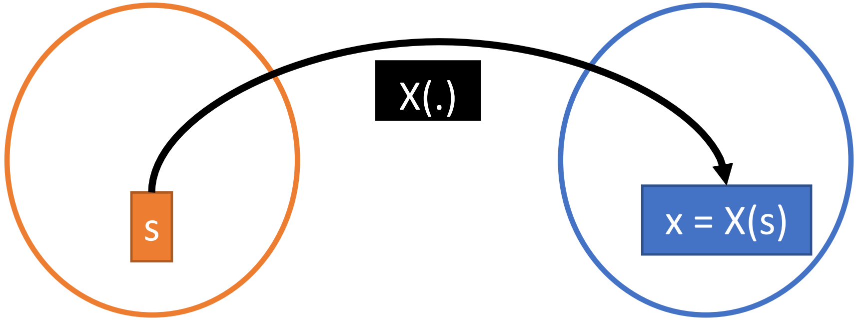
CS 215
Data Analysis and Interpretation

Random Variables

Suyash P. Awate

Random Variable

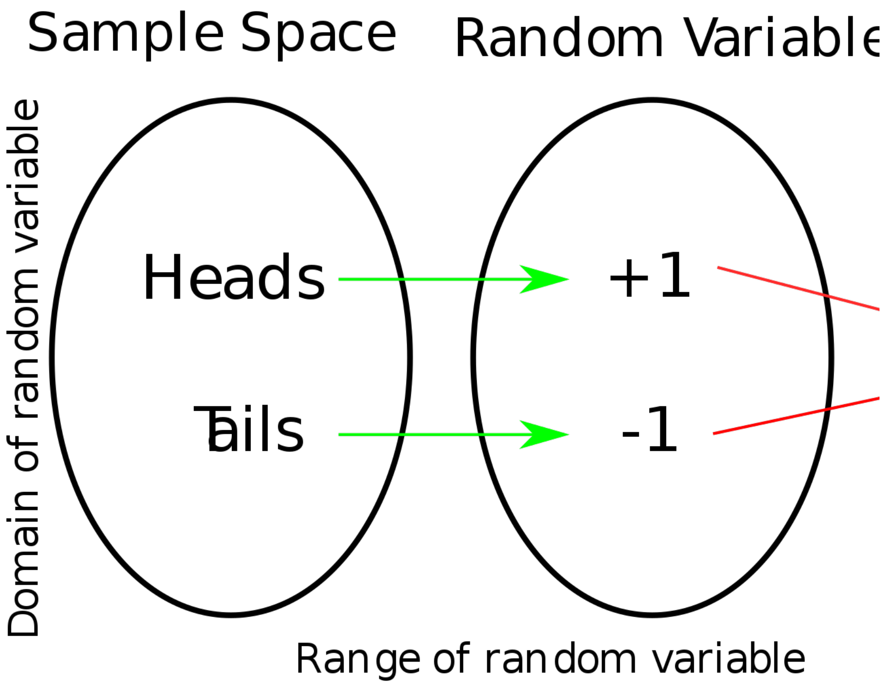
- Definition:
  Random variable X is a function defined on a probability space {Ω,ß,P}. Function X:Ω→R, maps each element in sample space Ω to a single numerical value belonging to the set of real numbers
- The name "random variable" is a misnomer
  - Random variables are functions



- Random variable is an abstraction
  - Used when we are more interested in analyzing the value associated with an outcome, rather than the outcome itself
  - In this way, the range of function X becomes the "sample space" of interest, induced by X(.)
  - X(.) also induces a probability function, say, $P_X(.)$, associated with the range of values that X takes

Random Variable

- How can we design sample space Ω, function X(.) for the following ?
  - Example
    - Coin toss



  - Example
    - Roll two dice. We want to analyse if the sum of the numbers is greater than 5.
  - Example
    - Take a random pixel location (say, (i,j)) in an image.
      We want to analyse if the pixel intensity (in a grayscale/black-and-white image) is 20.

Discrete Random Variable

- The definition will come soon. First, we build intuition.
- Values taken by the random variable form a discrete set
- Cardinality of the set of possible values is finite or countably infinite
  - Sample space is usually a discrete (countable) set, but not necessarily
- Example
  - Roll a die. Take number on top face (finite sample space).
- Example
  - Number of packets transmitted by a router on a network, in unit time
- Example
  - Number of photons hitting the detectors/sensors in a digital camera

Continuous Random Variable

- The definition will come soon. First, we build intuition.
- Random variable takes values anywhere within intervals on real line
- Cardinality of the set of possible values is uncountably infinite
  - Cardinality of the sample space must be uncountably infinite
- Example
  - Heights of children in school
- Example
  - Widths of leaves of a mango tree
- Example
  - Real-valued measurements in an image scan

Events via Random Variables

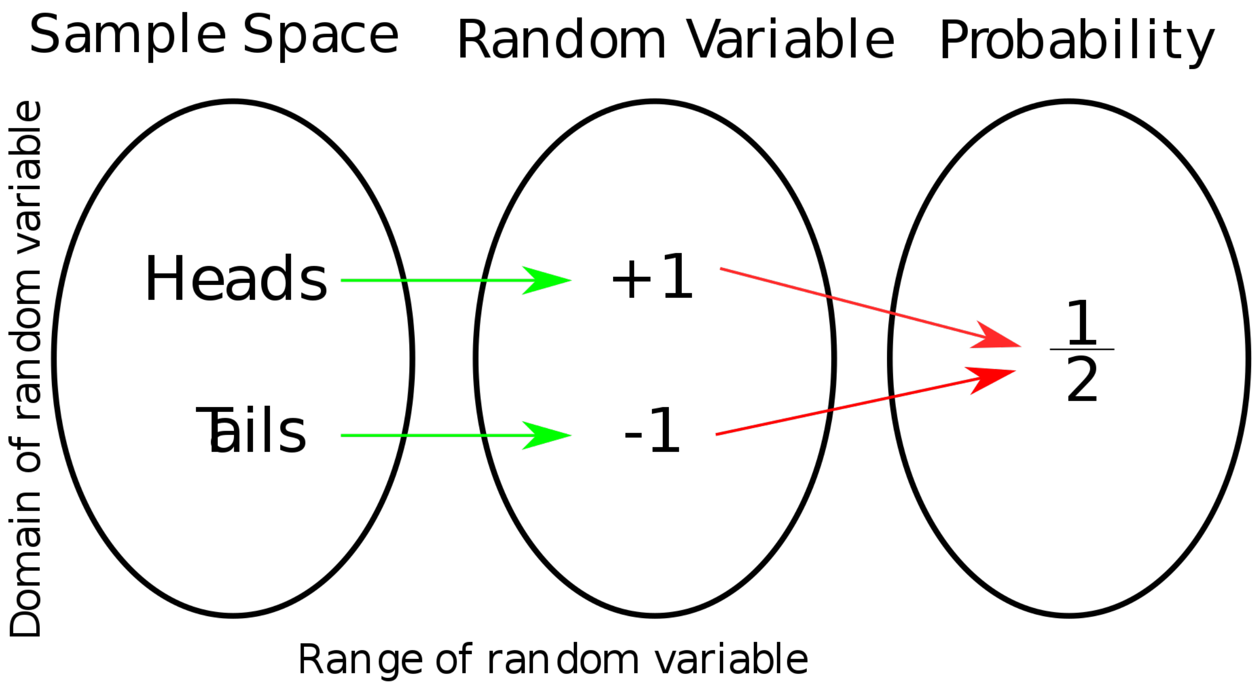
- Notation (legacy)
  - Upper case (e.g., X): random variable
    - So, X has an associated distribution of values
  - Lower case (e.g., x): a value taken by the random variable
    - So, x doesn't have an associated distribution
- Example events:

$$\{X = a\} = \{s \in \Omega : X(s) = a\}$$
$$\{X < a\} = \{s \in \Omega : X(s) < a\}$$
$$\{a < X < b\} = \{s \in \Omega : a < X(s) < b\}$$

- Example events
  - Consider a 8-bit grayscale image with intensities from 0 to 255
  - Event A = "pixel intensity is 100"
  - Event B = "pixel intensity between 100 and 200"

Event Probabilities via Random Variables

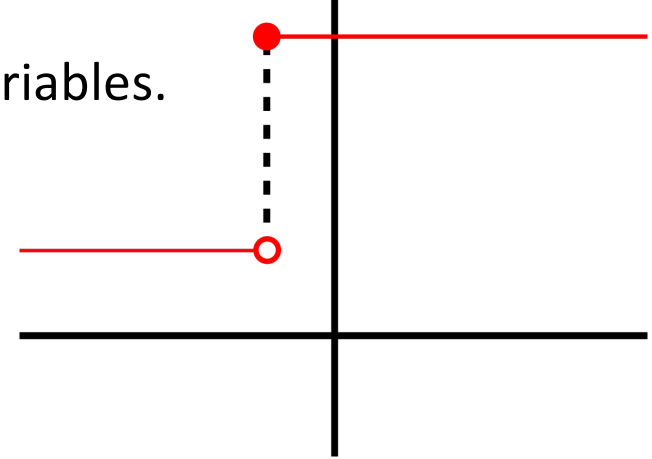
- Example: $P_X(\{a<X<b\})$:= $P(\{a < X < b\}) = P(\{s \in \Omega : a < X(s) < b\})$
  - For simplicity of notation, we can refer to the induced $P_X(.)$ by $P(.)$
- Example:



- Example
  - Consider a 8-bit grayscale image with intensities from 0 to 255
  - P("pixel intensity between 100 and 200")

Cumulative Distribution Function (CDF)

- Definition:
  For a real-valued random variable X, the **CDF** $f_X(x) := P_X(X \leq x)$
  - For all x, value $f_X(x) \in [0,1]$. We will omit the subscript, at times, for simplicity.
- Properties of CDFs: (1) f is monotonically non-decreasing
  (2) f is right continuous. $\lim_{\epsilon \to 0^+} f(x+\epsilon) = f(x), \forall x \in \mathbb{R}$
  (3) $\lim_{x \to -\infty} f(x) = 0$
  (4) $\lim_{x \to +\infty} f(x) = 1$
    - Note: A right-continuous function is where, if we approach a limit point from the right, then there isn't any jump discontinuity.
      This concept is especially relevant for discrete random variables.
  - $P(a<X\leq b) = f_X(b) - f_X(a)$
    - Proof follows from $\{-\infty<X\leq b\} = \{-\infty<X\leq a\} \cup \{a<X\leq b\}$
  - $P(X=c) = f_X(c) - f_X(c-)$
    - For proof, see Hogg-McKean-Craig



Discrete Random Variable. Probability Mass Function.

- Definition:
  A random variable X is "**discrete**" if
  the cardinality of the set of values X takes is countable,
  i.e., either finite or countably infinite
- Definition: For a discrete random variable the **probability mass function** (PMF) is the function $P_X(.)$,
  where $P_X(a) = P_X(X=a)$
- Properties of PMFs:
  - For all x, $0 \leq P_X(x) \leq 1$
  - $\sum_x P_X(x) = 1$

Continuous Random Variable. Probability Density Function.

- Definition:
  A random variable X is "**continuous**" if
  its CDF $f_X(x)$ is a continuous function (i.e., without jumps) for all $x \in R$
- Properties
  - For a continuous RV, $P(X=c) = f_X(c) – f_X(c\text{-})\ 0$, for all c

Continuous Random Variable. Probability Density Function.

- We can write the CDF as an integral of another function $P_X(.)$ called the **probability density function** (PDF)
  - $f_X(c) = \int_{-\infty}^{c} P_X(t)dt$
  - We assume 'absolute' continuity of CDF
    - No weird cases like Cantor function (devil's staircase), that isn't integral of its derivative
  - Thus, $P(a < X \leq b) = \int_a^b P_X(t)dt = P_X(a < X < b) = P_X(a \leq X \leq b) = P_X(a \leq X < b)$
- Because the CDF is non-decreasing, the PDF $P_X(x) \geq 0$, for all x
- Because the CDF evaluates to 1 at $+\infty$, the PDF integrates to 1 over $(-\infty, +\infty)$
- **Support** of a discrete/continuous random variable X comprises all points x for which PMF/PDF $P_X(x) > 0$

“Distribution”

- “Distribution” is an over-used/over-loaded term
  - It usually means PMF/PDF
    - As per https://en.wikipedia.org/wiki/Probability_distribution,
      - Usually, “a **probability distribution** is the mathematical function that gives the probabilities of occurrence of different possible **outcomes** for an experiment.”
    - https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture1.pdf
    - https://www.itl.nist.gov/div898/handbook/eda/section3/eda361.htm
  - But then the word “distribution” is part of the phrase “CDF”

Random Variables

- Discrete and Continuous

Discrete Random Variable: Examples

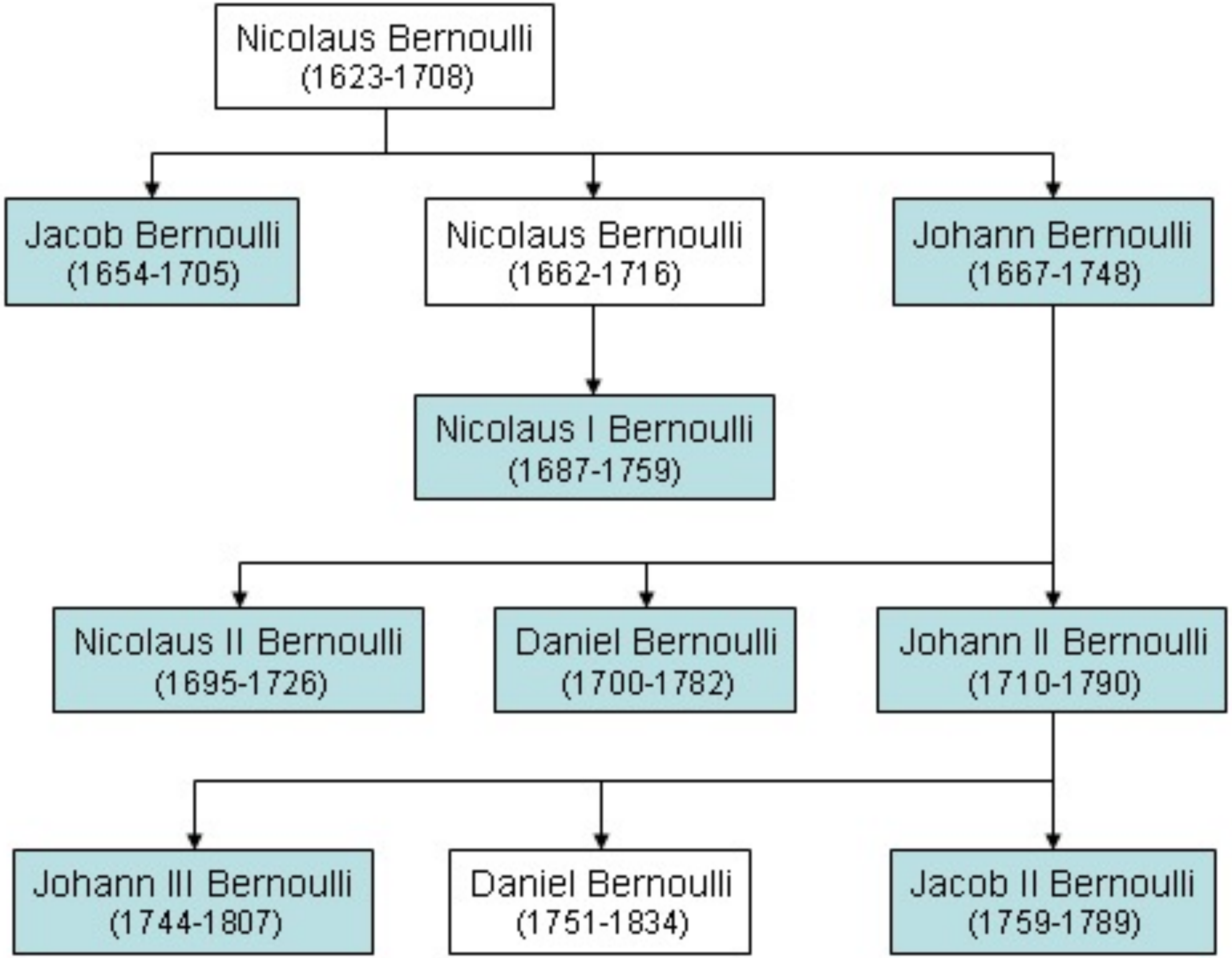
- **Bernoulli** Distribution
  - Random variable X can model success/failure at a task
  - $P(X=1; \alpha) = \alpha$, where $0 \leq \alpha \leq 1$
    - Notation:
      $g(x; \theta)$ means $g(.; .)$ is a function of variable x, and
      $g(.; .)$ is parametrized by constant θ.
      θ is called the parameter.
  - $P(X=0; \alpha) = 1-\alpha$
  - e.g.,
    (X=1) can model success in getting a head on a coin toss
    (X=0) can model failure in getting a head on a coin toss

