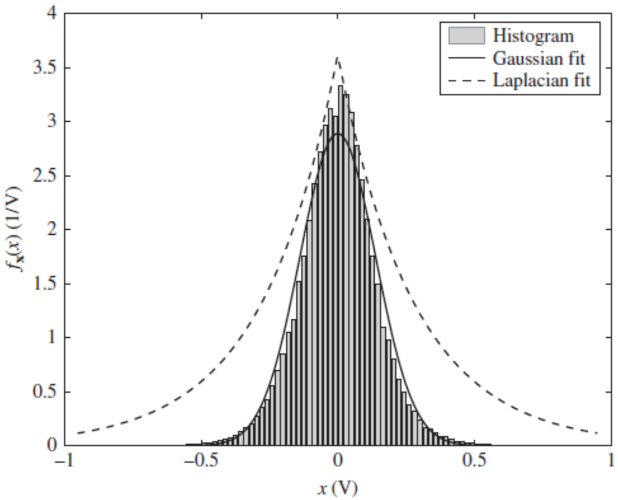
```
>> rand

ans =

    0.3816

>> rand(10,2)

ans =

    0.7655    0.6551
    0.7952    0.1626
    0.1869    0.1190
    0.4898    0.4984
    0.4456    0.9597
    0.6463    0.3404
    0.7094    0.5853
    0.7547    0.2238
    0.2760    0.7513
    0.6797    0.2551
```

Ex. Muscle Activity

- Look at electrical activity of skeletal muscle by recording a human electromyogram (EMG).

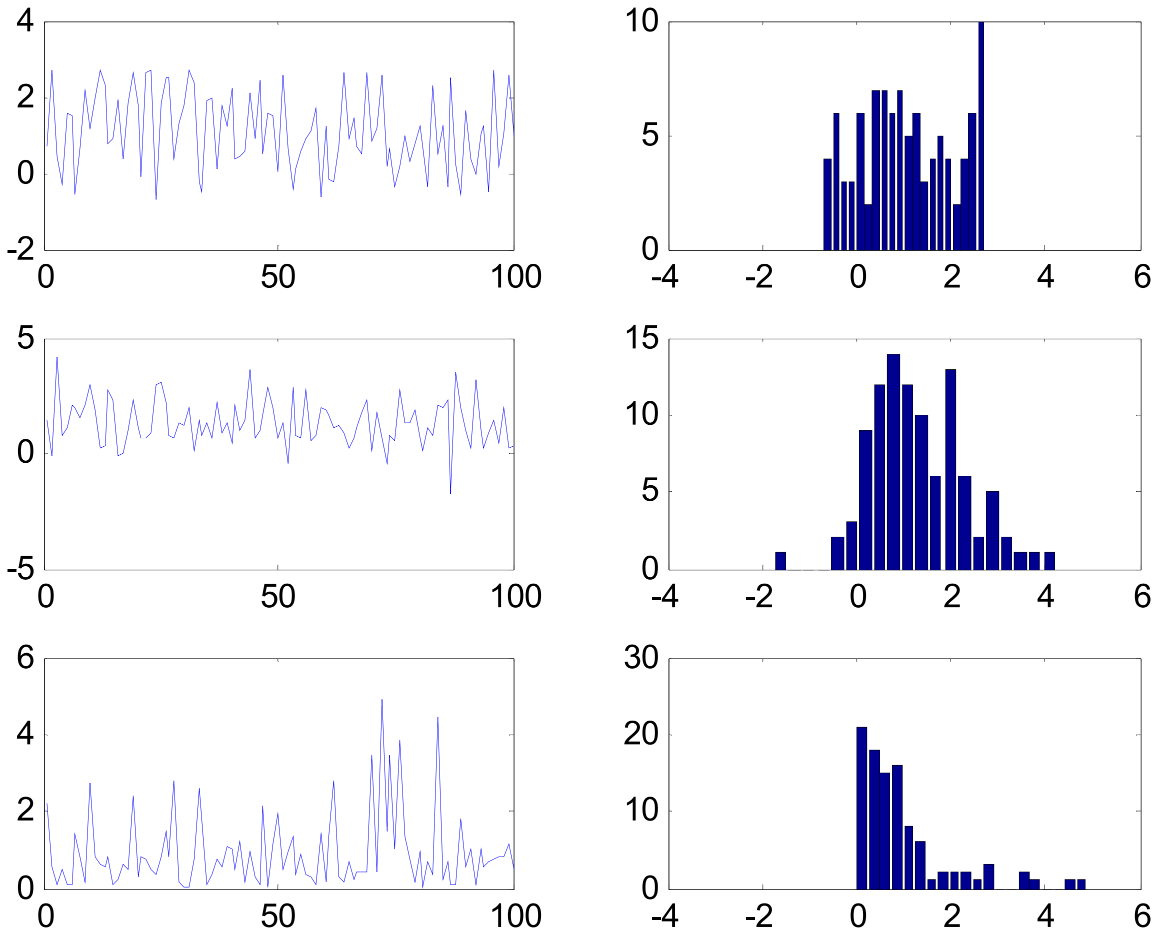






[http://www.adinstruments.com/solutions/education/ltexp/electromyography-emg-german]

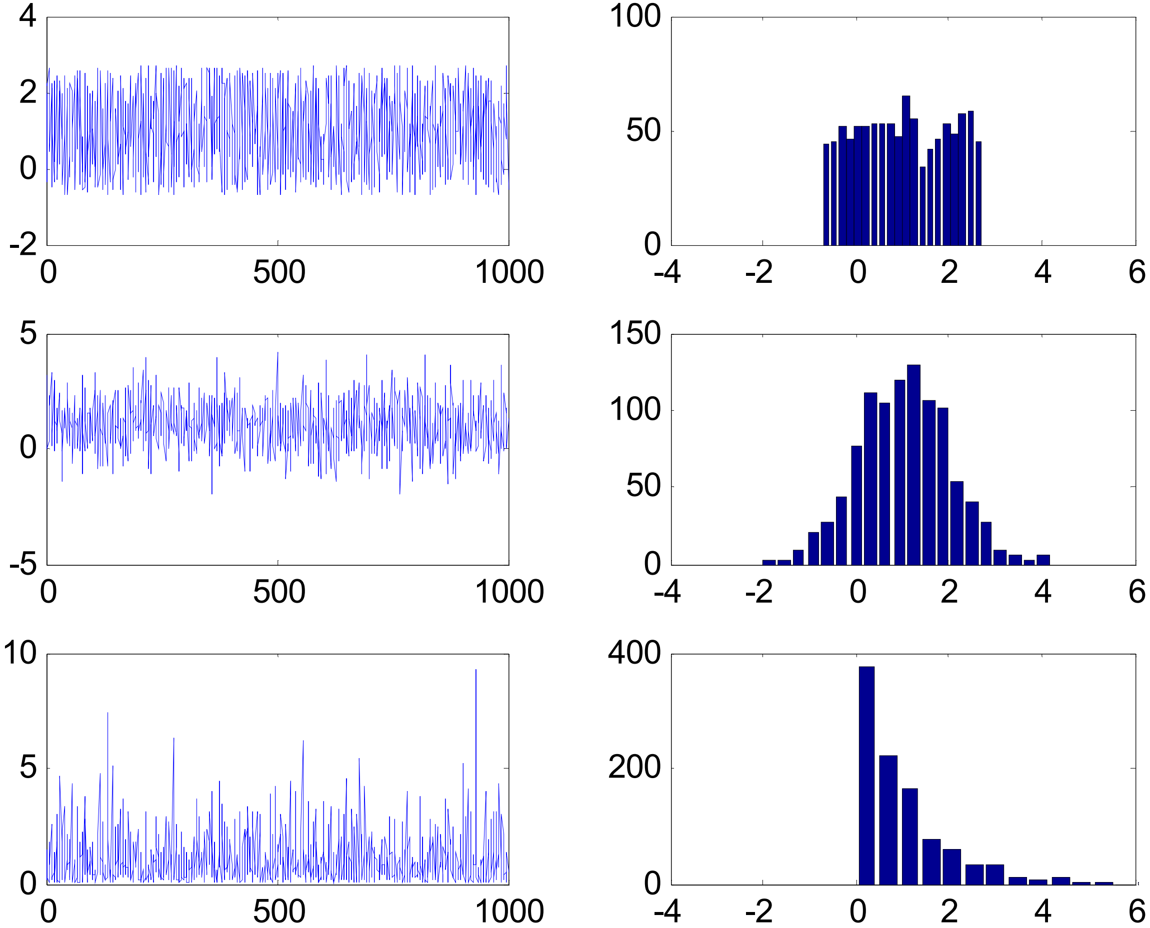
Three Important Continuous RVs



Mean = 1
Std = 1
N = 100

[IntroThreeContinuousRV.m]

