Class 7: Random Variables (Text: Section 4.3)

WHAT IS A RANDOM VARIABLE?

A **random variable** is a number whose value depends on random process. If X is a random variable (RV), it is a function whose input is an experiment and whose output is a number. The output number is represented by x.

Ex: Let X be the number of girls in family of three children. What values can X take? With what probabilities?
X can take the values,
We calculate the probability that X takes on each of these values, assuming that $P(G) = 1/2$ and the gender of births are independent:

$P(X = 0) =$

$P(X = 1) =$

$P(X = 2) =$

$P(X = 3) =$

We define the **probability distribution function, *pdf***, which gives the probability of each of the values, 0,1,2,3. We also define the **cumulative distribution function, *cdf***, which gives the cumulative probability of values up to 0, 1, 2, 3.

Ex: Fill the values of the pdf and the cdf in the table below.

X (# girls)
pdf
cdf

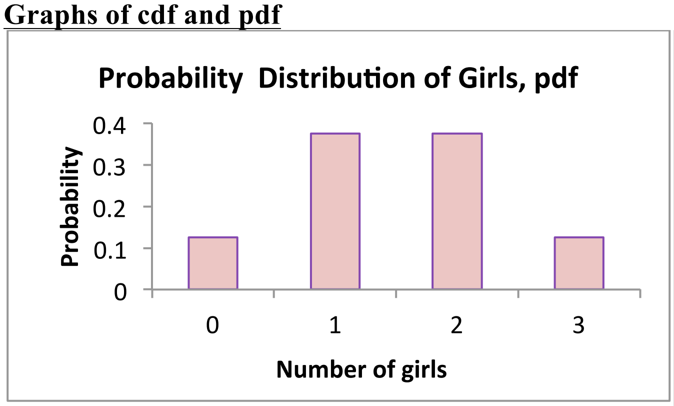
Ex: What does the statement $P(X = 2)$ mean and what is its value?

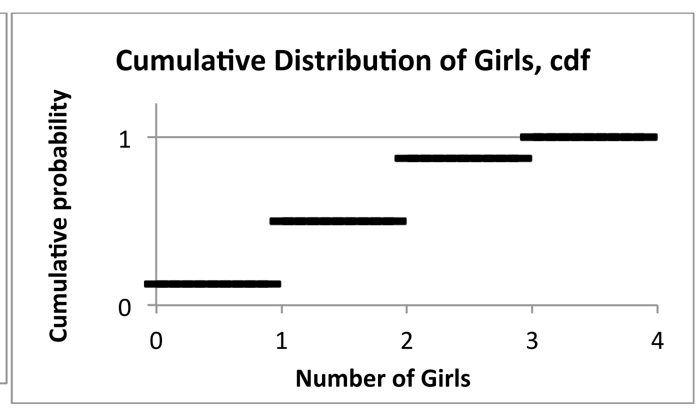
Ex: What does the statement $P(X \leq 2)$ mean and what is its value?

Ex: What do you notice about the values of the pdf? Why?

Ex: What do you notice about the values of the cdf? Why?

Graphs of cdf and pdf





<u>Two Types of Random Variable: Discrete and Continuous</u>

Random variables can be **discrete** or **continuous**.

<u>Ex</u>: The number of children in a family is a

<u>Ex</u>: The height of a child is

<u>DISCRETE RANDOM VARIABLES</u>: Have pdfs and cdfs that have values like this:

Values of X	x_1	x_2	x_3	$\cdots$	x_k
pdf	p_1	p_2	p_3	$\cdots$	p_k
cdf	p_1	$p_1 + p_2$	$p_1 + p_2 + p_3$	$\cdots$	$\sum_{i=1}^{i=k} p_i$

where $p_i = P(X = x_i)$ and $\Sigma p_i = 1$ and $p_i \geq 0$, all i

<u>Ex</u>: The distribution of times since children's last visit to the dentist. [1] (There is no overlap between categories.)

Length of time	Less than 6 mos	6 mos - 1 year	1 year - 2 years	2 years - 5 years	More than 5 years
pdf	0.57	0.18	0.08	0.03	
cdf					

<u>Ex</u>: Fire alarms: Each one has probability 95% of working in a fire. In a fire, assuming they behave independently, what is the probability of

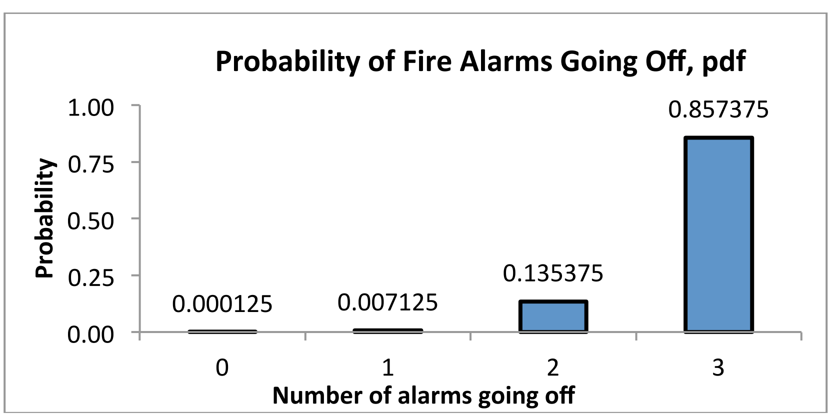
(a) None going off if there are 3?

(b) At least 1 going off if there are 3?

