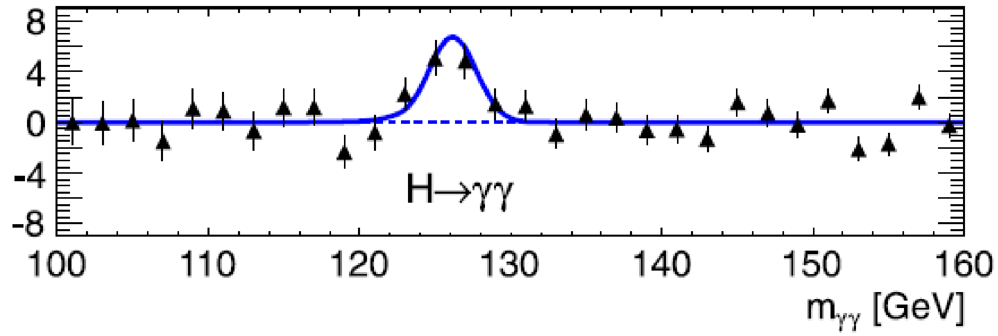
Introduction to Statistical Data Analysis

Chapter 2

Random Variables



Nov 2023

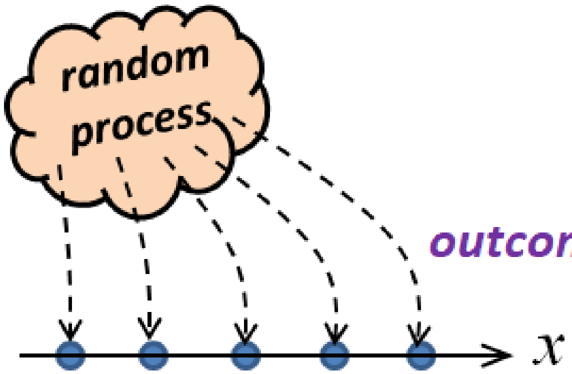
Content

- Random Variable (Discrete / Continuous)
- Expectation of a Random Variable
- Variance and Standard Deviation of Random Variable
- Qumulative Distribution Function
- Examples

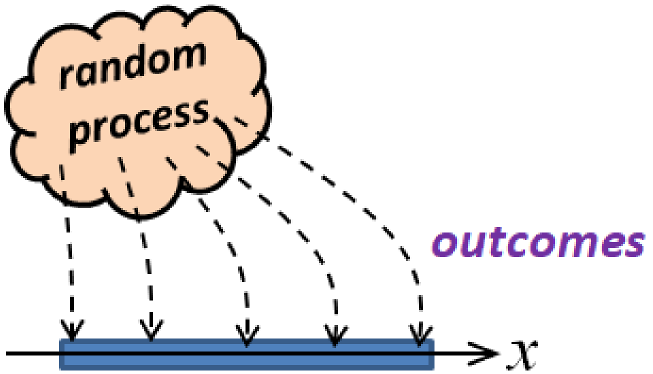
Random Variable

A random variable X is a function that assigns to every element in the sample space S a number representing the outcome of a random process.

We are interested in the range of values of X and their probabilities.



***Discrete** space; the number of outcomes in this is space is **countable**.*



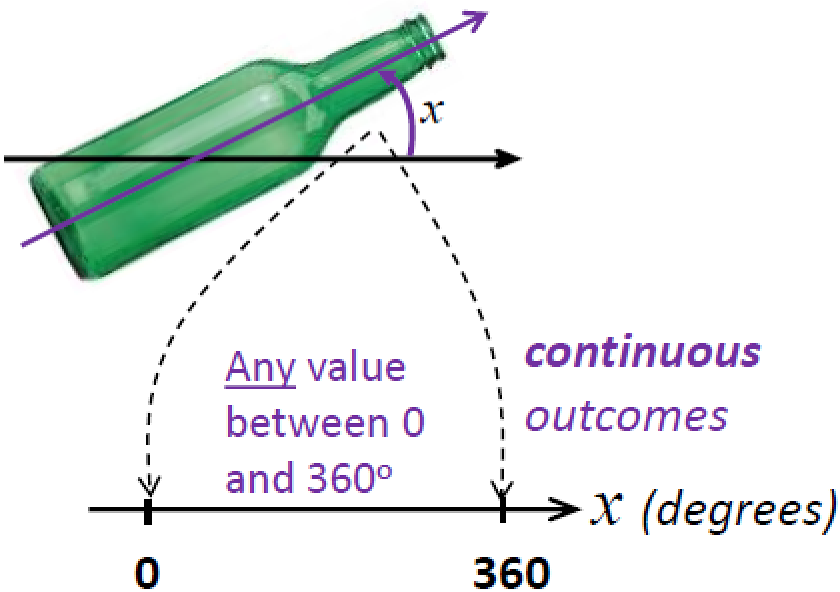
***Continuous** space; the number of outcomes in this space is **uncountable**.*

Note that x is a particular value in the collection X.

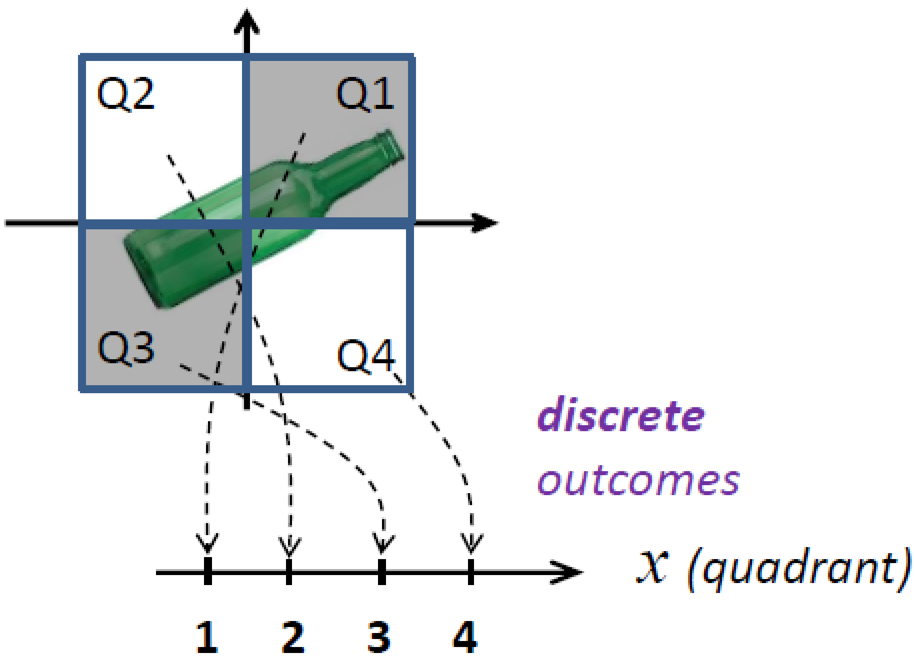
Example 1

The position at which a spinning bottle comes to rest on a smooth surface can be considered to be random.

Random variable X represents the ***angle*** at which the bottle points when it stops.



Random variable X represents the ***quadrant*** in which the bottle points when it stops.



