Independence

Definition (two events):

$$\begin{array}{rl} B \text{ and } C \text{ are independent if} & P(B \cap C) = P(B)\,P(C) \\ \text{or equivalently} & P(B\,|\,C) = P(B) \\ \text{or equivalently} & P(C\,|\,B) = P(C)\,. \end{array}$$

Definition (three events):
A, B, C are mutually independent if <u>all</u> of the following are true:

$$\begin{aligned} P(A \cap B) &= P(A)\,P(B) \\ P(A \cap C) &= P(A)\,P(C) \\ P(B \cap C) &= P(B)\,P(C) \\ P(A \cap B \cap C) &= P(A)\,P(B)\,P(C)\,. \end{aligned}$$

To verify mutual independence, you must check all of these.

General Definition:
A list of events is **mutually independent** if the probabilities of <u>all</u> possible intersections of these events are given by multiplying the individual event probabilities.

Note: For k events, there are $2^k - 1 - k$ possible intersections.

The word "mutual" is usually omitted. In this class, when we say "independent" events, we mean "mutually independent".

In our class, independence is usually an assumption; not something we verify.

Definition: Events $A_1, A_2, \ldots, A_k$ are **pairwise** independent if $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$.

To verify mutual independence, you must check <u>all</u> possible intersections, <u>not</u> just pairs.

<u>Example</u>: To show that pairwise independence does <u>not</u> imply mutual independence.

Suppose:

Coin with sides labeled $+1, -1$.

$P(\text{toss is } +1) = P(\text{toss is } -1) = 1/2$.

Toss the coin twice. (Tosses are independent.)

Let X_1 = result of first toss,

X_2 = result of second toss.

For this experiment

$$\Omega = \{(x_1, x_2) : x_i = \pm 1\},$$

$$\#(\Omega) = 4, \quad P(\omega) = \frac{1}{4} \text{ for all } \omega.$$

Define $X_3 = X_1 X_2$ (so that $X_3 = \pm 1$).

Define the events $A_i = \{X_i = +1\}$.

Fact: A_1, A_2, A_3 are pairwise indep., but <u>not</u> mutually independent.

Proof:

ω	$P(\omega)$	X_1	X_2	X_3
ω_1	1/4	+1	+1	+1
ω_2	1/4	+1	−1	−1
ω_3	1/4	−1	+1	−1
ω_4	1/4	−1	−1	+1

$A_1 = \{\omega_1, \omega_2\}$, $A_2 = \{\omega_1, \omega_3\}$, $A_3 = \{\omega_1, \omega_4\}$.

Thus $P(A_1) = P(A_2) = P(A_3) = 1/2$.

A_1 and A_3 are independent because $A_1 \cap A_3 = \{\omega_1\}$ so $P(A_1 \cap A_3) = 1/4$ which agrees with $P(A_1)P(A_3) = \frac{1}{2} \cdot \frac{1}{2}$
$= 1/4$.

Similarly A_1, A_2 are independent and A_2, A_3 are independent.

But $A_1 \cap A_2 \cap A_3 = \{\omega_1\}$ so that $P(A_1 \cap A_2 \cap A_3) = 1/4$ which <u>disagrees</u> with $P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$.

The events A_1, A_2, A_3 are <u>not</u> mutually independent.

Fact: If $A_1, \ldots, A_j, B_1, \ldots, B_K$ are mutually independent events, then any event defined in terms of $A_1, \ldots, A_j$ is independent of any event defined in terms of $B_1, \ldots, B_K$.

Example: If A, B, C, D are mut. indep., then $A \cup B$ is indep. of $C \cup D$. Also $A^c \cup B^c$ is indep. of $C \cap D$, etc.

How would you prove these facts? In the first, we must show

$$P((A \cup B) \cap (C \cup D)) = P(A \cup B)\, P(C \cup D).$$

This can be done in many ways. Here is one. By repeated use of the distributive property

$$(A \cup B) \cap (C \cup D) = AC \cup BC \cup AD \cup BD$$

where $AC \equiv (A \cap C)$, etc.