Three Important Continuous RVs

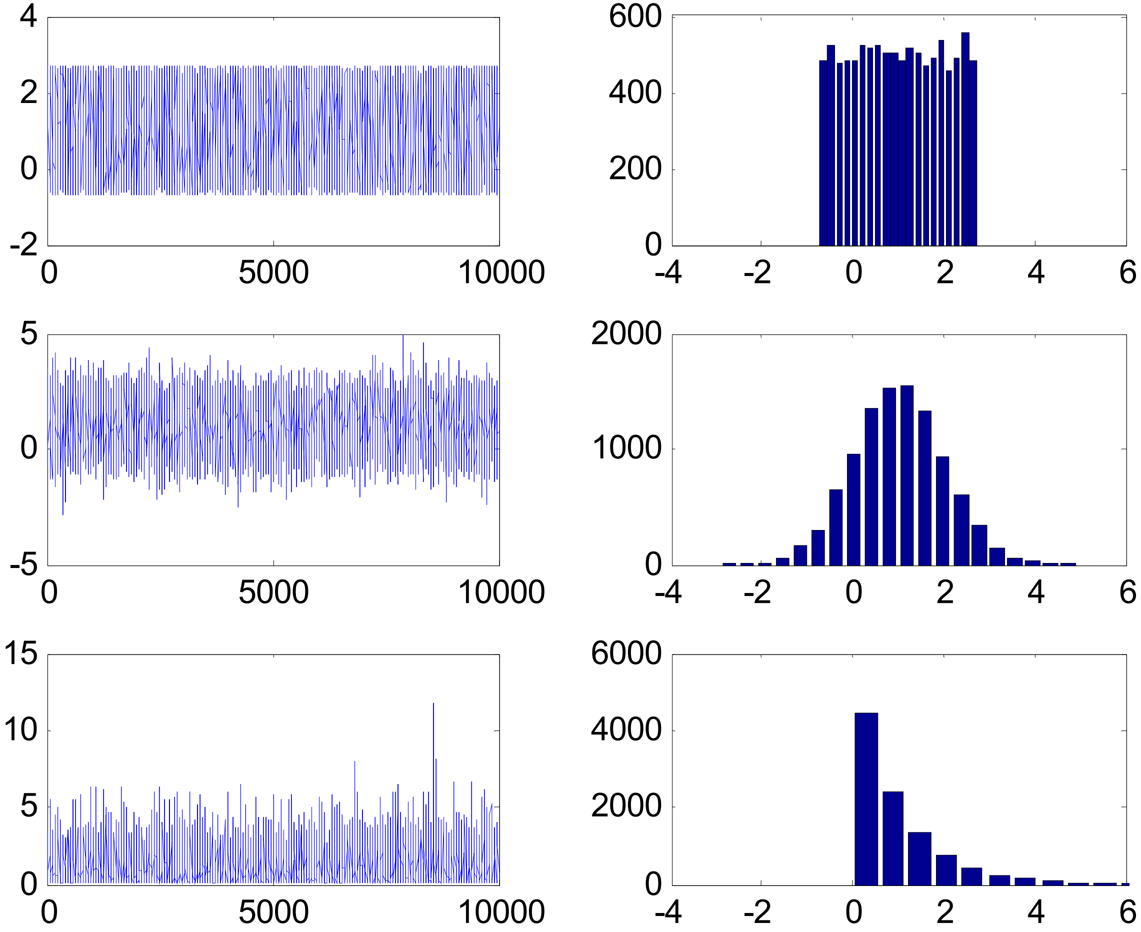


Mean = 1

Std = 1

N = 1,000

Three Important Continuous RVs



Mean = 1

Std = 1

N = 10,000

Review: P[some condition(s) on X]

For discrete random variable,

8.13. Steps to find probability of the form P [some condition(s) on X] when the pmf $p_X(x)$ is known.

(a) Find the support of X.

(b) Consider only the x inside the support. Find all values of x that satisfy the condition(s).

(c) Evaluate the pmf at x found in the previous step.

(d) Add the pmf values from the previous step.

$$P[\text{some condition(s) on } X] = \sum p_X(x)$$

Discrete RV

Sum over all the x values that satisfy the condition(s)

P[some condition(s) on X]

- **For discrete random variable,**

$$P[\text{some condition(s) on } \underset{\text{Discrete RV}}{X}] = \sum \overbrace{p_X(x)}^{\text{probability }\underline{\text{mass}}\text{ function (pmf)}}$$

Sum over all the x values that satisfy the condition(s)

- **For continuous random variable,**

pmf → pdf

$\sum \rightarrow \int$

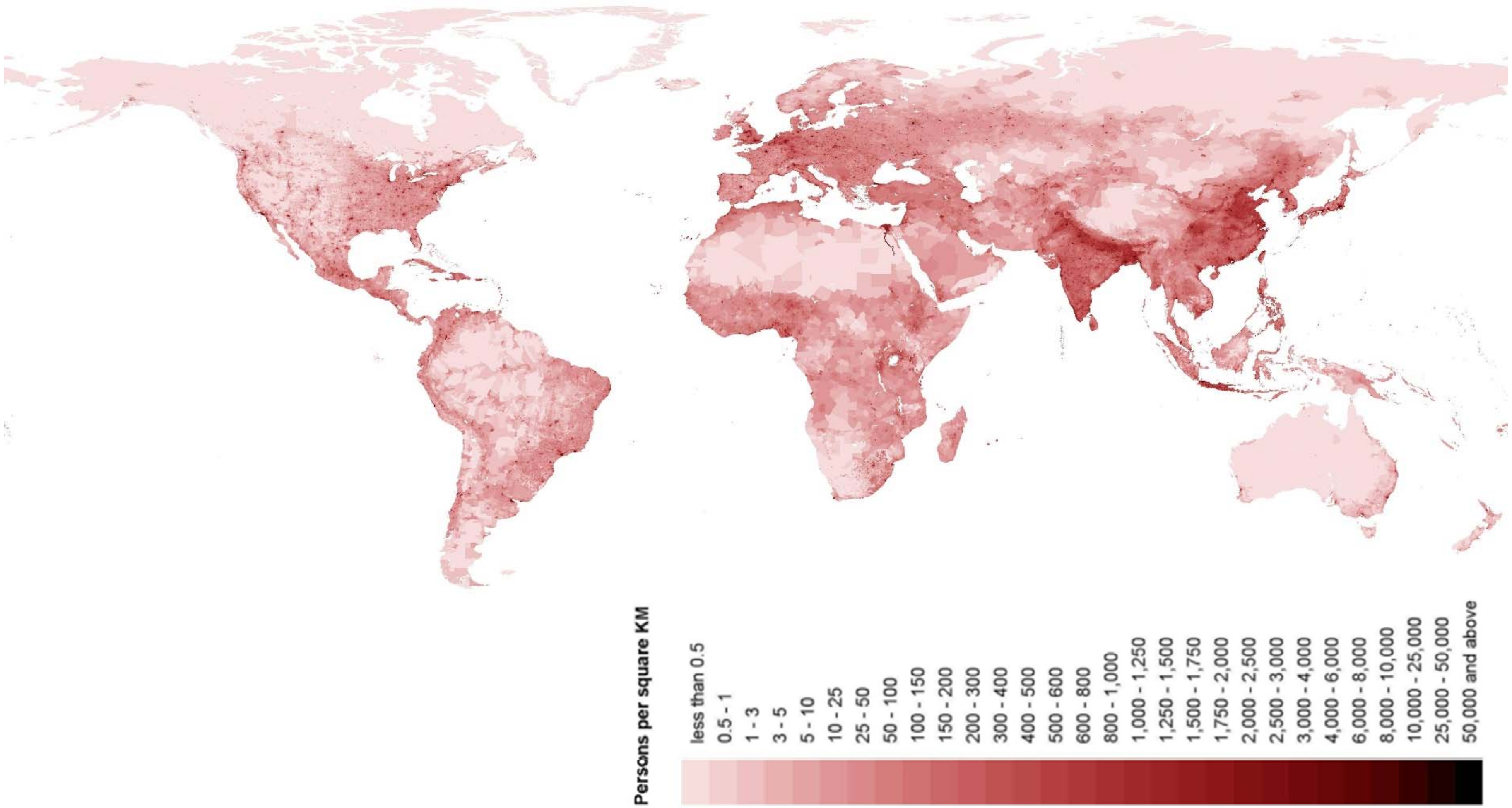
$$P[\text{some condition(s) on } \underset{\text{Continuous RV}}{X}] = \int \overbrace{f_X(x)}^{\text{probability }\underline{\text{density}}\text{ function (pdf)}} dx$$

