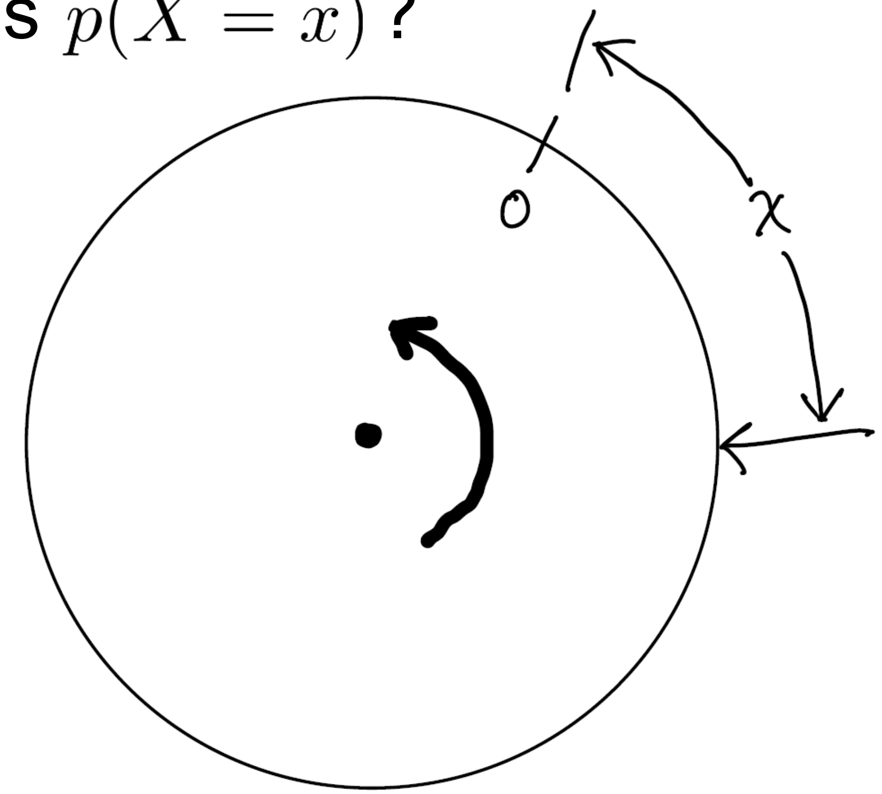
Probabilistic Graphical Models

Continuous RVs
WLLN
CLT

Continuous Probability

Experiment Spin continuous wheel and measure X displacement from 0

Question Assuming uniform probability, what is $p(X = x)$?



First, recall axioms of probability…

1. For any event E, $0 \leq P(E) \leq 1$
2. $P(\Omega) = 1 \text{ and } P(\emptyset) = 0$
3. For any *finite* or *countably infinite* sequence of <u>pairwise mutually disjoint events</u> $E_1, E_2, E_3, \ldots$

$$P\Big(\bigcup_{i \geq 1} E_i\Big) = \sum_{i \geq 1} P(E_i)$$

Sample space Ω is all points (real numbers) in $[0, 1)$

Continuous Probability

- Let $P(X = x) = \pi$ be the probability of any single outcome
- Let $S(k)$ be the set of k elements in $[0,1)$

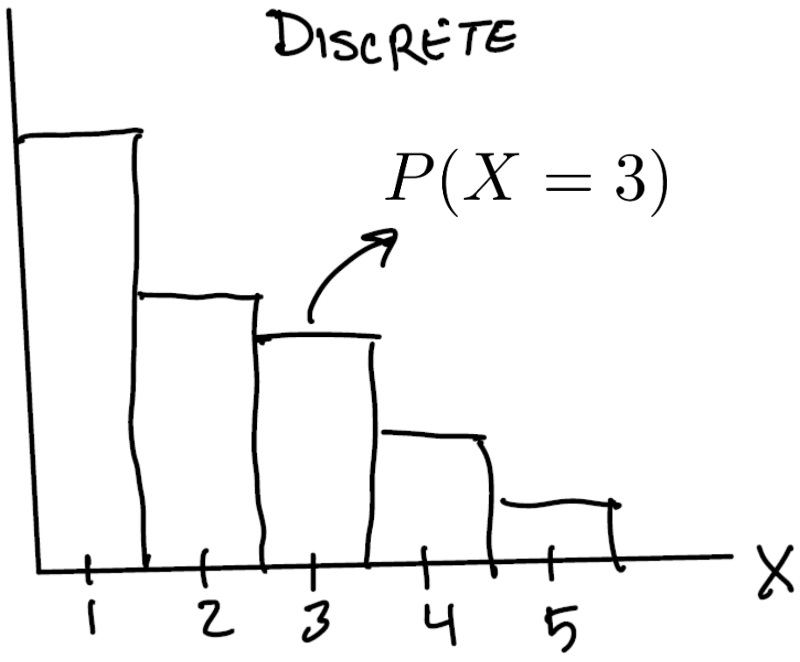
- Since $0 < P(x \in S(k)) < 1$ by axioms of probability, $k\pi < 1$ for any k
- Therefore: $\pi = 0$ and $P(x \in S(k)) = p(X = x) = 0$

