CS70: Lecture25.

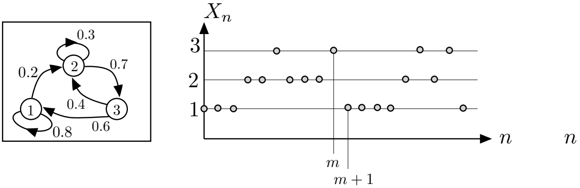
Markov Chains 1.5

1. Review
2. Distribution
3. Irreducibility
4. Convergence

Review

- Markov Chain:
  - Finite set $\mathscr{X}$; π_0; $P = \{P(i,j), i,j \in \mathscr{X}\}$;
  - $Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X}$
  - $Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i,j \in \mathscr{X}, n \geq 0$.
  - Note:
    $Pr[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \pi_0(i_0)P(i_0,i_1)\cdots P(i_{n-1},i_n)$.
- First Passage Time:
  - $A \cap B = \emptyset; \beta(i) = E[T_A | X_0 = i]; \alpha(i) = P[T_A < T_B | X_0 = i]$
  - $\beta(i) = 1 + \sum_j P(i,j)\beta(j)$;
  - $\alpha(i) = \sum_j P(i,j)\alpha(j)$. $\alpha(A) = 1, \alpha(B) = 0$.

Distribution of X_n



Recall π_n is a distribution over states for X_n.

Stationary distribution: $\pi = \pi P$.

Distribution over states is the same before/after transition.

probability entering i: $\sum_{i,j} P(j,i)\pi(j)$.

probability leaving i: π_i.

are Equal!

Distribution same after one step.

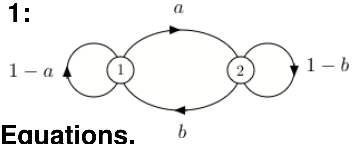
Questions? Does one exist? Is it unique?

If it exists and is unique. Then what?

Sometimes the distribution as $n \to \infty$

Stationary: Example

Example 1:



$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

Balance Equations.

$$\begin{aligned}
\pi P = \pi \quad & \Leftrightarrow \quad [\pi(1), \pi(2)] \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = [\pi(1), \pi(2)] \\
& \Leftrightarrow \quad \pi(1)(1-a) + \pi(2)b = \pi(1) \text{ and } \pi(1)a + \pi(2)(1-b) = \pi(2) \\
& \Leftrightarrow \quad \pi(1)a = \pi(2)b.
\end{aligned}$$

These equations are redundant! We have to add an equation: $\pi(1) + \pi(2) = 1$. Then we find

$$\pi = [\frac{b}{a+b}, \frac{a}{a+b}].$$

Stationary distributions: Example 2

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\pi P = \pi \Leftrightarrow [\pi(1), \pi(2)] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\pi(1), \pi(2)] \Leftrightarrow \pi(1) = \pi(1) \text{ and } \pi(2) = \pi(2).$$

Every distribution is invariant for this Markov chain. This is obvious, since $X_n = X_0$ for all n. Hence, $Pr[X_n = i] = Pr[X_0 = i], \forall (i, n)$.

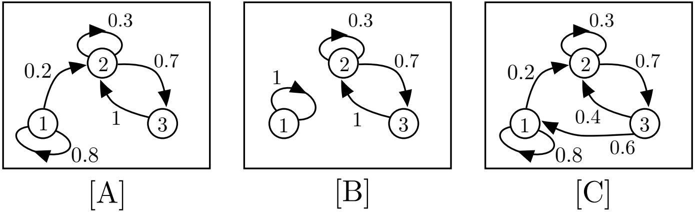
Discussion.
We have seen a chain with one stationary,
and a chain with many.

When is here just one?

Irreducibility.

Definition A Markov chain is irreducible if it can go from every state i to every state j (possibly in multiple steps).

Examples:



[A] is not irreducible. It cannot go from (2) to (1).

[B] is not irreducible. It cannot go from (2) to (1).

[C] is irreducible. It can go from every i to every j.

If you consider the graph with arrows when $P(i,j) > 0$, irreducible means that there is a single connected component.

Existence and uniqueness of Invariant Distribution

Theorem A finite irreducible Markov chain has one and only one invariant distribution.

That is, there is a unique positive vector $\pi = [\pi(1), \ldots, \pi(K)]$ such that $\pi P = \pi$ and $\sum_k \pi(k) = 1$.