Bernoulli

- Jacob Bernoulli
  - Swiss mathematician
  - Law of large numbers
  - Discovered ‘e’
    - While studying compound interest
      $$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$
    - Interest of R% per year,
      when compounded continuously,
      yields exp(R/100) after 1 year

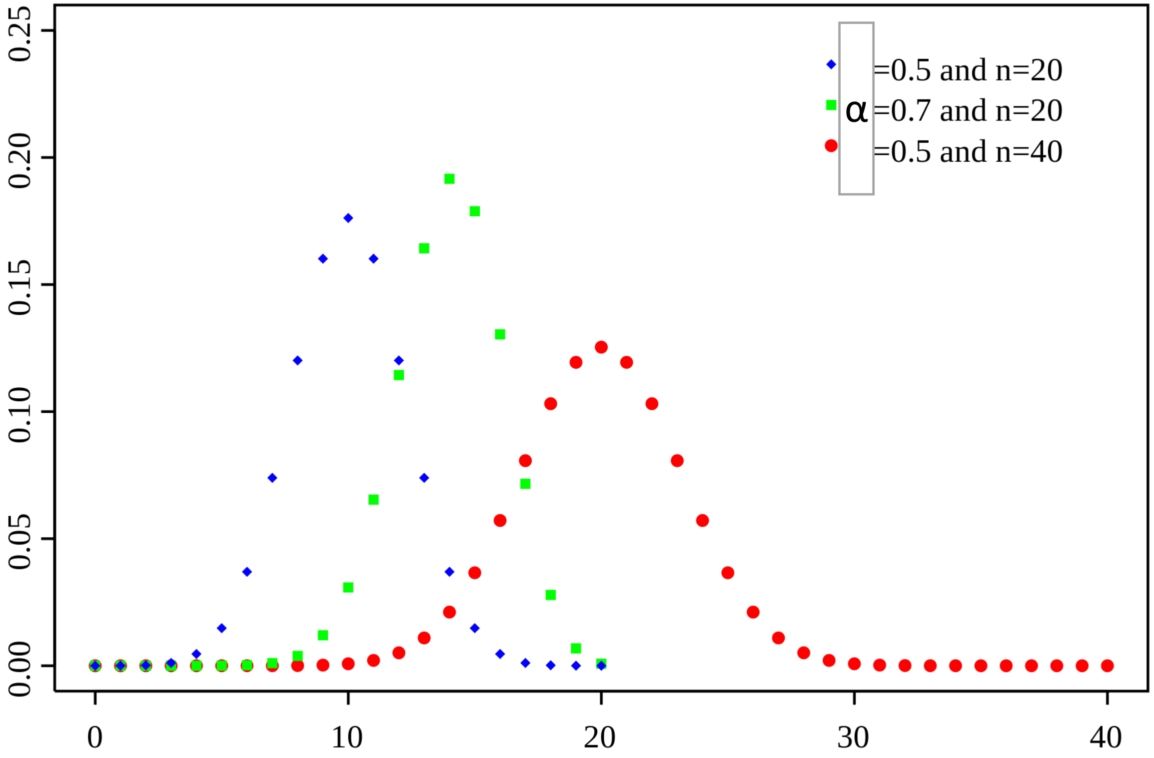
Bernoulli

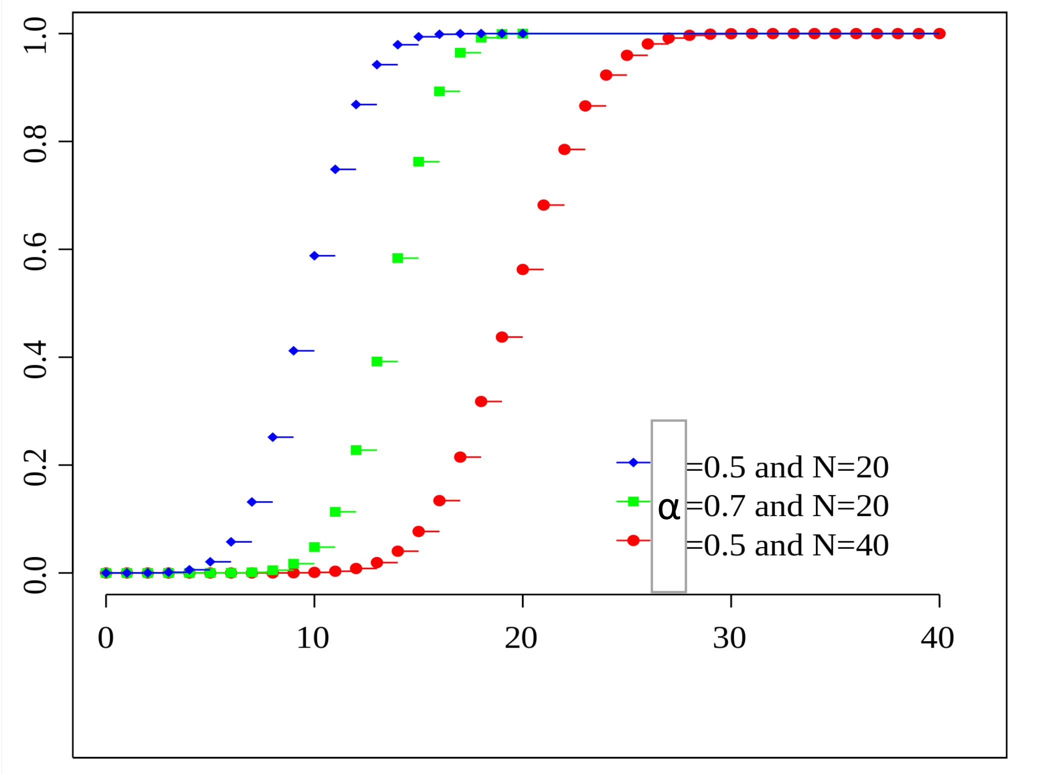
- Bernoulli family
  - Among the greatest in the history of mathematics
  - https://www.jstor.org/stable/27958158



Discrete Random Variable: Examples

- **Binomial** Distribution: Repeated Bernoulli Trials
  - Random variable X can model **number of successes** observed in multiple tries
  - Probability of getting k successes from n trials is $P(X=k;\alpha,n) = {}^{n}C_{k}\, \alpha^{k}\, (1-\alpha)^{(n-k)}$
  - Left: PMF. Right: CDF.





Discrete Random Variable: Examples

- Binomial Distribution: Repeated Bernoulli Trials
  - Check that the sum of all values in the PMF is 1
    - We can apply the binomial theorem: $$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

      with x:=α and y:=1-α

      to the binomial PMF P(X=k;α,n) = $^{n}C_{k}\ \alpha^{k}\ (1-\alpha)^{(n-k)}$

Discrete Random Variable: Examples

- **Geometric** Distribution: Repeated Bernoulli Trials
  - Definition 1:
    Random variable **X** models number of **trials** made to get the **first success**
    - **P(X=k) := P(first success in trial k) := $(1–\alpha)^{(k–1)}\alpha$**, for k = 1,2,3,…,∞
  - Definition 2 (NOT identical to definition 1):
    Random variable **Y** models number of **failures before first success**
    - **P(Y=k) := P(k failures before first success) := $(1–\alpha)^{k}\alpha$**, for k = 0,1,2,…,∞
  - Note
    - P(X=1) = P(Y=0). In general, P(X=n) = P(Y=n–1), for n = 1, 2, 3, …, ∞
    - $\sum_{k=1,2,3,\ldots,\infty}$ P(X=k) = 1
    - $\sum_{k=0,1,2,\ldots,\infty}$ P(Y=k) = 1

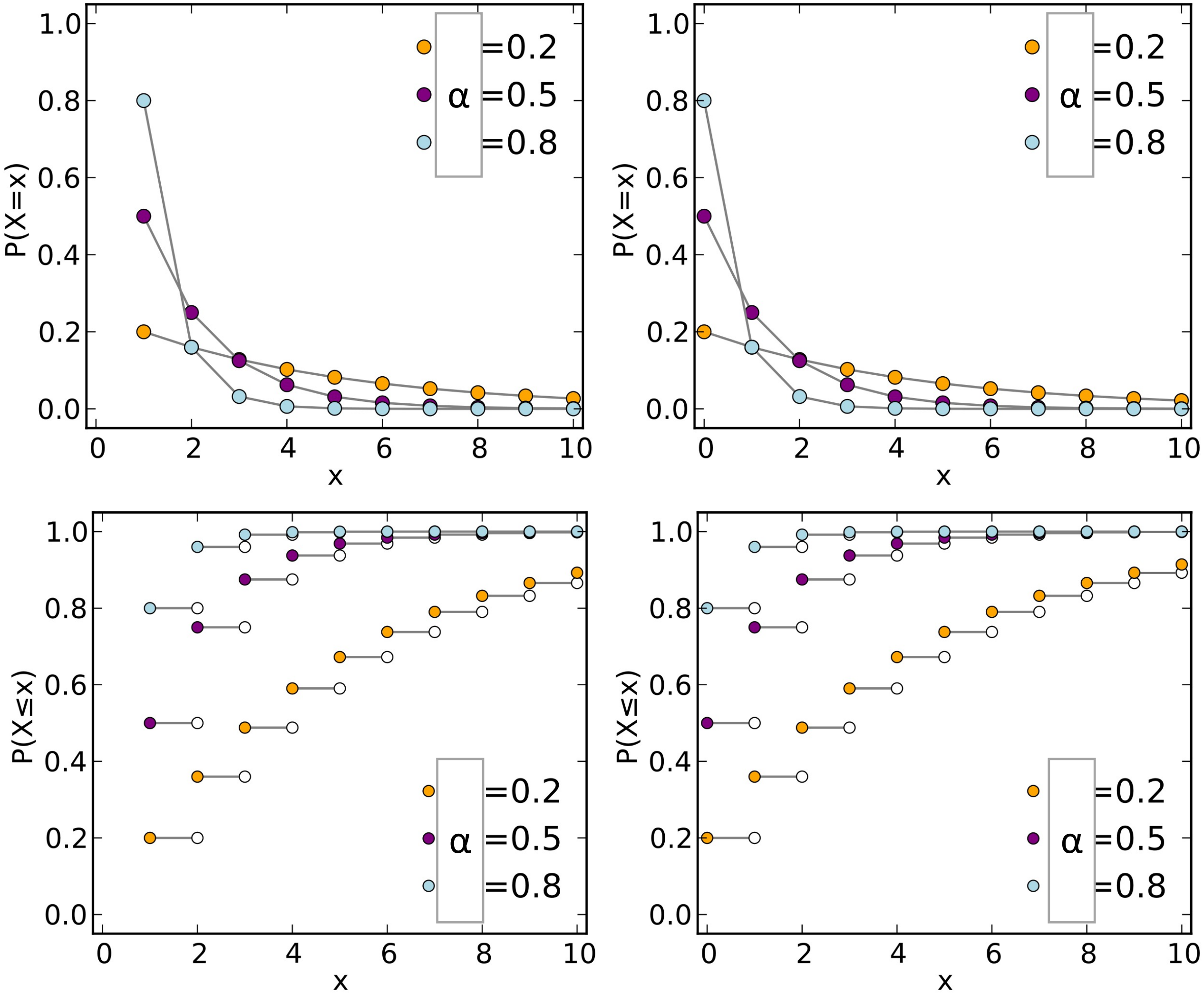
Discrete Random Variable: Examples

- **Geometric** Distribution: Repeated Bernoulli Trials
  - **Cumulative distribution function**
  - Definition 1
    - CDF f(k) = P(X ≤ k) = 1 – P(X > k)
      - P(first success in trial 1) + P(first success in trial 2) + … + P(first success in trial k)
    - $P(X > k) = 1 - [\alpha + \alpha(1-\alpha) + \cdots + \alpha(1-\alpha)^{k-1}] = 1 - \alpha\,\frac{1-(1-\alpha)^k}{1-(1-\alpha)} = (1-\alpha)^k$
    - Then, **CDF f(k) = 1 – $(1–\alpha)^k$**
  - Definition 2
    - CDF f'(k) = P(Y ≤ k) = 1 – P(Y > k)
      - P(0 failures before first success) + … + P(k failures before first success)
    - $P(Y > k) = 1 - [\alpha + \alpha(1-\alpha) + \cdots + \alpha(1-\alpha)^{k}] = 1 - \alpha\,\frac{1-(1-\alpha)^{k+1}}{1-(1-\alpha)} = (1-\alpha)^{k+1}$
    - Then, **CDF f'(k) = 1 – $(1–\alpha)^{k+1}$**

$$a + ar + ar^2 + ar^3 + \cdots + ar^n = \sum_{k=0}^{n} ar^k = a\left(\frac{1-r^{n+1}}{1-r}\right)$$

Discrete Random Variable: Examples

- Geometric Distribution: Repeated Bernoulli Trials
  - Left column: Definition 1
  - Right column: Definition 2
  - Top row: PMFs
  - Bottom row: CDFs



Discrete Random Variable: Examples

- Geometric Distribution: Repeated Bernoulli Trials
  - **Definition 1**
  - CDF: $f(k) = P(X \leq k) = 1 - P(X > k) = 1 - (1-\alpha)^k$
  - **Memoryless Property: $P(X > k+m \mid X > k) = P(X > m)$**
    - Intuition
      - Suppose, for person A, first k trials don't give success
      - Then, for person A, probability that it takes at least m more trials to get success is the same as
        probability, for person B, performing first m trials and not getting success
    - Proof:
      - $P(X > k+m \mid X > k)$

      $= P(X > k+m, X > k) / P(X > k)$

      $= P(X > k+m) / P(X > k)$

      $= (1-\alpha)^{k+m} / (1-\alpha)^k$

      $= (1-\alpha)^m$

      $= P(X > m)$

Discrete Random Variable: Examples

- Geometric Distribution: Repeated Bernoulli Trials
  - **Definition 2**
  - CDF: $f'(k) = P(Y \leq k) = 1 - P(Y > k) = 1 - (1-\alpha)^{k+1}$
  - **Memoryless Property: $P(Y \geq k+m \mid Y \geq k) = P(Y \geq m)$**
    - Intuition
      - Suppose, for person A, at least first k trials were failures
      - Then, for person A, probability that at least m more trials were failures is the same as probability, for person B, probability of at least first m trials were failures
    - Proof:
      - $P(Y \geq k+m \mid Y \geq k)$