Now use Inclusion–Exclusion:

$P(AC \cup BC \cup AD \cup BD)$

$= P(AC) + P(BC) + P(AD) + P(BD)$

$\quad - P(ABC) - P(ACD) - P(ABCD)$

$\quad - P(ABCD) - P(BCD) - P(ABD)$

$\quad + P(ABCD) + P(ABCD) + P(ABCD)$

$\quad + P(ABCD) - P(ABCD)$

$= P(AC) + P(BC) + P(AD) + P(BD)$

$\quad - P(ABC) - P(ACD) - P(BCD) - P(ABD)$

$\quad + P(ABCD)$

Now note that $P(A \cup B)\,P(C \cup D)$

$= (P(A) + P(B) - P(AB))(P(C) + P(D) - P(CD))$

Expand the product and compare with earlier expression. They are the same since independence implies

$P(ABCD) = P(A)P(B)P(C)P(D)$

$\qquad = P(AB)P(CD)$, etc.

Independence versus Conditional Independence.

Example: Hat with two coins.

Coin #1 has $P(\text{heads}) = 1/4 = \pi_1$

Coin #2 " " $= 3/4 = \pi_2$

Pick a coin at random.
Toss it repeatedly.
Let $A_i = \{ i^{th} \text{ toss is heads} \}$.

Are A_1 and A_2 independent?

Answer: No!

Discussion: Let $B_j = \{\text{pick coin } j\}$.

B_1, B_2 is a partition of Ω.

Problem implicitly assumes

$$P(A_1 \cap A_2 \mid B_j) = P(A_1 \mid B_j)\, P(A_2 \mid B_j)$$
$$= \pi_j \cdot \pi_j = \pi_j^2 ,$$

$$P(A_1 \cap A_2 \cap A_3 \mid B_j)$$
$$= P(A_1 \mid B_j)\, P(A_2 \mid B_j)\, P(A_3 \mid B_j)$$
$$= \pi_j^3 ,$$

and more generally

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k} \mid B_j)$$

$$= \prod_{l=1}^{k} P(A_{i_l} \mid B_j) = \pi_j^{k} .$$

We say that the events $A_1, A_2, A_3, \ldots$ are <u>conditionally</u> <u>independent</u> given B_j.

Back to the original question:

$$P(A_1) = P(B_1)\,P(A_1 \mid B_1) + P(B_2)\,P(A_1 \mid B_2)$$

$$= \frac{1}{2} \cdot \underbrace{\pi_1}_{1/4} + \frac{1}{2} \cdot \underbrace{\pi_2}_{3/4} = \frac{1}{2}$$

Similarly $P(A_2) = \frac{1}{2}$.

$$P(A_1 \cap A_2) = P(B_1)\,P(A_1 \cap A_2 \mid B_1) + P(B_2)\,P(A_1 \cap A_2 \mid B_2)$$

$$= \frac{1}{2} \cdot \pi_1^2 + \frac{1}{2} \cdot \pi_2^2 = \frac{5}{16}$$

Since $\frac{5}{16} \neq \frac{1}{2} \cdot \frac{1}{2} = P(A_1)\,P(A_2)$, we see A_1 and A_2 are <u>not</u> independent.

Note that

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{5/16}{1/2}$$

$$= \frac{5}{8} > \frac{1}{2} = P(A_2).$$

Intuitively, $P(A_2|A_1) > P(A_2)$ because getting a head on the first toss makes it more likely you have drawn coin #2 which is biased towards heads.

Random Variables (r.v.'s)

Recall: A (probabilistic) **Experiment** is described by giving (Ω, P) where:

$$\begin{aligned} \Omega &= \text{sample space} \quad \text{(consisting of outcomes } \omega\text{)} \\ P &= \text{probability function} \end{aligned}$$

Informally: A random variable (rv) X is a quantity that achieves a value determined by the outcome of the experiment.

Formally: A **random variable** X is a rule which assigns a value $X(\omega)$ to each outcome ω in the sample space.

That is, a random variable (rv) is a function:

$$X : \Omega \to \mathbb{R}$$

Usually, the ω's are omitted and the value of the rv is written as X (instead of $X(\omega)$).