Ok. Now.
Only one stationary distribution if irreducible (or connected.)

Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π.

Then, for all i,

$$\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \rightarrow \pi(i), \text{ as } n \rightarrow \infty.$$

The left-hand side is the fraction of time that $X_m = i$ during steps $0, 1, \ldots, n-1$. Thus, this fraction of time approaches $\pi(i)$.

Proof: Lecture note 24 gives a plausibility argument.

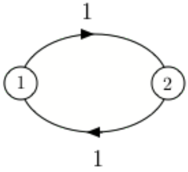
□

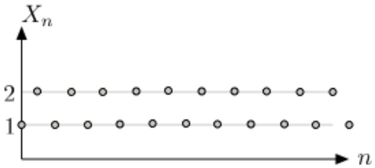
Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π. Then, for all i, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

Example 1:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \pi P = \pi \Rightarrow \pi = [1/2, 1/2]$$



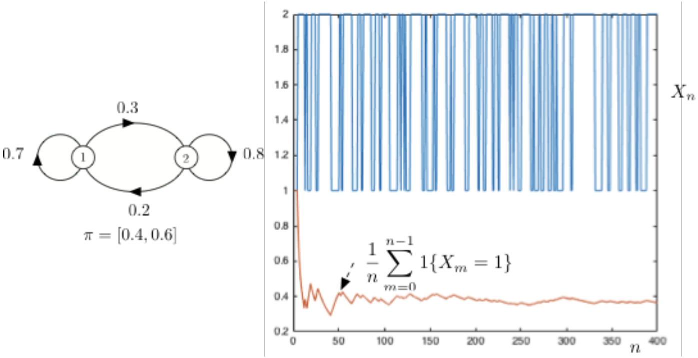


The fraction of time in state 1 converges to $1/2$, which is $\pi(1)$.

Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π. Then, for all i, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

Example 2:

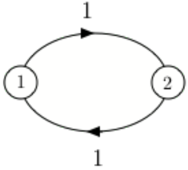


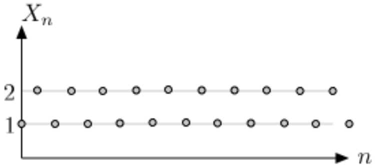
Convergence to Invariant Distribution

Question: Assume that the MC is irreducible. Does π_n approach the unique invariant distribution π?

Answer: Not necessarily. Here is an example:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \pi P = \pi \Rightarrow \pi = [1/2, 1/2]$$





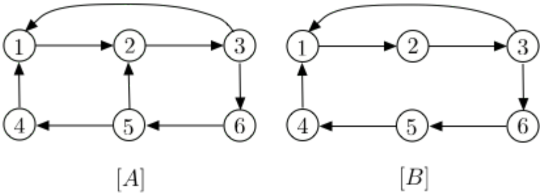
Assume $X_0 = 1$. Then $X_1 = 2, X_2 = 1, X_3 = 2, \ldots$.

Thus, if $\pi_0 = [1,0]$, $\pi_1 = [0,1], \pi_2 = [1,0], \pi_3 = [0,1]$, etc.

Hence, π_n does not converge to $\pi = [1/2, 1/2]$.
Notice, all cycles or closed walks have even length.

Periodicity

Definition: Periodicity is gcd of the lengths of all closed walks.
Previous example: 2.
Definition If periodicity is 1, Markov chain is said to be aperiodic.
Otherwise, it is periodic.
Example



[A]: Closed walks of length 3 and length 4 $\implies$ periodicity = 1.

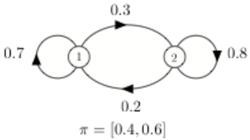
[B]: All closed walks multiple of 3 $\implies$ periodicity =2.