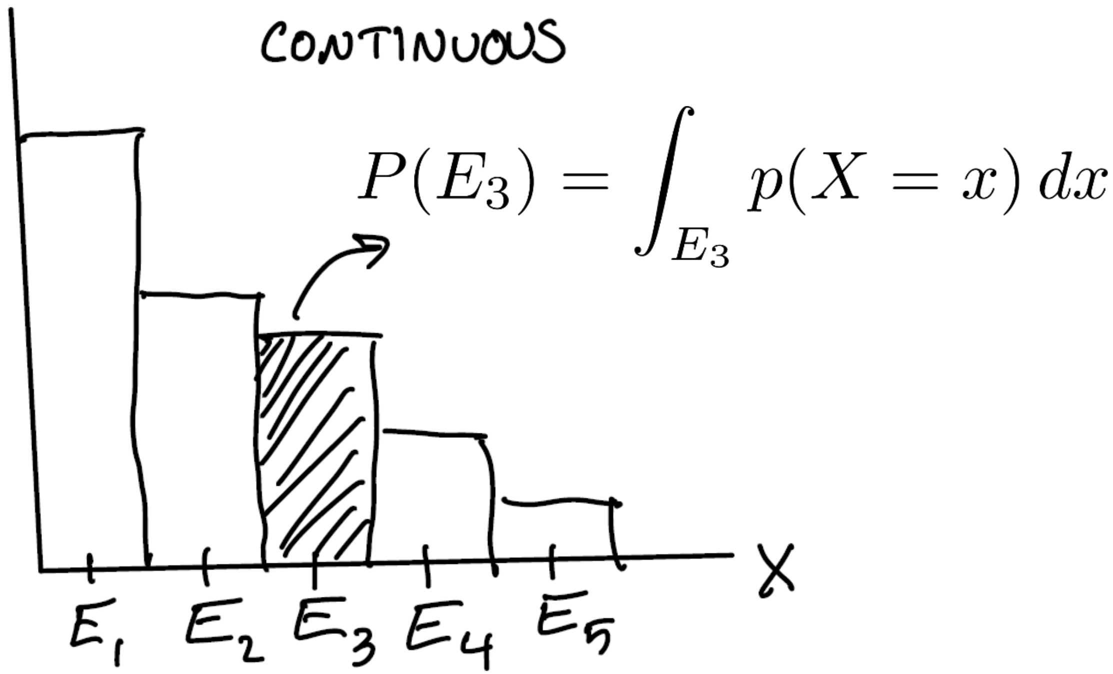
What does this mean?

- Let E be event that $x \in S(k)$
- In infinite sample space, an event may be **possible** but have zero "probability"
- Since $P(\bar{E}) = 1 - P(E) = 1$ events may have "probability" 1 but **not certain**

Assign probability to intervals, not individual values

Continuous Probability



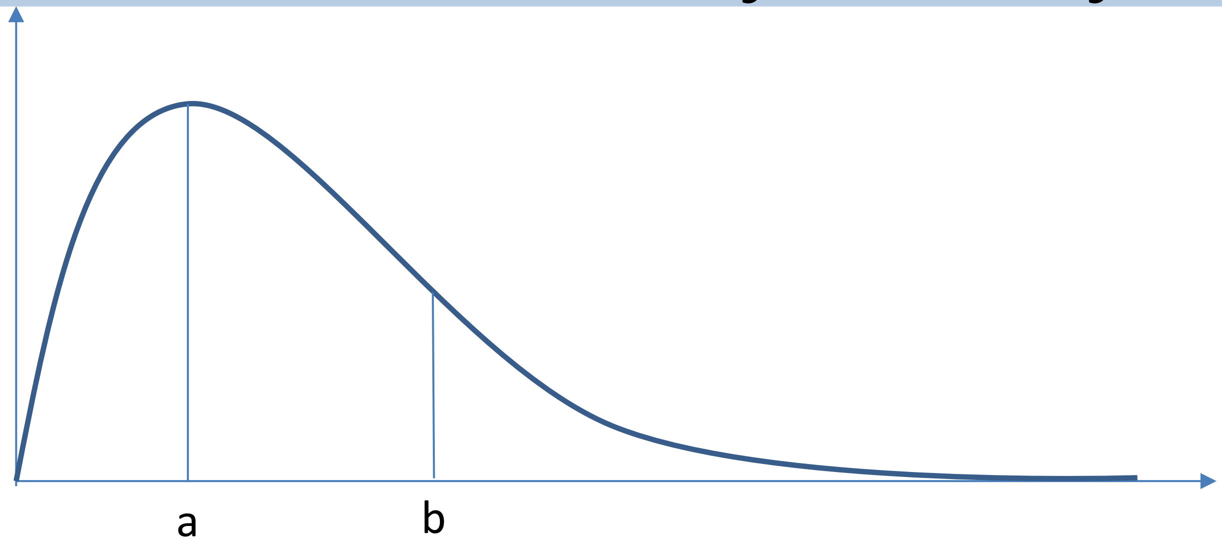


- Events represented as intervals $a \leq X < b$ with probability,

$$P(a \leq X < b) = \int_a^b p(X = x)\, dx$$

- Specific outcomes have zero probability $P(X = 0) = P(x \leq X < x) = 0$
- But may have nonzero *probability density* $p(X = x)$

Probability Density Function



$$P(a \leq x \leq a+\epsilon) = \int_a^{a+\epsilon} p(x)dx = p(a) * \epsilon$$

$$P(\beta \leq x \leq \beta+\epsilon) = \int_\beta^{\beta+\epsilon} p(x)dx = p(\beta) * \epsilon$$

Continuous Probability

Most definitions for discrete RVs hold, replacing PMF with PDF

Two RVs X & Y are **independent** if and only if,

$$p(x, y) = p(x)p(y) \quad \text{or} \quad P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

Conditionally independent given Z iff,

Shorthand: $P(x) = P(X \le x)$

$$p(x, y \mid z) = p(x \mid z)p(y \mid z) \quad \text{or} \quad P(x, y \mid z) = P(x \mid z)P(y \mid z)$$

Probability chain rule,

$$p(x, y) = p(x)p(y \mid x) \quad \text{and} \quad P(x, y) = P(x)P(y \mid x)$$

Continuous Probability

…and by replacing summation with integration…

Law of Total Probability for continuous distributions,

$$p(x) = \int_{\mathcal{Y}} p(x, y)\, dy$$

Expectation of a continuous random variable,

$$\mathbf{E}[X] = \int_{\mathcal{X}} x \cdot p(x)\, dx$$

Covariance of two continuous random variables X & Y,

$$\mathbf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \int_{\mathcal{X}} \int_{\mathcal{Y}} (x - \mathbf{E}[X])(y - \mathbf{E}[Y]) p(x,y)\, dx dy$$

Continuous Probability

***Caution** Some technical subtleties arise in continuous spaces…*

For **discrete** RVs X & Y, the conditional

P(Y=y)=0 means impossible

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

is **undefined** when P(Y=y) = 0 … no problem.