(c) At least 1 going off if there are 10? If there are n?

<u>Ex</u>: For three fire alarms, what is pdf for the number going off? Graph the pdf.

Number going off	
Probability $P(X = i)$	



Graph shows

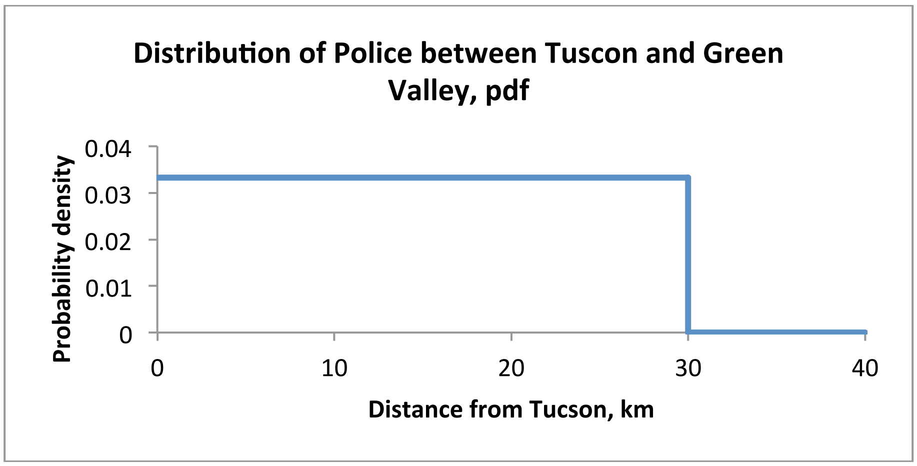
[1] "Summary Health Statistics for US Children: National Health Interview Survey, 2003" by A.D. Det et al, 2005, Table 17. Reported by Baldi and Moore in IPS Life Sciences.

CONTINUOUS RANDOM VARIABLE: UNIFORM DISTRIBUTION

Ex: Police are equally likely to be anywhere between Tucson and Green Valley, 30 km away. What is the probability that the police are some particular fixed number x km from Tucson, say $x = 20$?

We cannot use a probability distribution function as before (because it would be 0 everywhere), so we use a continuous curve, as we did for the normal distribution. The curve is called a **probability *density* function, pdf**. The area under the graph of a probability density function gives ***probability***.

Ex: Since the police are equally likely to be anywhere, the probability density function is constant and its graph is a horizontal line. What is its height?

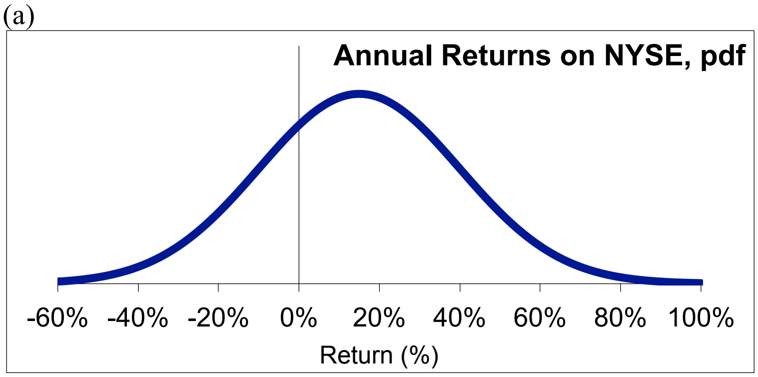


Height of curve is

Ex: Find the probability that police are
(a) Between 20 and 30 km from Tucson (b) Less that 20 km from Tucson.

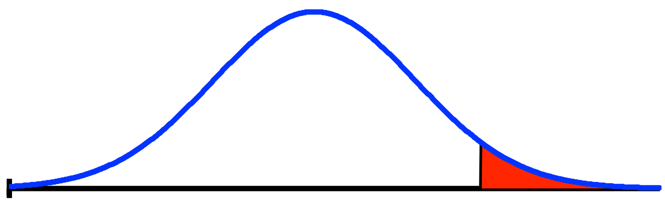
Ex: In "normal" years, the distribution of returns on the New York Stock Exchange (NYSE) are approximately normal, with mean 15% and $SD = 25\%$. (a) Sketch the pdf, and (b) Find the probability of a loss.

(a)



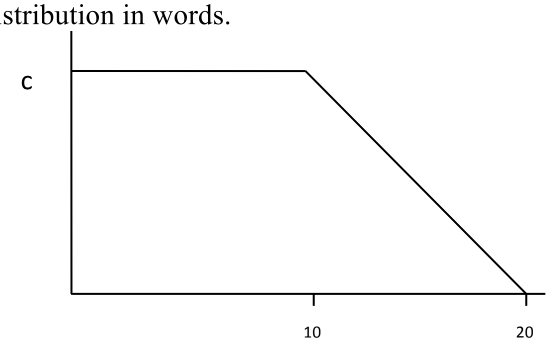
(b)

Ex: In a normal year on the NYSE, what returns are earned on the top 5% of stocks?



Ex: What are the bottom 5% of returns on the NYSE in a normal year?

Ex: For the continuous distribution whose pdf is graphed below, what is the value of c? Describe the distribution in words.



Ex: A continuous random variable X takes only values from 10 to 20, and with steadily decreasing likelihood. (a) Sketch the pdf and label the axes. (b) Find $P(15 < X < 20)$ and $P(10 < X < 15)$.