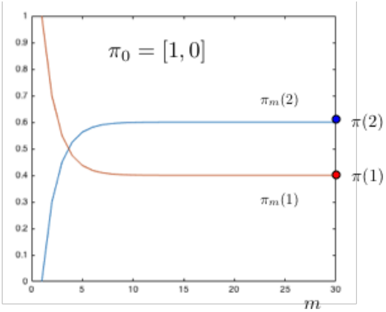
Convergence of π_n

Theorem Let X_n be an irreducible and aperiodic Markov chain with invariant distribution π. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

Example



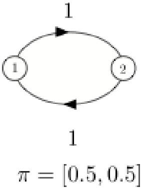


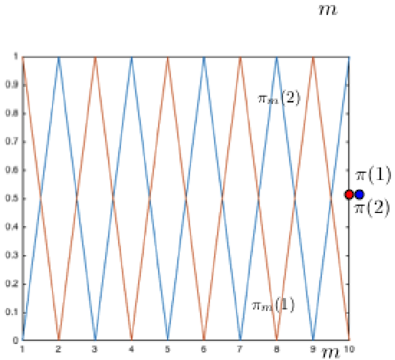
Convergence of π_n

Theorem Let X_n be an irreducible and aperiodic Markov chain with invariant distribution π. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

Example





Summary

Markov Chains

- Markov Chain: $Pr[X_{n+1} = j | X_0, \ldots, X_n = i] = P(i, j)$
- FSE: $\beta(i) = 1 + \sum_j P(i,j)\beta(j); \alpha(i) = \sum_j P(i,j)\alpha(j)$.
- $\pi_n = \pi_0 P^n$
- π is invariant iff $\pi P = \pi$
- Irreducible $\Rightarrow$ one and only one invariant distribution π
- Irreducible $\Rightarrow$ fraction of time in state i approaches $\pi(i)$
- Irreducible + Aperiodic $\Rightarrow \pi_n \to \pi$.
- Calculating π: One finds $\pi = [0, 0 \ldots, 1]Q^{-1}$ where $Q = \cdots$.

CS70: Continuous Probability.

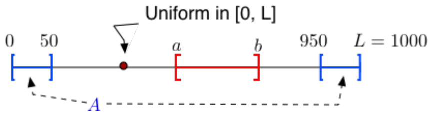
Continuous Probability 1

1. Examples
2. Events
3. Continuous Random Variables

Uniformly at Random in $[0,1]$.

Choose a real number X, uniformly at random in $[0,1]$.

What is the probability that X is exactly equal to $1/3$? Well, ..., 0.



What is the probability that X is exactly equal to 0.6? Again, 0.

In fact, for any $x \in [0,1]$, one has $Pr[X = x] = 0$.

How should we then describe 'choosing uniformly at random in $[0,1]$'?

Here is the way to do it:

$$Pr[X \in [a,b]] = b - a, \forall 0 \leq a \leq b \leq 1.$$

Makes sense: $b - a$ is the fraction of $[0,1]$ that $[a,b]$ covers.

Uniformly at Random in $[0,1]$.

Let $[a,b]$ denote the **event** that the point X is in the interval $[a,b]$.

$$Pr[[a,b]] = \frac{\text{length of } [a,b]}{\text{length of } [0,1]} = \frac{b-a}{1} = b-a.$$

Intervals like $[a,b] \subseteq \Omega = [0,1]$ are **events.**
More generally, events in this space are unions of intervals.
Example: the event A - "within 0.2 of 0 or 1" is $A = [0,0.2] \cup [0.8,1]$.
Thus,

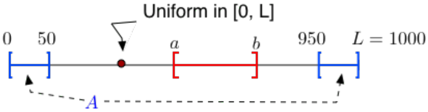
$$Pr[A] = Pr[[0,0.2]] + Pr[[0.8,1]] = 0.4.$$

More generally, if A_n are pairwise disjoint intervals in $[0,1]$, then

$$Pr[\cup_n A_n] := \sum_n Pr[A_n].$$

Many subsets of $[0,1]$ are of this form. Thus, the probability of those sets is well defined. We call such sets events.

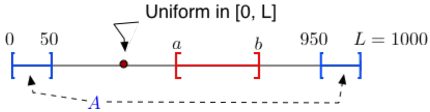
Uniformly at Random in $[0,1]$.



Note: A **radical** change in approach.
Finite prob. space: $\Omega = \{1,2,\ldots,N\}$, with $Pr[\omega] = p_\omega$.
$\implies$ $Pr[A] = \sum_{\omega \in A} p_\omega$ for $A \subset \Omega$.

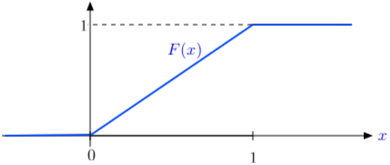
Continuous space: e.g., $\Omega = [0,1]$,
$Pr[\omega]$ is typically 0.
Instead, start with $Pr[A]$ for some events A.
Event A = interval, or union of intervals.