      $= P(Y \geq k+m, Y \geq k) / P(Y \geq k)$

      $= P(Y > k+m-1) / P(Y > k-1)$

      $= (1-\alpha)^{k+m} / (1-\alpha)^{k}$

      $= (1-\alpha)^{m}$

      $= P(Y > m\text{-}1) = P(Y \geq m)$

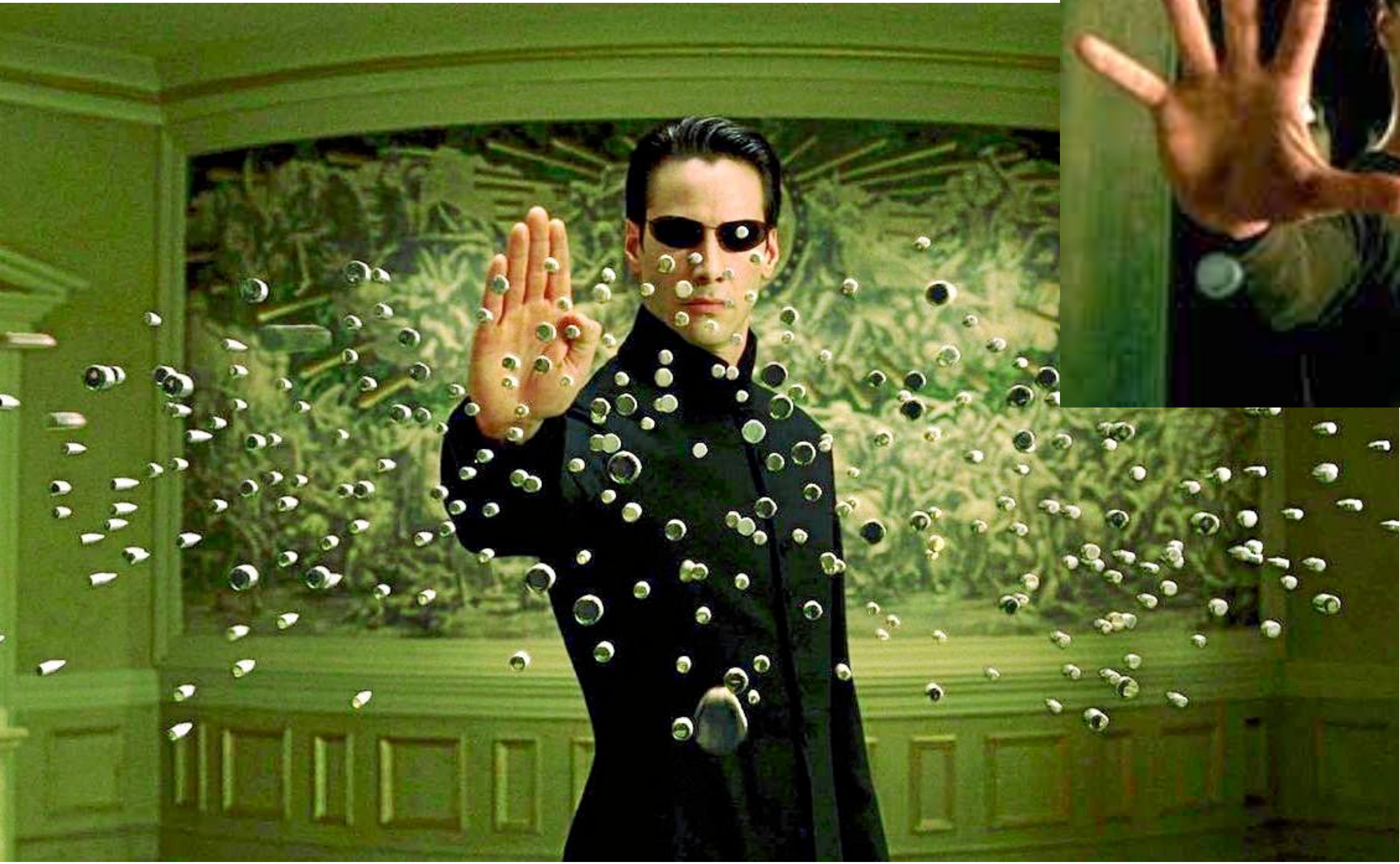
Discrete Random Variable: Examples

- **Poisson** Distribution
  - Consider random events/arrivals/hits occurring at a constant average rate $\lambda>0$, i.e., λ arrivals/hits (typically) per unit time
  - Examples
    - Vehicles passing an intersection
    - Photons hitting a sensor inside a photographic camera

  - The arrivals/hits occur independently of each other, i.e., occurrence of new arrivals/hits doesn't depend on time of previous one
  - Poisson random variable X models number of arrivals/hits occurring in unit time
  - $P(X=k; \lambda) = \lambda^k e^{-\lambda} / k!$, for $k=0,1,2,\ldots$

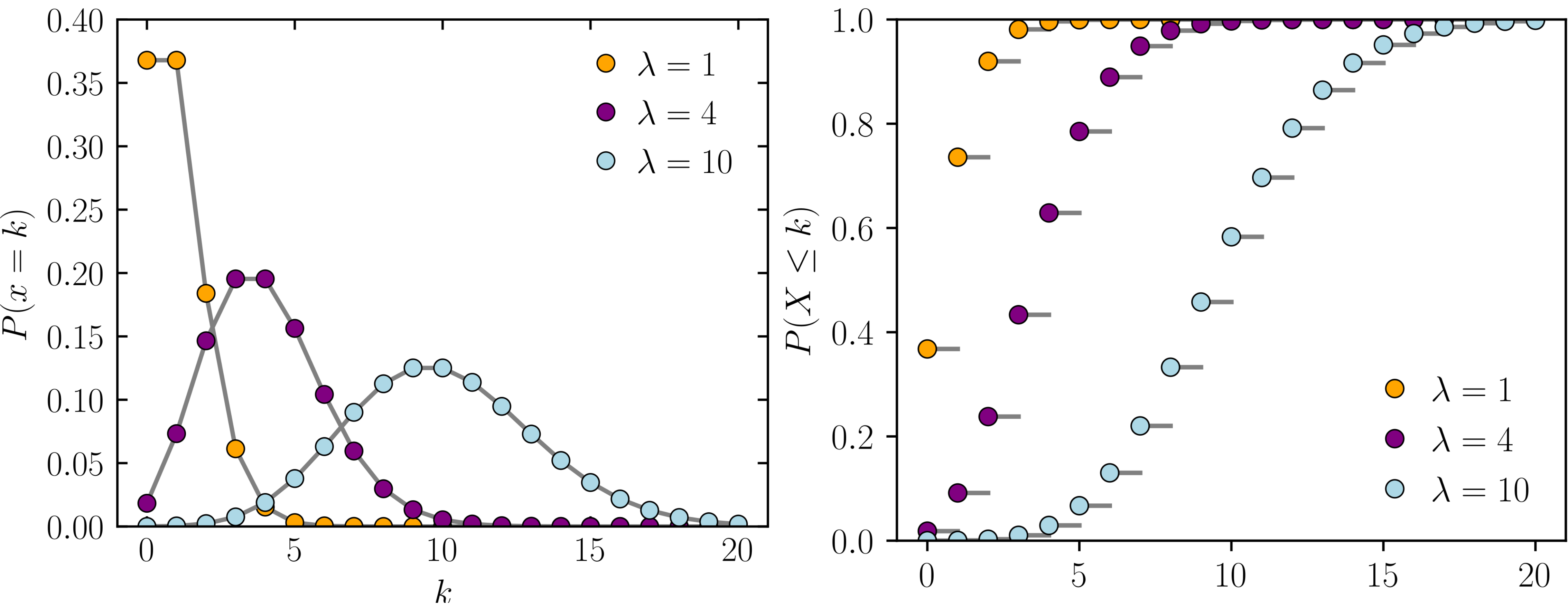
Discrete Random Variable: Example

- **Poisson** Distribution

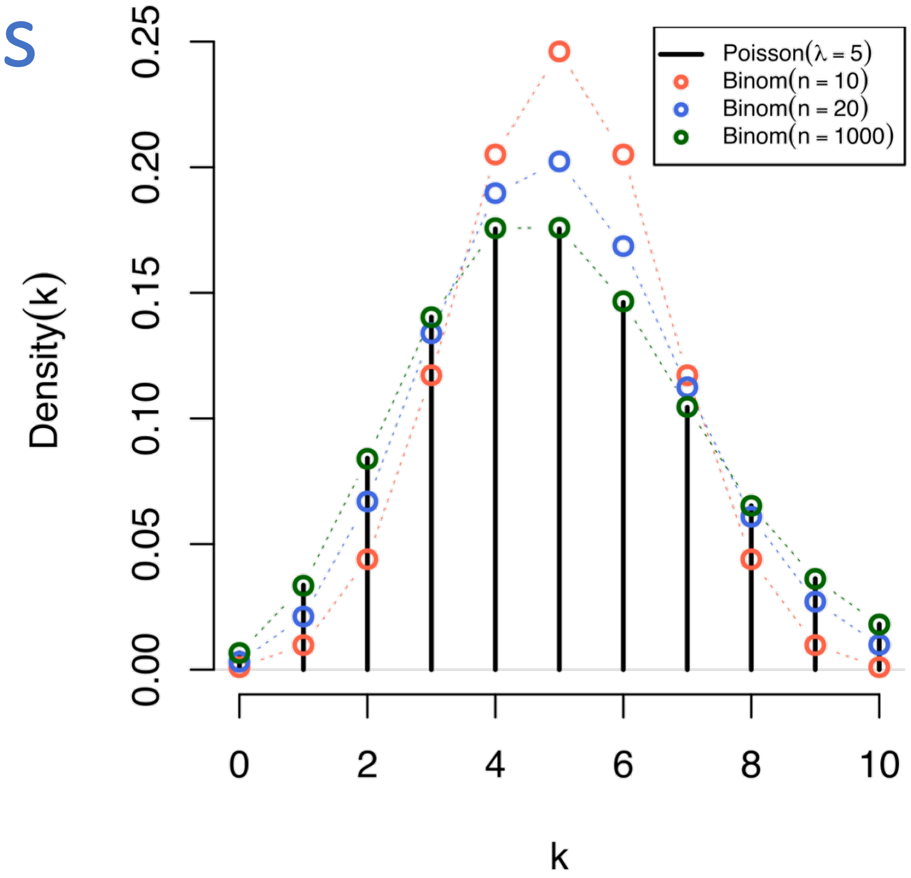


Discrete Random Variable: Examples

- Poisson Distribution



Discrete Random Variable: Examples

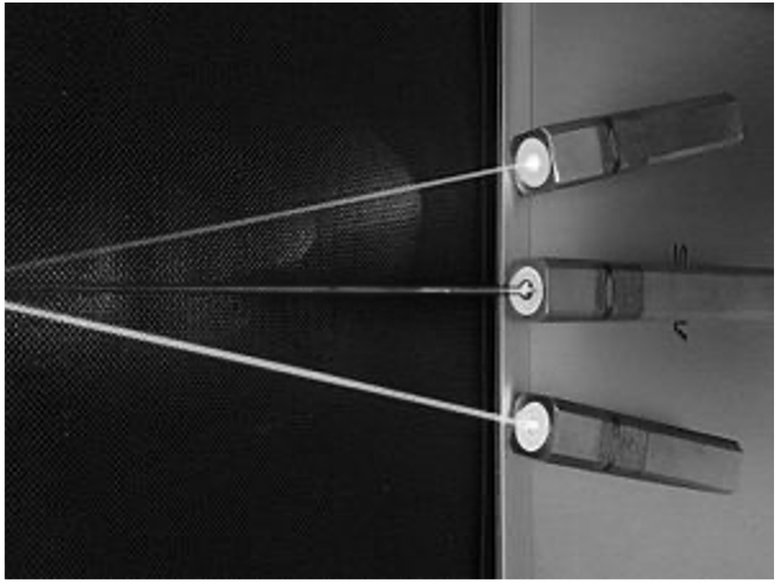


- Poisson Distribution
  - **A limiting case of the Binomial distribution**
    - $\text{Limit}_{n\to\infty} P_{\text{Binomial}}(X=k;\ \alpha=\lambda/n,\ n)$
      $= P_{\text{Poisson}}(X=k;\ \lambda)$
    - Proof:
      - $P_{\text{Binomial}}(X=k;\ \alpha,\ n) = \frac{n!}{(n-k)!k!}\alpha^k(1-\alpha)^{n-k}$
      - $= \frac{n!}{(n-k)!k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$
      - $= \frac{\lambda^k}{k!}\left(\frac{n!}{(n-k)!n^k}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$
      - $= \frac{\lambda^k}{k!}(1)e^{-\lambda}(1)$ when $n\to\infty$
  - Counts “successes” in a unit-length interval by modeling the interval as infinitely-many ($n\to\infty$) infinitesimal-length ($1/n\to 0$) intervals that each have an infinitesimal success probability $\alpha=\lambda/n$ underlying an approx.-Bernoulli PMF
    - This gives the interpretation for parameter $\lambda=n\alpha$ as
      $n*$(probability of success, or “hits”, in an interval). More analysis on this later.

Random Variables

- **Sum of Random Variables**
  - Consider two random variables X and Y,
    and a joint probability space $\{\Omega_{12},ß_{12},P_{12}\}$
    and an associated joint distribution $P_{XY}(x,y)$
  - Defining Z := X + Y means the following:
    - Conduct joint random experiment, underlying the pair (X,Y),
      leading to outcome $(s_1,s_2) \in \Omega_{12}$
    - Get $x := X(s_1,s_2)$ and $y := Y(s_1,s_2)$
    - Define z := x + y
    - So, Z is a random variable, taking values z,
      with an associated distribution $P_Z(z)$
      and a probability space $\{\Omega_{12},ß_{12},P_{12}\}$
    - Note that $P_Z(t)$ isn't necessarily equal to $P_X(t)+P_Y(t)$,
      i.e., adding random variables isn't the same as superposing their PDFs/PMFs

Discrete Random Variable: Examples



- Poisson Distribution
  - **Sums of Independent Poisson Random Variables**
    - Let X~Poisson(λ) and Y~Poisson(μ) be independent
    - Define random variable Z := X + Y
    - Then, P(Z) = P_{Poisson}(Z; λ+μ)
    - Proof:
      - P(Z=k) = $\sum_{j=0}^{k} P(X=j, Y=k-j) = \sum_{j=0}^{k} \frac{e^{-\lambda}\lambda^j}{j!}\frac{e^{-\mu}\mu^{k-j}}{(k-j)!}$

$$= \frac{e^{-(\lambda+\mu)}}{k!}\sum_{j=0}^{k}\frac{k!}{j!\,(k-j)!}\lambda^j\mu^{k-j}$$

$$= \frac{e^{-(\lambda+\mu)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\lambda^j\mu^{k-j}$$

$$= \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^k}{k!}$$

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$$

Discrete Random Variable: Examples

- Poisson Distribution
  - **Poisson Thinning**
    - Let X ~ Poisson(λ)
    - Consider joint PMF P(X,Y), where P(Y|X=j) = $P_{Binomial}$(Y;j,p)
      - This is a "thinning" process that selects a subset of the j arrivals
      - Similar to selecting each of the j arrivals/hits (stochastically) with (Bernoulli) probability p
    - Then, P(Y) = $P_{Poisson}$(Y; λp)
    - Proof:
      - P(Y=k) = $\sum_{j=k}^{\infty} P(X = j, Y = k) =$
      - $\sum_{j=k}^{\infty} P(y = k \mid X = j)P(X = j) =$

$$= \sum_{j=k}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \binom{j}{k} p^k (1-p)^{j-k}$$

$$= e^{-\lambda} \sum_{j=k} \frac{\lambda^j}{j!} \frac{j!}{k!\,(j-k)!} p^k (1-p)^{j-k}$$

$$= \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{j=k} \frac{(\lambda(1-p))^{j-k}}{(j-k)!}$$

$$= \frac{e^{-\lambda}(\lambda p)^k}{k!} e^{\lambda(1-p)}$$

$$= \frac{e^{-\lambda p}(\lambda p)^k}{k!}$$

Poisson

- Siméon Denis Poisson
  - French mathematician, engineer, physicist
  - Advisors
    - Lagrange, Laplace
  - Students (non-PhD)
    - Dirichlet
    - Carnot (father of thermodynamics)
  - One of 72 names inscribed on Eiffel Tower

Poisson

"I don't know if you're a mathematician but
my wife's not happy with her Poisson distribution."