Example 2: Two Body Decay Angles in CM

Consider the two-body decay in center of mass frame: $0 \to 1 + 2$

$$E_1' = (m_0^2 + m_1^2 - m_2^2)/2m_0$$

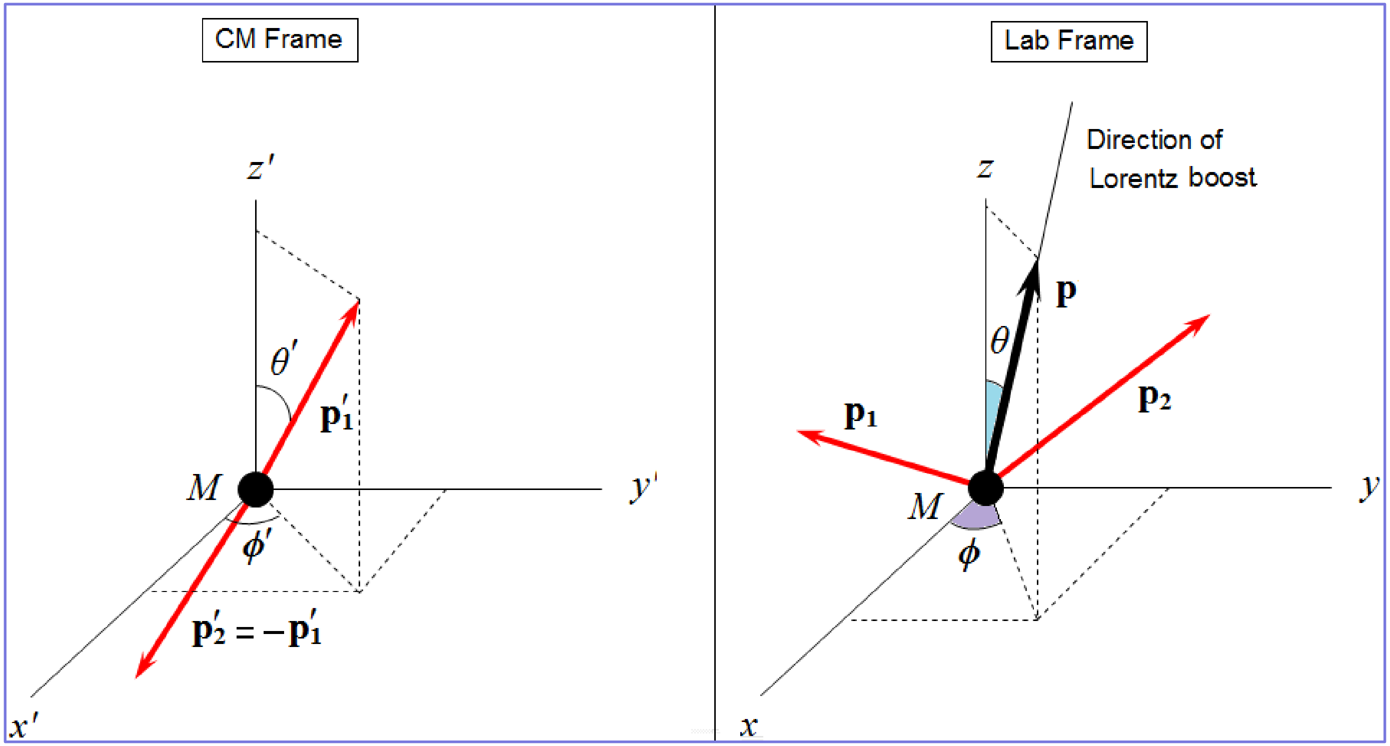
$$E_2' = (m_0^2 + m_2^2 - m_1^2)/2m_0$$

$$|\mathbf{p}_1'| = |\mathbf{p}_2'| = \sqrt{E_1'^2 - m_1^2}$$

Decay angles are continues RV:

$$\phi' = [0, 2\pi]$$

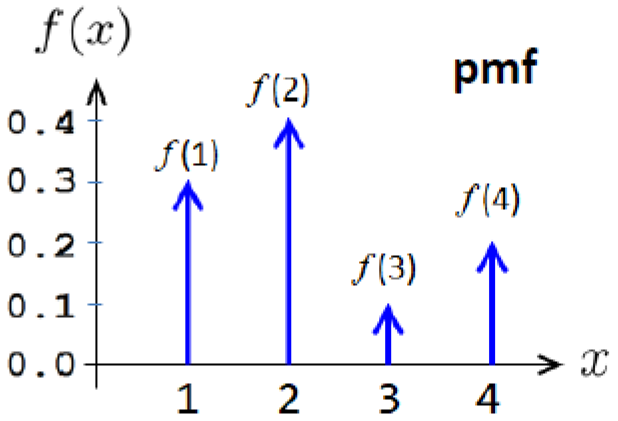
$$\cos\theta' = [-1, 1]$$



Probability Distribution Function (PDF)

Definition: If X is a random variable, the function given by $f(x) = P(X=x)$ for each x in the range of X is called the probability distribution of X.

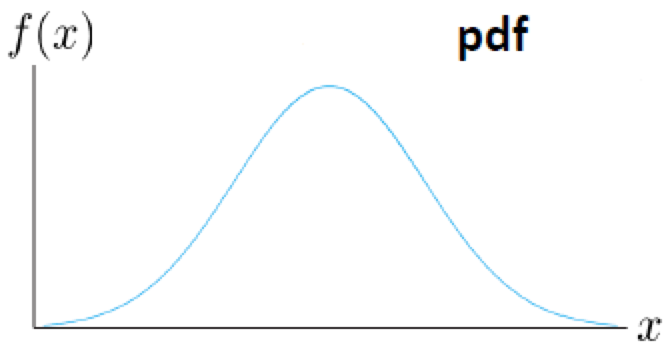
If X is ***discrete*** then $f(x)$ is called the ***probability mass function*** (**pmf**).



$$P(X = x) = f(x)$$

$f(x)$ represents *probability*

If X is ***continuous*** then $f(x)$ is called the ***probability density function*** (**pdf**).



$$P(a < X < b) = \int_a^b f(x)dx$$

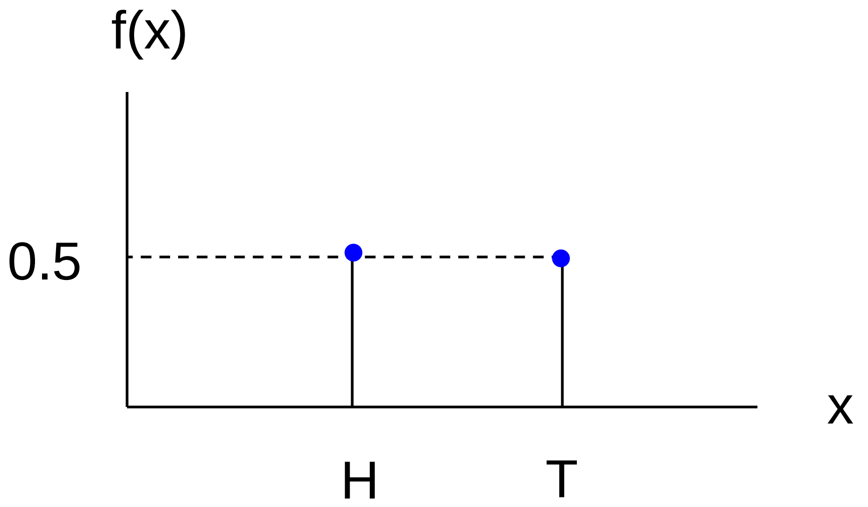
$f(x)$ represents *probability* *density*

Discrete RV Examples:

Distribution (pmf)

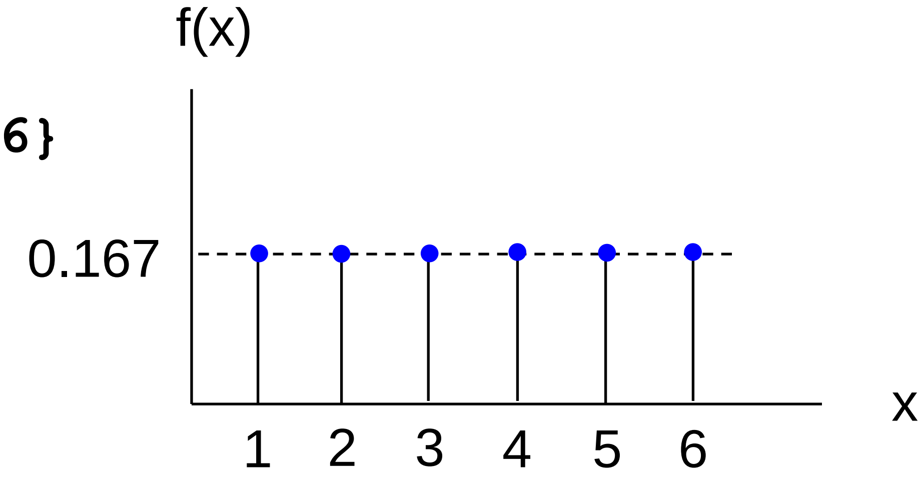
Throwing a coin:

```
   X = {  H,   T}
f(x) = {1/2, 1/2}
```



Thorowing a dice:

```
  X = {1,2,3,4,5,6}
f(x)= {1/6,1/6,1/6,1/6,1/6,1/6}
```

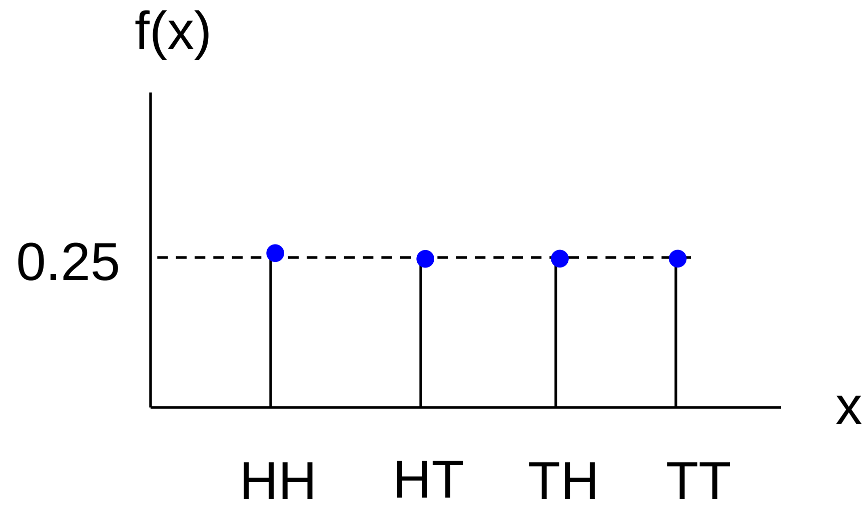


Discrete RV Examples:

Throwing two coins:

```
    X = {HH,   HT,  TH,  TT}
f(x) = {1/4, 1/4, 1/4, 1/4}
```

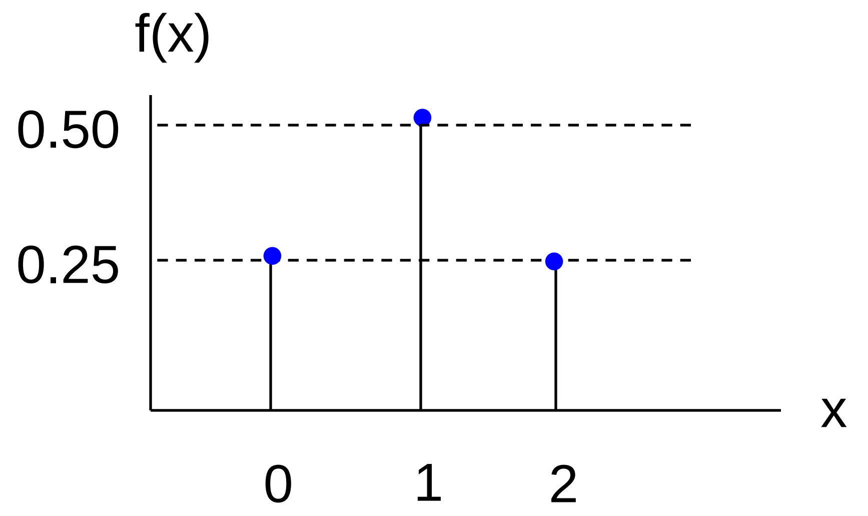
P(HH) = P (HT) = P(TH) = P(TT) = 0.25



Throwing two coins:

Let X = {number of heads}

```
    X = {  0,   1,   2}
f(x) = {1/4, 1/2,  1/4}
```



Discrete RV Examples:

Two dice are thrown. Let X = {sum of number}, Namely

```
    X = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
 f(x) = ?
```

Probabilities:

```
P(X=2) = P(1)P(1) = P11                                    = 1/36
P(X=3) = P(1)P(2)+P(2)P(1) = P12+P21 = 1/36+1/36           = 2/36
P(X=4) = P(1)P(3)+P(3)P(1)+P(2)P(2)  = P13+P31+P22         = 3/36
. . .
```

$P(X=2) = P11 = 1/36 = 0.0278$

$P(X=3) = P12+P21 = 2/36 = 0.0556$

$P(X=4) = P13+P31+P22 = 3/36 = 0.0833$

$P(X=5) = P14+P41+P23+P32 = 4/36 = 0.1111$

$P(X=6) = P15+P51+P24+P42+P33 = 5/36 = 0.1389$

$P(X=7) = P16+P61+P25+P52+P34+P43 = 6/36 = 0.1667$

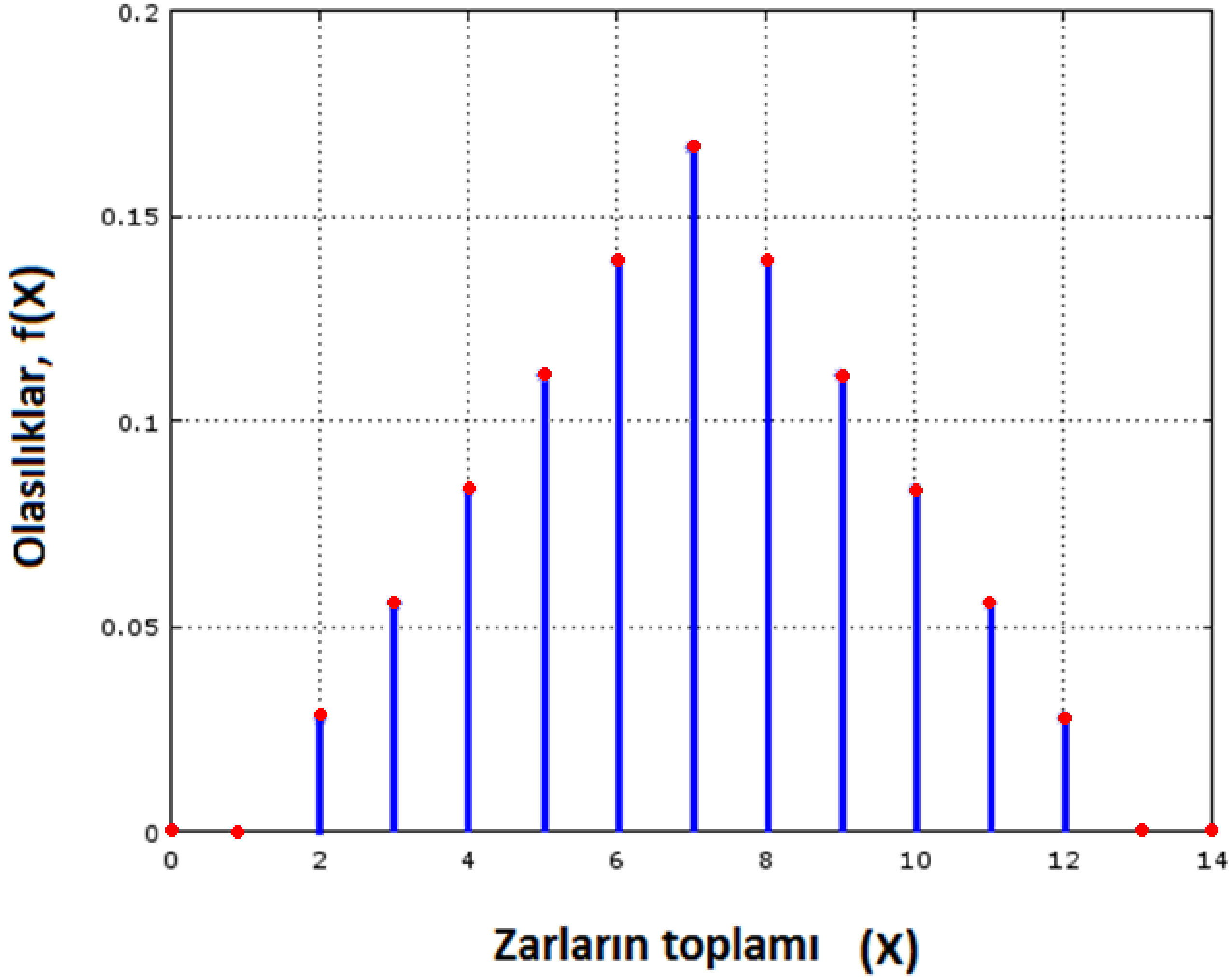
$P(X=8) = P26+P62+P35+P53+P44 = 5/36 = 0.1389$

$P(X=9) = P36+P63+P45+P54 = 4/36 = 0.1111$

$P(X=10) = P46+P64+P55 = 3/36 = 0.0833$

$P(X=11) = P56+P65 = 2/36 = 0.0556$

$P(X=12) = P66 = 1/36 = 0.0278$



X	f(X)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Notice that: $\sum_i f(x_i) = 1$

Continues RV Examples:

Two-body decay angles:

$\Phi = [0, 2\pi]$

$f(\varphi) = 1 / 2\pi$

$\cos\Theta = [-1, 1]$

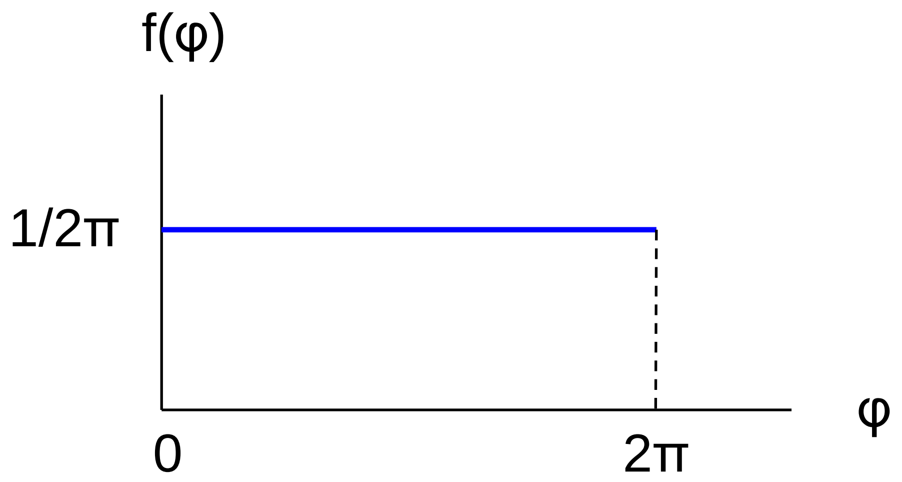
$f(\theta) = \frac{1}{2}\sin(\theta)$

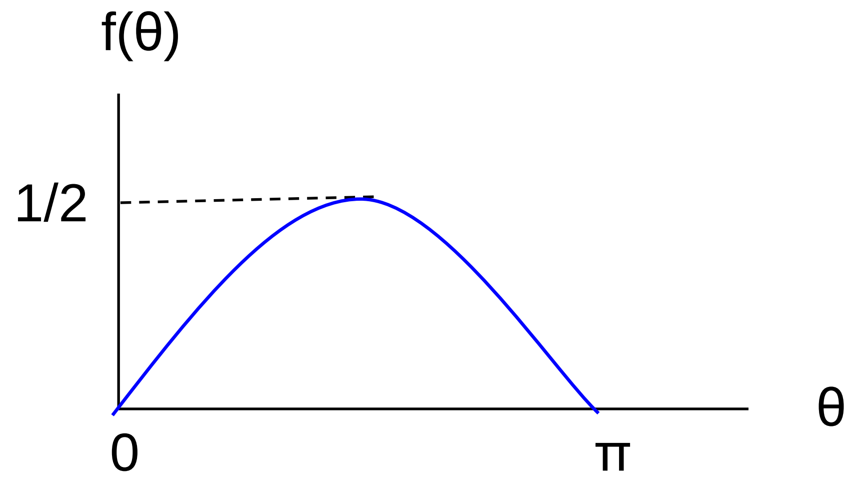
Last eqn. is obtained from:

$$f_\theta(\theta)d\theta = f_x(x)dx \rightarrow f_\theta(\theta) = f_x(x)\left|\frac{dx}{d\theta}\right| = \frac{1}{2}\frac{d}{d\theta}[\cos(\theta)]$$

where $f_x(x)$ is a uniform pdf in the range [-1,1].

pdf

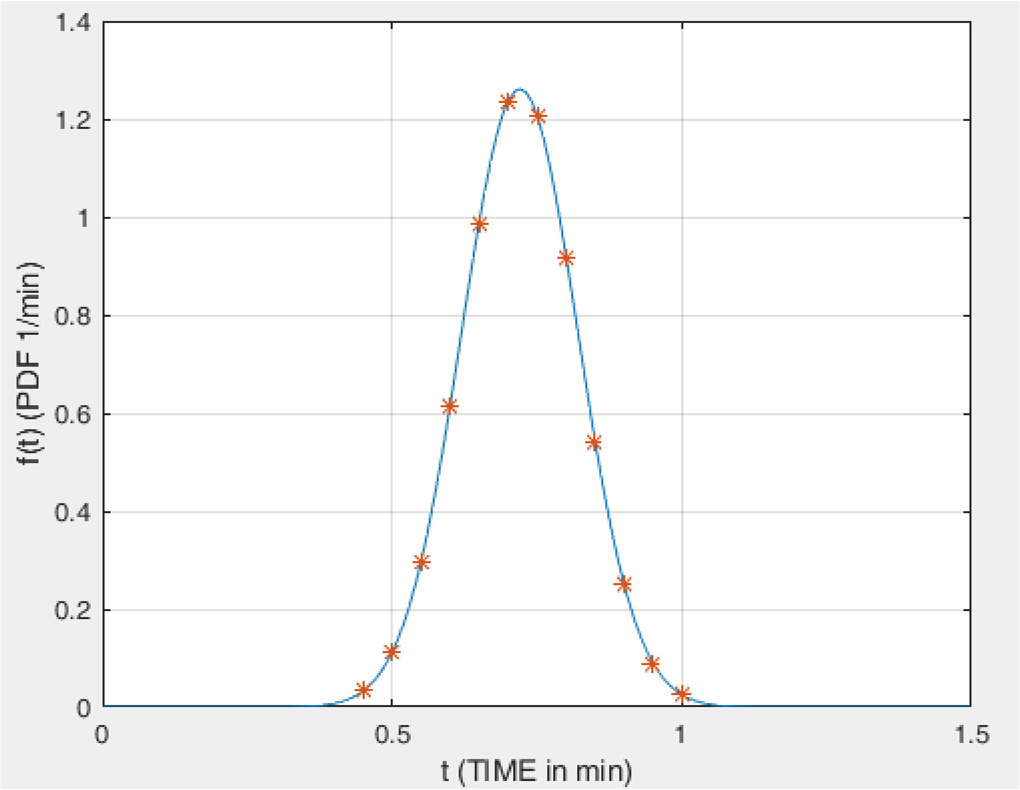




Continues RV Examples:

T = {Run-time of a program in a server}

f(t) = {measured discrete values} → *might be continues*!



Some Properties

Discrete

$$f(x_i) \geq 0$$

$$\sum_i f(x_i) = 1$$

$$\sum_{i=a}^{b} f(x_i) = P(a \leq x \leq b)$$

Continues

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\int_a^b f(x)dx = P(a \leq x \leq b)$$

f(*x*) is NOT a probability! It has dimension 1/x

Expectation Values

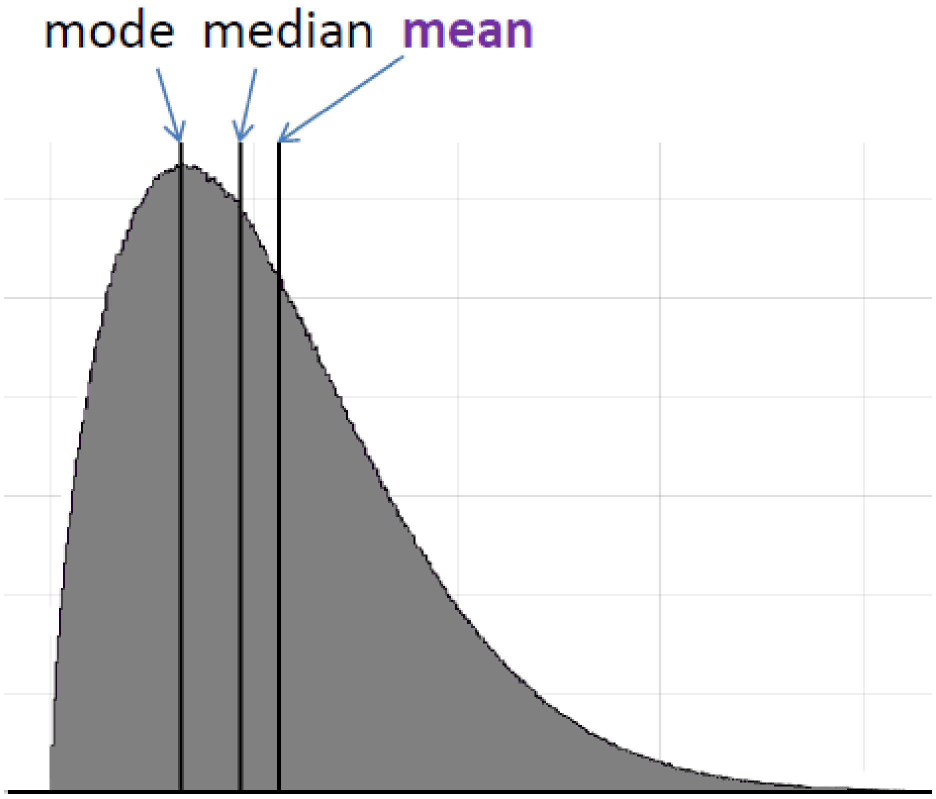
Expectation values are mathematical objects that summarize a distribution in the form of a few descriptive values (*descriptive statistics*).

In this course we focus on two objects:

1. The ***mean***, μ_X
This is a measure of the center of the distribution (the center of mass).

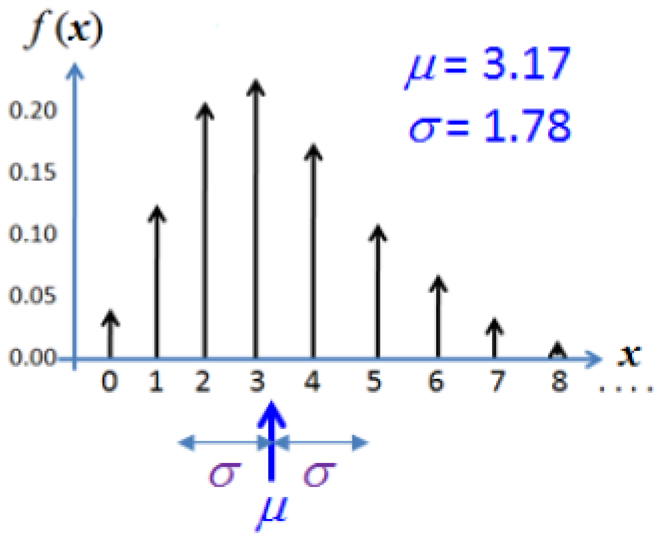
2. The ***standard deviation***, σ_X
This is a measure of how much the distribution is spread about the mean.