Integrate over all the x values that satisfy the condition(s)

Support of a RV

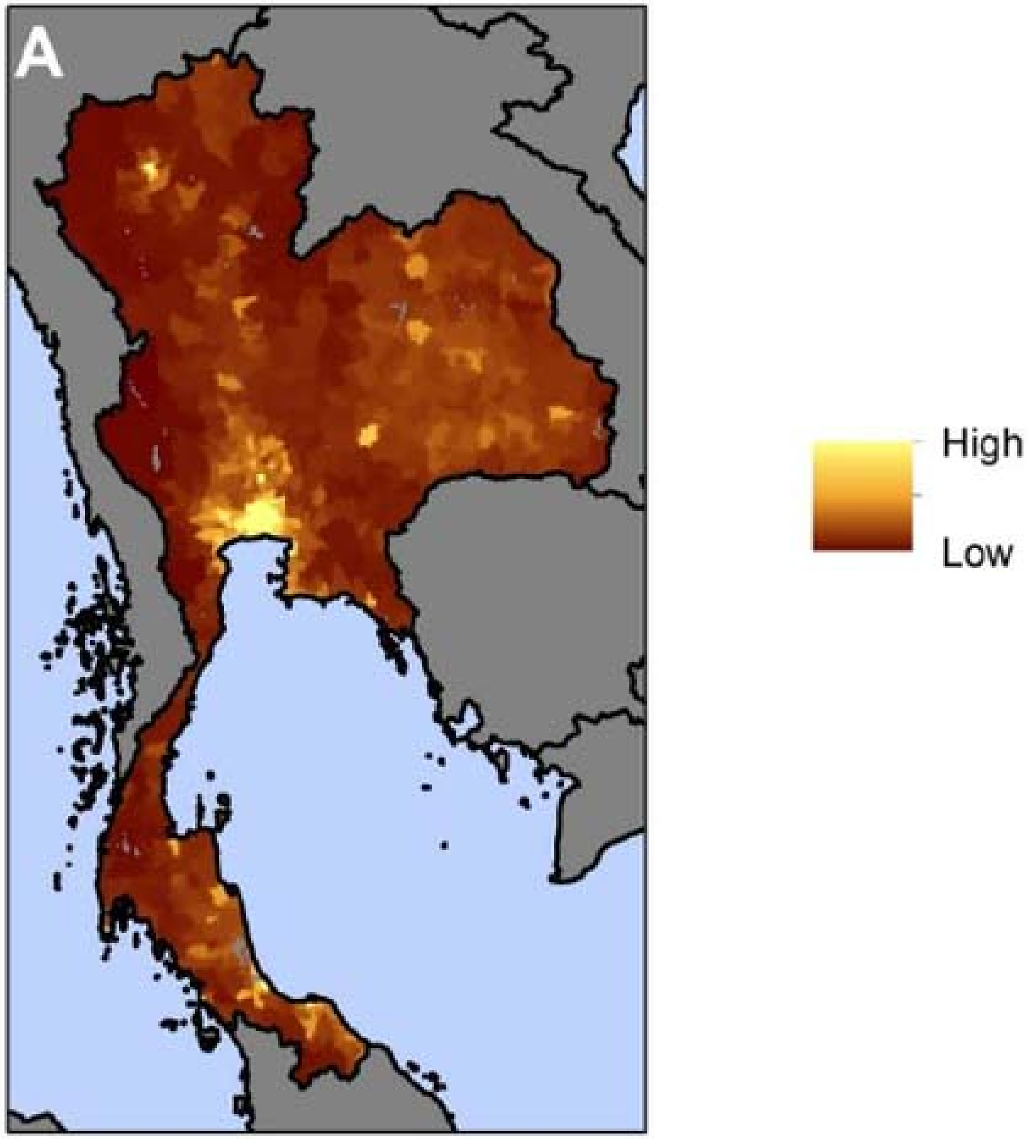
- In general, the **support** of a RV X is any set S such that
$$P[X \in S] = 1.$$
- In this class, we try to find the smallest (minimal) set that works as a support.
- **For discrete random variable,**
$$S_X = \{x: p_X(x) > 0\}$$
- **For continuous random variable,**
$$S_X = \{x: f_X(x) > 0\}$$

World Map of Population Density



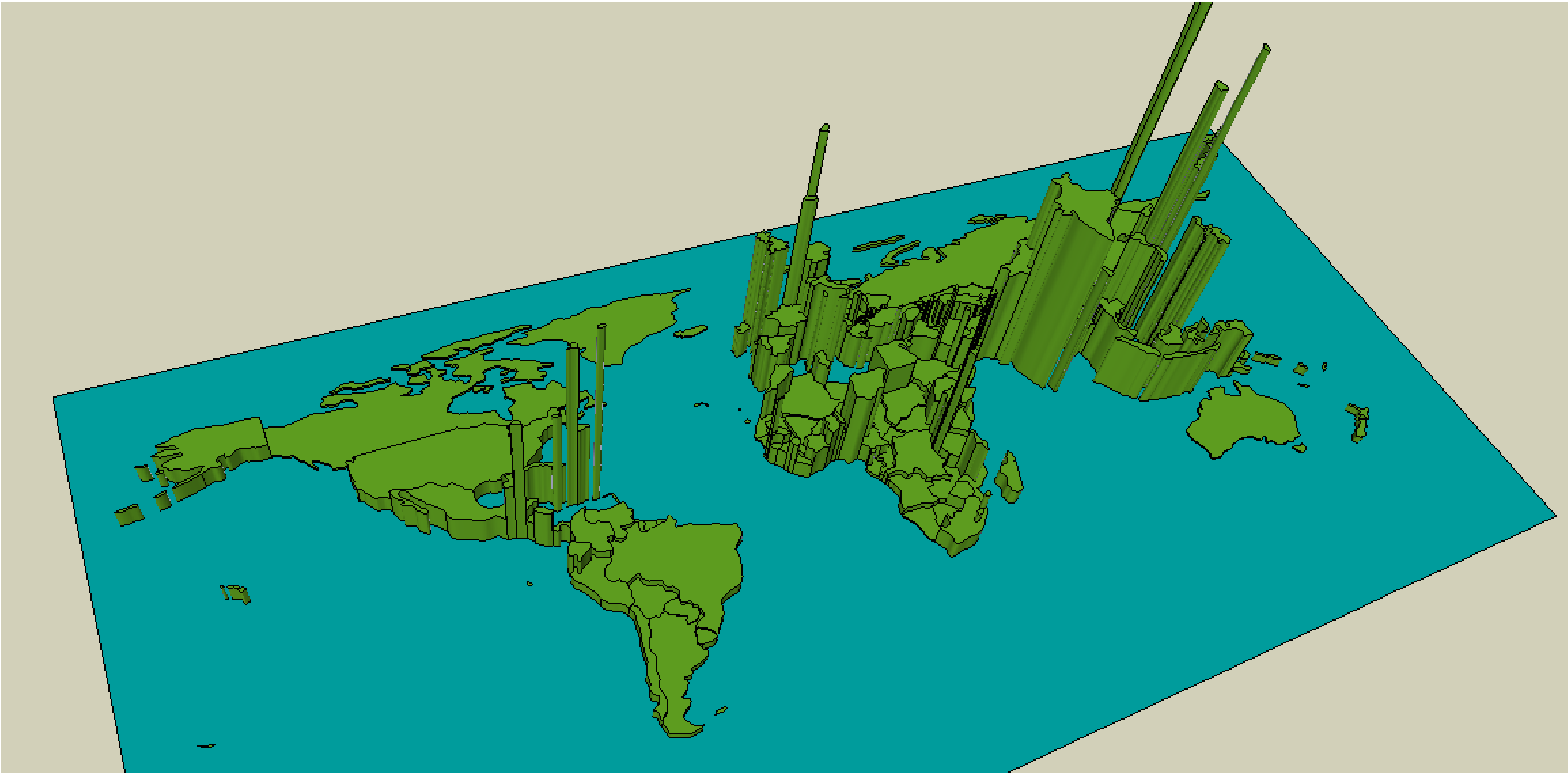
[http://i.imgur.com/gBYMfWO.jpg]