For **continuous** RVs we have,

$$P(X \leq x \mid Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}$$

but numerator and denominator are 0/0.

P(Y=y)=0 means improbable, but not impossible

Continuous Probability

Defining the conditional distribution as a limit fixes this…

$$P(X \leq x \mid Y = y) = \lim_{\delta \to 0} P(X \leq x \mid y \leq Y \leq y + \delta)$$

$$= \lim_{\delta \to 0} \frac{P(X \leq x, y \leq Y \leq y + \delta)}{P(y \leq Y \leq y + \delta)}$$

$$= \lim_{\delta \to 0} \frac{P(X \leq x, Y \leq y + \delta) - P(X \leq x, Y \leq y)}{P(Y \leq y + \delta) - P(Y \leq y)}$$

$$= \int_{-\infty}^{x} \lim_{\delta \to 0} \frac{\frac{\partial}{\partial x} P(u, y + \delta) - \frac{\partial}{\partial x} P(u, y)}{P(y + \delta) - P(y)} \, du$$

(Fundamental theorem of calculus)

(Assume interchange limit / integral)

$$= \int_{-\infty}^{x} \lim_{\delta \to 0} \frac{\left(\frac{\partial}{\partial x} P(u, y + \delta) - \frac{\partial}{\partial x} P(u, y)\right) / \delta}{\left(P(y + \delta) - P(y)\right) / \delta} \, du$$

(Multiply by $\frac{\delta}{\delta} = 1$)

$$= \int_{-\infty}^{x} \frac{\frac{\partial^2}{\partial x \partial y} P(u, y)}{\frac{\partial}{\partial y} P(y)} \, du \quad = \int_{-\infty}^{x} \frac{p(u, y)}{p(y)} \, du$$

(Definition of partial derivative)

(Definition PDF)

Definition The conditional PDF is given by,

$$p(x \mid y) = \frac{p(x, y)}{p(y)}$$

Mixed Variables

Let X be the consumption of Soda (in ml)

What does this look like?

Is this a pdf or a pmf?

CDF

Definition *The* <u>*cumulative distribution function*</u> *(CDF) of a real-valued continuous RV X is the function given by,*

$$F(x) = P(X \leq x)$$

- Can easily measure probability of closed intervals,

$$P(a \leq X < b) = F(b) - F(a)$$

- If *X* is *absolutely continuous* (i.e. differentiable) then,

$$f(x) = \frac{dF(x)}{dx} \quad \text{and} \quad F(t) = \int_{-\infty}^{t} f(x)\, dx$$

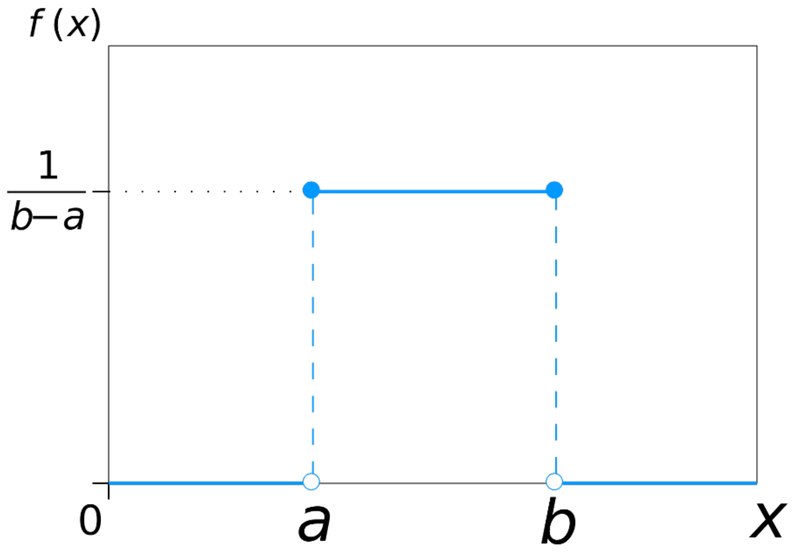
Where $f(x)$ is the *probability density function* (PDF)

- Typically use shorthand P for CDF and p for PDF instead of F and f

Useful Continuous Distributions

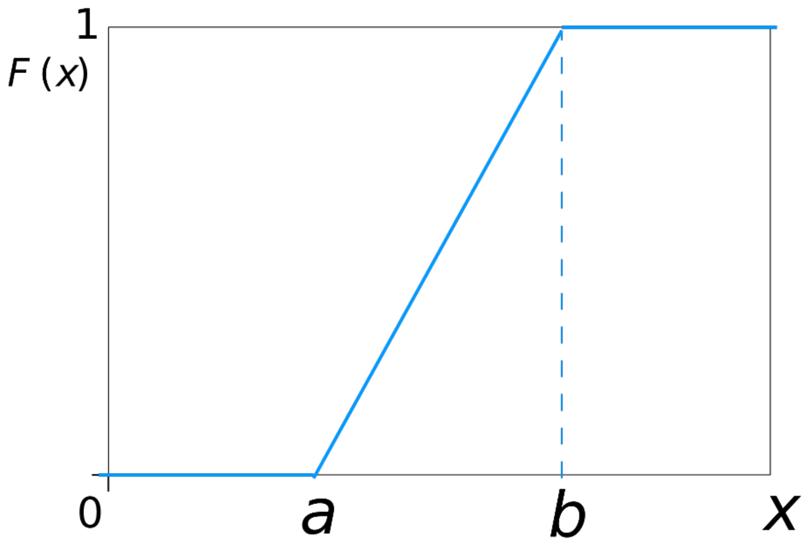
Uniform distribution on interval $[a, b]$,

$$p(x) = \begin{cases} 0 & \text{if } x \le a, \\ \frac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{if } b \le x \end{cases} \qquad P(X \le x) = \begin{cases} 0 & \text{if } x \le a, \\ \frac{x-a}{b-a} & \text{if } a \le x \le b, \\ 1 & \text{if } b \le x \end{cases}$$