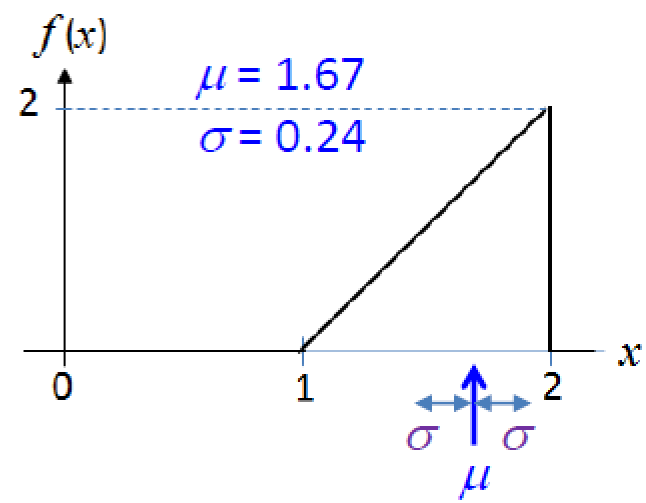
See also skewness and kurtosis (not covered in this course).



Examples of the mean μ and standard deviation σ

(we will learn how to calculate μ and σ later).





Estimation

$$\begin{cases} \mu = \text{center of mass} \\ \mu \pm \sigma \text{ covers} \approx \dfrac{2}{3} \text{ of the distribution} \end{cases}$$

Definition of Mean

Let X be a random variable with probability distribution $f(x)$.

The ***mean*** of X is

$$\mu_X = E[X] = \sum_x x f(x)$$ if X is a **discrete** random variable

$$\mu_X = E[X] = \int_x x\, f(x) dx$$ if X is a **continuous** random variable

This is a measure of the *location* of the distribution (*the center of mass*).

NOTE1: Notation for mean can be $E[X]$ or $< X >$ or $\bar{x}$

NOTE2: In Quantum Mechanics, expectation value of position is postulated/defined as follows:

$$< x >= E[X] = \int_{-\infty}^{+\infty} \Psi^* x \Psi \, dx$$

In general:

$$< g >= E[g] = \int_{-\infty}^{+\infty} \Psi^* g \Psi \, dx$$

The expectation of the function $g(X)$ is:

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x) f(x)$$ if X is **discrete**

$$\mu_{g(X)} = E[g(X)] = \int_x g(x) f(x) dx$$ if X is **continuous**

Simply replace x with $g(x)$

Some properties

$$E[aX^n + b] = aE[X^n] + b$$ *a*, *b*, *n* are constants

$$E[g(X)] \sim g(E[X])$$ This approximation improves as $\sigma \to 0$

Mean of Squares

Definition:

$$E[X^2] = \sum_i x_i^2 \; f(x_i)$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \; f(x) dx$$

Root Mean Square:

$$RMS = \sqrt{E[X^2]}$$

Variance and Standard Deviation

The variance (σ^2) of a random variable X is defined by:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x) \qquad \text{discrete}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_x (x-\mu)^2 f(x)dx \qquad \text{continuous}$$

*This is a measure of by how much the distributution **varies** about the mean value.*

$$\begin{aligned}\sigma_X^2 &= E[(X-\mu_X)^2] = E[X^2 - 2X\mu_X + \mu_X^2] \\ &= E[X^2] - 2\mu_X E[X] + \mu_X^2 \\ &= E[X^2] - \mu_X^2 \;\Rightarrow\; \sigma_X^2 = E[X^2] - E[X]^2\end{aligned}$$

This is generally easier to calculate

The standard deviation is then $\sigma_X = \sqrt{\sigma_X^2}$

Property: $\sigma_{(aX+b)} = a\ \sigma_X$

$$\begin{aligned}\sigma_X &= \sqrt{E[X^2] - (E[X])^2} \\ &= \sqrt{<X^2> - <X>^2}\end{aligned}$$

Example 3: Tossing 3 coins:

What is the mean number of heads obtained when three coins are tossed?

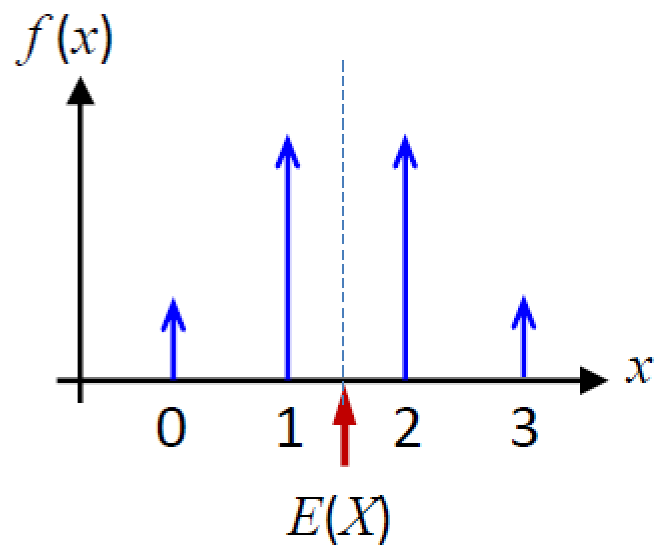
Solution

X	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

$$\mu_X = E[X] = \sum_x x f(x)$$

= 0×1/8 + 1×3/8 + 2×3/8 + 3×1/8 = **1.5**

The mean outcome of the random process is "1.5 heads".



Note that in this case, since μ is the center of mass we can directly say, "by symmetry the mean is 1.5"

Example 4: Tossing a dice:

$$X = \{1, 2, 3, 4, 5, 6\}$$

$$f(x) = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

Find:

(a) $\langle X \rangle = 1 * 1/6 + 2 * 1/6 + \ldots + 6 * 1/6 \quad = 3.50$

(b) $\langle X^2 \rangle = 1^2 * 1/6 + 2^2 * 1/6 + \ldots + 6^2 * 1/6 \quad = 15.17$

(c) $RMS = sqrt(\langle X^2 \rangle) \quad = 3.89$

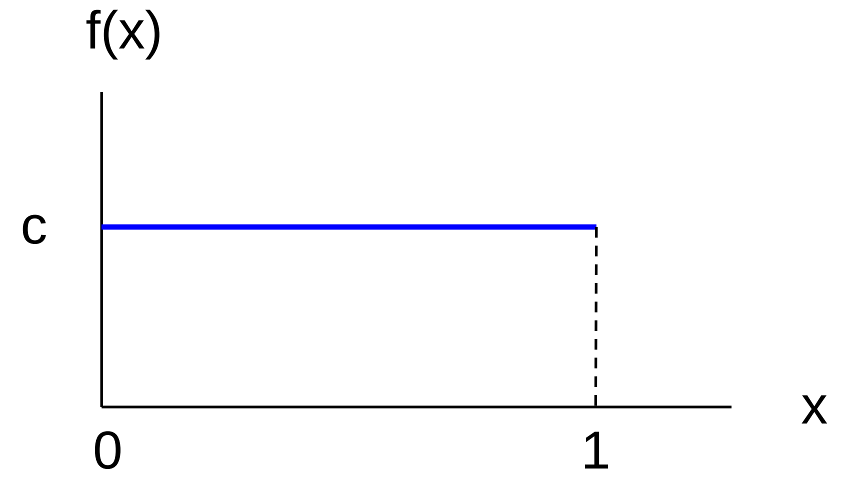
(d) $\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 = 15.17 - (3.5)^2 \quad = 2.92$

(e) $\sigma = sqrt(2.92) \quad = 1.71$

Example 5: Uniform Dist.

X = [0, 1]

f(x) = c = constant



(a) Find c

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \longrightarrow \int_{0}^{1} c\,dx = c(1-0) = 1 \longrightarrow c = 1$$

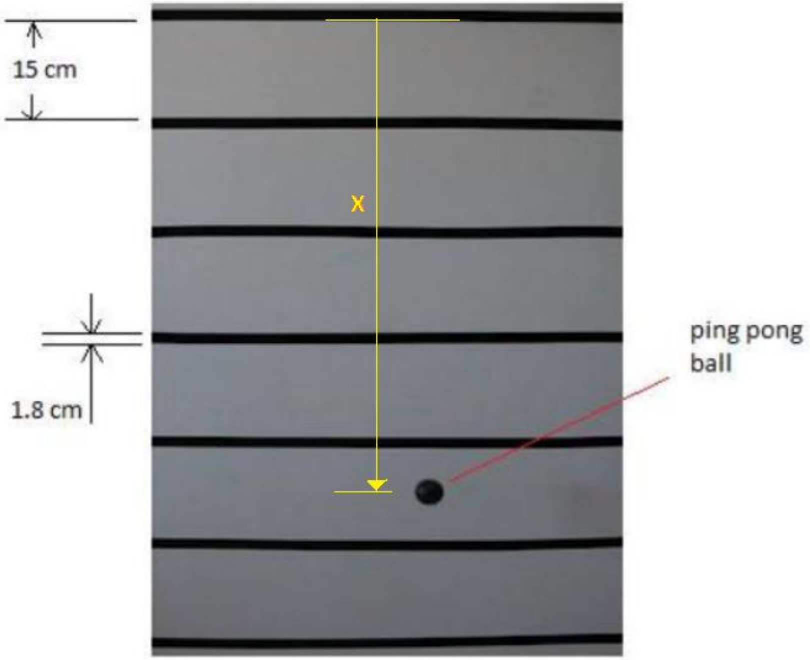
(b) Mean and std.dev

$$< X > = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{1} x\,dx = [x^2/2]_0^1 = 1^2/2 - 0^2/2 = 1/2 = 0.5$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\bar{x})^2 \; f(x)dx = \int_{0}^{1} (x-1/2)^2 \; dx = 1/12$$

$$\sigma = \sqrt{\sigma^2} = 1/\sqrt{12} = 0.289$$

Example 6: Freefall Experiment



We took a video of a freely falling ping-pong ball.
As it falls, we snap totally N = 59 frames (photos). Frame rate of the camera was F = 120 fps. I drop it from an initial height h = 1.3 m.