Thailand’s Population Density

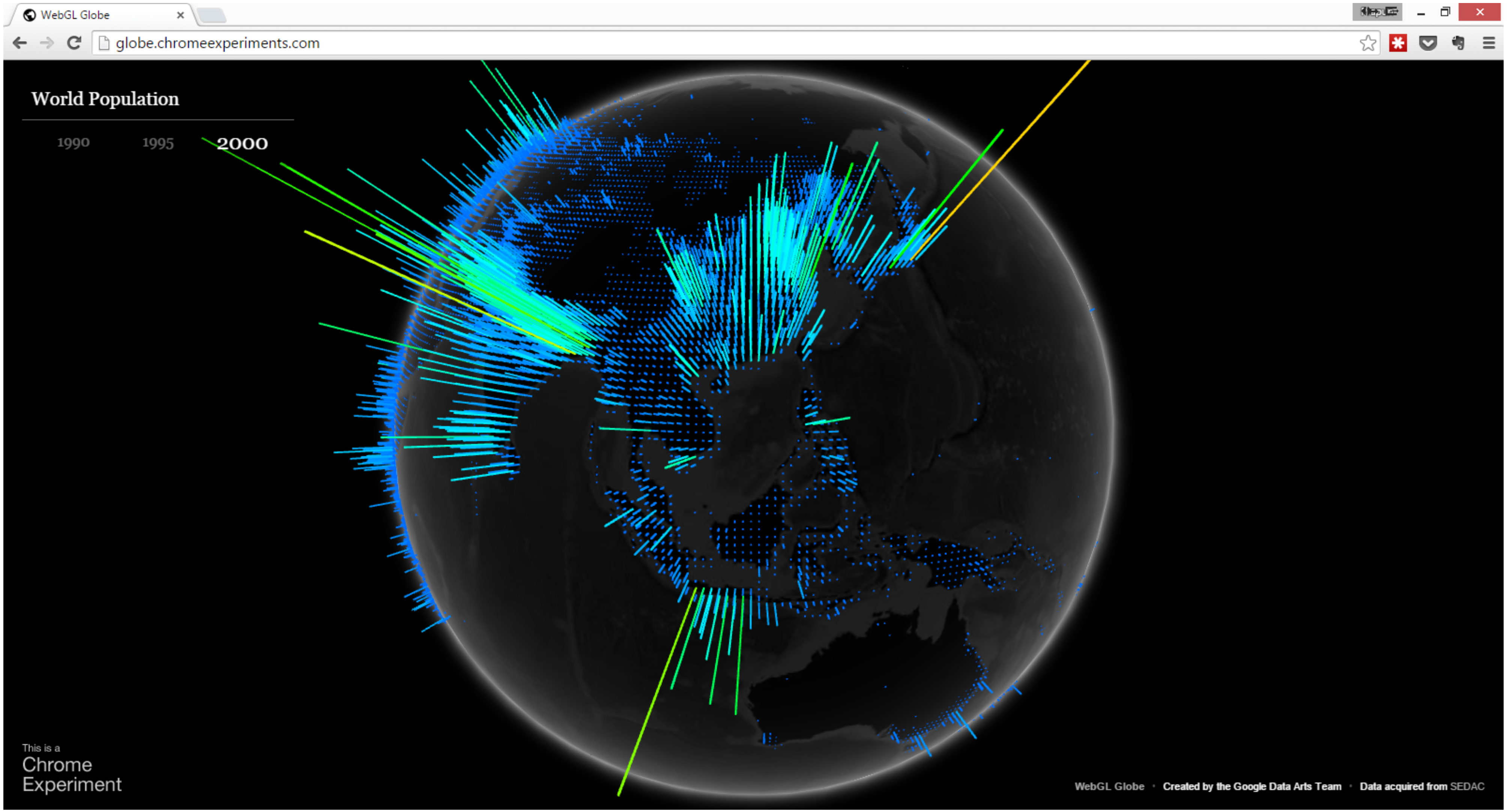


https://www.researchgate.net/publication/260378246_Climate-Related_Hazards_A_Method_for_Global_Assessment_of_Urban_and_Rural_Population_Exposure_to_Cyclones_Droughts_and_Floods/figures?lo=1

World Map of Population Density



World Map of Population Density



http://globe.chromeexperiments.com/

“Density”

- Density = quantity per unit of measure.
- Population Density = number of people per unit area
  - Location with high density value means there are a lot of people around that location.
  - Given a region, we integrate the density over that region to get the number of people residing in that region.
- Probability Density = probability per unit “length”.
  - Given an interval, we integrate the density over that interval to get the probability that the RV will be in that interval.

pdf and cdf for continuous RV

$$F_X(b) - F_X(a)$$

$$P[a \underset{-}{<} X \underset{-}{<} b] \longleftarrow F_X(x) \equiv P[X \leq x]$$

$$P[a < X < b]$$

$$P[a \leq X < b]$$

$$P[a < X \leq b]$$

$$P[a \leq X \leq b]$$

$$\text{“}\frac{d}{dx}\text{”} F_X(x)$$

$$\int_{-\infty}^{x} f_X(t)\, dt$$

$$\int_{a}^{b} f_X(x)\, dx$$

$$f_X(x)$$

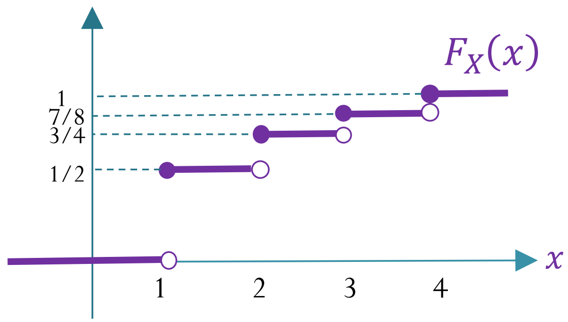
Sections 10.1-10.2

Discrete RV

- **pmf**: $p_X(x) \equiv P[X = x]$
  - Two characterizing properties:
    - $p_X(x) \geq 0$
    - $\sum_x p_X(x) = 1$
- $S_X = \{x: p_X(x) > 0\}$
- $P[\text{some condition(s) on } X]$

$$= \sum_{\substack{\text{\{all the } x \text{ values that} \\ \text{satisfy the condition(s)\}}}} p_X(x)$$

- **cdf** is a staircase function with jumps whose size at $x = c$ gives $P[X = c]$.

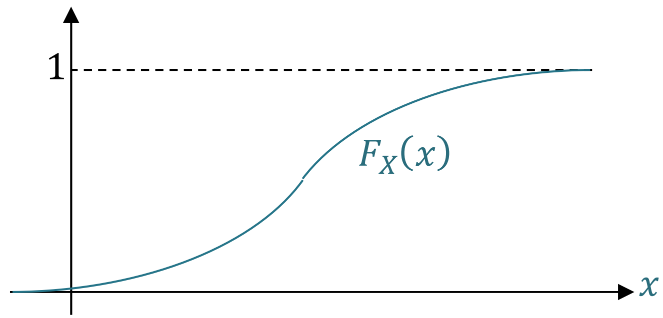


Continuous RV

- $P[X = x] = 0$
- **pdf** : $P[x_0 \leq x \leq x_0 + \Delta x] \approx f_X(x_0)\Delta x$
  - Two characterizing properties:
    - $f_X(x) \geq 0$
    - $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- $S_X = \{x: f_X(x) > 0\}$
- $P[\text{some condition(s) on } X] =$

$$\int_{\substack{\text{\{all the } x \text{ values that} \\ \text{satisfy the condition(s)\}}}} f_X(x)dx$$

- **cdf** is a continuous function.



