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, 0,1,2,3, corresponding to the fact that there can be 0,1,2, or 3 girls in the family.
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:

$$P(X = 0) = (1/2)^3 = 0.125$$
$$P(X = 1) = 3/8 = 0.375$$
$$P(X = 2) = 3/8 = 0.375$$
$$P(X = 3) = 1/8 = 0.125$$

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)	0	1	2	3
pdf	0.125	0.375	0.375	0.125
cdf	0.125	0.5	0.875	1

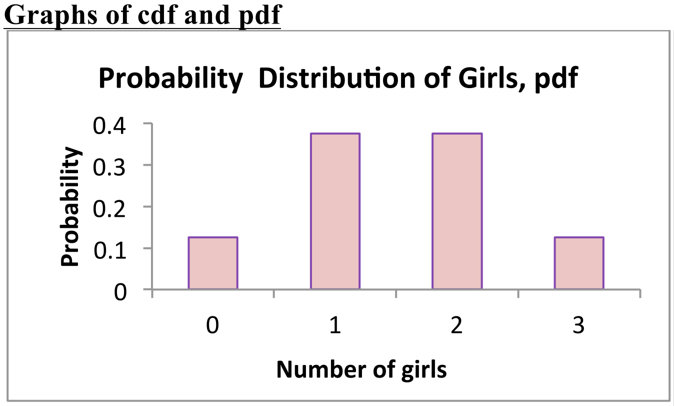
Ex: What does the statement $P(X = 2)$ mean and what is its value?
The probability that a family has exactly 2 girls; its value is 0.375.

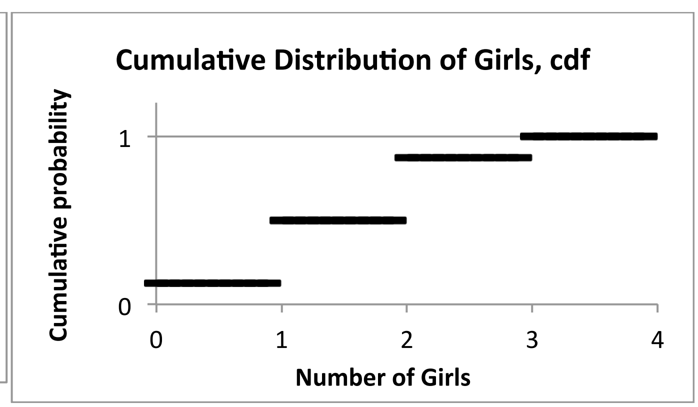
Ex: What does the statement $P(X \leq 2)$ mean and what is its value?
The probability that a family has less than or equal to 2 girls; its value is 0.875.

Ex: What do you notice about the values of the pdf? Why?
They add to 1 because the values 0,1,2,3 exhaust all the possibilities—-one of them has to happen.

Ex: What do you notice about the values of the cdf? Why?
They level off to 1 because every family of 3 children must have less than or equal to 3 girls.

Graphs of cdf and pdf





Two Types of Random Variable: Discrete and Continuous

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

Ex: The number of children in a family is a discrete RV

Ex: The height of a child is a continuous RV

DISCRETE RANDOM VARIABLES: 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

Ex: 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	0.14
cdf	0.57	0.75	0.83	0.86	1

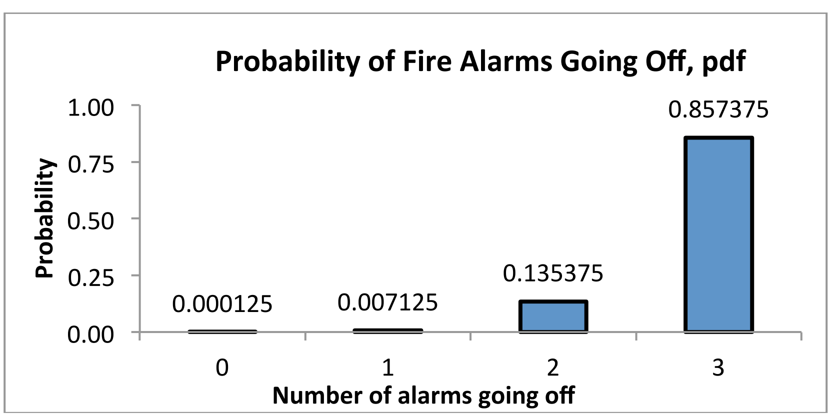
Ex: Fire alarms: Each one has probability 95% of working in a fire. In a fire, assuming they behave independently, what is the probability of

(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?