Say that $X \sim U(a, b)$ whose moments are,

$$\mathbf{E}[X] = \frac{b+a}{2} \qquad \mathbf{Var}[X] = \frac{(b-a)^2}{12}$$



Suppose $X \sim U(0, 1)$ and we are told $X \le \frac{1}{2}$ what is the conditional distribution?

$$P(X \le x \mid X \le \tfrac{1}{2}) = U(0, \tfrac{1}{2})$$

Holds generally: Uniform closed under conditioning

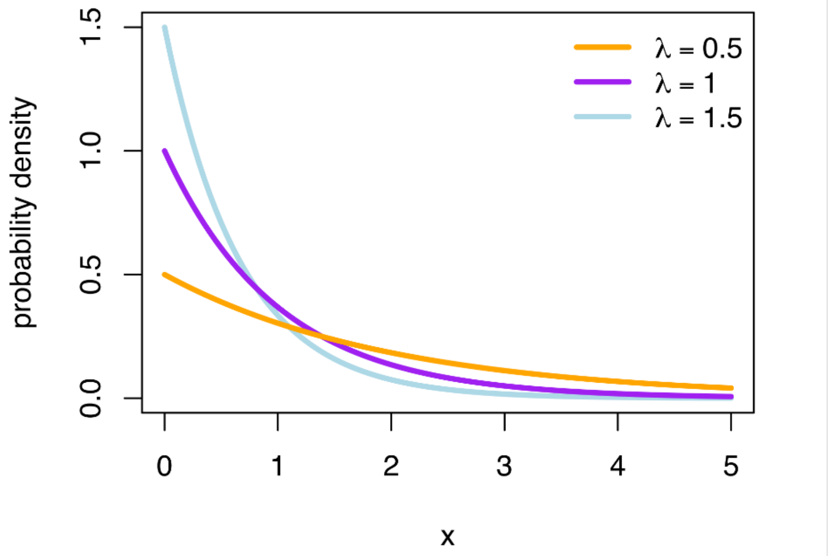
Useful Continuous Distributions

Exponential distribution with scale λ,

$$p(x) = \lambda e^{-\lambda x} \qquad P(x) = 1 - e^{-\lambda x}$$

for X>0. Moments given by,

$$\mathbf{E}[X] = \frac{1}{\lambda} \qquad \mathbf{Var}[X] = \frac{2}{\lambda^2}$$



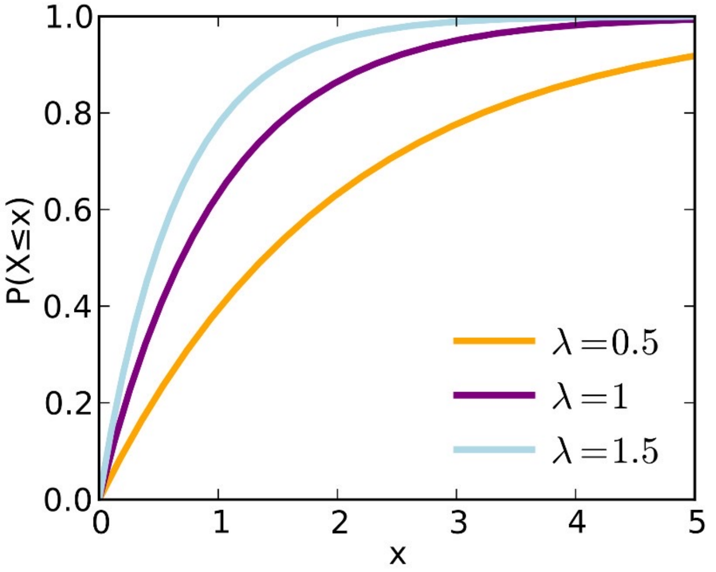
Useful properties

- **Closed under conditioning** If $X \sim \text{Exponential}(\lambda)$ then,

$$P(X \geq s + t \mid X \geq s) = P(X \geq s) = e^{-\lambda s}$$

- **Minimum** Let $X_1, X_2, \ldots, X_N$ be i.i.d. exponentially distributed with scale parameters $\lambda_1, \lambda_2, \ldots, \lambda_N$ then,

$$P(\min(X_1, X_2, \ldots, X_N)) = \text{Exponential}(\textstyle\sum_i \lambda_i)$$