In our class, rv's are usually upper case letters late in the alphabet (U, V, W, X, Y, Z), but they don't have to be.

Example 1

Experiment: Toss pair of dice. (Record results.)

$$\Omega = \{(i,j) : 1 \le i \le 6, 1 \le j \le 6\}$$

$$P(A) = \frac{\#(A)}{\#(\Omega)}.$$

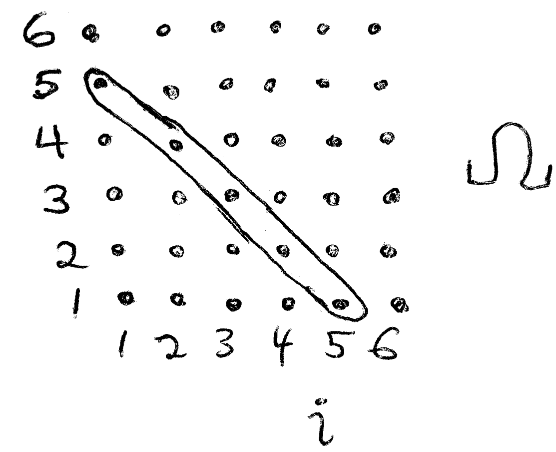
Some rv's $\qquad \omega = (i,j)$

X = # on first die $\qquad X(\omega) = i$

Y = # on second die $\qquad Y(\omega) = j$

Z = total on both die $\qquad Z(\omega) = i + j$

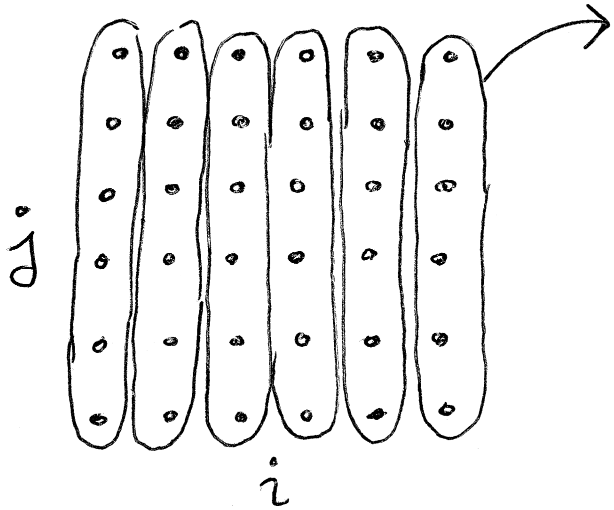
Notation for events



$\{Z = 6\}$ means $\{\omega : Z(\omega) = 6\} \subset \Omega$

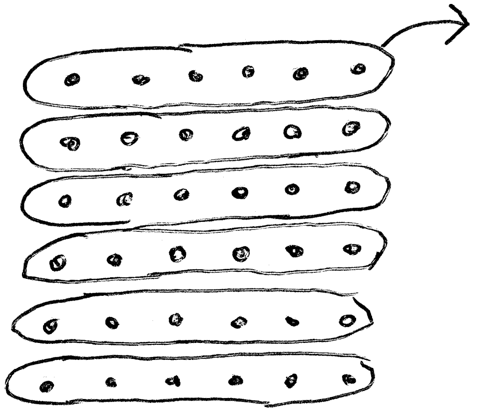
Every rv gives a partition of Ω.

In our example,



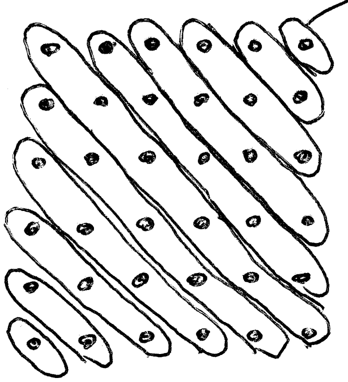
$\{X=6\}$

partition induced by X,



$\{Y=6\}$

partition induced by Y,



$\{Z=12\}$

partition induced by Z.