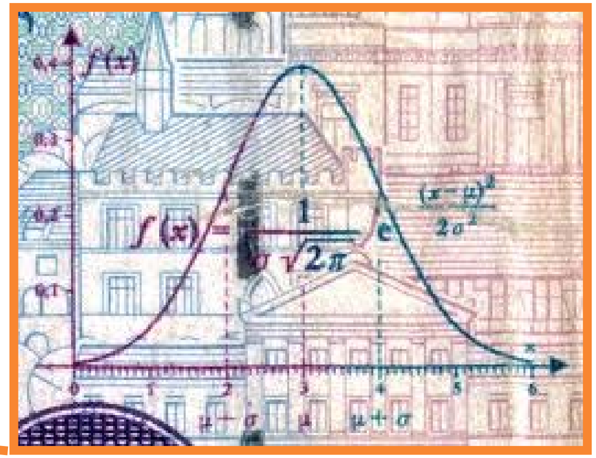
Chapter 9 vs. Section 10.3

Discrete RV	Continuous RV
$\mathbb{E}X = \sum_{x} x p_X(x)$	$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx$
$\mathbb{E}[g(X)] = \sum_{x} g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
$\mathbb{E}[X^2] = \sum_{x} x^2 p_X(x)$	$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

$$\sigma_X = \sqrt{\mathrm{Var}[X]}$$

Johann Carl Friedrich Gauss

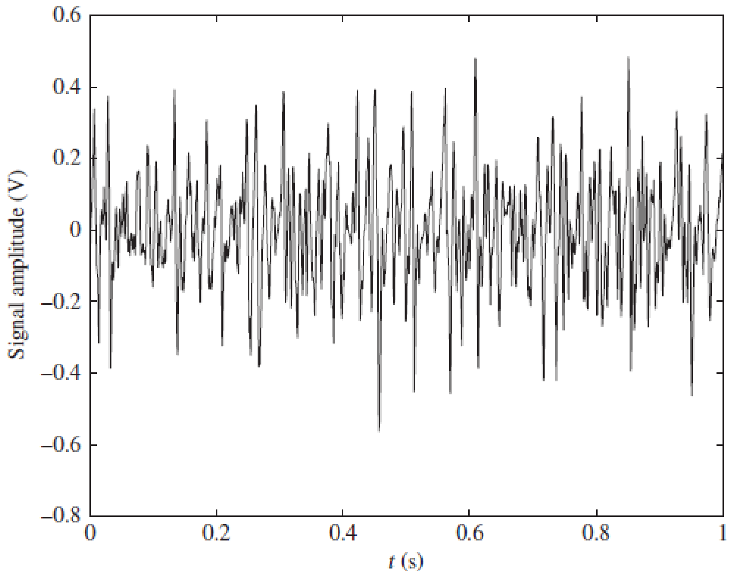


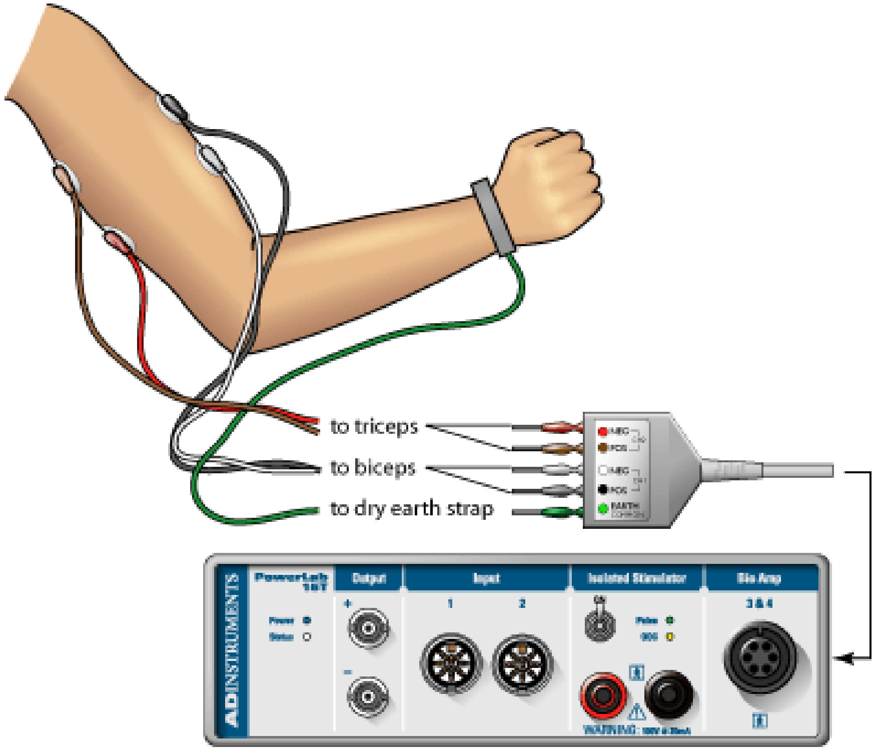
German 10-Deutsche Mark Banknote (1993; discontinued)

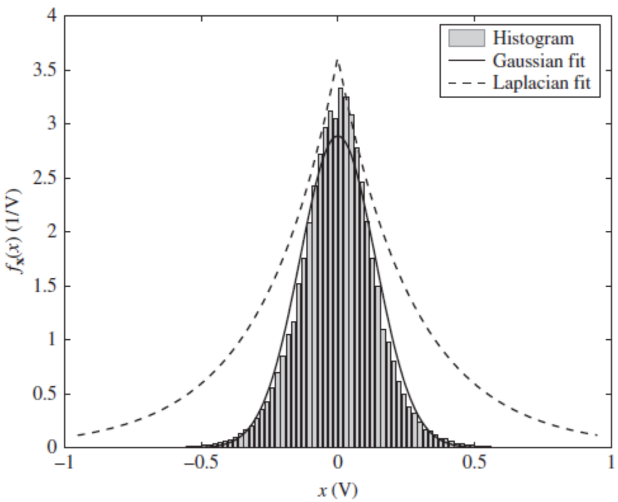
- 1777 –1855
- A German mathematician

Ex. Muscle Activity

- Look at electrical activity of skeletal muscle by recording a human electromyogram (EMG).



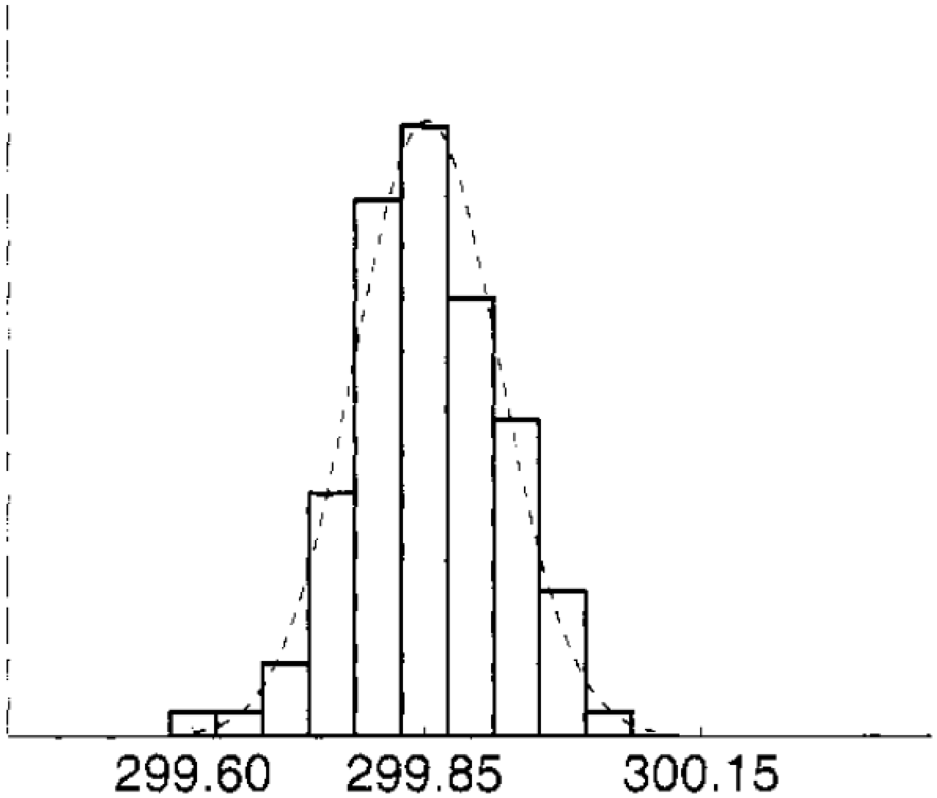




[http://www.adinstruments.com/solutions/education/ltexp/electromyography-emg-german]

Ex. Measuring the speed of light

- 100 measurements of the speed of light (×1,000 km/second), conducted by Albert Abraham Michelson in 1879.