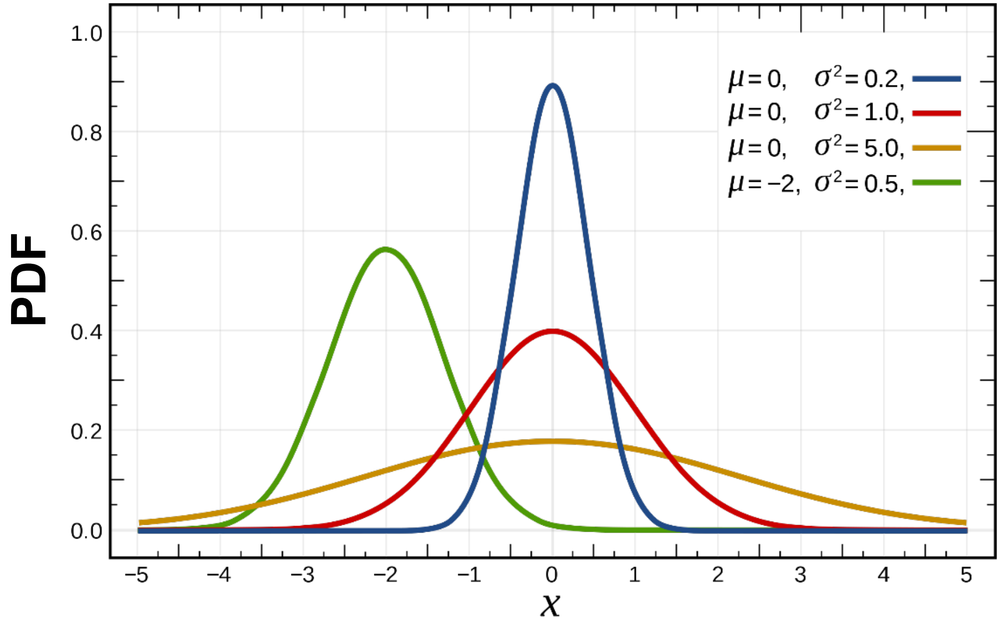


Useful Continuous Distributions

Gaussian (a.k.a. Normal) distribution with mean mean (location) μ and variance (scale) σ^2 parameters,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2}(x-\mu)^2/\sigma^2$$

We say $X \sim \mathcal{N}(\mu, \sigma^2)$.



Useful Properties

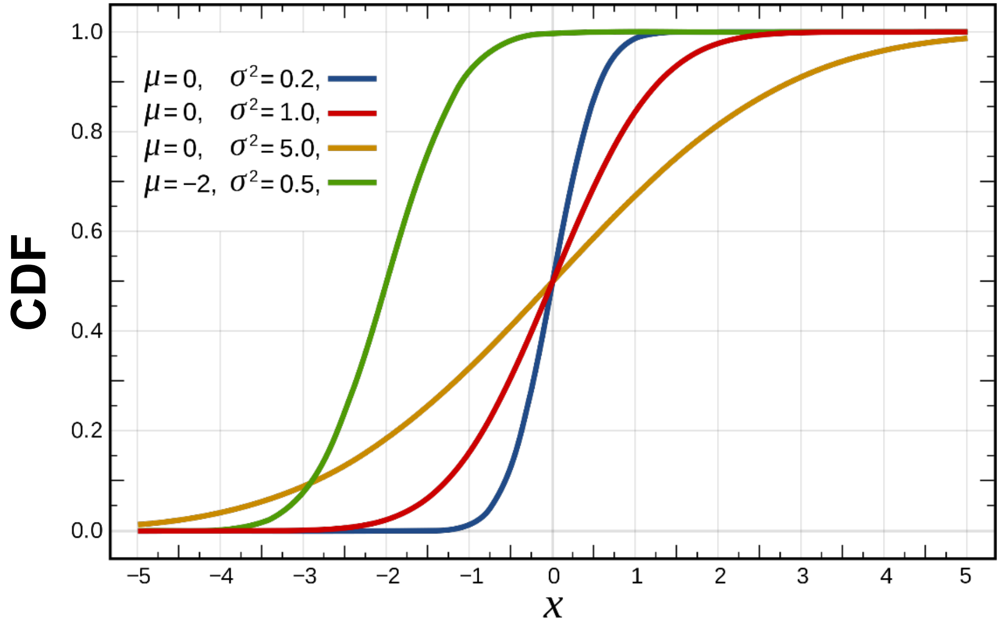
- Closed under additivity:

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

- Closed under linear functions (a and b constant):

$$aX + b \sim \mathcal{N}(a\mu_x + b, a^2\sigma_x^2)$$

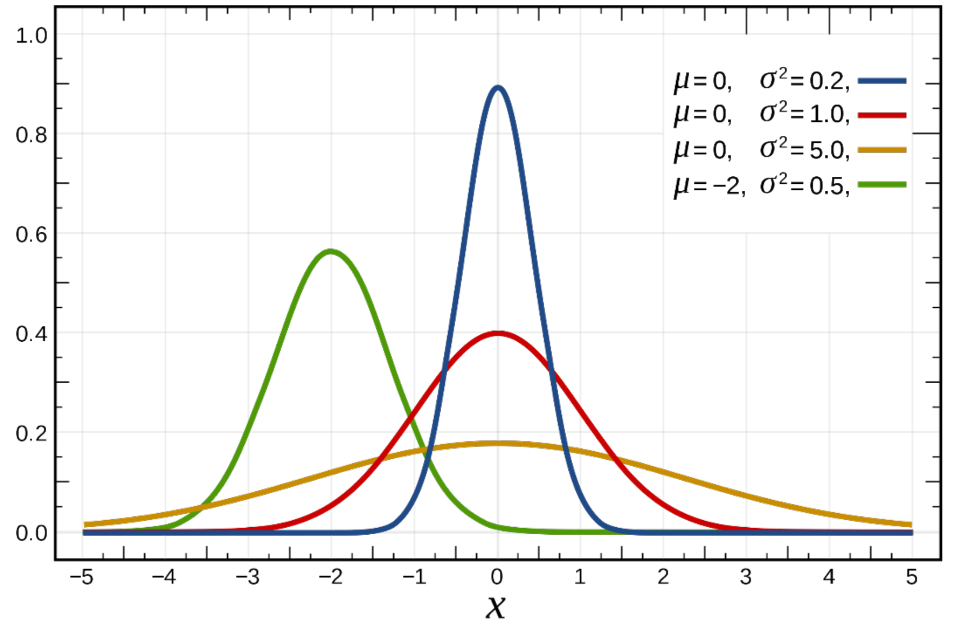


Computing Probabilities Using the Normal CDF

Normal Variables are very common (for reasons we will discuss very soon)