(a) None going off means all three fail, each with probability 0.05: Probability = $(0.05)^3 = 0.000125$
(b) At least one goes off means not all fail: Probability $= 1 - (0.05)^3 = 1 - 0.000125 = 0.999875$
(c) For 10 alarms: Probability $= 1 - (0.05)^{10}$. For n alarms: Probability $= 1 - (0.05)^n$

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

Number going off	0	1	2	3
Probability	$(0.05)^5$	$3(0.05)^2(0.95)$	$3(0.05)(0.95)^2$	$(0.95)^2$
$P(X = i)$	0.000125	0.007125	0.135375	0.857375



Graph shows—as the numbers do—that the most likely outcome by far is that all three alarms go off.

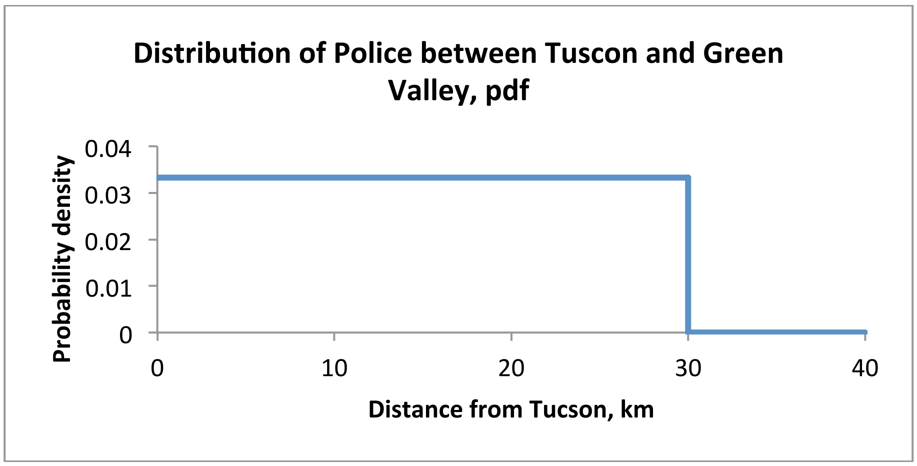
[1] "Summary Health Statistics for US Children: National Health Interview Survey, 2003" by A.D. Det et al, 2005, Table 17. Reported by Baldi and Moore in IPS Life Sciences.

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 $x = 20$?
Here x is a continuous random variable. For any fixed x, the probability that the police are exactly that distance away is 0. For example, $P(X = 20) = 0$. There are infinitely many values that x can take and the probability of any one of them is 0.

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?

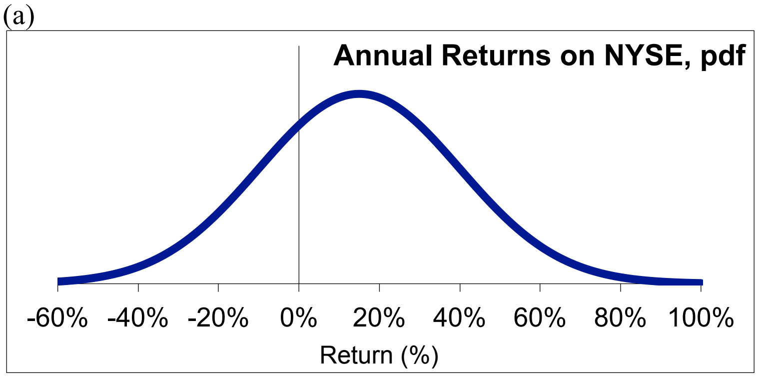


Height of curve is found from the fact that area under the curve must be 1. Since length is 30, the height is = $1/30$