Random Variables

- **Sum of Random Variables**
  - Consider two **independent** random variables X and Y, and a joint probability space $\{\Omega_{12}, ß_{12}, P_{12}\}$ and an associated joint distribution $P_{XY}(X=x, Y=y)$
  - Define $Z := X + Y$. What is $P_Z(Z)$?
  - **Case 1: Discrete** random variables
    - $P_Z(Z = z)$
    $= P_{XY}(\{ (t, z-t) : X = t, Y = z-t \})$
    $= \sum_t P_{XY}(X = t, Y = z-t)$
    $= \sum_t P_X(X = t)\, P_Y(Y = z-t)$
    - This operation, $h(a) := \sum_t f(t)\, g(a-t)$ is the **convolution** of functions $f(t)$ and $g(t)$, producing the function $h(t)$. This is typically written in the notation $h := f*g$
    - Thus, addition of two independent discrete random variables X and Y leads to a third random variable Z whose PMF $P_Z(.)$ is the convolution of the PMFs $P_X(.)$ and $P_Y(.)$

Random Variables

- **Sum of Random Variables**
  - Consider two **independent** random variables X and Y, with PDFs A(.) and B(.)
  - Consider and a joint probability space $\{\Omega_{12}, ß_{12}, P_{12}\}$ and an associated joint distribution $P_{XY}(x,y)$
  - Define Z := X + Y. What is its PDF C(.) ?
  - **Case 2: Continuous** random variables
    - CDF $P(X+Y \leq z) = \int_{x=-\infty}^{\infty} \left[\int_{y=-\infty}^{z-x} p(x,y) dy \right] dx$
    - PDF $C(z) = d/dz\,(P(X+Y \leq z)) = \int_{x=-\infty}^{\infty} \left[\frac{d}{dz} \int_{y=-\infty}^{z-x} p(x,y) dy \right] dx$

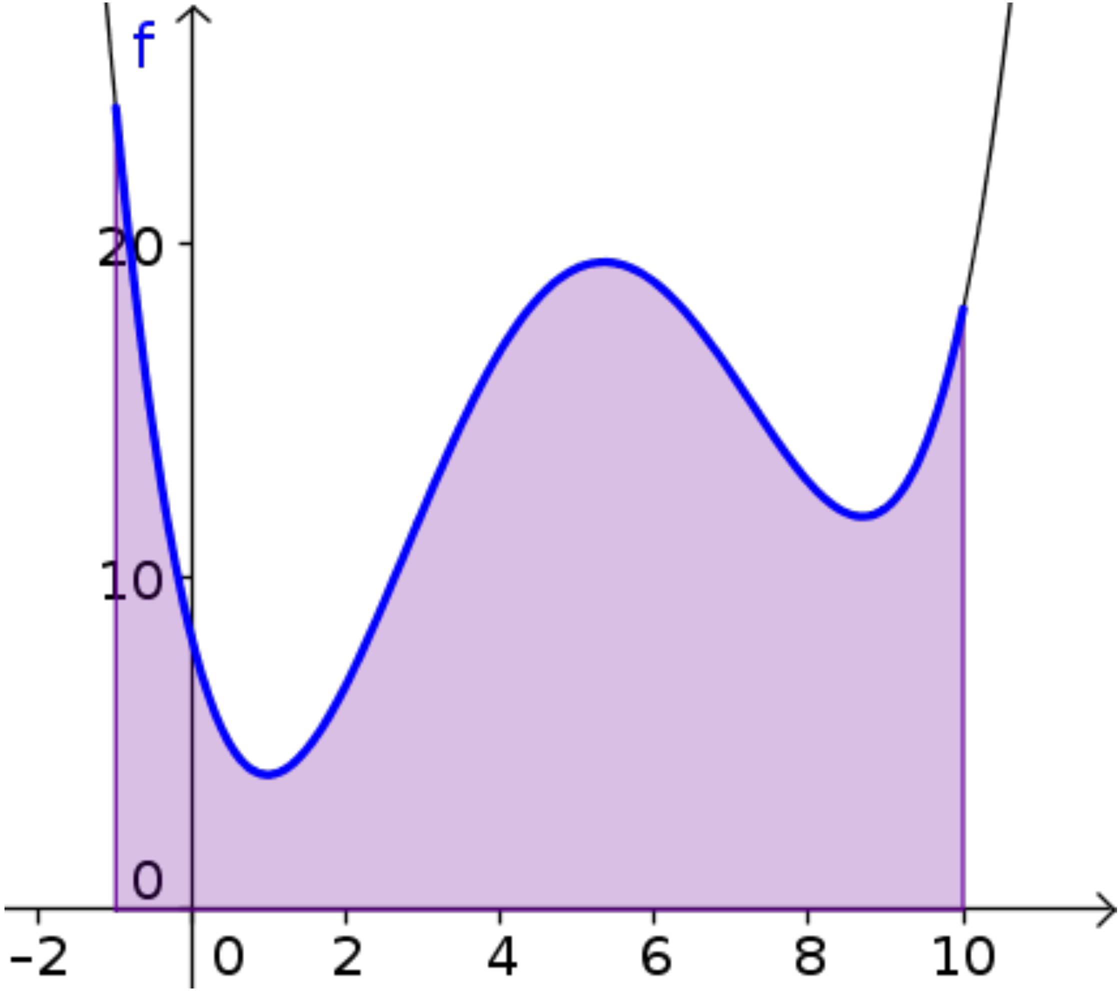
$$= \int_{x=-\infty}^{\infty} p(x, z-x) dx$$

    - Thus, $C(z) = \int_x A(x)B(z-x)dx$ (because of independence of X and Y)
    - This is the convolution operation for functions on the continuous domain (C := A*B)

Random Variables

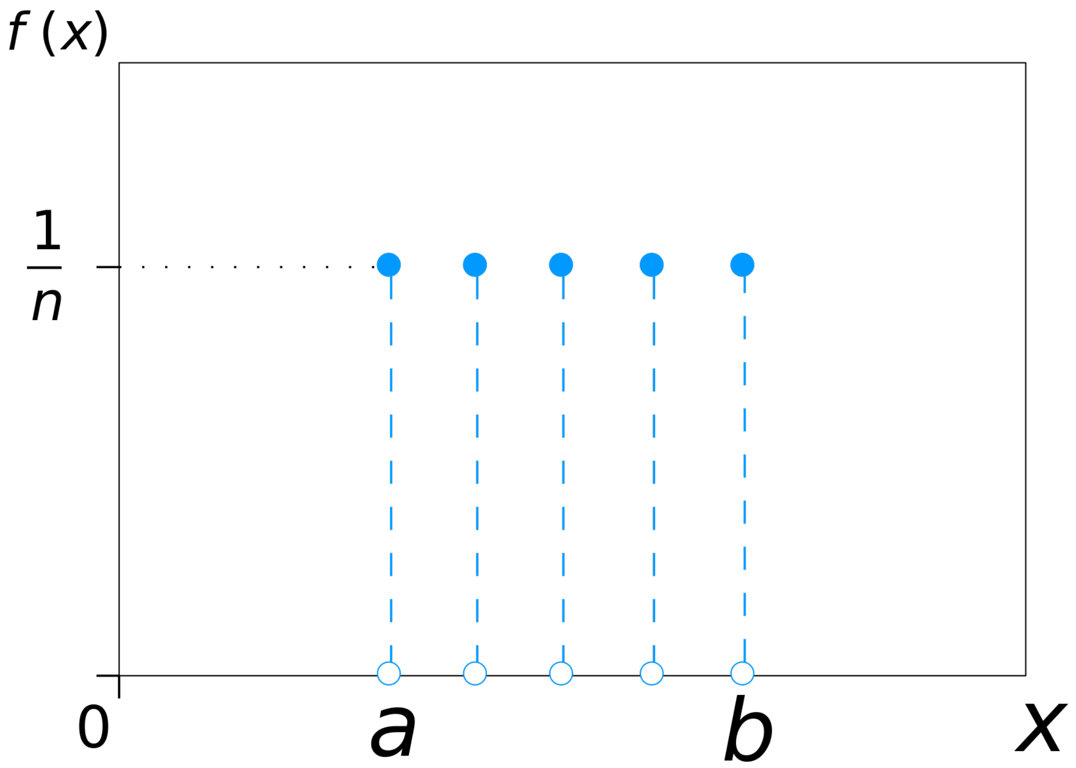
- Leibniz's integral rule

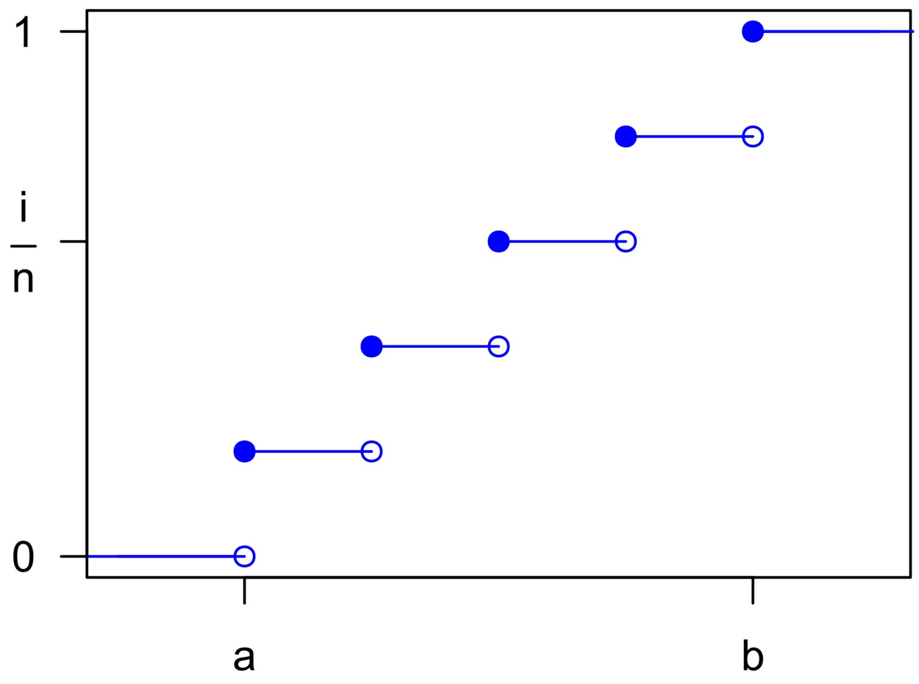
$$\frac{\partial}{\partial a}\int_{l(a)}^{u(a)} f(z,a)dz = \int_{l(a)}^{u(a)} \frac{\partial f}{\partial a}dz + f(z=u(a),a)\frac{\partial u}{\partial a} - f(z=l(a),a)\frac{\partial l}{\partial a}$$



Discrete Random Variable: Examples

- **Uniform** distribution
  - Sample space has finite cardinality (say, n < ∞)
  - Within sample space, each outcome has equal probability (1/n)
  - Examples: Coin toss. Die roll.





Continuous Random Variable: Examples

- **Uniform** Distribution
  - Random variable X takes values within a finite interval that can be closed, i.e., [a,b], or open, i.e., (a,b) where -∞ < a < b < ∞
  - Within [a,b] or (a,b), all sub-intervals of same size/measure have equal probability

  PDF:

  $p(x) = 0, \forall x < a$ or $x > b$

  $p(x) = 1/|b-a|, \forall a \leq x \leq b$

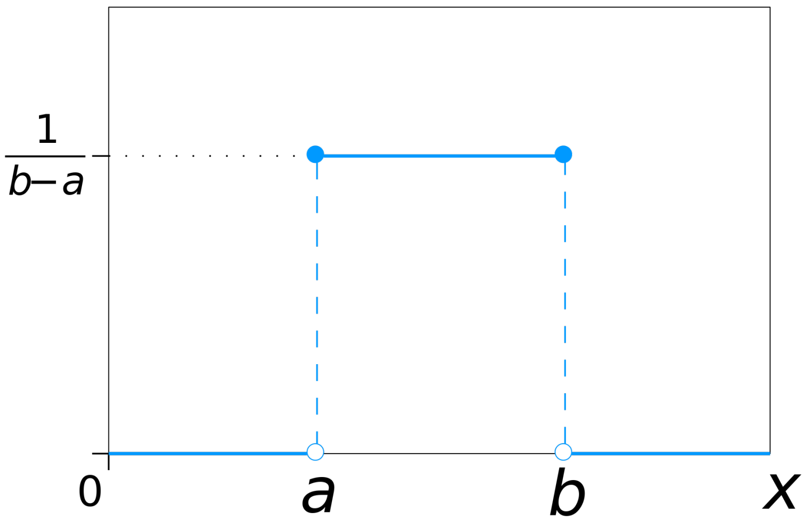
  CDF:

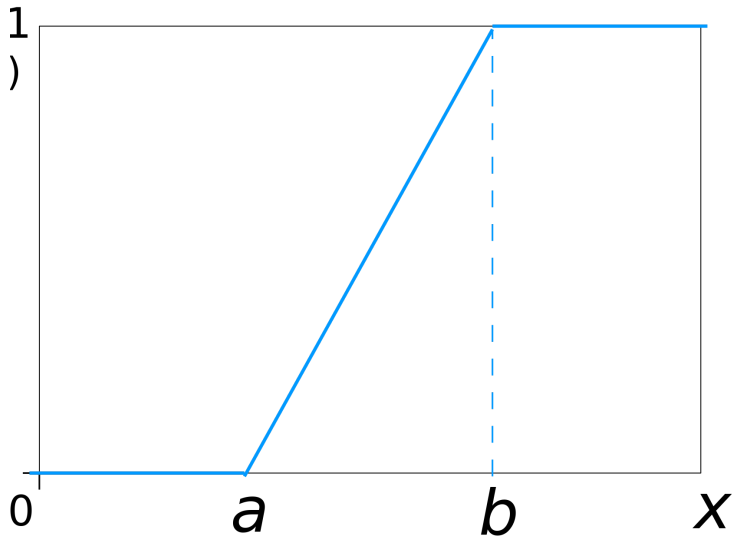
  $f(x) = 0, x < a$

  $f(x) = 1, x > b$

  $f(x) = (x-a)/(b-a), a \leq x \leq b$

  Note: a and b are called parameters.