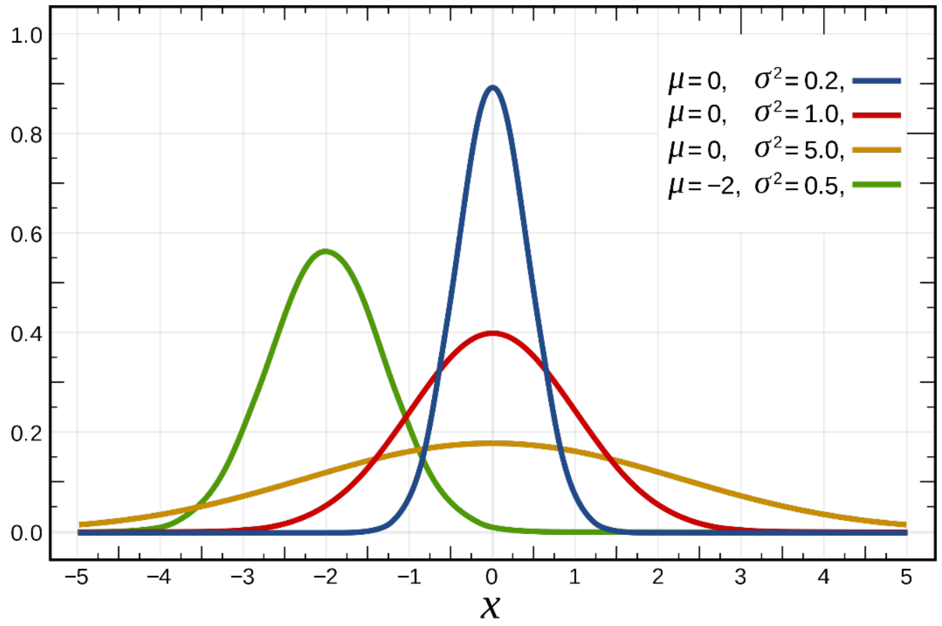
We often want to compute $P(X \leq a)$ *for* $X \sim N(\mu, \sigma^2)$

$\Phi(\alpha)$ *is tabulated at the end of all statistical books.*



Computing Probabilities Using the Normal CDF

Normal Variables are very common (for reasons we will discuss very soon)



We often want to compute $P(X \leq a)$ *for* $X \sim N(\mu, \sigma^2)$

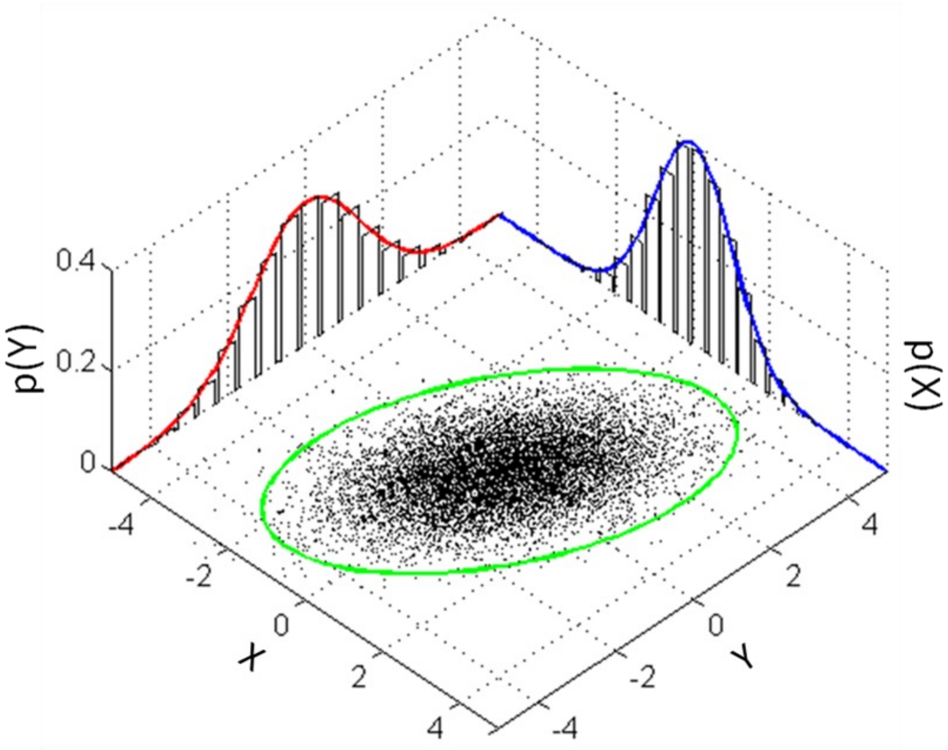
$\Phi(\alpha)$ *is tabulated at the end of all statistical books.*

$$P(X \leq a) = P\left(\frac{X-\mu}{\sigma^2} \leq \frac{a-\mu}{\sigma^2}\right) = P\left(Z \leq \frac{a-\mu}{\sigma^2}\right) = \Phi\left(\frac{a-\mu}{\sigma^2}\right)$$

Useful Continuous Distributions

Multivariate Gaussian On RV $X \in \mathcal{R}^d$ with mean $\mu \in \mathcal{R}^d$ and positive semidefinite covariance matrix $\Sigma \in \mathcal{R}^{d \times d}$,

$$p(x) = |2\pi\Sigma|^{-1/2} \exp -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$$



A matrix $A \in \mathbb{R}^{n \times n}$ is called symmetric positive (semi-) definite if $A = A^\top$, and

$$v^\top A v \geq 0 \forall v \in \mathbb{R}^n.$$

Equivalent statement: All eigenvalues of the symmetric matrix A are non-negative.