Definition: The induced probability function of r.v. X is P_X defined by

$$P_X(\underbrace{A}_{\substack{\text{a set of} \\ \text{real numbers}}}) = P(\underbrace{X \in A}_{\substack{\text{event} \\ \text{subset of } \Omega}}) = P(\omega : X(\omega) \in A).$$

In particular
$$P_X(\{x\}) = P(X = x).$$

Notation:

$\mathcal{X}$ = range of X = set of all possible values of X.

$P_X(\cdot)$ is a probability function defined for $A \subset \mathcal{X}$. (or $A \subset \mathbb{R}$)

$P(\cdot)$ is a prob. fn. defined for $A \subset \Omega$.

Example 1 (Dice)

X		r.v. Y		r.v. Z	
x	$P_X(\{x\})$	y	$P_Y(\{y\})$	z	$P_Z(\{z\})$
1	1/6	1	1/6	2	1/36
⋮	⋮	⋮	⋮	3	2/36
6	1/6	6	1/6	⋮	
				11	2/36
				12	1/36

<u>Definition</u> If $P(X \in A) = P(Y \in A)$ for all $A \subset \mathbb{R}$, (that is, $P_X = P_Y$), we say X and Y are <u>identically distributed</u> (or have <u>same distribution</u>).

X and Y above have $P_X = P_Y$.

$$P_Z((3,\infty)) = P(Z>3) = 33/36$$

Example 2

Experiment: Toss dart (uniformly) at random on circle of radius R.

$$\Omega = \{(x,y): x^2+y^2 \leq R^2\}$$

$$P(A) = \frac{\text{Area}(A)}{\text{Area}(\Omega)} \quad (\text{for } A \subset \Omega).$$

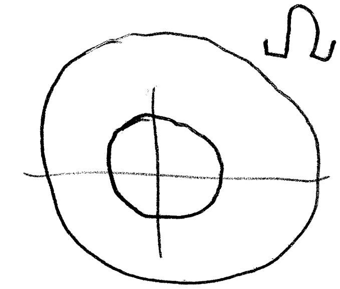
some RV's

$$\omega = (x,y)$$

Z = distance from center $\qquad Z(\omega) = \sqrt{x^2+y^2}$

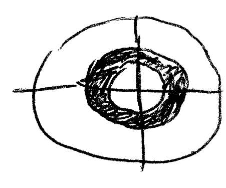
Q = quadrant dart lies in $\qquad Q(\omega) = \begin{cases} 1 & \text{if } x \geq 0, y \geq 0 \\ \vdots & \\ 4 & \text{if } x \geq 0, y < 0 \end{cases}$

some events

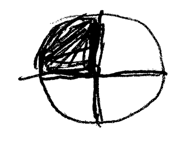


$$\{Z=1\} = \{\omega : Z(\omega) = 1\}$$

$$\{1 \leq Z \leq 2\} = \{\omega : 1 \leq Z(\omega) \leq 2\}$$



$$\{Q=2\} = \{\omega : Q(\omega) = 2\}$$



Example 2 (Dart)

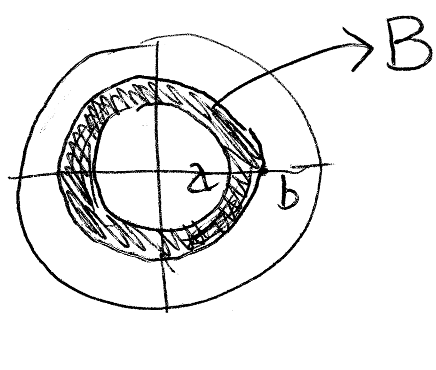
$$Q$$

q	$P_Q(\{q\})$
1	1/4
2	1/4
3	1/4
4	1/4

$P_Q(A)$ computed from table.

$$Z$$

$$P_Z(\underbrace{(a,b)}_{\subset \mathbb{R}}) = P(\underbrace{a<Z<b}_{B\subset\Omega}) = \frac{\text{Area}(B)}{\text{Area}(\Omega)}$$



$$= \frac{\pi b^2 - \pi a^2}{\pi R^2}$$

The distribution of <u>any</u> r.v. may be described by giving its

<u>induced prob. fn.</u> $P_X(A) = P(X \in A)$,

or its

<u>cdf</u> (cumulative dist fn.) $F_X(t) = P(X \leq t)$

$(= P_X((-\infty, t]))$.