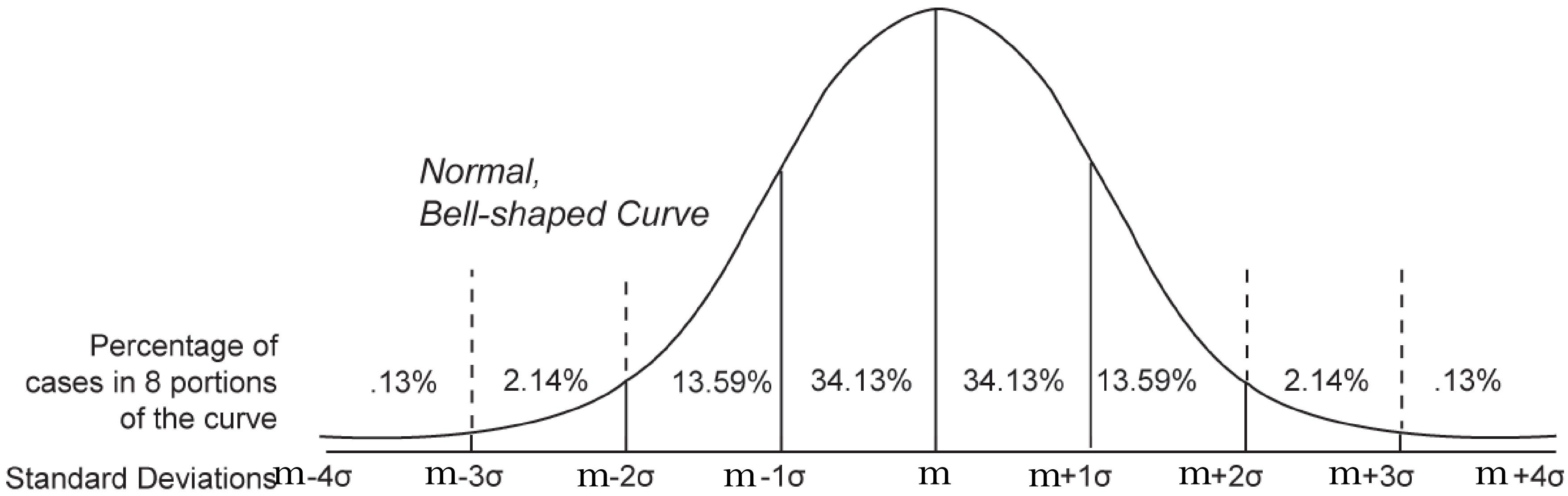


Expected Value and Variance

"Proof" by MATLAB's symbolic calculation

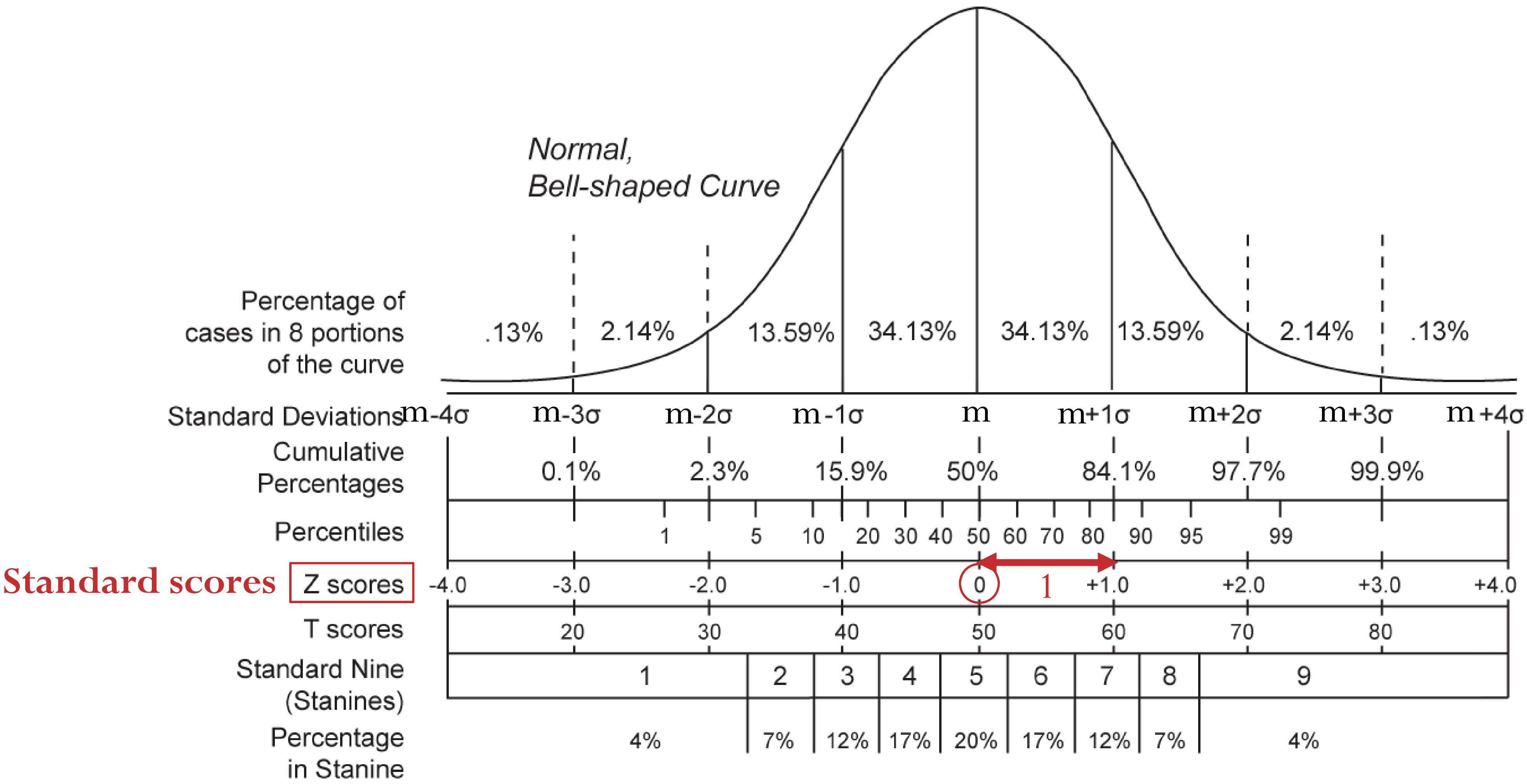
```
>> syms x
>> syms m real
>> syms sigma positive
>> int(1/(sqrt(sym(2)*pi)*sigma)*exp(-(x-m)^2/(2*sigma^2)),x,-inf,inf)
ans =
1
>> EX = int(x/(sqrt(sym(2)*pi)*sigma)*exp(-(x-m)^2/(2*sigma^2)),x,-inf,inf)
EX =
m
>> EX2 = int(x^2/(sqrt(sym(2)*pi)*sigma)*exp(-(x-m)^2/(2*sigma^2)),x,-inf,inf)
EX2 =
-(2^(1/2)*(limit(- x*sigma^2*exp((x*m)/sigma^2 - m^2/(2*sigma^2) - x^2/(2*sigma^2)) - m*sigma^2*exp((x*m)/sigma^2 - m^2/(2*sigma^2) - x^2/(2*sigma^2)) -
(2^(1/2)*pi^(1/2)*sigma*erfi((2^(1/2)*(x - m)*i)/(2*sigma))*(m^2 + sigma^2)*i)/2, x == -Inf) - limit(- x*sigma^2*exp((x*m)/sigma^2 - m^2/(2*sigma^2) -
x^2/(2*sigma^2)) - m*sigma^2*exp((x*m)/sigma^2 - m^2/(2*sigma^2) - x^2/(2*sigma^2)) - (2^(1/2)*pi^(1/2)*sigma*erfi((2^(1/2)*(x - m)*i)/(2*sigma))*(m^2 +
sigma^2)*i)/2, x == Inf)))/(2*pi^(1/2)*sigma)
>> EX2 = simplify(EX2)
EX2 =
m^2 + sigma^2
>> VarX = EX2 - (EX)^2
VarX =
sigma^2
```