Useful Continuous Distributions

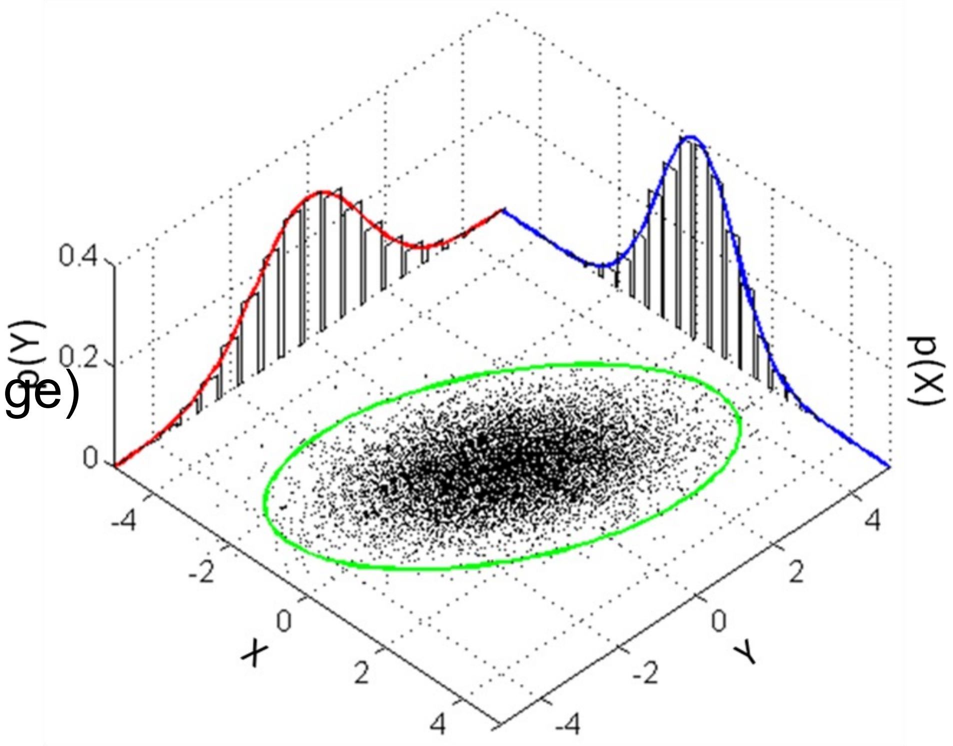
Useful Properties

- Closed under additivity (same as univariate case)
- Closed under linear functions,

$$AX + b \sim \mathcal{N}(A\mu_x + b, A\Sigma A^T)$$

Where $A \in \mathcal{R}^{m \times d}$ and $b \in \mathcal{R}^m$(output dimensions may change)

- Closed under conditioning and marginalization



Sampling

Assume that we have a random variable X *representing height of male people in Crete.*

Let's say that $X \sim N(\mu, \sigma^2)$

We draw N ***random*** *samples from that distribution –i.i.d.*

This is like drawing $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$

Problems with sampling (i.i.d)

- *-Sampling without replacement*
- *-Sampling biases*

Sample Mean

Sample mean:

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

What is the expectation of the sample mean?
What is the variance of the sample mean?

What is the distribution of the sample mean?

Sequence of Random Variables

A sequence of random variables is a sequence of functions.

Example: Toss a fair coin once. Define

$$X_n = \begin{cases} \frac{1}{n+1}, & if\ heads \\ 1, & otherwise \end{cases}$$

a. Find the PMF and CDF of X_n for $n = 1, 2, 3, \ldots$
b. As $n \to \infty$, what does $F(X_n)$ look like?
c. Are they independent?

Sequence of Random Variables

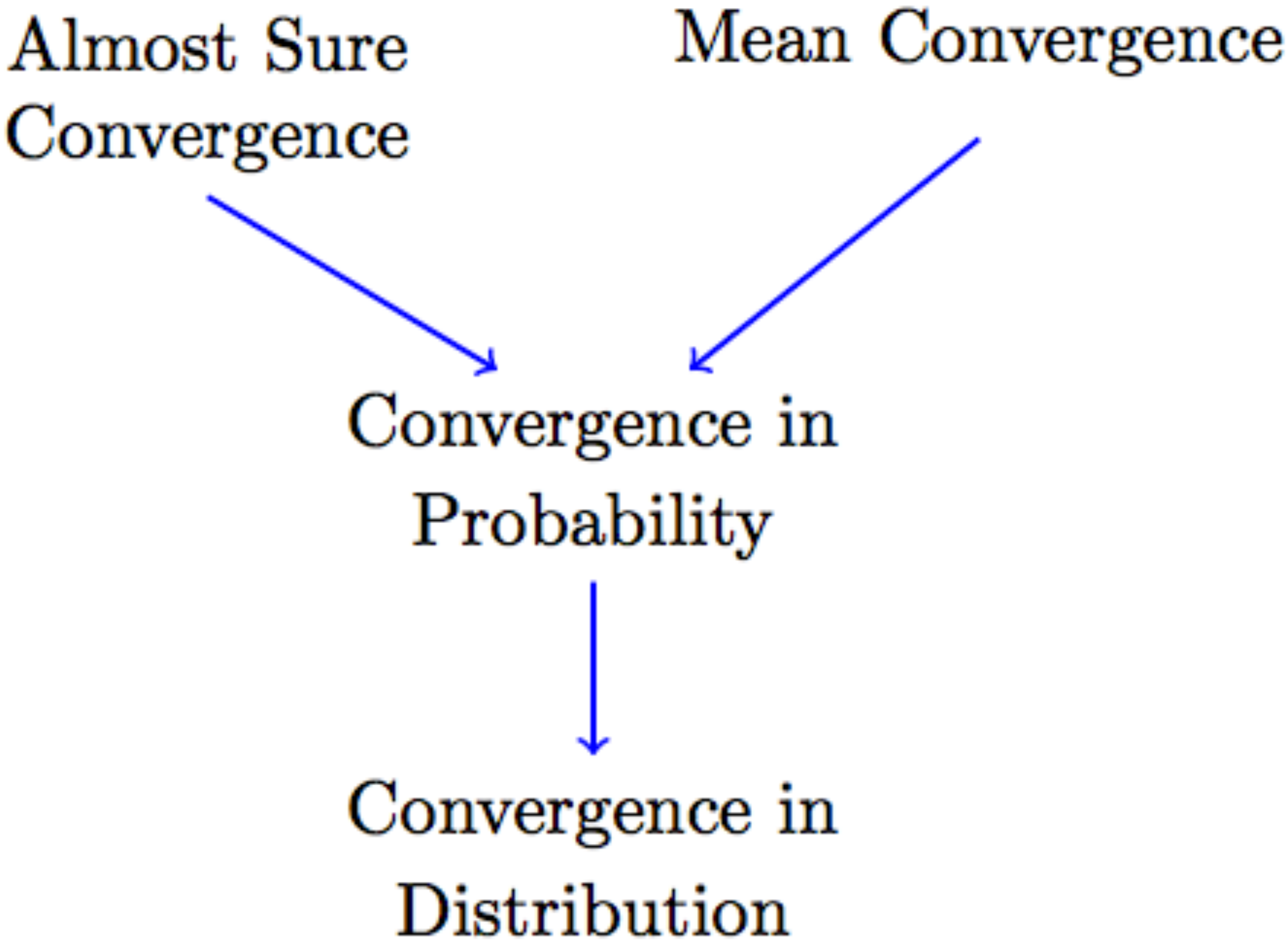
A sequence of random variables is a sequence of functions.