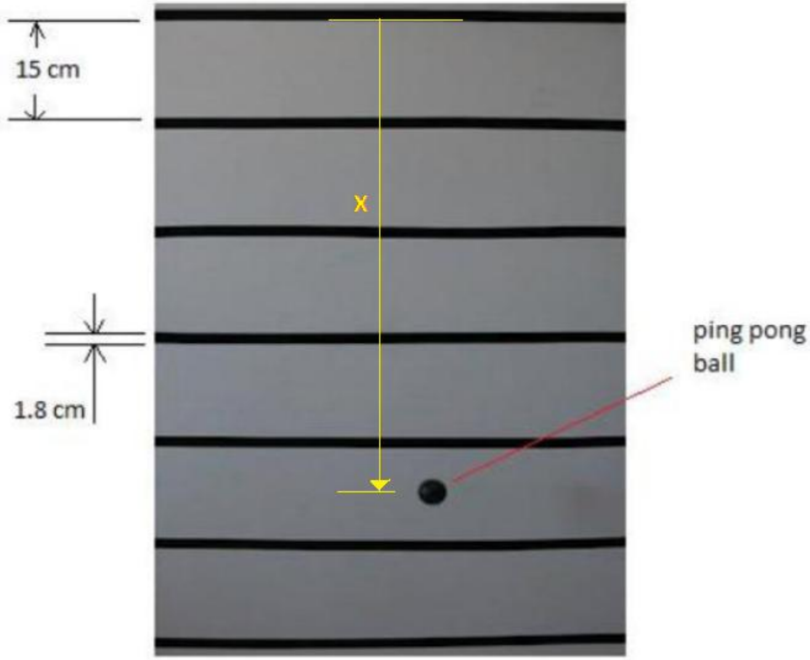
(a) Determine the probability density function, f(x), of the ball where x is the distance from the intial height.

(b) Plot the distribution of ball's position extracted randomly from 1000 photos using image processing tool. (*Namely, you have 59 photos. You select 1000 random photos from the pool of 59 photos and evaluate ball position*).

Position: $$x = \frac{1}{2} g t^2$$

Velocity: $$v = \frac{dx}{dt} = gt$$

Total time: $$T = \sqrt{\frac{2h}{g}}$$



Probability of getting any image is $P_i = 1/N = 1/59$.

Probability that the camera flashes in time dt is $P_c = dt / T = P_i$ (dt = $1/F = 1/120$).

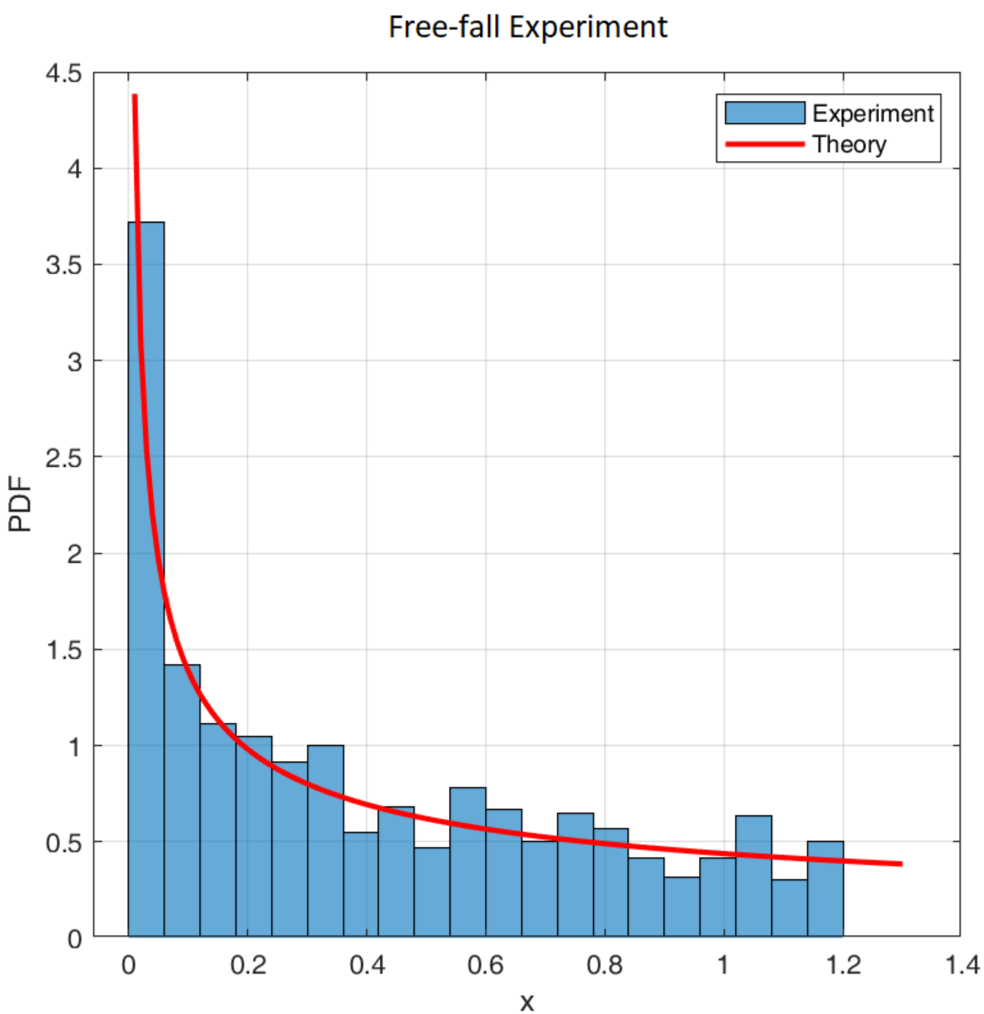
$$P = \frac{dt}{T} = \frac{dx/v}{\sqrt{2h/g}} = \frac{dx/gt}{\sqrt{2h/g}} = \frac{dx/gt}{\sqrt{2h/g}} = \frac{1}{2\sqrt{hx}} dx = f(x)dx$$

Therefore p.d.f. is $\boxed{f(x) = \frac{1}{2\sqrt{hx}}}$ $\quad 0 < x < h$

$$f(x) = \frac{1}{2\sqrt{hx}} = \frac{1}{2\sqrt{1.3x}} = \frac{0.4385}{\sqrt{x}}$$

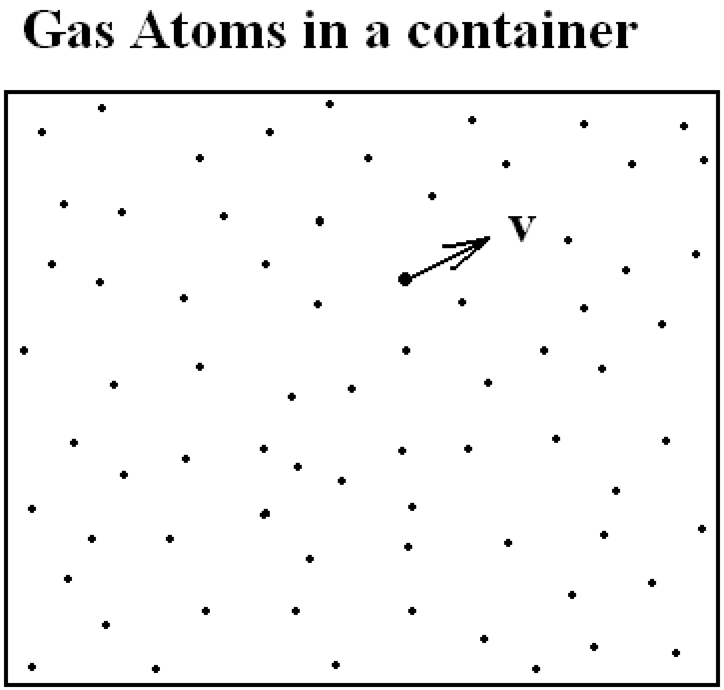
$$\int_0^h f(x)dx = 1$$

$$\langle x \rangle = \int_0^h x\, f(x)dx = 0.43\,\mathrm{m}$$



Example 7:

Maxwell Speed Distribution Function



$$f(v) = \sqrt{\frac{2}{\pi}\left(\frac{m}{kT}\right)^3} v^2 \exp\left(\frac{-mv^2}{2kT}\right)$$

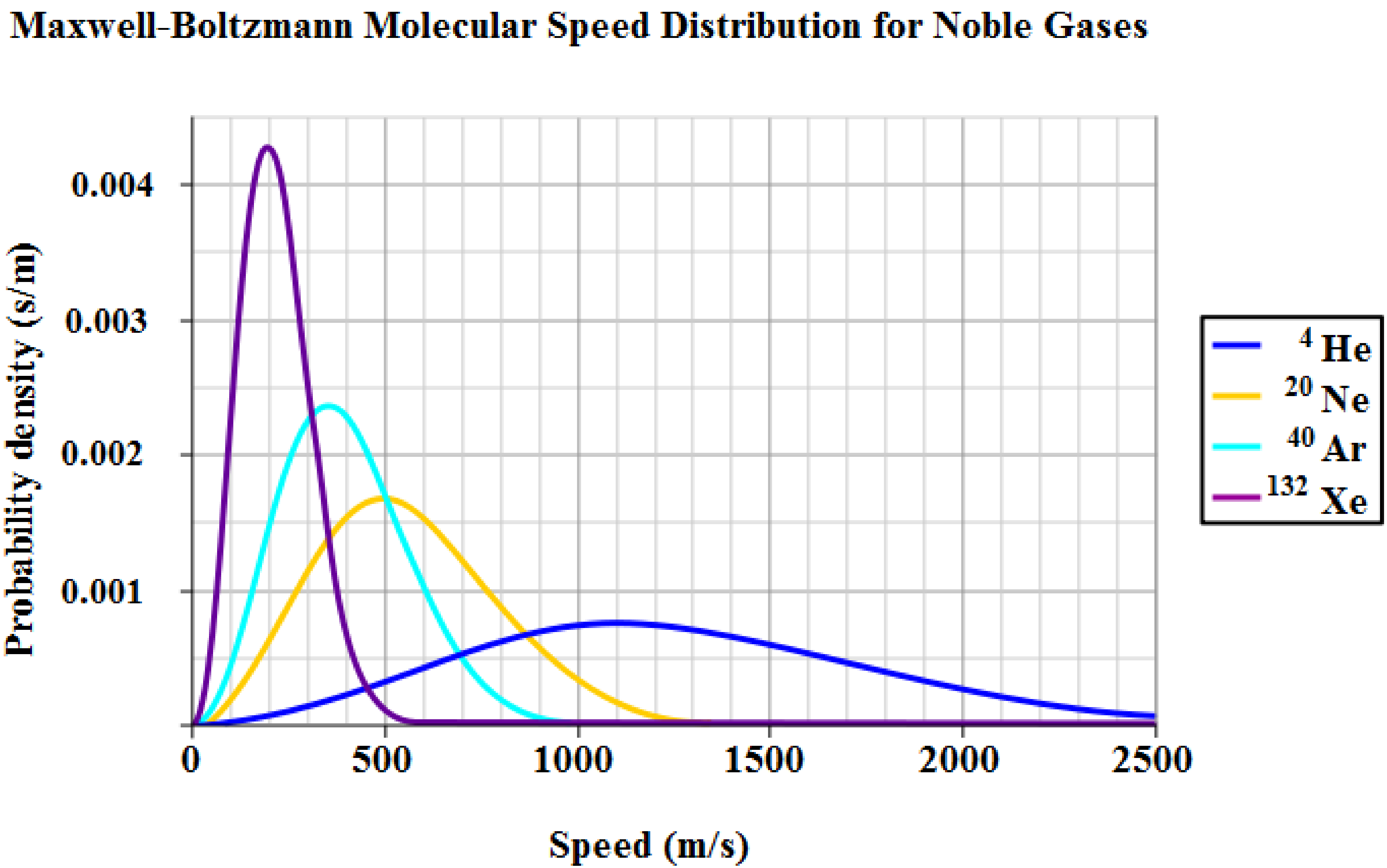
here

m = mass of the particle (gas atom)

T = temperature (Kelvin)

k = Boltzman const. (k = 1.38065×10^{-23} J/K)

- It is first defined and used for describing particle speeds (v) in idealized gases.
- f(*v*) is pdf = probability per unit speed of finding the particle with a speed near (*v*).



Maxwell-Boltzmann Molecular Speed Distribution for Noble Gases
0.004
0.003
0.002
0.001
Probability density (s/m)
0
500
1000
1500
2000
2500
Speed (m/s)
4 He
20 Ne
40 Ar
132 Xe

$$f(v) = \sqrt{\frac{2}{\pi}\left(\frac{m}{kT}\right)^3} v^2 \exp\left(\frac{-mv^2}{2kT}\right)$$

Mean values: $< v > = \int_0^\infty v \; f(v) dv$ $\qquad$ $< v^2 > = \int_0^\infty v^2 \; f(v) dv$

Expectation value of K.E. $< K > = \frac{1}{2} m < v^2 >$

Pressure (V = volume of the container): $p = \frac{2}{3V} < K >$

Probability of having speed greater than c / 1000 = 3×10^5 m/s:

$$P(v > c/1000) = \int_{c/1000}^{\infty} f(v) dv$$

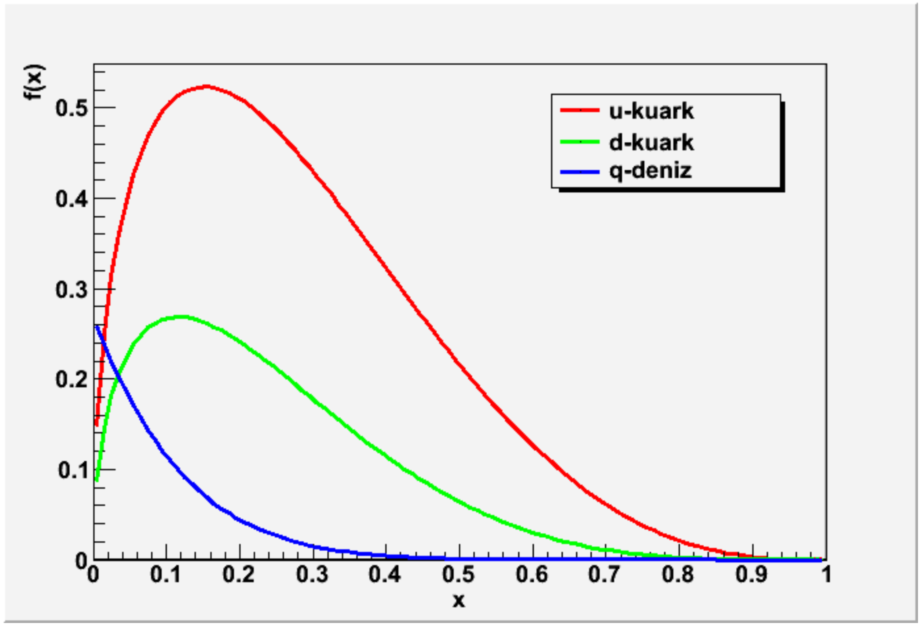
Example 8:

Parton Distribution Functions in proton

$$u_v(x) = 2.13\sqrt{x}(1-x)^{2.8}$$

$$d_v(x) = 1.26\sqrt{x}(1-x)^{3.8}$$

$$q_s(x) = 0.27(1-x)^{8.1}$$



$$< u >= \int_0^1 x\, u_v(x) dx = 0.283$$

Cumulative Distribution Function (CDF)