Continuous Random Variable: Examples

- **Exponential** Distribution
  - Consider random events/arrivals/hits occurring at a constant average rate $\lambda>0$, i.e., λ arrivals/hits (typically) per unit time
  - Example: packets arriving at a router
  - The arrivals/hits occur independently of each other
  - Exponential random variable models the probability density of the **time elapsed between two consecutive** arrivals/hits
  - **PDF**
    - P(x) = 0, for all x < 0
    - P(x) = $\lambda \exp(-\lambda x), \forall x \geq 0$
  - **CDF**
    - f(x) = 0, for all x < 0
    - f(x) = $1 - \exp(-\lambda x), \forall x \geq 0$

Continuous Random Variable: Examples

- Exponential Distribution

PDF

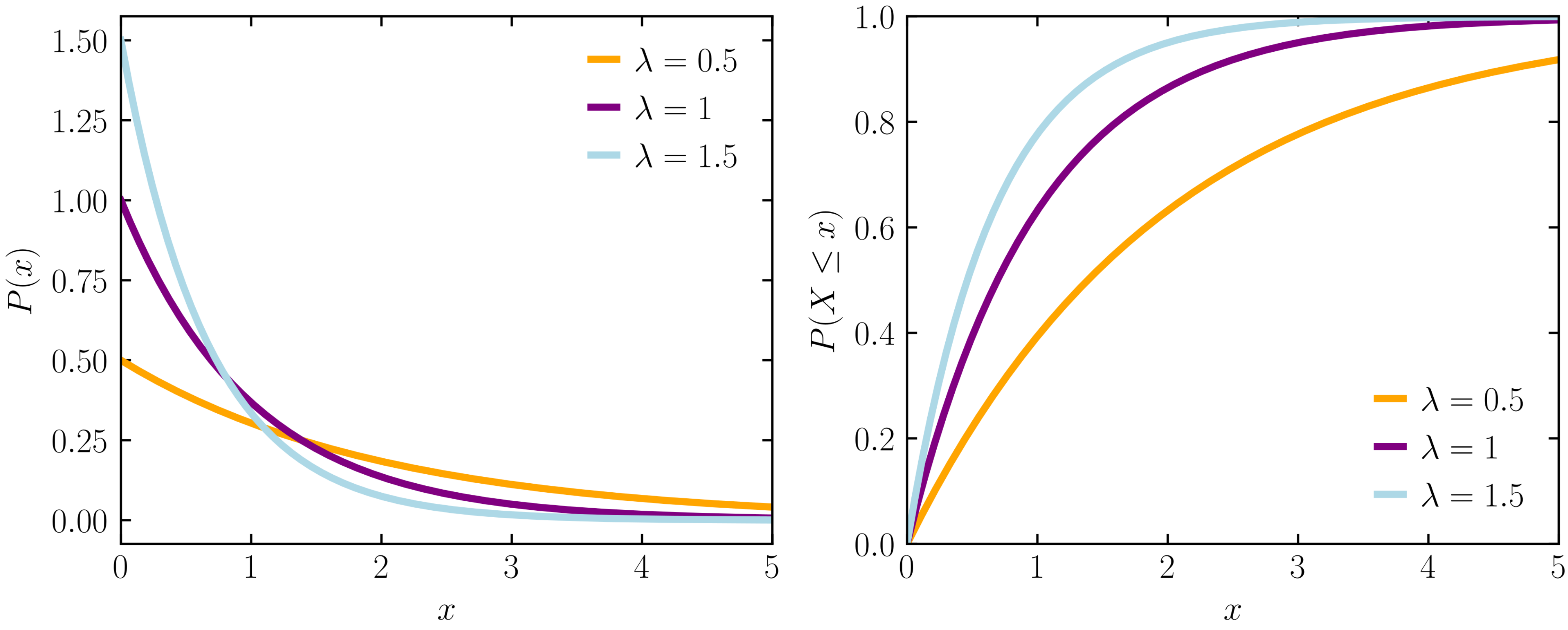
P(x) = 0, for all x < 0

P(x) = $\lambda \exp(-\lambda x), \forall x \geq 0$

CDF

f(x) = 0, for all x < 0

f(x) = $1 - \exp(-\lambda x), \forall x \geq 0$



Continuous Random Variable: Examples

- **Exponential Distribution relates to Poisson Distribution**
  - Let average arrival/hit-occurrence rate be λ per unit time
    - Let random variable Y model the number of arrivals/hits occurring in unit time
    - So, $P_{Poisson}(Y=k; \lambda) = \lambda^k e^{-\lambda} / k!$
    - Also, in time T units, average arrival/hit-occurrence rate will be λT
  - Within time interval [0,t], let N_t := number of arrivals/hits occurred
  - Let random variable X_t model **time taken for next arrival/hit to occur**
  - Consider the case where the next occurrence is at any time after 'x'
  - $P_{exponential}(X_t > x) = P(N_t = N_{t+x})$
  - So, $P_{exponential}(X_t \leq x) = 1 - P(N_t = N_{t+x})$
  - By assumptions underlying the Poisson distribution with parameter λx, $P(N_t = N_{t+x}) = P_{Poisson}(Y=0; \lambda x) = e^{-\lambda x}$
  - Thus, $P_{exponential}(X_t \leq x) = 1 - e^{-\lambda x}$ (this is the **CDF**)
  - Thus, $P_{exponential}(x) = \lambda e^{-\lambda x}$ (this is the **PDF**)

Continuous Random Variable: Examples

- **Exponential Distribution satisfies the memoryless property**
  - Proof:
    - Let X be an exponential random variable, with parameter λ
      - Its CDF is: $P_{exponential}(X_t \le x) = 1 – e^{-\lambda x}$
    - Then, we want to show that $P_{exponential}(X > x+t \mid X > x) = P_{exponential}(X > t)$
    - $P_{exponential}(X > x+t \mid X > x)$

    $= P_{exponential}(X > x+t, X > x) / P_{exponential}(X > x)$

    $= P_{exponential}(X > x+t) / P_{exponential}(X > x)$

    $= e^{-\lambda(x+t)} / e^{-\lambda x}$

    $= e^{-\lambda t}$

    $= P_{exponential}(X > t)$

Continuous Random Variable: Examples

- Exponential Distribution
  - **The only continuous random variable satisfying the memoryless property is the exponential random variable**
  - Proof:
    - Let exponential random variable X have PDF P(.)
    - Define **survival function** $S(x) := P(X > x)$
    - $P(x)dx$

      $\approx P(x < X < x+dx)$

      $= P(X < x+dx \mid X > x)\, P(X > x)$

      $= P(X > x)\, (1 - P(X > x+dx \mid X > x))$

      $= P(X > x)\, (1 - P(X > dx))$ assuming memoryless property

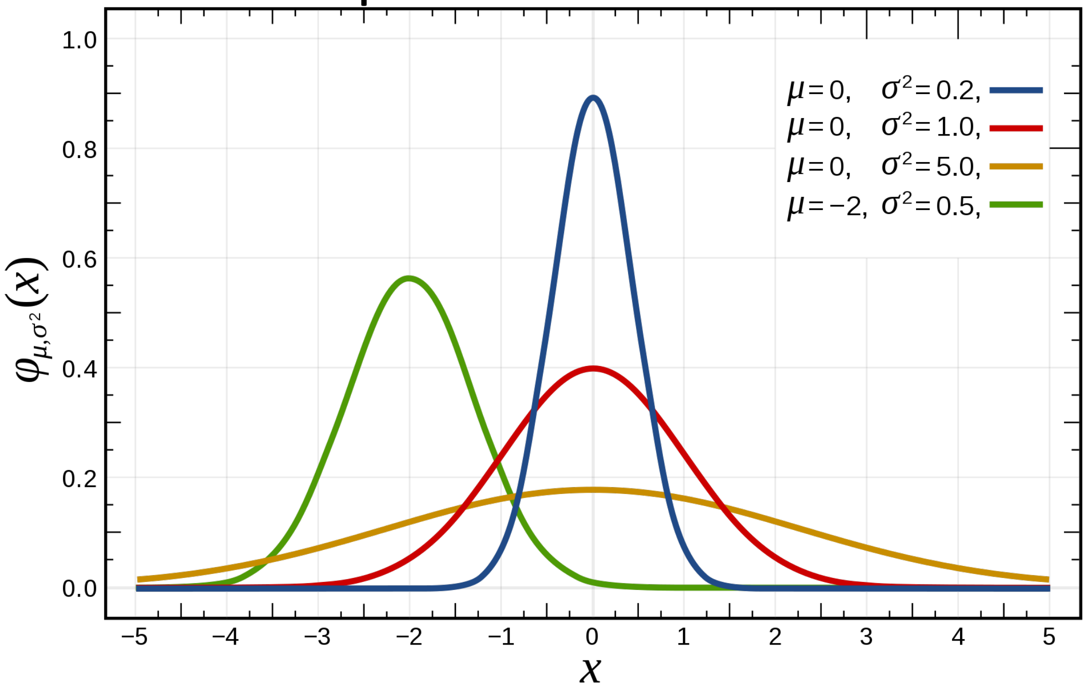
      $= S(x)\, (1 - S(dx))$
    - By Taylor series expansion around x=0, we get $S(dx) \approx S(0) + S'(0)dx = 1 + S'(0)dx$
      - Because $S(0)= P(X > 0) = 1$
    - So, $P(x)dx \approx S(x)\, (-S'(0)dx)$

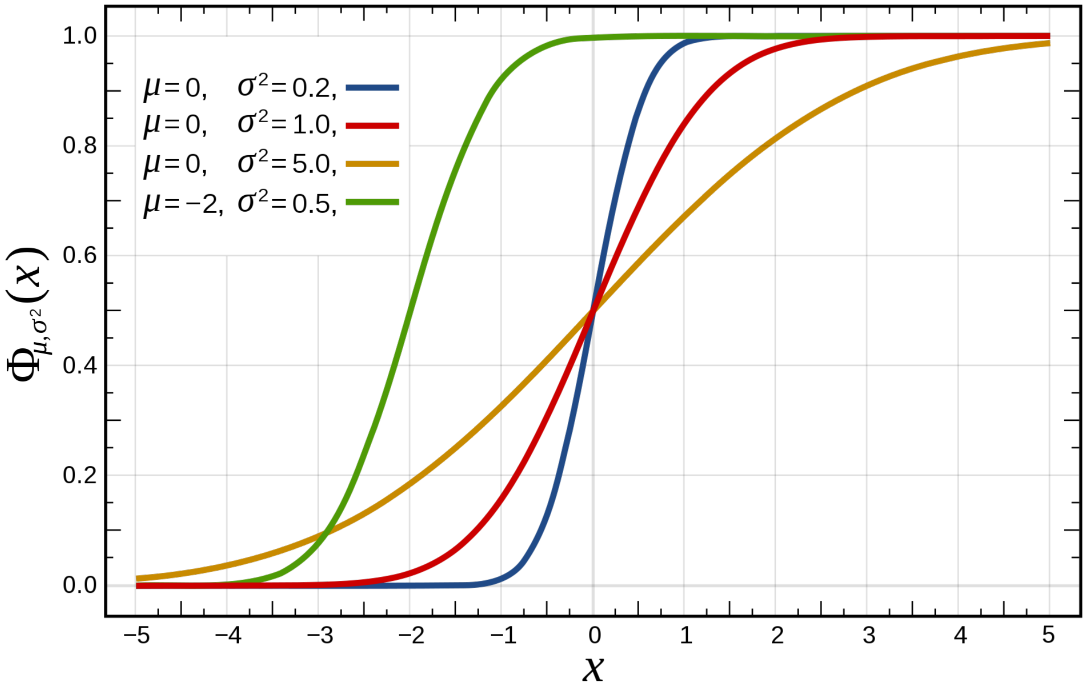
Continuous Random Variable: Examples

- Exponential Distribution
  - The only continuous random variable satisfying the memoryless property is the exponential random variable
  - Proof:
    - Let exponential random variable X have PDF P(.)
    - Define **survival function** $S(x) := P(X > x)$
    - $P(x)dx \approx S(x)\,(-S'(0)dx)$
    - Because $1-CDF_X(x) = S(x)$, we get: $-P(x) = S'(x)$
    - Thus, substituting $-S'(x)$ on the left hand side, we get: $S'(x)dx \approx S(x)\,S'(0)\,dx$
    - So, $S'(x) \approx S(x)\,S'(0)$
    - So, $d/dx\,[S(x)] \approx S'(0)\,S(x)$
    - Given $S(0) = 1$, the solution is given by: $S(x) = \exp(S'(0)\,x)$
    - Defining $\lambda := -S'(0)$, we get $1-CDF_X(x) = \exp(-\lambda x)$, or $P(x) = \lambda \exp(-\lambda x)$
      - λ must be positive for $P(x)$ to be a valid PDF

Continuous Random Variable: Examples

- **Gaussian (Normal) Distribution**
  - PDF is P(x) $= \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$, with parameters μ, σ
  - Why called “normal” ?
    - https://condor.depaul.edu/ntiourir/NormalOrigin.htm
  - PDF shape is called the “bell curve”





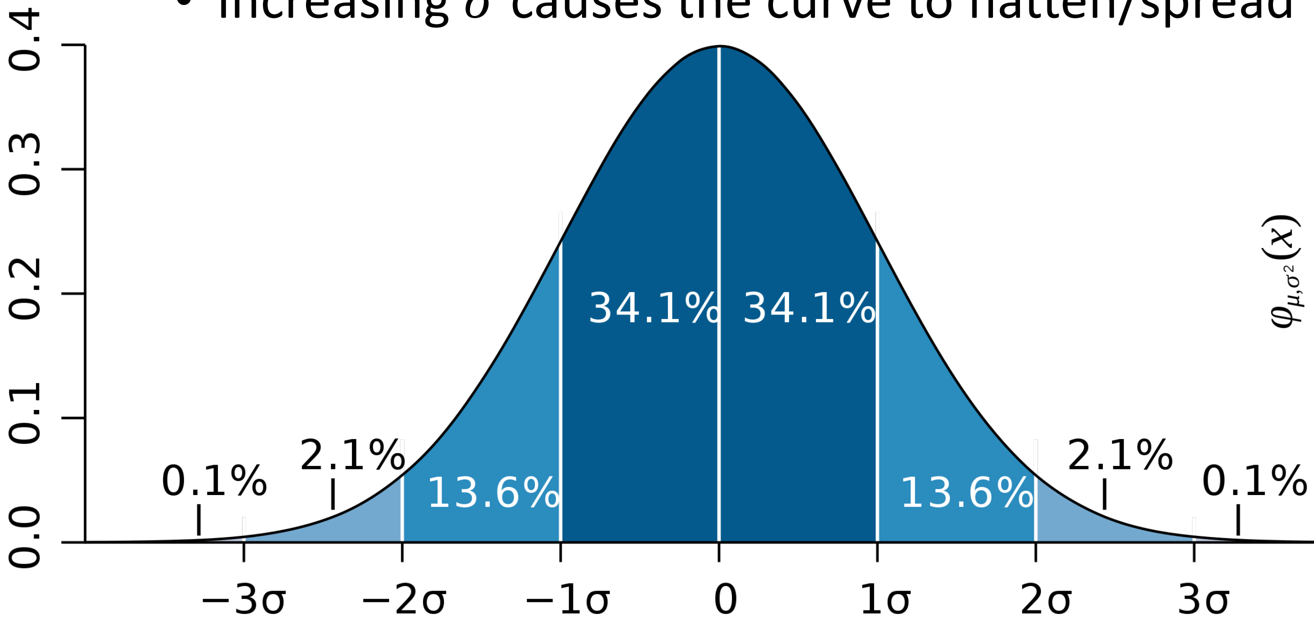
Continuous Random Variable: Examples

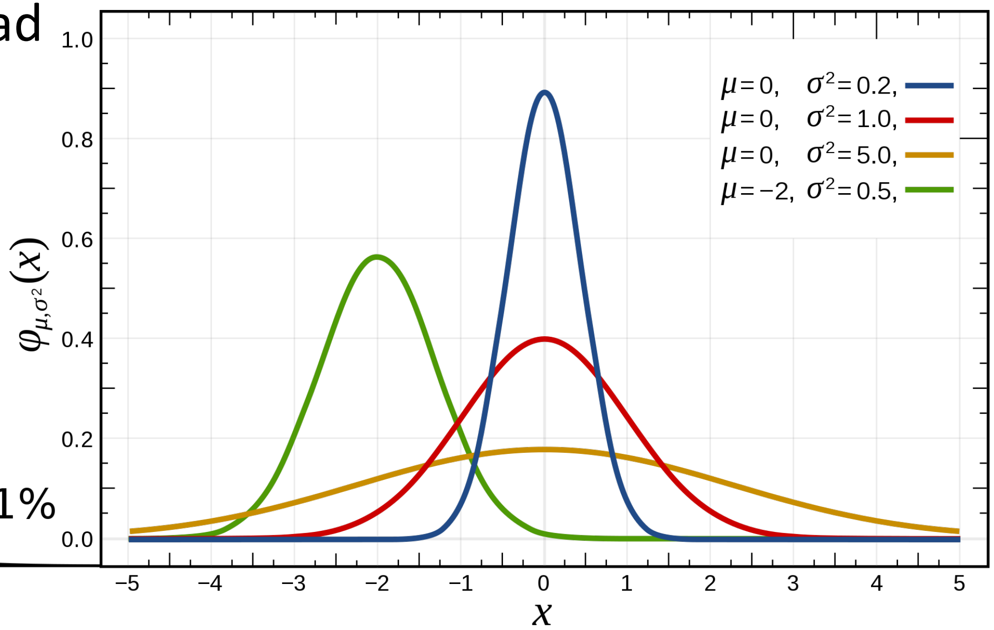
- Gaussian (Normal) Distribution
  - PDF is P(x) $= \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$, with parameters μ, σ
  - This PDF does integrate to 1
    - Proof:

1) Substitute $t = (x - \mu)/\sigma$
2) Then, Gaussian integral is $J/\sqrt{2\pi}$, where $J = \int_{-\infty}^{\infty} \exp(-t^2/2)dt$
3) Then $J^2 = (\int_{-\infty}^{\infty} \exp(-u^2/2)du)(\int_{-\infty}^{\infty} \exp(-v^2/2)dv) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-(u^2 + v^2)/2)dudv$
4) Change to polar coordinates: $u^2 + v^2 = r^2; dudv = (rd\theta)(dr)$
5) Then $J^2 = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} \exp(-r^2/2)rdrd\theta = 2\pi \int_{r=0}^{\infty} \exp(-r^2/2)rdr$
6) Substitute $w = r^2/2; dw = rdr$
7) Then $J^2 = 2\pi \int_{r=0}^{\infty} \exp(-w)dw = 2\pi[-\exp(-w)]_0^{\infty} = 2\pi$
8) So $J = \sqrt{2\pi}$. QED.