Example: Toss a fair coin forever. Define

$$X_n = \begin{cases} 1, & if\ n-th\ toss\ is\ heads \\ 0, & otherwise \end{cases}$$

a. Find the PMF and CDF of X_n *for* $n = 1, 2, 3, \ldots$

Convergence of RVs



Convergence In Probability

Arithmetic Convergence	Convergence in Probability
Sequence a_n of numbers converges to number l	Sequence X_n of random variables converges to random variable X
$\lim_{n\to\infty} \alpha_n = l$ or $\alpha_n \to l$	$X_n \xrightarrow{p} X$
a_n gets arbitrarily close to l	The probability distribution of $\lvert X_n - X \rvert$ gets more concentrated around 0.
$\forall \epsilon > 0, \exists n_o : \forall n > n_0 \quad \lvert \alpha_n - l \rvert < \epsilon$	$\forall \epsilon > 0, \lim_{n\to\infty} P(\lvert X_n - X \rvert < \epsilon) = 1$

Convergence in distribution

A sequence of random variables $X_1, X_2, X_3, \cdots$ *converges in distribution*

to a random variable X*, shown by* $X_n \xrightarrow{d} X$ *, then*

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x),$$

for all x *at which* $F_X(x)$ *is continuous.*

Weak Law of Large Number

Draw n samples from a distribution with $E[X] = \mu$

Let $\overline{\mathrm{X}}_{\mathrm{n}}$ be the sample mean

The sample mean is an RV

The sample means for $\mathrm{n} = 1, 2, \ldots$, form a sequence of random variables.

Weak law of large numbers:

$$\overline{\mathrm{X}}_{\mathrm{n}} \xrightarrow{p} \mu$$

Central Limit Theorem

- $S_n = \sum_{i=1}^{n} X_i$, mean $n\mu$, variance $n\sigma^2$.
- $\bar{X}_n = \frac{S_n}{n}$, mean μ variance $\frac{\sigma^2}{n}$.
- $\frac{S_n}{\sqrt{n}}$, mean $\mu\sqrt{n}$, variance σ^2
- $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$, mean 0 , variance 1 .

Central Limit Theorem

Let $X_1, X_2, \ldots, X\_n$ *be i.i.d. random variables with expected value* $E[X_i\ ] = \mu < \infty$ *and variance* $0 < Var(X_i) = \sigma^2 < \infty$*.*

Then, the random variable $Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$ *converges in distribution to the standard normal random variable as* n *goes to infinity:*

$$\lim_{n \to \infty} P(Z_n \le x) = \Phi(x), \text{ for all } x \in \mathbb{R}$$

CLT - Applications

- *You are doing a poll on "ratio of people agreeing with s".*
- *True ratio:* p, *estimate:* $\bar{X}_n$ *after asking* n *people.*
- *You want* $|\bar{X}_n - p|$ *to be small (e.g., less than 1%)*
- *You want to estimate*

$$P(|\bar{X}_n - p| \geq 0.01) \leq 0.05$$

Apply CLT:

Recap

Random Variables can be discrete or continuous

Often, we describe their distributions using families of parametric probability distributions $P(x; \boldsymbol{\theta})$

- *The height of a student is approximately normal with mean* θ_1 *and some variance* θ_2

- *The number of people that have a disease out of a group of* N *people follows the Binomial(*N, θ*) distribution.*

In practice, we do not know $\boldsymbol{\theta}$

Statistical Inference

What can we infer about $\boldsymbol{\theta}$ *given the observed data?*

Assuming we observe random variables $X_1, \ldots, X_n$ *following some distribution with parameter* θ*, what conclusions can we draw about parameter* θ *?*

Example

Say I take a random sample of 100 people and test them all for a disease. If 3 of them have the disease, what can I say about $\theta =$ *the prevalence of the disease in the population?*

Tasks

- *Point Estimation (estimate* $\boldsymbol{\theta}$*)*
- *Estimate confidence bounds for my estimate*
- *Hypothesis testing (is* $\boldsymbol{\theta}$*>0.2)?*
- *Prediction: For a given person, if know that their age/comorbidities/lifestyle habbits, can I predict their probability of having the disease?*

Approaches to Statistical Inferences

Frequentist approach

Parameters are numbers, I will try to identify the most likely number given my data.

Bayesian Approach

Parameters are numbers, but I have uncertainty about them, so I will treat them like random variables, that have distributions.

Next: Point estimation with Frequentist/Bayesian approach