<u>Types of r.v.'s</u>

X is <u>discrete</u> r.v. $\Rightarrow$ range of X (called $\mathcal{X}$) is <u>finite</u> or $\underbrace{\text{countable}}$.

infinite, but can be arranged in a sequence

$\mathcal{X} = \{x_1, x_2, x_3, \dots\}$.

The distn. of a discrete r.v. is usually described by its <u>probability</u> <u>mass</u> <u>function</u> (pmf):

$f_X(x) = P(X = x) = P_X(\{x\})$.

The cdf of a discrete rv is a

step function which has steps (jumps) at the values in $\mathcal{X}$ and is flat between these values.

X is a (absolutely) continuous r.v. means

X has a probability density function (pdf)

denoted $f_X(x)$ satisfying

$$P(a<X<b) = P_X((a,b)) = \int_a^b f_X(x)\,dx,$$

$$P(X \in A) = P_X(A) = \int_A f_X(x)\,dx.$$

For a continuous r.v., the cdf is a continuous function (no jumps),

Range $\mathcal{X}$ is uncountable, and $P(X=x)=0$ for all x.

Other types of r.v.'s

Mixed Continuous/Discrete : the distn. has both a discrete and a continuous component.

Exotic cases such as rv's which are $\underbrace{\text{continuous}}_{P(X=x)=0 \text{ for all } x}$, but $\underbrace{\text{not absolutely continuous}}_{X \text{ does not have a pdf.}}$.

Synopsis

Experiment (Ω, P)

Random Variable $X : \Omega \to \mathbb{R}$ (real numbers)

$$\omega \mapsto X(\omega)$$

$$P_X(A) \equiv P\{X \in A\}$$

$$F_X(t) \equiv P_X((-\infty, t]) = P\{X \le t\} \quad \underline{\underline{\text{cdf}}}$$