Uniformly at Random in $[0, 1]$.



$Pr[X \leq x] = x$ for $x \in [0, 1]$. Also, $Pr[X \leq x] = 0$ for $x < 0$.
$Pr[X \leq x] = 1$ for $.2x > 1$.

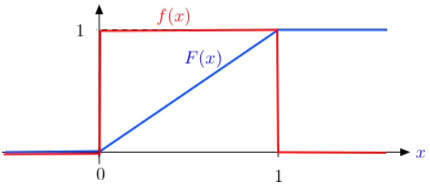
Define $F(x) = Pr[X \leq x]$.



Then we have $Pr[X \in (a, b]] = Pr[X \leq b] - Pr[X \leq a] = F(b) - F(a)$.

Thus, $F(\cdot)$ specifies the probability of all the events!

Uniformly at Random in $[0,1]$.



$$Pr[X \in (a,b]] = Pr[X \le b] - Pr[X \le a] = F(b) - F(a).$$

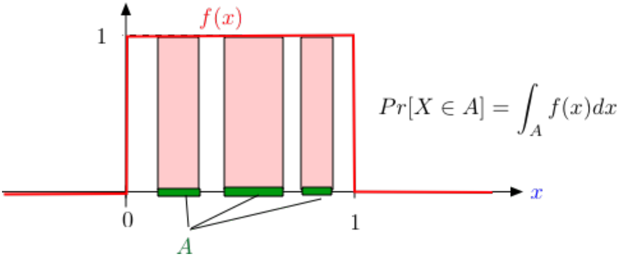
An alternative view is to define $f(x) = \frac{d}{dx}F(x) = 1\{x \in [0,1]\}$. Then

$$F(b) - F(a) = \int_a^b f(x)dx.$$

Thus, the probability of an event is the integral of $f(x)$ over the event:

$$Pr[X \in A] = \int_A f(x)dx.$$

Uniformly at Random in $[0, 1]$.



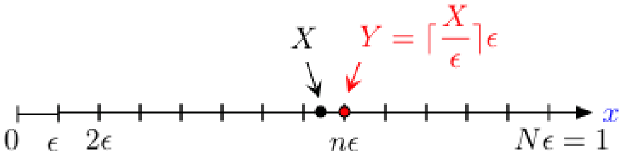
Think of $f(x)$ as describing how
one unit of probability is spread over $[0, 1]$: uniformly!

Then $Pr[X \in A]$ is the probability mass over A.

Observe:

- This makes the probability automatically additive.
- We need $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Uniformly at Random in $[0,1]$.



Discrete Approximation: Fix $N \gg 1$ and let $\varepsilon = 1/N$.

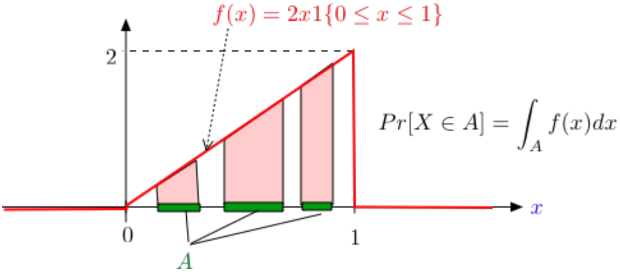
Define $Y = n\varepsilon$ if $(n-1)\varepsilon < X \leq n\varepsilon$ for $n = 1, \ldots, N$.

Then $|X - Y| \leq \varepsilon$ and Y is discrete: $Y \in \{\varepsilon, 2\varepsilon, \ldots, N\varepsilon\}$.

Also, $Pr[Y = n\varepsilon] = \frac{1}{N}$ for $n = 1, \ldots, N$.

Thus, X is ‘almost discrete.’

Nonuniformly at Random in $[0,1]$.



This figure shows a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x)dx = 1$.

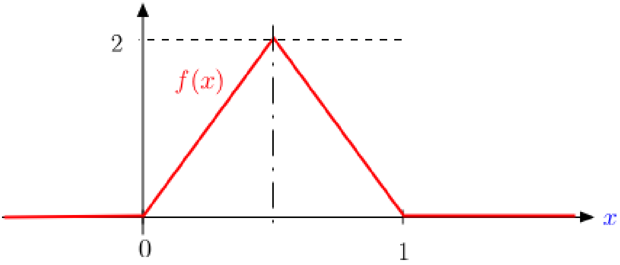
It defines another way of choosing X at random in $[0,1]$.