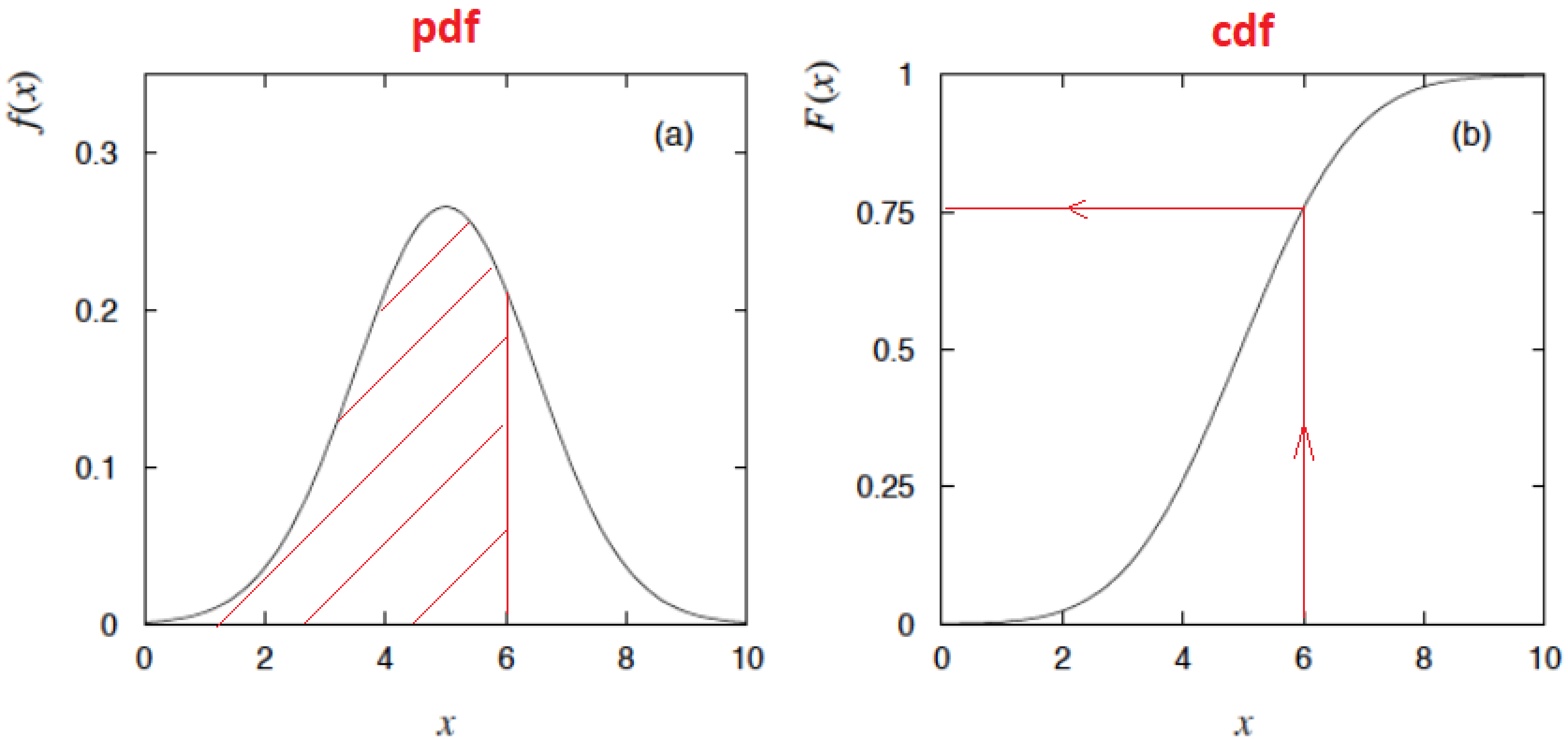
Given a pdf, $f(x')$, probability to have outcome less than or equal to **x** is:

$$\sum_{i=1}^{x} f(x'_i) = F(x) \qquad \int_{-\infty}^{x} f(x')dx' = F(x)$$

- F(x) is called cdf.
- F(-∞) = 0 and F(+∞) = 1.
- If cdf is given, then pdf can be calculated from:

$$f(x) = \frac{\Delta F}{\Delta x} \qquad f(x) = \frac{dF}{dx}$$

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$

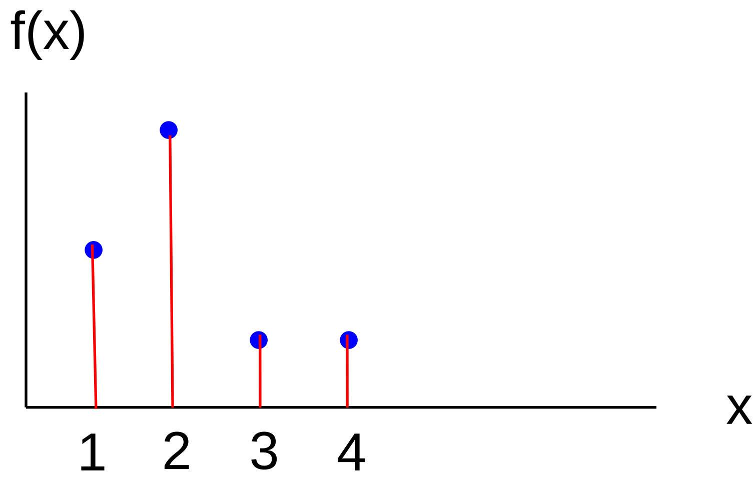


Example 9:

Given X = {1.0, 2.0, 3.0, 4.0} and

f(x) = {0.3, 0.5, 0.1, 0.1}

Determine cdf



$$\sum_{i=1}^{x} f(x'_i) = F(x)$$

$$F(0) = 0.0$$

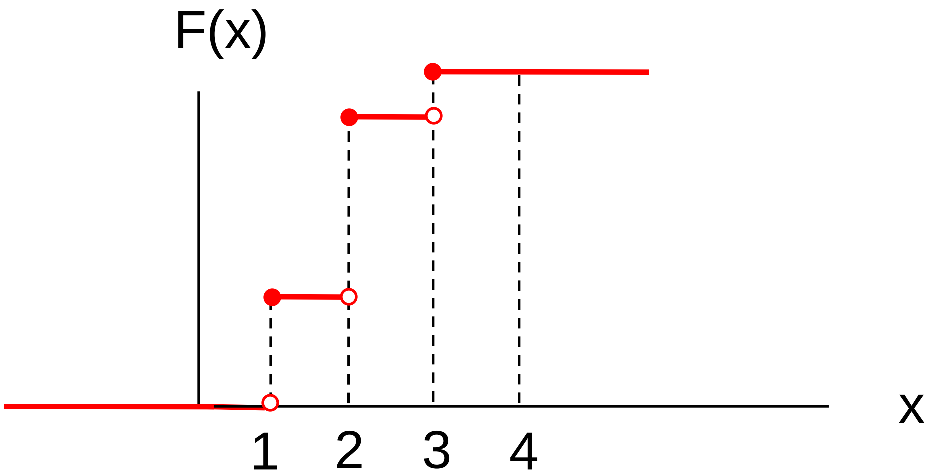
$$F(1) = 0.3$$

$$F(2) = 0.8$$

$$F(3) = 0.9$$

$$F(4) = 1.0$$

$$F(5) = 1.0$$



Example 10:

We can describe an unstable particle, that decays with a lifetime τ. In this case total probability of finding the particle in space is not a constant and decrease with time (*t*):

$$P(t) = \int_{-\infty}^{+\infty} |\Psi(x,t)|^2 \, dx \propto e^{-t/\tau}$$

(a) Determine cdf and pdf.

(b) Determine mean lifetime of the particle.

(c) For muons at τ = 2.2 μs. What is the probability that a muon can survive 200 μs

i. if it is at rest?

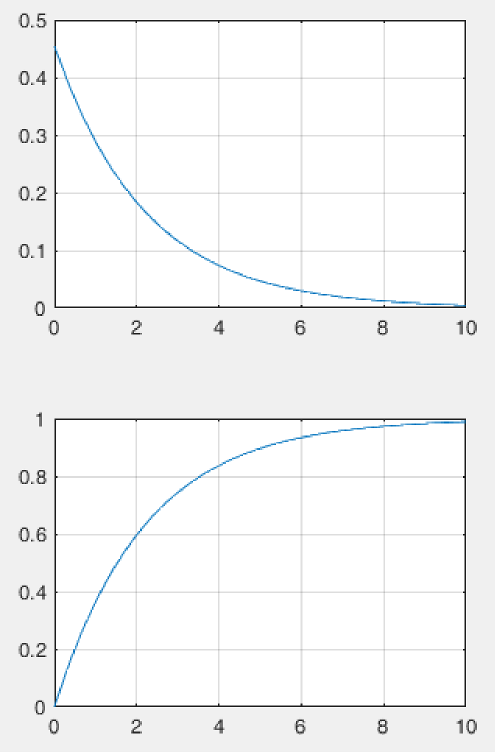
ii. if it moving at p = 10 GeV/c?

iii. if it moving at p = 100 GeV/c?

Answers:

(a) $f(t) = \frac{1}{\tau} e^{-t/\tau}$ $\quad F(t,\tau) = 1 - e^{-t/\tau}$

(b) $< t >= \tau$



(c) τ = 2.2 μs, t = 1 ms = 200 us

$$P_{survive} = 1 - F(t)$$

$$\tau' = \frac{\tau}{\sqrt{1-\beta^2}} = \frac{\tau}{\sqrt{1 - p/\sqrt{p^2+m^2}}}$$

i. $P_{survive}(t) = 0$ [p = 0]

ii. $P_{survive}(t) = 0.3815$ [p = 10 GeV]

iii. $P_{survive}(t) = 0.9081$ [p = 100 GeV]