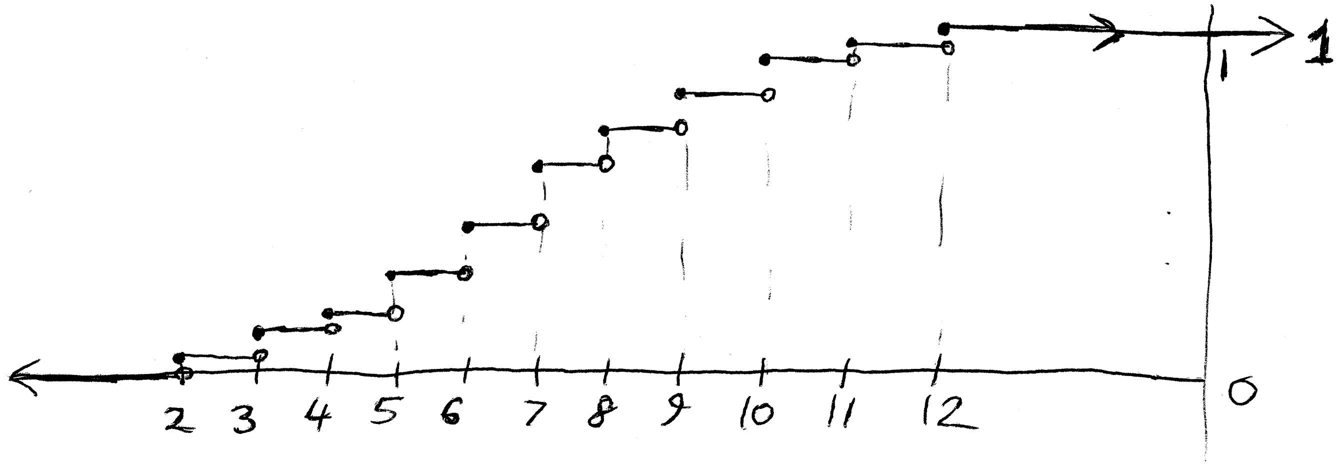
Examples

Z = total on 2 dice

pmf

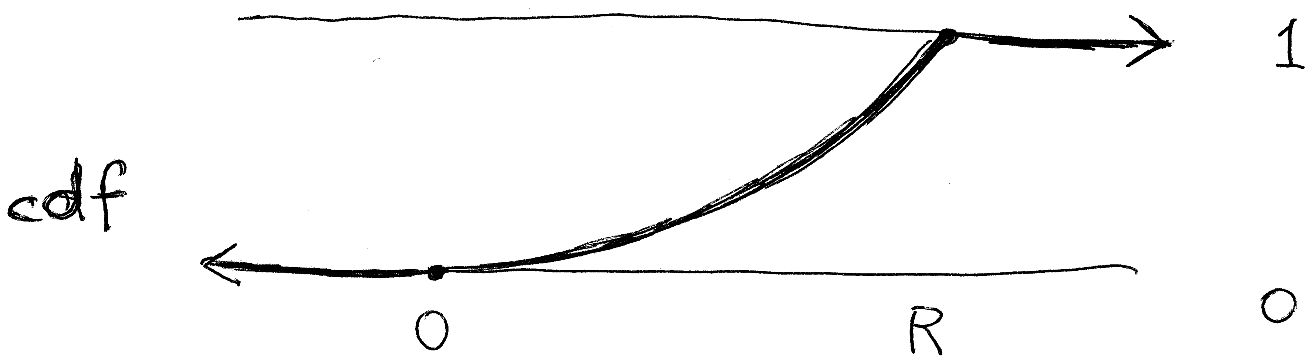
z	2	3	4	5	6	7	8	9	10	11	12
$f_Z(z)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

cdf



$X =$ distance of ω from center

dart location



X and Z are <u>bounded</u> random variables.

(range of X is bounded $[0,R]$
range of $Z = \{2,3,\dots,12\}$ is bounded.)

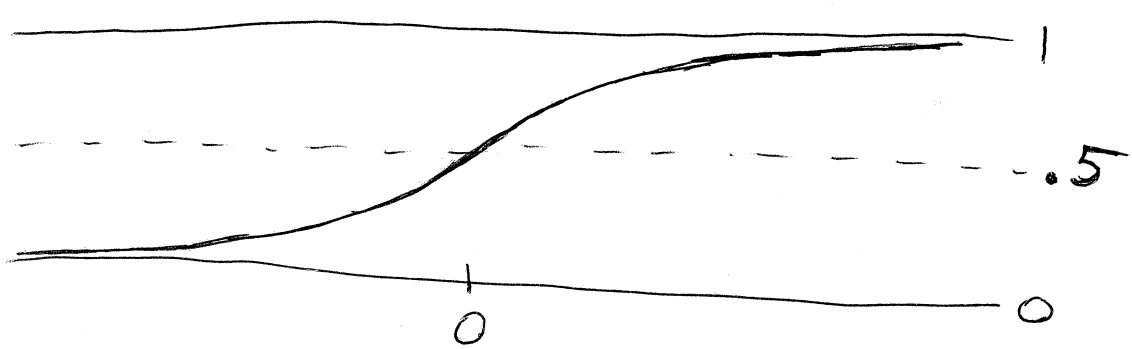
some unbounded examples are:

Standard Normal Random Variable

density (pdf) $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(\equiv \varphi(x) \right)$