Note that X is more likely to be closer to 1 than to 0.

One has $Pr[X \leq x] = \int_{-\infty}^{x} f(u)du = x^2$ for $x \in [0,1]$.

Also, $Pr[X \in (x, x+\varepsilon)] = \int_{x}^{x+\varepsilon} f(u)du \approx f(x)\varepsilon$.

Another Nonuniform Choice at Random in $[0, 1]$.



This figure shows yet a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x)dx = 1$.

It defines another way of choosing X at random in $[0, 1]$.

Note that X is more likely to be closer to $1/2$ than to 0 or 1.

For instance, $Pr[X \in [0, 1/3]] = \int_0^{1/3} 4xdx = 2\left[x^2\right]_0^{1/3} = \frac{2}{9}$.

Thus, $Pr[X \in [0, 1/3]] = Pr[X \in [2/3, 1]] = \frac{2}{9}$ and $Pr[X \in [1/3, 2/3]] = \frac{5}{9}$.

General Random Choice in $\Re$

Let $F(x)$ be a nondecreasing function with $F(-\infty) = 0$ and $F(+\infty) = 1$.

Define X by $Pr[X \in (a, b]] = F(b) - F(a)$ for $a < b$. Also, for $a_1 < b_1 < a_2 < b_2 < \cdots < b_n$,

$$\begin{aligned} Pr[X \in (a_1, b_1] \cup (a_2, b_2] \cup (a_n, b_n]] \\ &= Pr[X \in (a_1, b_1]] + \cdots + Pr[X \in (a_n, b_n]] \\ &= F(b_1) - F(a_1) + \cdots + F(b_n) - F(a_n). \end{aligned}$$

Let $f(x) = \frac{d}{dx} F(x)$. Then,

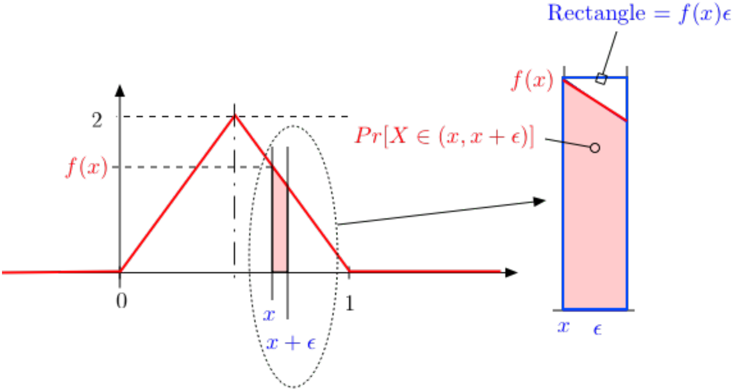
$$Pr[X \in (x, x + \varepsilon]] = F(x + \varepsilon) - F(x) \approx f(x)\varepsilon.$$

Here, $F(x)$ is called the cumulative distribution function (cdf) of X and $f(x)$ is the probability density function (pdf) of X.

To indicate that F and f correspond to the RV X, we will write them $F_X(x)$ and $f_X(x)$.

$Pr[X \in (x, x+\varepsilon)]$

An illustration of $Pr[X \in (x, x+\varepsilon)] \approx f_X(x)\varepsilon$:



Thus, the pdf is the ‘local probability by unit length.’

It is the ‘probability density.’

Discrete Approximation

Fix $\varepsilon \ll 1$ and let $Y = n\varepsilon$ if $X \in (n\varepsilon, (n+1)\varepsilon]$.

Thus, $Pr[Y = n\varepsilon] = F_X((n+1)\varepsilon) - F_X(n\varepsilon)$.

Note that $|X - Y| \leq \varepsilon$ and Y is a discrete random variable.

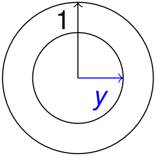
Also, if $f_X(x) = \frac{d}{dx} F_X(x)$, then $F_X(x+\varepsilon) - F_X(x) \approx f_X(x)\varepsilon$.

Hence, $Pr[Y = n\varepsilon] \approx f_X(n\varepsilon)\varepsilon$.

Thus, we can think of X of being almost discrete with $Pr[X = n\varepsilon] \approx f_X(n\varepsilon)\varepsilon$.