Gaussian Random Variable



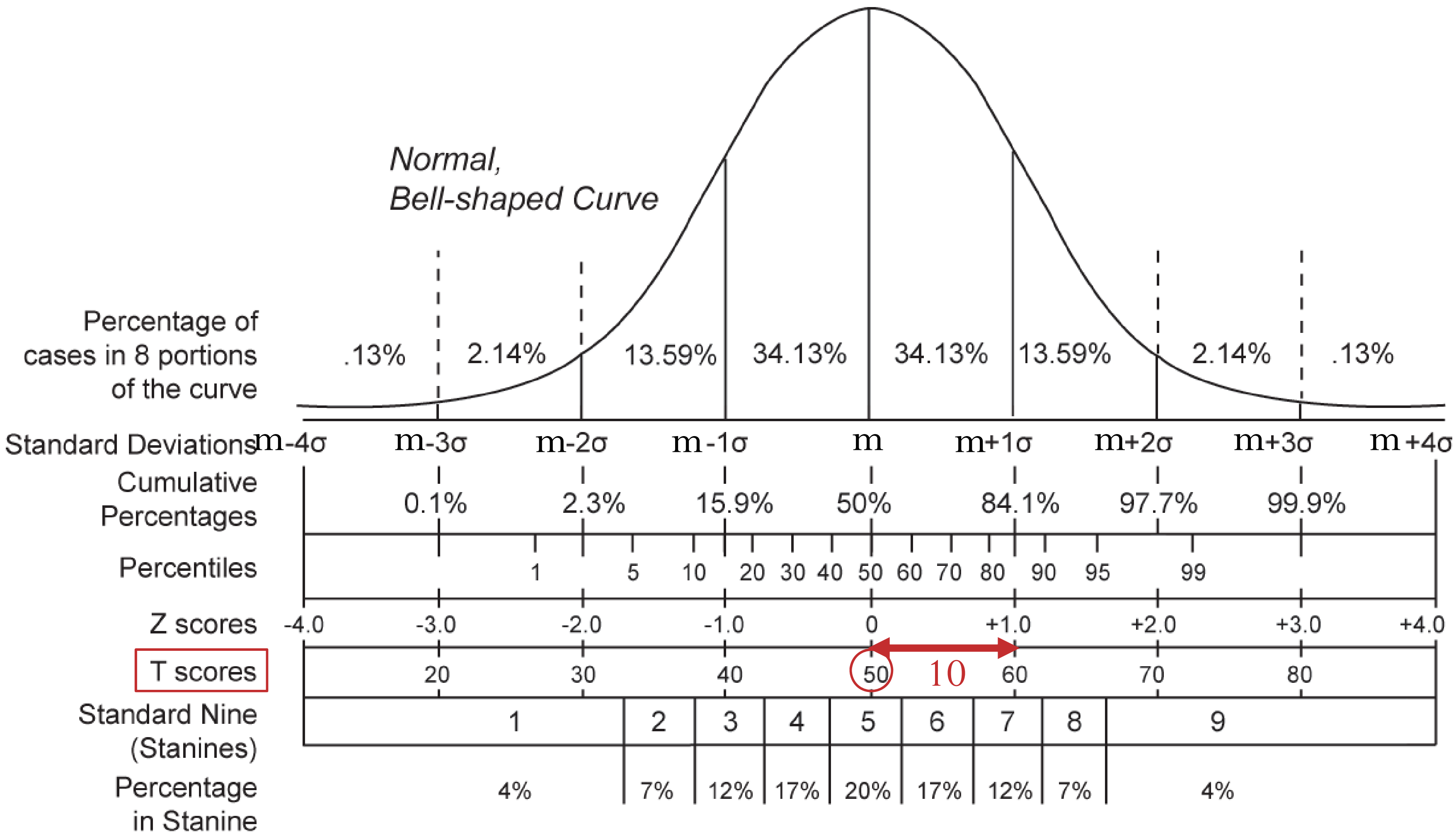
[Wikipedia.org]

Gaussian Random Variable



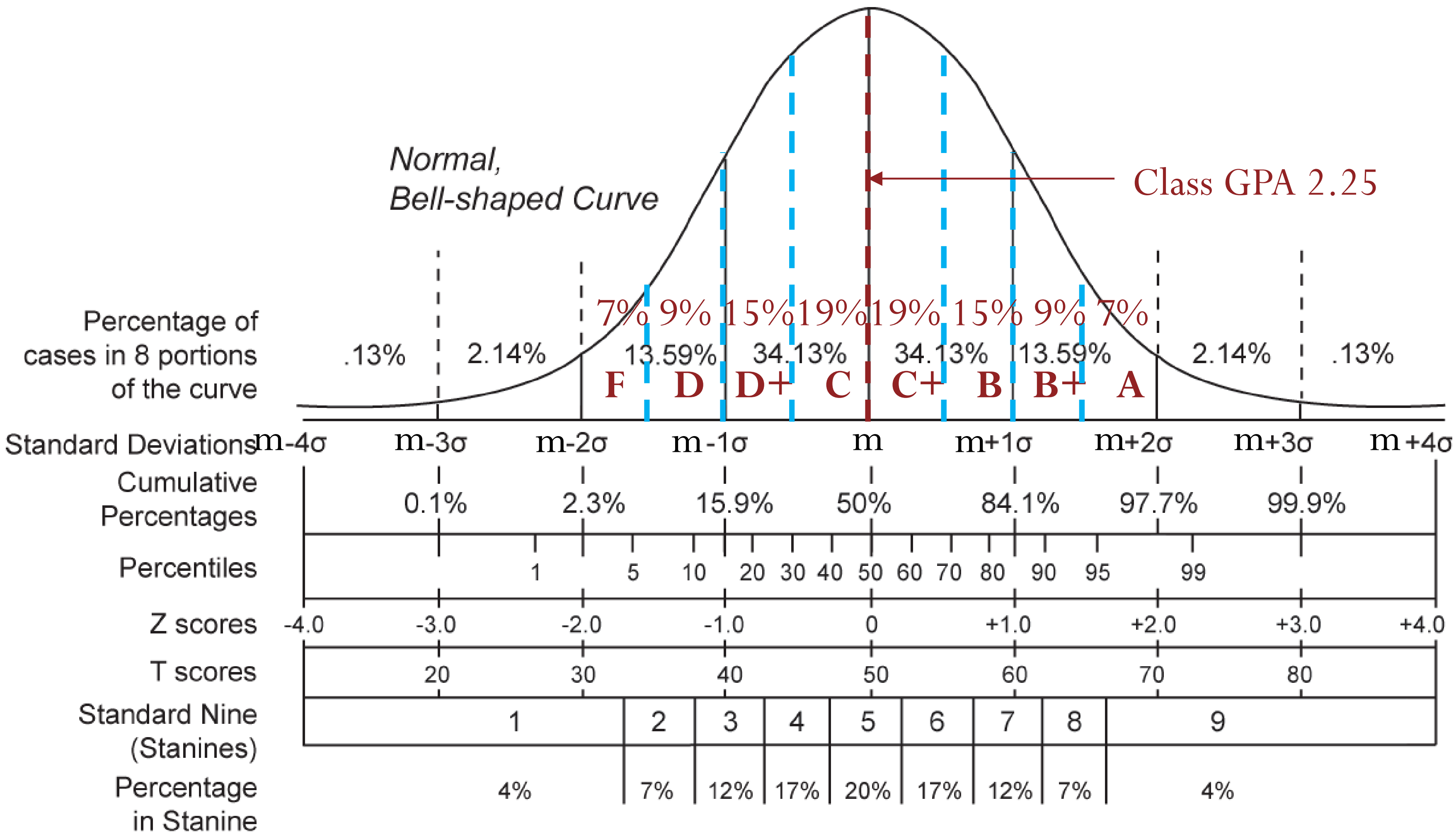
[Wikipedia.org]