Continuous Random Variable: Examples

- Gaussian (Normal) Distribution $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$, with parameters μ, σ
  - Parameter μ is called the "**location** parameter"
    - Changing μ simply shifts the curve left/right
  - Areas under curve
    - Most of the area (around 99.8%) is within a distance of $\pm 3\sigma$ around location μ
  - Parameter σ is called the "**scale** parameter"
    - Increasing σ causes the curve to flatten/spread

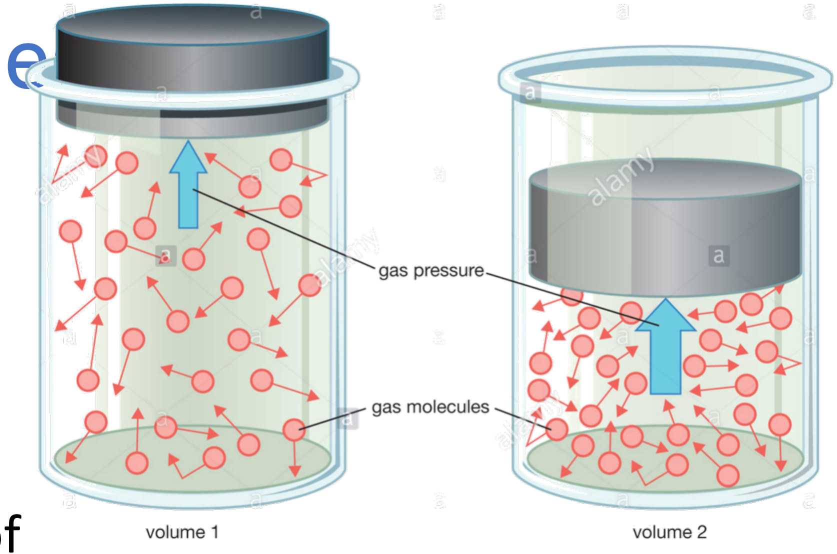




Gauss

- Carl Friedrich Gauss
  - German mathematician and physicist
  - “Greatest mathematician since antiquity”
  - Child prodigy
  - Refusing to publish work which he didn’t consider complete and above criticism
  - Declined to present intuition behind his often very elegant proofs
  - Mathematician Eric Temple Bell said that if Gauss had published all his discoveries in a timely manner, then he would have advanced mathematics by 50 years
  - Disliked teaching

Continuous Random Variable: Example



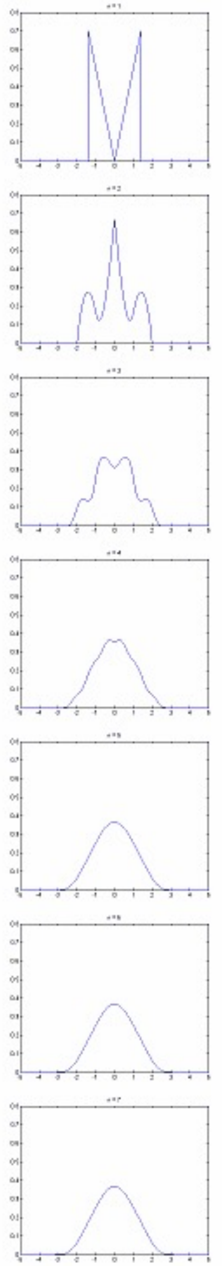
- Gaussian (Normal) Distribution **in nature**
  - **Velocities of molecules in the ideal gas**
    - Maxwell–Boltzmann distribution
      in statistical mechanics
    - Wikipedia:
      "An ideal gas is a theoretical gas composed of a set of randomly moving, **non-interacting** point particles."
      - "A gas behaves more like an ideal gas at **higher temperature and lower pressure**, as:
        the work which is against intermolecular forces becomes less significant compared with the particles' kinetic energy, and
        the size of the molecules becomes less significant compared to the empty space between them."
      - "More generally, velocities of the particles in any system in thermodynamic equilibrium will have a Gaussian distribution, due to the maximum-entropy principle."
    - In the ideal case,
      each component of the velocity vector will have a Gaussian distribution, and
      each component will be independent of all other components

Continuous Random Variable: Examp

- Gaussian (Normal) Distribution **in nature**
  - **Position of a particle undergoing diffusion**
    - If initially the particle is located at the origin,
      then after time t,
      its displacement (signed distance) traversed
      is described by
      a Gaussian distribution with $\sigma^2=t$ and $\mu=0$

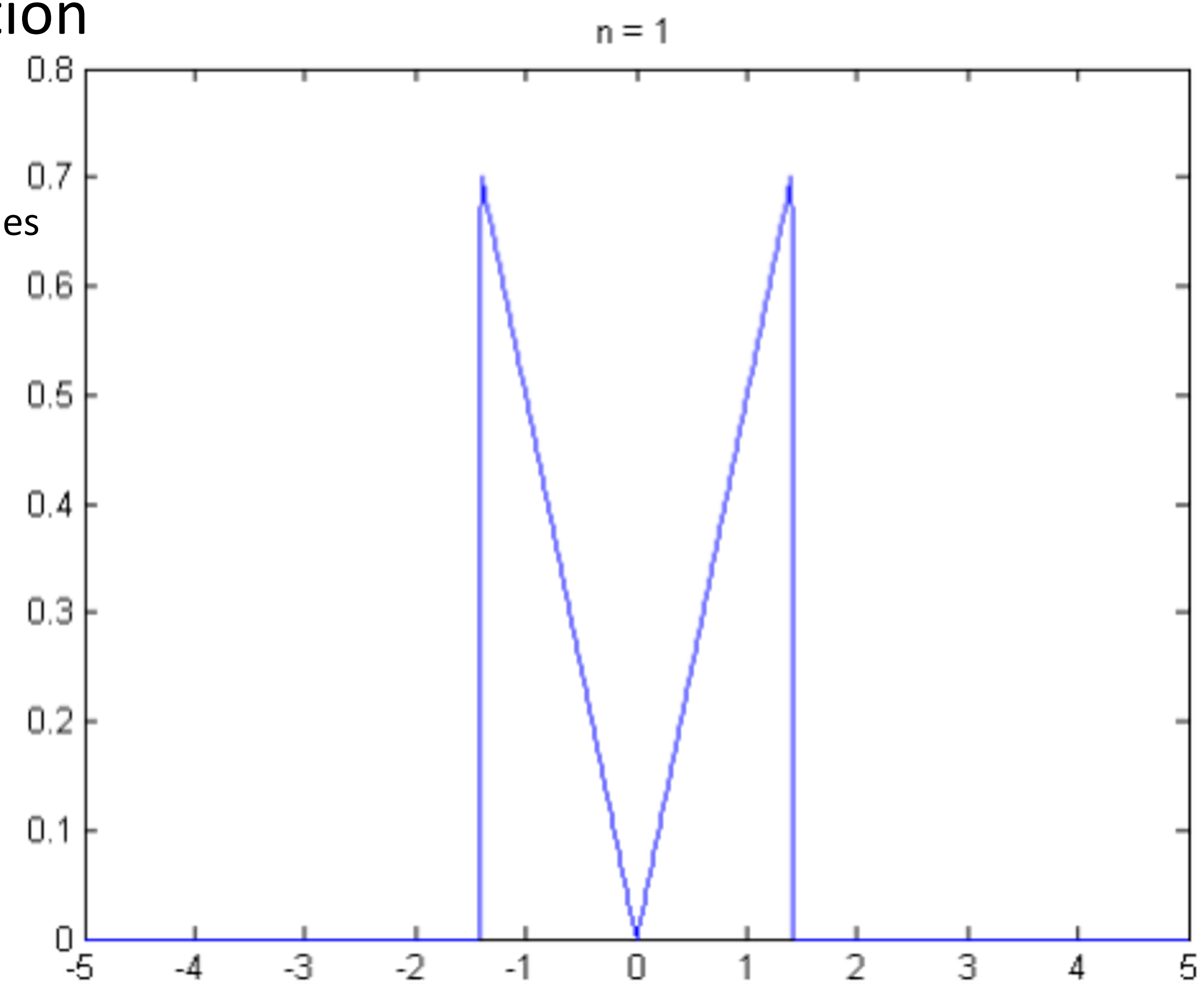
Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Central limit theorem
    - Consider a large number of RVs $X_1, X_2, \ldots, X_n$ that are:
      - Independent and Identically Distributed (i.i.d.) (with finite scale parameter)
    - Let random variable $\overline{X_n} := (X_1 + X_2 + \ldots + X_n) / n$
    - Then, $\lim_{n \to \infty} P(\overline{X}_n)$ = Gaussian Distribution
  - Examples
    - Average of Bernoulli random variables → Gaussian
    - Average of Binomial random variables → Gaussian
    - Average of Geometric random variables → Gaussian
    - Average of Poisson random variables → Gaussian
    - Average of Uniform (discrete/continuous) random variables → Gaussian
    - Average of Exponential random variables → Gaussian
    - Average of Gaussian random variables → Gaussian
    - Average of "M" random variables → Gaussian

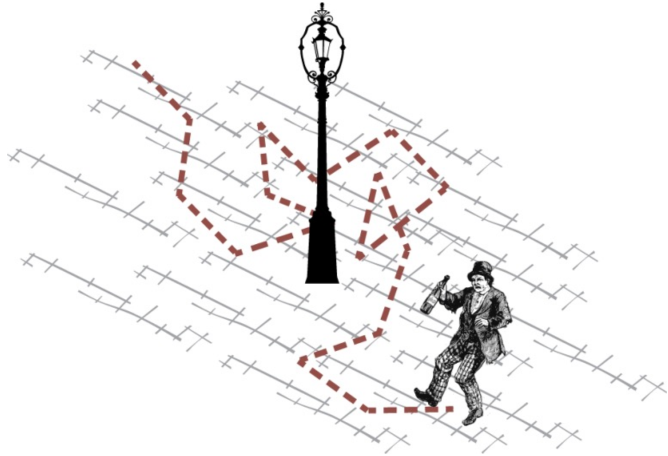


Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Central limit theorem
    - Example
      - Average of “M” random variables → Gaussian



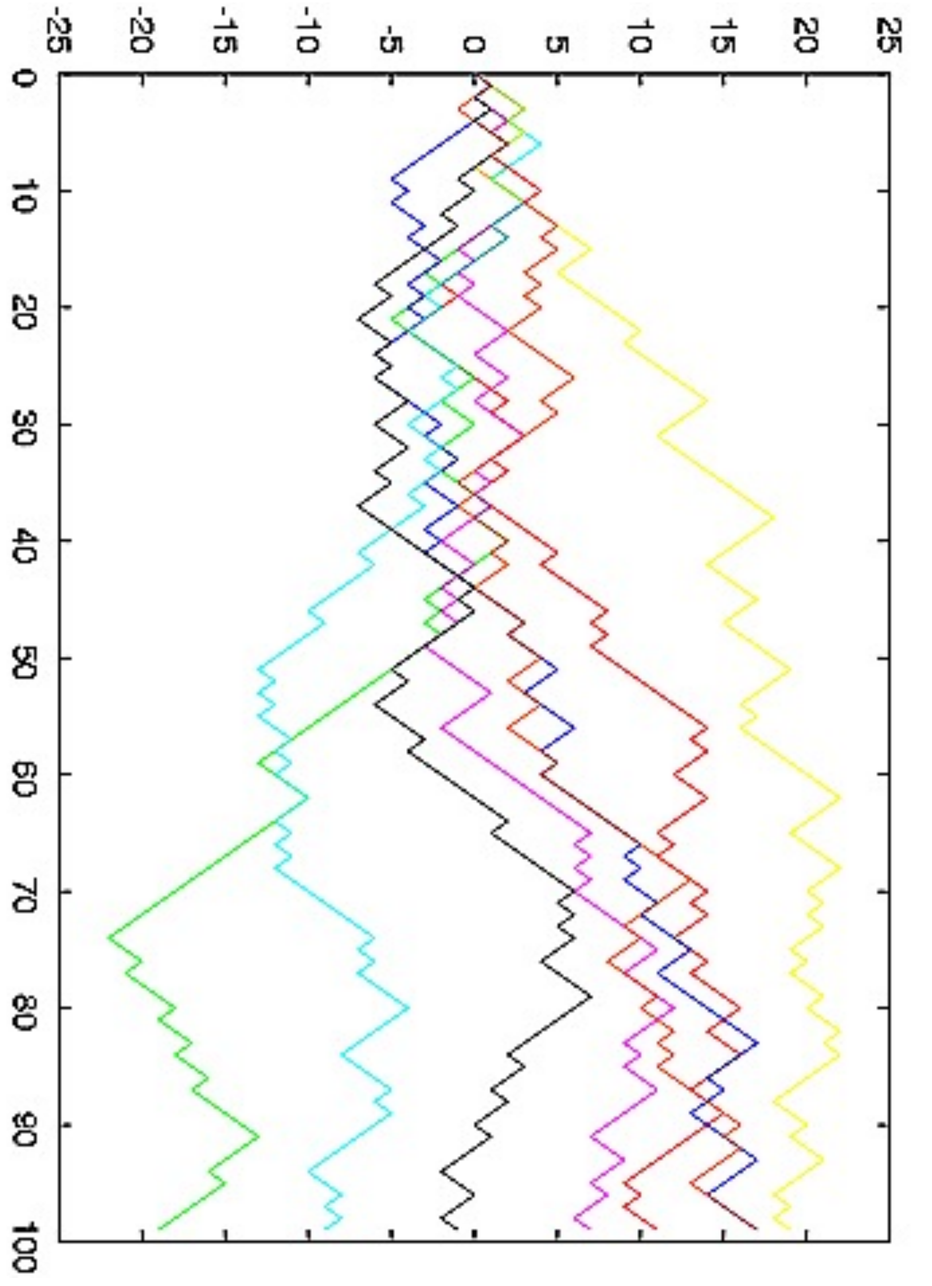
Continuous Random Variable: Examples



- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
    - De Moivre worked on this before Gauss:
      De Moivre–Laplace theorem
  - Imagine a "random walker" on the real line, starting at origin z=0 at time t=0
  - Within each timeframe Δt, the walker takes either
    a Δz step right with probability p, OR a Δz step left with probability q:=1-p
  - Let total number of steps taken = n
  - Consider an event where walker took 'x' steps right; 'n-x' steps left
    - Random walker's location is at z = Δz(2x-n)
    - Probability of this event is P(x) $= \frac{n!}{x!(n-x)!} p^x q^{n-x}$
    - We want to make the walker's location a continuous (real-valued) random variable.
      So, we want to make n, x, (n-x) (infinitely) large, and Δt and Δz (infinitesimally) small