Example: CDF

Example: hitting random location on gas tank.
Random location on circle.



Random Variable: Y distance from center.
Probability within y of center:

$$
\begin{aligned}
Pr[Y \leq y] &= \frac{\text{area of small circle}}{\text{area of dartboard}} \\
&= \frac{\pi y^2}{\pi} = y^2.
\end{aligned}
$$

Hence,

$$
F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}
$$

Calculation of event with dartboard..

Probability between .5 and .6 of center?
Recall CDF.

$$F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

$$\begin{aligned} Pr[0.5 < Y \leq 0.6] &= Pr[Y \leq 0.6] - Pr[Y \leq 0.5] \\ &= F_Y(0.6) - F_Y(0.5) \\ &= .36 - .25 \\ &= .11 \end{aligned}$$

PDF.

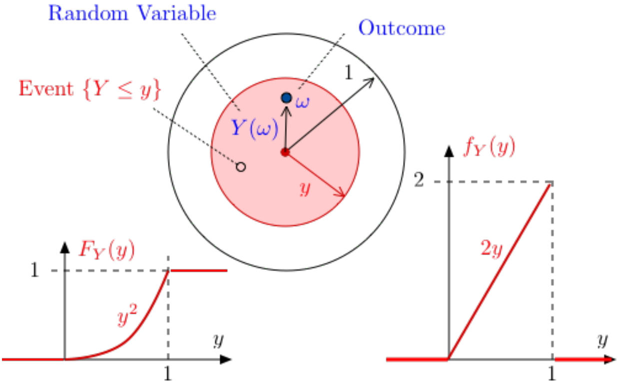
Example: “Dart” board.
Recall that

$$F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

$$f_Y(y) = F'_Y(y) = \begin{cases} 0 & \text{for } y < 0 \\ 2y & \text{for } 0 \leq y \leq 1 \\ 0 & \text{for } y > 1 \end{cases}$$

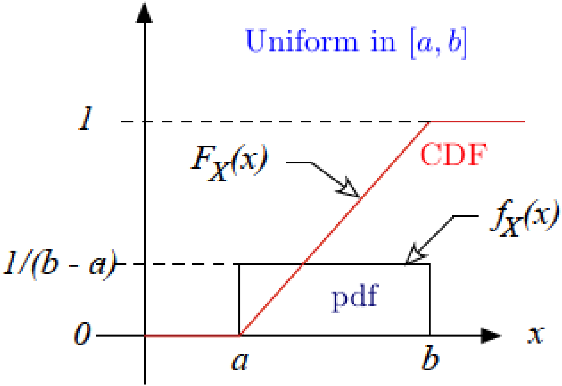
The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.
Use whichever is convenient.

Target



Random Variable
Outcome
Event $\{Y \leq y\}$
1
ω
$Y(\omega)$
y
$f_Y(y)$
2
$2y$
y
1
$F_Y(y)$
1
y^2
y
1

$U[a, b]$

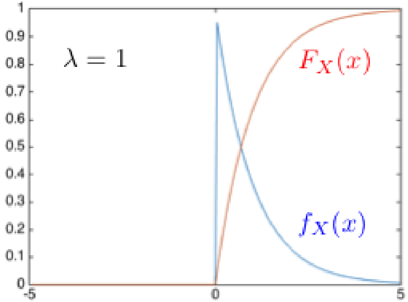


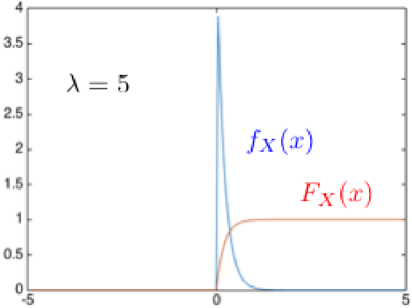
$Expo(\lambda)$

The exponential distribution with parameter $\lambda > 0$ is defined by

$$f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$





Note that $Pr[X > t] = e^{-\lambda t}$ for $t > 0$.