Ex: Find the probability that police are
(a) Between 20 and 30 km from Tucson (b) Less that 20 km from Tucson.
(a) $P(20 < X < 30) = (1/30) \cdot (30 - 20) = 1/3$
(b) $P(X < 20) = (1/30) \cdot 20 = 2/3$

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) Find the z-value:

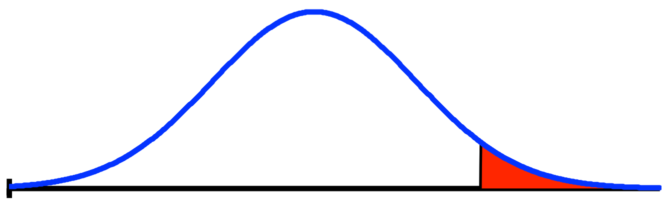
$$z = (0 - 0.15)/0.25 = -0.6$$

So

$$P(X < 0) = P(Z < -0.6) = 0.274.$$

The probability of a loss is 27.4%.

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



We are interested in what z-value has 5% of the returns to the right of it. Looking up 0.95 in the table, we find $z = 1.645$, so we solve

$$1.645 = \frac{x - 0.15}{0.25}$$

$$x = 0.15 + 1.645 \cdot 0.25 = 0.56125.$$

So the top 5% of the returns are above 56.125%.

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

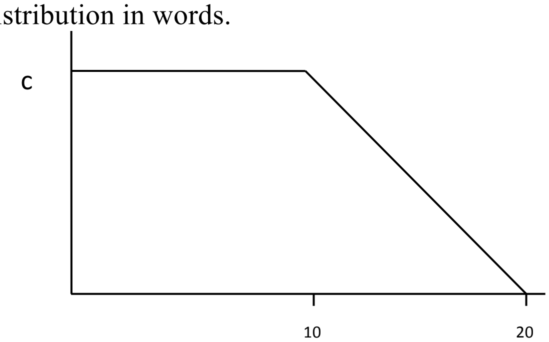
By symmetry, the bottom 5% are the same number of standard deviations below the mean as the previous answer was above, namely 1.645. So we solve

$$-1.645 = \frac{x - 0.15}{0.25}$$

$$x = 0.15 - 1.645 \cdot 0.25 = -0.26125.$$

So the bottom 5% of the returns lose more than 26.125%.

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

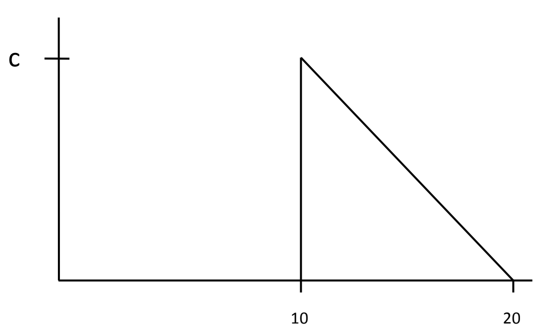


Since the area under the curve is 1, we have

$$\text{Area} = c \cdot 10 + \frac{1}{2}c \cdot 10 = 15c = 1 \text{ so } c = {}^{1}/_{15}$$

The random variable X is equally likely to take on any value $0 \leq X \leq 10$; progressively less likely to take on values between 10 and 20.

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 $P(10 < X < 15)$.



(a) For the pdf: $\frac{1}{2}c \cdot 10 = 1$ so $c = {}^{1}/_{5}$

(b) Area of triangle = ½ Base · Height. Probability is the area under the curve, so

$$P(15 < X < 20) = {}^{1}/_{2} \cdot {}^{1}/_{10} \cdot 5 = {}^{1}/_{4}$$

and

$$P(10 < X < 15) = 1 - {}^{1}/_{4} = {}^{3}/_{4}$$