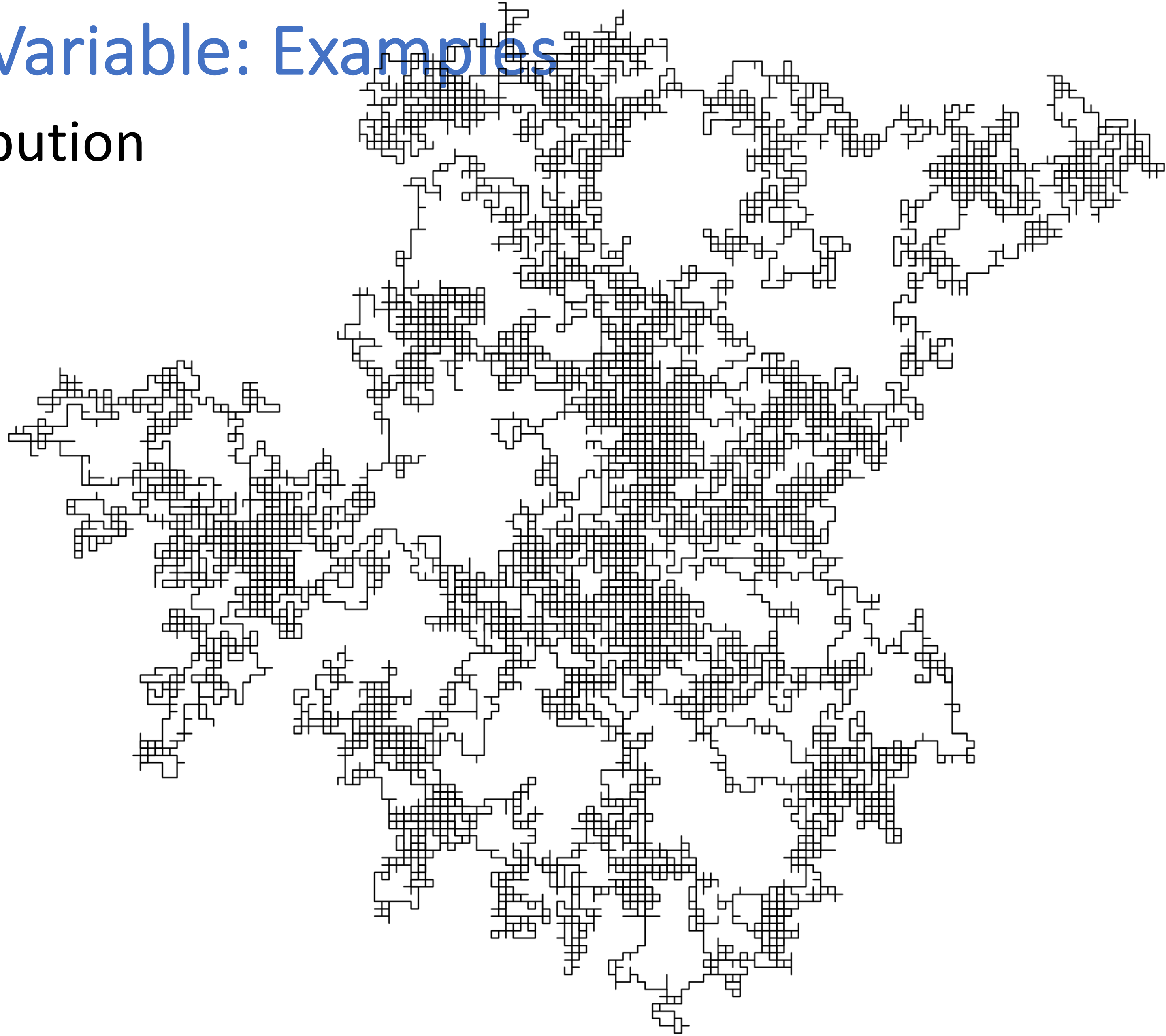
Continuous Random Variable: Exan

- Gaussian (Normal) Distribution
  - Random walk on 1D real line
    - Horizontal axis: location
    - Vertical axis: time



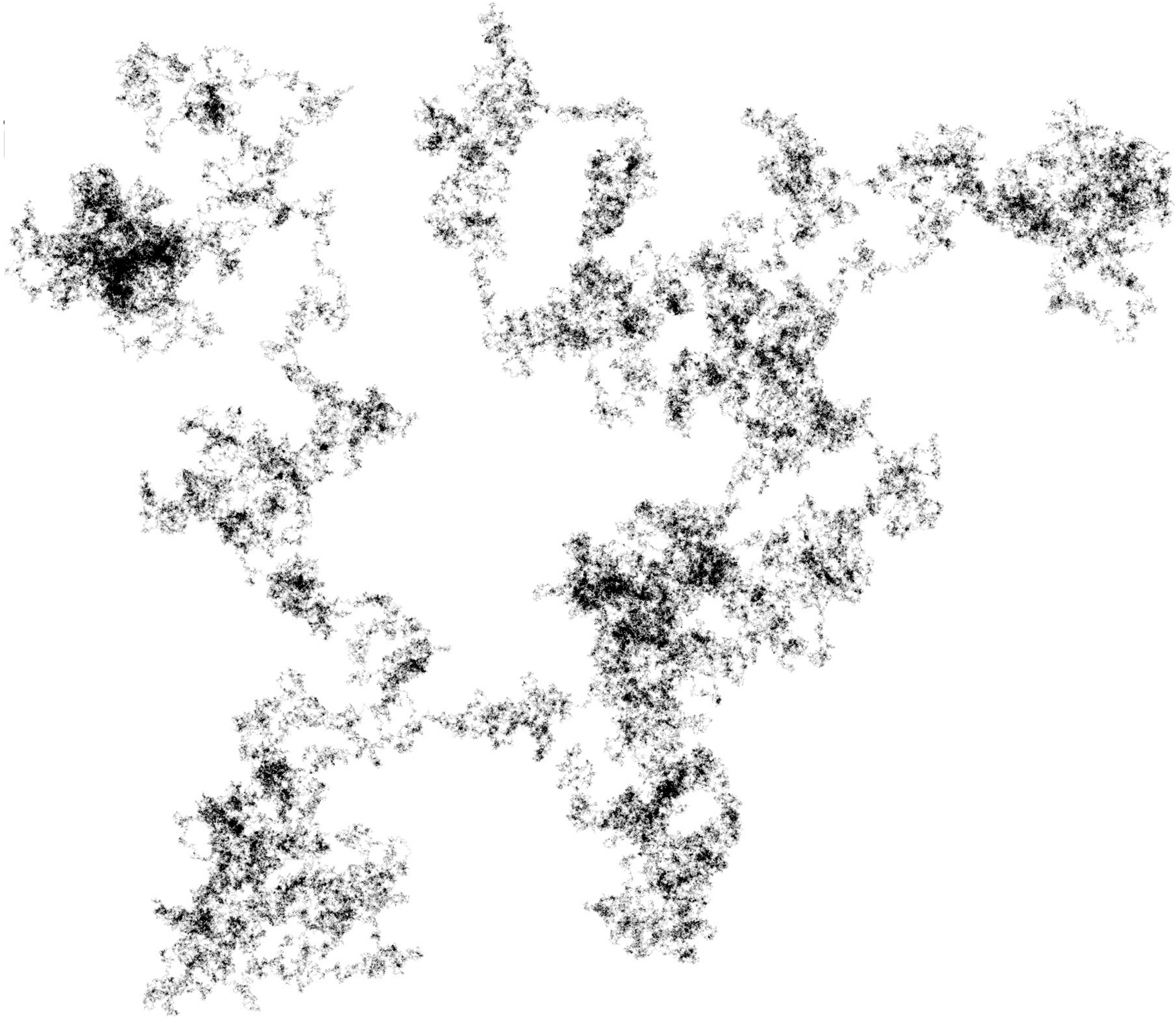
Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Random walk in 2D
    - Small steps, numbering 25,000



Continuous Random

- Gaussian (Normal) Distri
  - Random walk in 2D
    - Tiny steps, numbering 2 million

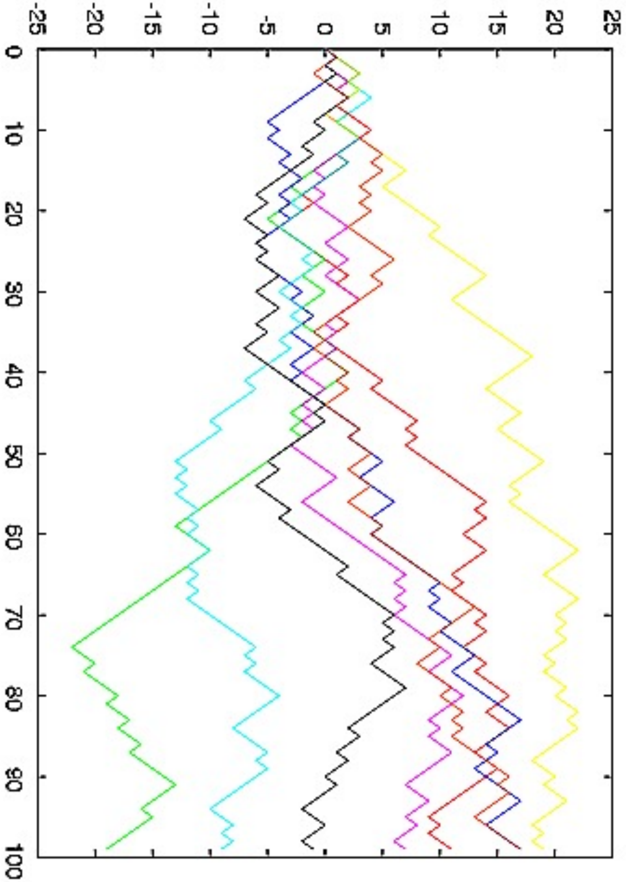


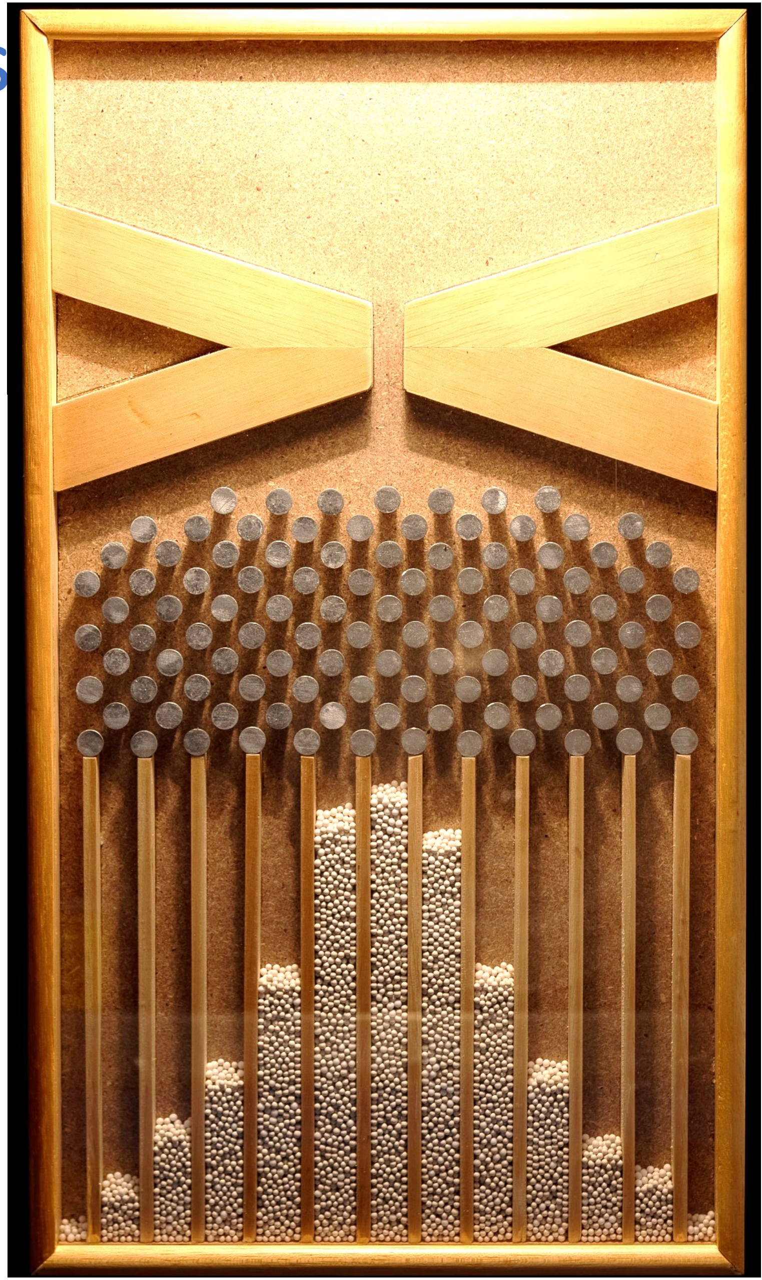
Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Brownian motion in 2D
    - https://www.youtube.com/watch?v=FAdxd2Iv-UA

Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Galton Board (also called bean machine)
  - Device invented by Sir Francis Galton
  - Simulates a random walk
  - Shows that,
    with sufficient sample size,
    binomial distribution
    tends to a Gaussian
  - https://www.youtube.com
  - https://www.youtube.com





Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
  - Factorial values can be approximated for large ‘n’ using Stirling’s formula

$$n! = n^n e^{-n}\sqrt{2\pi n}\left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

- So, P(x) $= \dfrac{n^n \cancel{e^{-n}}\sqrt{\cancel{2\pi} n}}{x^x \cancel{e^{-x}}\sqrt{2\pi x}(n-x)^{n-x}\cancel{e^{-(n-x)}}\sqrt{\cancel{2\pi}(n-x)}} p^x q^{n-x}\left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$

$$= (p/x)^x (q/(n-x))^{n-x} n^n \sqrt{\frac{n}{2\pi x(n-x)}}\left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

$$= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x}\sqrt{\frac{n}{2\pi x(n-x)}}\left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

Continuous Random Variable:

$$= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}}\left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
  - Let deviation of x from np be $\delta = x - np$, so that $x = \delta + np$ and $n - x = nq - \delta$
    - When n is large, then x is close to np by our design/assignment of "probability function"
      - Probability of success is (intended/designed to be) the ratio of number of successes 'x' to number of tries 'n'. More analysis on this later.
  - So $\boxed{\ln\left(\frac{np}{x}\right) = \ln\left(\frac{np}{np+\delta}\right) = -\ln\left(1+\frac{\delta}{np}\right)}$ $\boxed{\ln\left(\frac{nq}{n-x}\right) = \ln\left(\frac{nq}{nq-\delta}\right) = -\ln\left(1-\frac{\delta}{nq}\right)}$
  - Then, $\ln\left[\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x}\right] = x\ln\left(\frac{np}{x}\right) + (n-x)\ln\left(\frac{nq}{n-x}\right)$

$$= -(\delta+np)\left[\frac{\delta}{np} - \frac{1}{2}\frac{\delta^2}{n^2p^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right)\right] - (nq-\delta)\left[-\frac{\delta}{nq} - \frac{1}{2}\frac{\delta^2}{n^2q^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right)\right]$$

$$= -\frac{\delta^2}{2npq} + \mathcal{O}\left(\frac{\delta^3}{n^2}\right)$$

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \mathcal{O}(x^3)$$

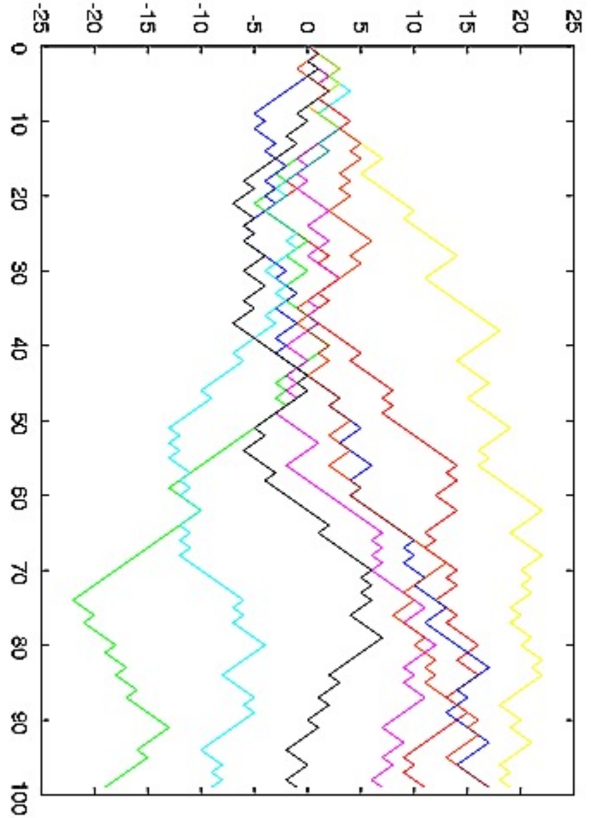
Continuous Random Variable: $= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}} \left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$

- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
  - Also, $\sqrt{\frac{n}{2\pi x(n-x)}} = \sqrt{\frac{n}{2\pi(np+\delta)(nq-\delta)}} = ?$ $\sqrt{\frac{1}{2\pi npq}}\left[1+\mathcal{O}\left(\frac{\delta}{n}\right)\right]$
    - Write $\frac{1}{\sqrt{np+\delta}} = \frac{1}{\sqrt{np}} \frac{1}{\sqrt{1+\frac{\delta}{np}}}$ and write $\frac{1}{\sqrt{nq-\delta}} = \frac{1}{\sqrt{nq}} \frac{1}{\sqrt{1-\frac{\delta}{nq}}}$
    - As $n \to \infty$, the "deviation" $\delta \to 0$. So, ratio $\frac{\delta}{n} \to 0$
    - We want to evaluate $\mathrm{g(t)} \coloneqq \frac{1}{\sqrt{1+\mathrm{t}}}$ at a small displacement $\mathrm{t} = \frac{\delta}{np}$ around 0
    - Taylor series: $g(a+\mathrm{t}) \coloneqq g(a) + \frac{g'(a)(\mathrm{t}-a)}{1!} + \frac{g''(a)(t-a)^2}{2!} + \cdots$
      - Evaluate around a=0
    - We get the Maclaurin series for g(t) := $\frac{1}{\sqrt{1+t}} = 1 - \frac{t}{2} + O(t^2)$

Continuous Random Variable:

$$= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}} \left[1+\mathcal{O}\left(\frac{1}{n}\right)\right]$$

$\delta = x - np$, so that $x = \delta + np$ and $n - x = nq - \delta$



- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
  - So, for large 'n' and 'x', P(x) tends to:

$$e^{-\delta^2/2npq}\left[1+\mathcal{O}\left(\frac{\delta^3}{n^2}\right)\right]\sqrt{\frac{1}{2\pi npq}}\left[1+\mathcal{O}\left(\frac{\delta}{n}\right)\right]$$

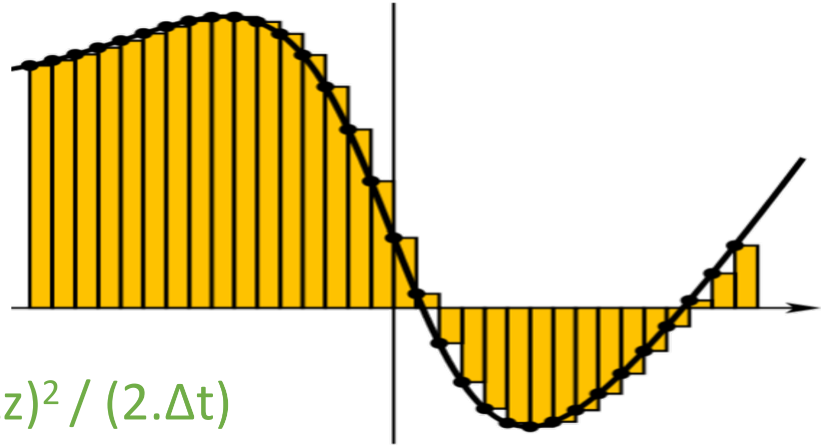
$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

  - For this binomial distribution
    - When δ is large, i.e., when x is far from np, then P(x) is close to 0 (unlikely event)
    - For P(x) to be not-close to 0, value x needs to be close to 'np', i.e., δ << n
  - Thus, functional form of P(x) → $\frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq}$

Continuous Random Variable: Examples

$$\frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq}$$

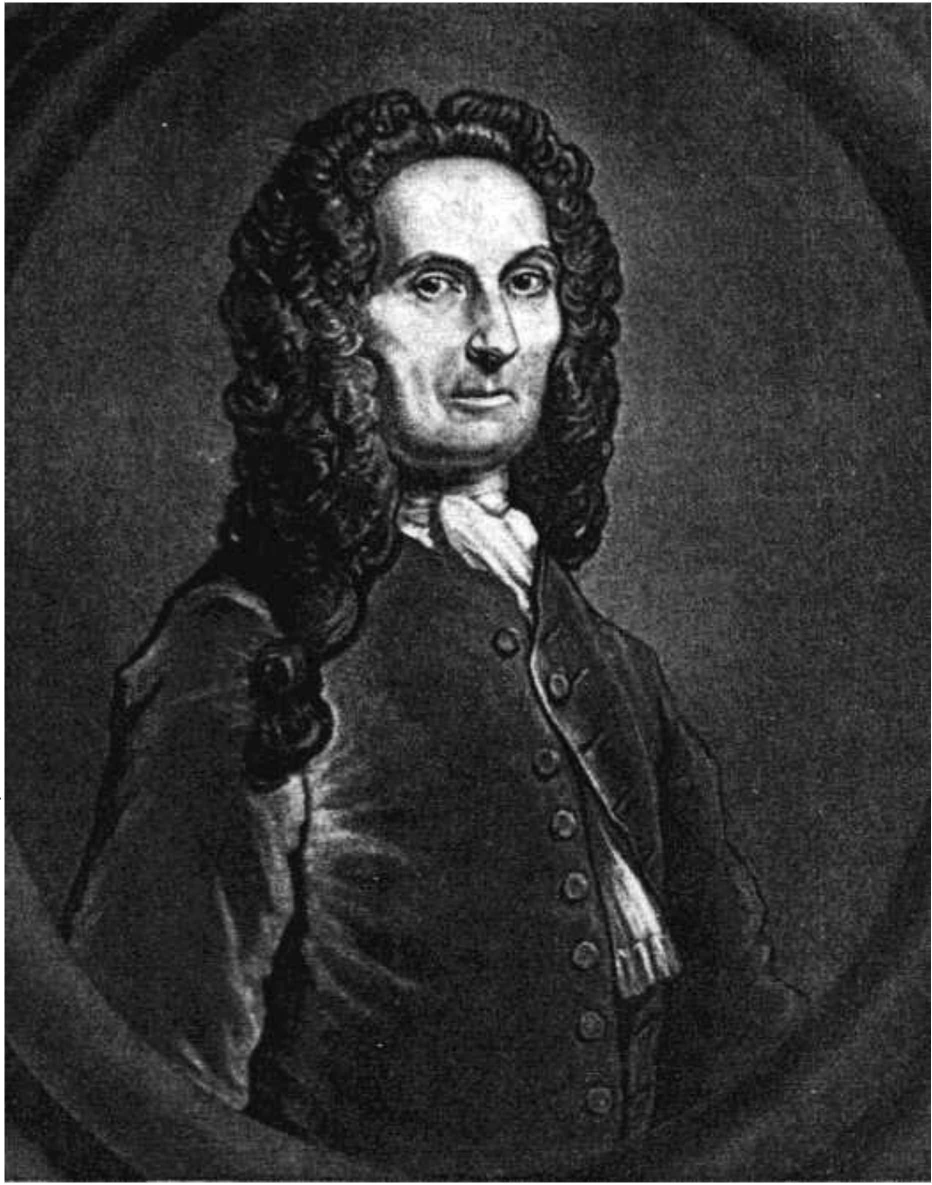
- Gaussian (Normal) Distribution
  - Limiting case of a Binomial distribution
  - After x steps towards right (out of n steps) in time $t = n.\Delta t$, random walker's location is at $z = \Delta z(2x-n)$ that implies $x = n/2 + z/(2.\Delta z)$
  - Consider distributions: P(x) on natural numbers x, P(z) on real coordinates z
    - As $\Delta z \rightarrow 0$ and $n \rightarrow \infty$, P(z) becomes continuous
    - Consider a mapping between the distributions
      - A change of $2.\Delta z$ in z leads to a change of 1 in x
      - So, $P(z)\, 2.\Delta z = P(x)$, where z corresponds to its associated x



  - Define
    - For simplicity, p=q=0.5. "Diffusion coefficient" $D := (\Delta z)^2 / (2.\Delta t)$
  - Then, $P(z;t) = \frac{1}{2\Delta z} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\Delta t}}{\sqrt{0.25t}} \exp\left(-\frac{1}{2}\frac{z^2}{(\Delta z)^2}\frac{\Delta t}{t}\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2Dt}} \exp\left(-\frac{z^2}{4Dt}\right)$;
    - A Gaussian with location parameter 0 (because p=q) and scale-parameter² ∝ t (spread increase with time)

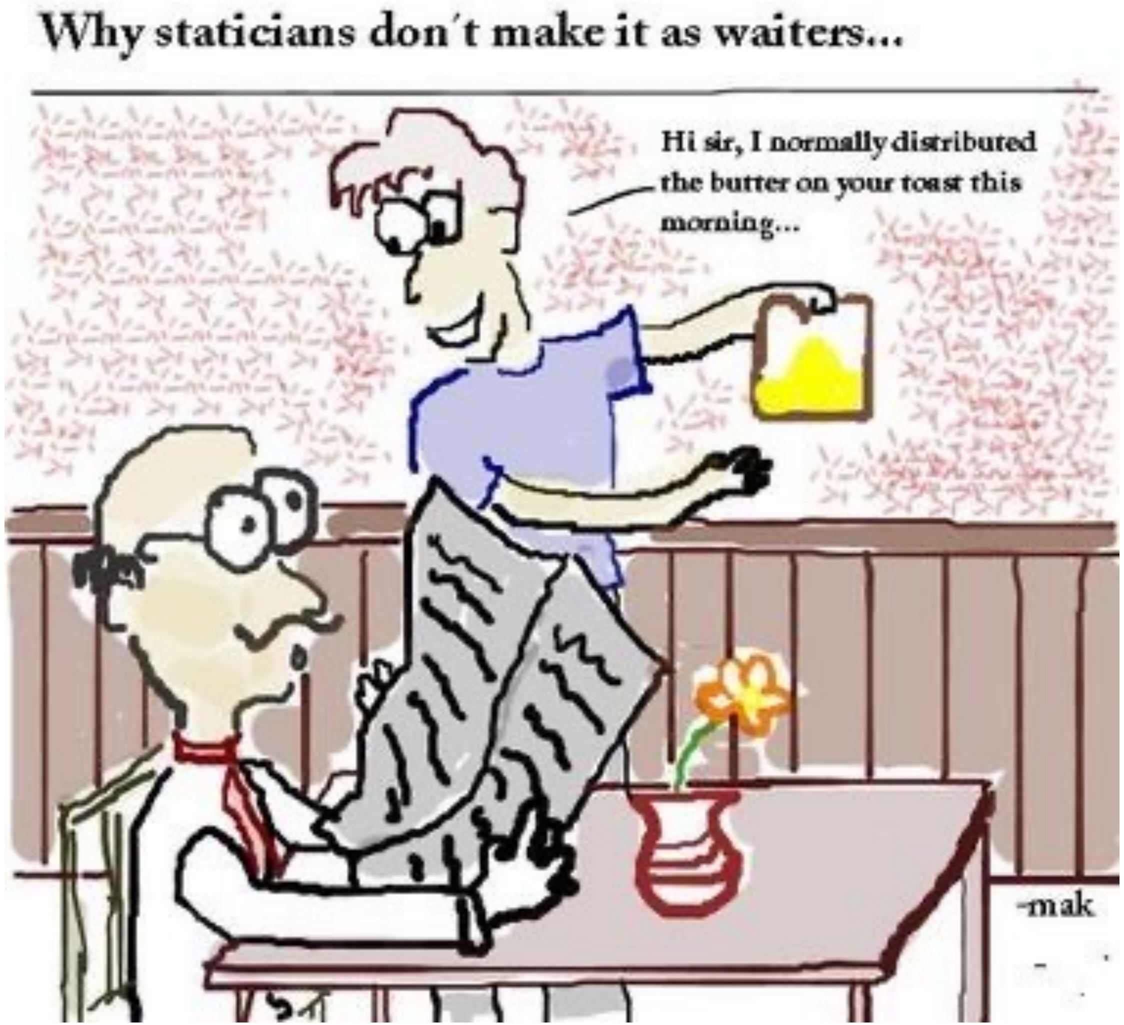
De Moivre

- Abraham de Moivre
  - French mathematician
  - Persecution from France led to London where he met Newton and came across Principia Mathematica
  - Used to work as private math tutor, taking long walks to travel from one student to another, carrying pages torn out of Principia Mathematica to read during walks
  - Was accepted into Royal Society for his work in math, but couldn't get a university job in London due to his French origins
  - Remained poor throughout his life



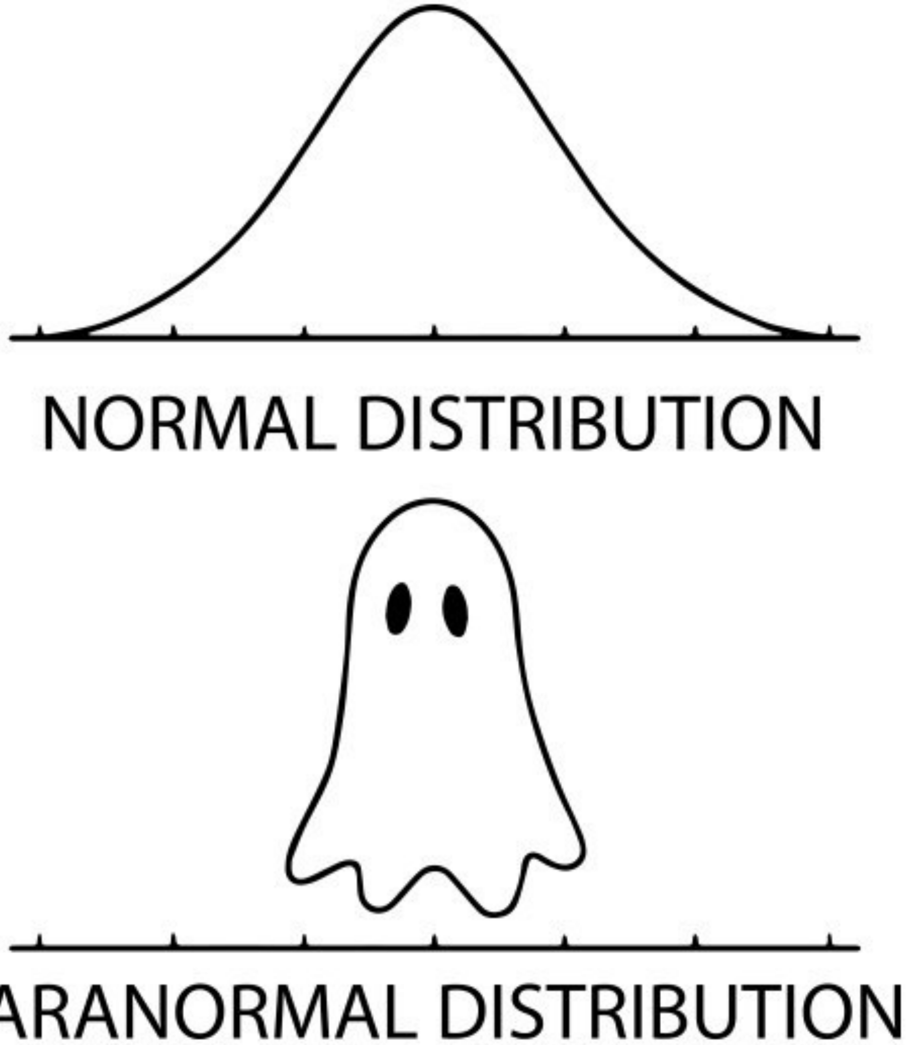
Continuous Random Variable: Examples

- Gaussian (Normal) Distribution



Continuous Random Variable: Examples

- Gaussian (Normal) Distribution



Continuous Random Variable: Examples

- Gaussian (Normal) Distribution

Continuous Random Variable: Examples

- Gaussian (Normal) Distribution
  - "Everyone believes in the normal law,
    the experimenters because they imagine that it is a mathematical theorem,
    and the mathematicians because they think it is an experimental fact."
    - J.F. Gabriel Lippmann (1845 – 1921)
      - Nobel laureate in physics (1908)
        for inventing a method for capturing
        color photographs
      - Held positions as
        Professor of Mathematical Physics,
        Professor of Experimental Physics