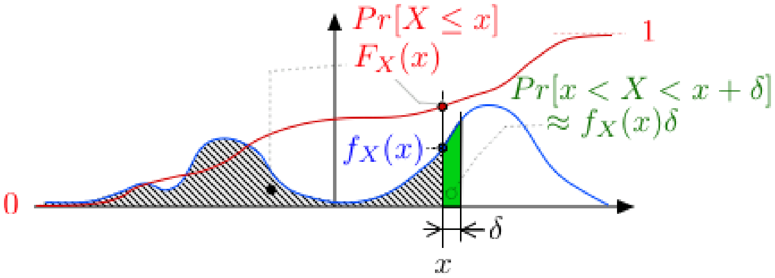
Continuous Random Variables

Continuous random variable X, specified by

1. $F_X(x) = Pr[X \leq x]$ for all x.
   Cumulative Distribution Function (cdf).
   $Pr[a < X \leq b] = F_X(b) - F_X(a)$
   1.1 $0 \leq F_X(x) \leq 1$ for all $x \in \Re$.
   1.2 $F_X(x) \leq F_X(y)$ if $x \leq y$.
2. Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ or $f_X(x) = \frac{d(F_X(x))}{dx}$.
   Probability Density Function (pdf).
   $Pr[a < X \leq b] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$
   2.1 $f_X(x) \geq 0$ for all $x \in \Re$.
   2.2 $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Recall that $Pr[X \in (x, x+\delta)] \approx f_X(x)\delta$.
X “takes” value $n\delta$, for $n \in Z$, with $Pr[X = n\delta] = f_X(n\delta)\delta$

A Picture



The pdf $f_X(x)$ is a nonnegative function that integrates to 1.

The cdf $F_X(x)$ is the integral of f_X.

$$Pr[x < X < x + \delta] \approx f_X(x)\delta$$

$$Pr[X \leq x] = F_x(x) = \int_{-\infty}^{x} f_X(u)du$$

Multiple Continuous Random Variables

One defines a pair (X, Y) of continuous RVs by specifying $f_{X,Y}(x, y)$ for $x, y \in \Re$ where

$$f_{X,Y}(x,y)dxdy = Pr[X \in (x, x+dx), Y \in (y+dy)].$$

The function $f_{X,Y}(x, y)$ is called the joint pdf of X and Y.

Example: Choose a point (X, Y) uniformly in the set $A \subset \Re^2$. Then

$$f_{X,Y}(x,y) = \frac{1}{|A|} 1\{(x,y) \in A\}$$

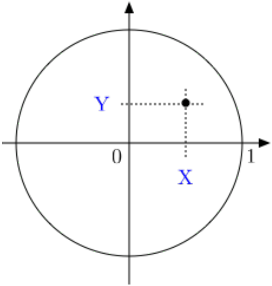
where $|A|$ is the area of A.

Interpretation. Think of (X, Y) as being discrete on a grid with mesh size ε and $Pr[X = m\varepsilon, Y = n\varepsilon] = f_{X,Y}(m\varepsilon, n\varepsilon)\varepsilon^2$.

Extension: $\mathbf{X} = (X_1, \ldots, X_n)$ with $f_{\mathbf{X}}(\mathbf{x})$.

Example of Continuous (X, Y)

Pick a point (X, Y) uniformly in the unit circle.



Thus, $f_{X,Y}(x,y) = \frac{1}{\pi} 1\{x^2 + y^2 \leq 1\}$.

Consequently,

$$Pr[X > 0, Y > 0] = \frac{1}{4}$$

$$Pr[X < 0, Y > 0] = \frac{1}{4}$$

$$Pr[X^2 + Y^2 \leq r^2] = \frac{r^2}{\pi}$$

$$Pr[X > Y] = \frac{1}{2}.$$

Summary

Continuous Probability 1

1. pdf: $Pr[X \in (x, x+\delta]] = f_X(x)\delta$.
2. CDF: $Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$: $f_X(x) = \frac{1}{b-a} 1\{a \leq x \leq b\}; F_X(x) = \frac{x-a}{b-a}$ for $a \leq x \leq b$.
4. $Expo(\lambda)$:
   $f_X(x) = \lambda \exp\{-\lambda x\} 1\{x \geq 0\}; F_X(x) = 1 - \exp\{-\lambda x\}$ for $x \leq 0$.
5. Target: $f_X(x) = 2x1\{0 \leq x \leq 1\}; F_X(x) = x^2$ for $0 \leq x \leq 1$.
6. Joints: Is this 4/20?
   Joint pdf: $Pr[X \in (x, x+\delta), Y = (y, y+\delta)) = f_{X,Y}(x,y)\delta^2$.