cdf $F(x) = \int_{-\infty}^{x} f(y)\,dy \left(\equiv \Phi(x) \right)$

cdf

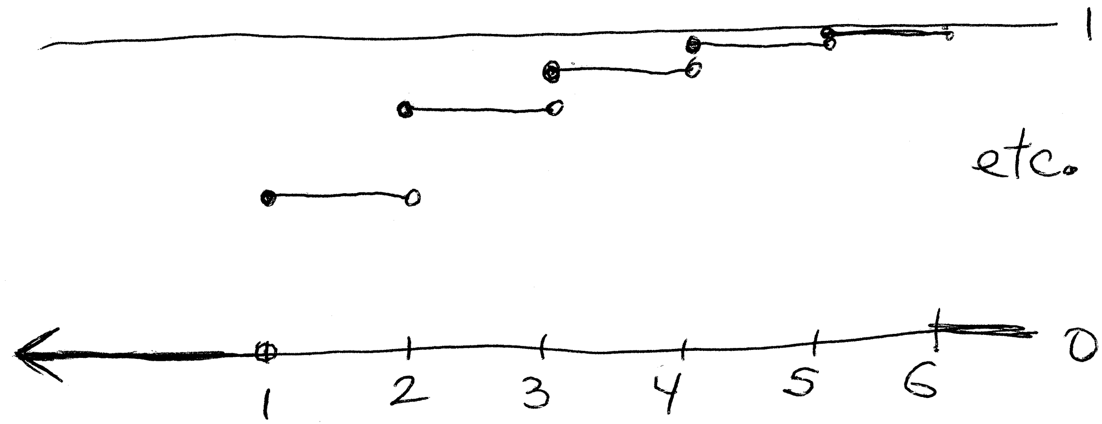


Geometric Random Variable

pmf $f(x) = \begin{cases} (1-p)^{x-1} p & \text{for } x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$

For $p = 1/2$, $f(x) = \left(\frac{1}{2}\right)^x$.

cdf



More on cumulative distribution functions (cdf's)

Recall the definition:

For a r.v. X, the cdf $F_X(\cdot)$ is

$$F_X(t) = P_X((-\infty, t]) = P(X \leq t).$$

Properties of cdf's

If F is the cdf of some r.v., then

(1) $F(-\infty) = 0$, $F(+\infty) = 1$.

(2) F is nondecreasing.

(3) F is right continuous.

Conversely, any function with properties 1, 2, 3 is the cdf of some random variable.

Details on 1, 2, 3

(1) says that

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1.$$

(2) says that

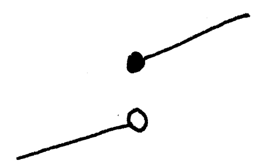
$x \le y$ implies $F(x) \le F(y)$.

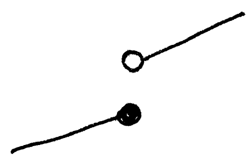
F can<u>not</u> decrease, but can have flat spots.

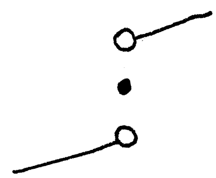
(3) says that

$$\lim_{x \downarrow y} F(x) = F(y) \quad \text{for all } y.$$

$\underbrace{x \downarrow y}$ → x approaches y from above (from the right)

Jumps must look like this. 

<u>Not</u> like this. 

<u>Not</u> like this. 

$$F(y-) \le F(y) = F(y+) \quad \text{for all } y$$

$$\lim_{x \uparrow y} F(x) \qquad\qquad \lim_{x \downarrow y} F(x).$$

The cdf $F_X(\cdot)$ $\underbrace{\text{contains all the info in}}_{\text{determines}}$ $P_X(\cdot)$

$$F_X(t) = P_X((-\infty, t]) = P(X \leq t)$$

$$F_X(t-) = P_X((-\infty, t)) = P(X < t)$$

$$F_X(b) - F_X(a) = P_X((a, b])$$
$$= P(a < X \leq b)$$

$$F_X(b-) - F(a) = P_X((a, b)) = P(a < X < b)$$

$$F_X(a) - F_X(a-) = P_X(\{a\}) = P(X = a)$$

Thus $F_X(\cdot)$ determines the probability assigned to any interval or point.
Can build up from there.

$$1 - F_X(t) = P_X((t, \infty)) = P(X > t)$$

Properties of Densities:

If a r.v. X has a pdf $f(x)$, then

① $f(x) \geq 0$ for all x,

② $\int_{-\infty}^{\infty} f(x)dx = 1$.

Conversely, any function $f(x)$ satisfying ① and ② can serve as a pdf (that is, there exists a r.v. X with $f(x)$ as its pdf).