Gaussian Random Variable



[Wikipedia.org]

SIIT Grading Scheme (Option 3)



[Wikipedia.org]

From the News

Higgs boson-like particle discovery claimed at LHC

COMMENTS (1665)

By Paul Rincon
Science editor, BBC News website, Geneva

4 July 2012

$$\frac{1}{1-\Phi(5)} \approx 3.5\times10^{6}$$

$$\frac{1}{1-\Phi(4.9)} \approx 2\times10^{6}$$

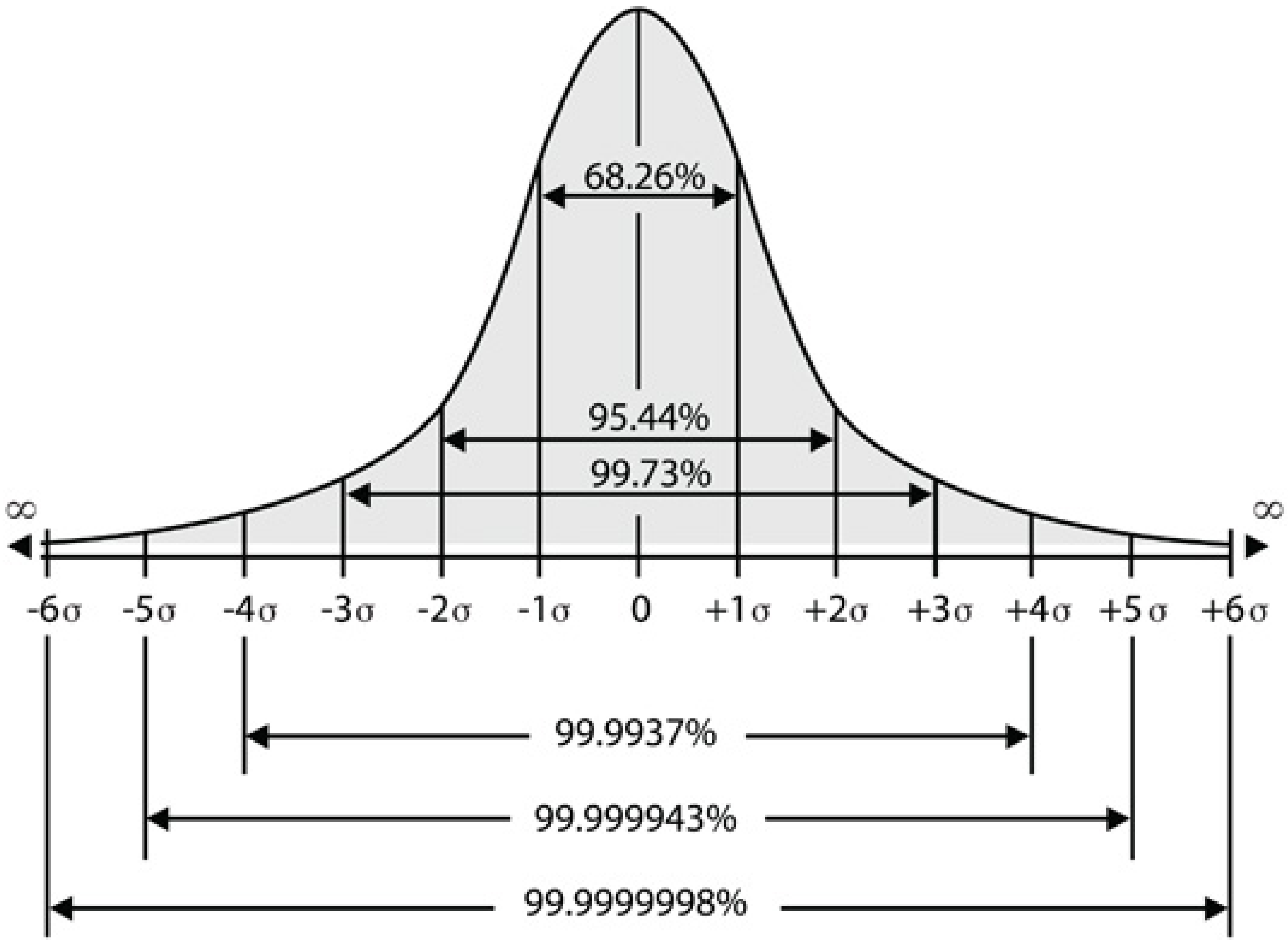
Particle physics has an accepted definition for a **discovery**: a "five-sigma" (or five standard-deviation) level of certainty

The number of sigmas measures how unlikely it is to get a certain experimental result as a matter of chance rather than due to a real effect

They claimed that by combining two data sets, they had attained a confidence level just at the "five-sigma" point - about a **one-in-3.5 million chance** that the signal they see would appear if there were no Higgs particle.

However, a full combination of the CMS data brings that number just back to **4.9 sigma** - a one-in-two million chance.

Six Sigma

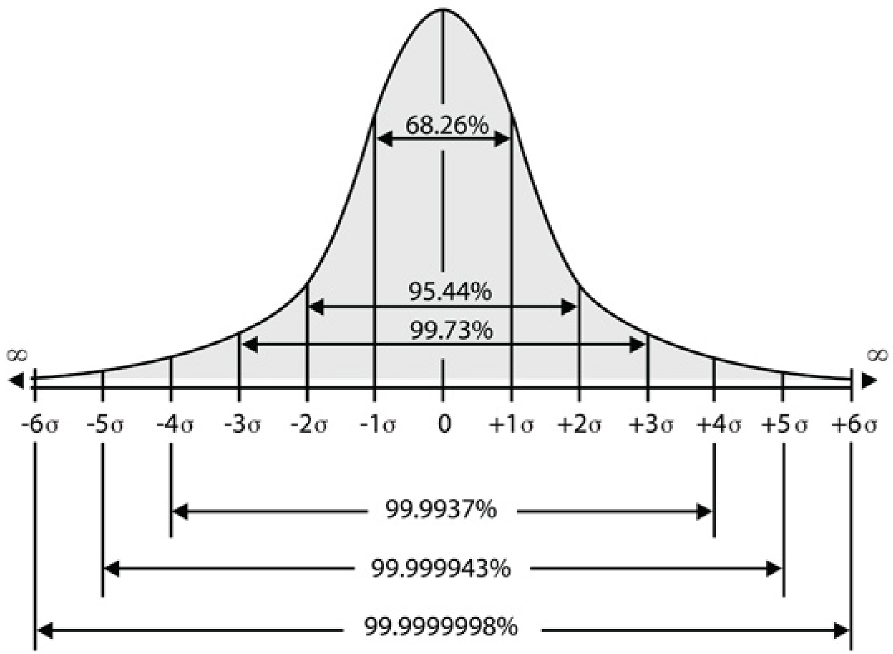

68.26%
95.44%
99.73%
∞
∞
-6σ
-5σ
-4σ
-3σ
-2σ
-1σ
0
+1σ
+2σ
+3σ
+4σ
+5σ
+6σ
99.9937%
99.999943%
99.9999998%

Six Sigma

- If you **manufacture** something that has a normal distribution and get an observation outside six σ of μ, you have either seen something extremely unlikely or there is something wrong with your manufacturing process. You'd better look it over.
- This approach is an example of **statistical quality control**, which has been used extensively and saved companies a lot of money in the last couple of decades.
- The term **Six Sigma**, a registered trademark of **Motorola**, has evolved to denote a methodology to monitor, control, and improve products and processes.
- There are Six Sigma societies, institutes, and conferences.
- Whatever Six Sigma has grown into, it all started with considerations regarding the normal distribution.

[Olofsson, 2006, p. 168]

Six Sigma



Range around μ	Percentage of products in conformance	Percentage of nonconforming products
-1σ to $+1\sigma$	68.26	31.74
-2σ to $+2\sigma$	95.46	4.54
-3σ to $+3\sigma$	99.73	0.27
-4σ to $+4\sigma$	99.9937	0.0063
-5σ to $+5\sigma$	99.999943	0.000057
-6σ to $+6\sigma$	99.9999998	0.00000002

[Bass, 2007, p. 20]