Relationship between cdf and pdf:

Let $F(x)$ be the cdf of X.
If X has a pdf $f(x)$, then

(a) $F(x) = \int_{-\infty}^{x} f(y)dy$ for all x,

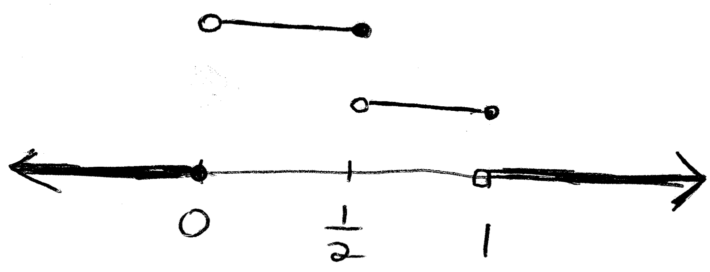
(b) $F'(x) = f(x)$ at all points x where $f(\cdot)$ is continuous.

(The text takes (a) as the definition of the pdf.)

Example:

The pdf $f(x) = \begin{cases} \frac{3}{2} & \text{for } 0 < x \le 1/2, \\ \frac{1}{2} & \text{for } \frac{1}{2} < x \le 1, \\ 0 & \text{otherwise} \end{cases}$

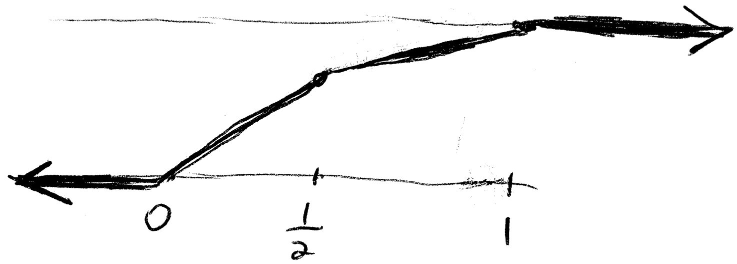
with graph



corresponds to the cdf

$$F(x) = \begin{cases} \frac{3}{2}x & \text{for } 0 \le x < 1/2, \\ \frac{1}{2}x + \frac{1}{2} & \text{for } \frac{1}{2} \le x < 1, \\ 1 & \text{for } x \ge 1, \\ 0 & \text{for } x < 0. \end{cases}$$

with graph

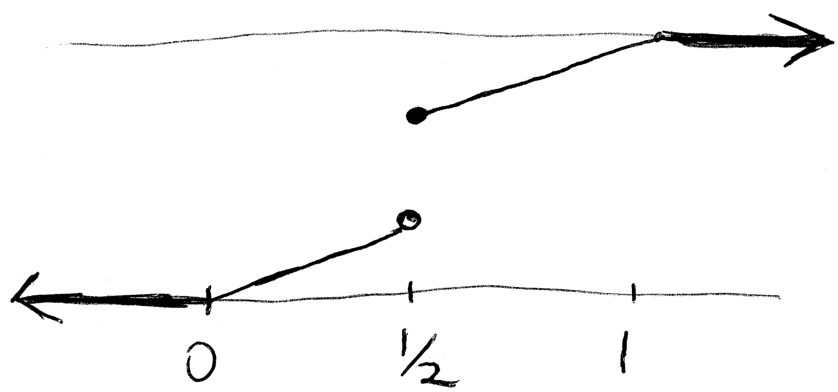


$F'(x) = f(x)$ except at $x = 0, \frac{1}{2}, 1$ where $F'(x)$ does not exist and $f(x)$ is not continuous.

Comment: If a cdf $F(x)$ has a jump (at any value of x), then there is no pdf.

Example:

The cdf $F(x) = \begin{cases} x/2 & \text{for } 0 \le x < 1/2 \\ x/2 + 1/2 & \text{for } 1/2 \le x < 1 \\ 1 & \text{for } x \ge 1 \\ 0 & \text{for } x < 0 \end{cases}$



has $F'(x) = \frac{1}{2}$ for $0 < x < 1$ except at the point $x = \frac{1}{2}$ where the derivative does not exist.

But $f(x) = \begin{cases} \frac{1}{2} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

is not a pdf.