Chapter 7.2 & 7.3

Random Variables

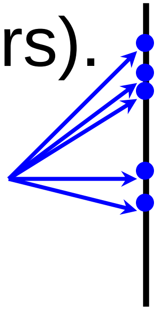
&

Probability Distributions

Discrete and Continuous Random Variables

A random variable is **discrete** if its set of possible values is a collection of isolated points on the number line (usually integers).

Possible values of a discrete random variable



A random variable is **continuous** if its set of possible values includes an entire interval on the number line.

Possible values of a continuous random variable

We will use lowercase letters, such as x and y, to represent random variables.

Common Distributions for Statistics

Discrete Distributions

- Binomial Distributions
- Geometric Distributions
- Poission Distributions (*future stats classes*)

Continuous Distributions

- Normal Distributions
- Uniform distributions
- Chi-Square Distributions
- Student's *t* Distributions (*t* Distributions)

Notation for Random Variables

- For a probability $P(X \le x)$, what do x and X mean here?

A chosen constant

- Consider X to be the random variable which represents the outcome of a single roll of a die, so that X can take on values of $\{1,2,3,4,5,6\}$
- $P(X \le 2)$ means what is the probability that the outcome will be 1 or 2.
- $P(X \le 5)$ means what is the probability that the outcome will be 1,2, 3, 4, or 5.

Notation: In general $P(\, X \ \leq x \,)$

…means the probability that the random variable X is less than or equal to the realization x .

Our textbook might show the following: Given two common dice, the random variable is the sum of the two dice $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

What is the probability that the sum is six or less?

$$\boldsymbol{P(\, X \ \leq 6 \,) = \ p\,(x \leq 6)}$$

What is the probability that the sum is 9?

$$\boldsymbol{P(\, X = 9 \,)\ or\ \ p\,(x = 9)\ or\ \ p(9)}$$

7.2: Probability Distributions for Discrete Random Variables

The **probability distribution of a discrete random variable x** gives the probability associated with each possible x value.

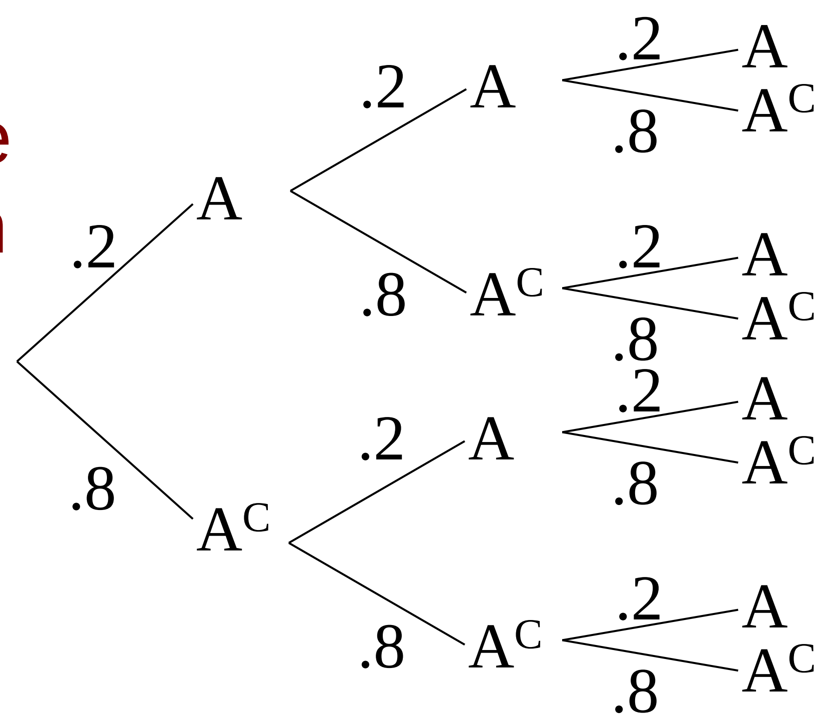
Each probability is the limiting relative frequency of occurrence of the corresponding x value when the experiment is repeatedly performed (LOLn).

Roll	1	2	3	4	5	6
p =	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
p =	**0.167**	**0.167**	**0.167**	**0.167**	**0.167**	**0.167**

Example

Suppose that 20% of the apples sent to a sorting line are Grade A. If 3 of the apples sent to this plant are chosen randomly, determine the probability distribution of the number of Grade A apples in a sample of 3 apples.

Consider the tree diagram



	X
$.2 \bullet .2 \bullet .2 = .008$	**3**
$.2 \bullet .2 \bullet .8 = .032$	**2**
$.2 \bullet .8 \bullet .2 = .032$	**2**
$.2 \bullet .8 \bullet .8 = .128$	**1**
$.8 \bullet .2 \bullet .2 = .032$	**2**
$.8 \bullet .2 \bullet .8 = .128$	**1**
$.8 \bullet .8 \bullet .2 = .128$	**1**
$.8 \bullet .8 \bullet .8 = .512$	**0**

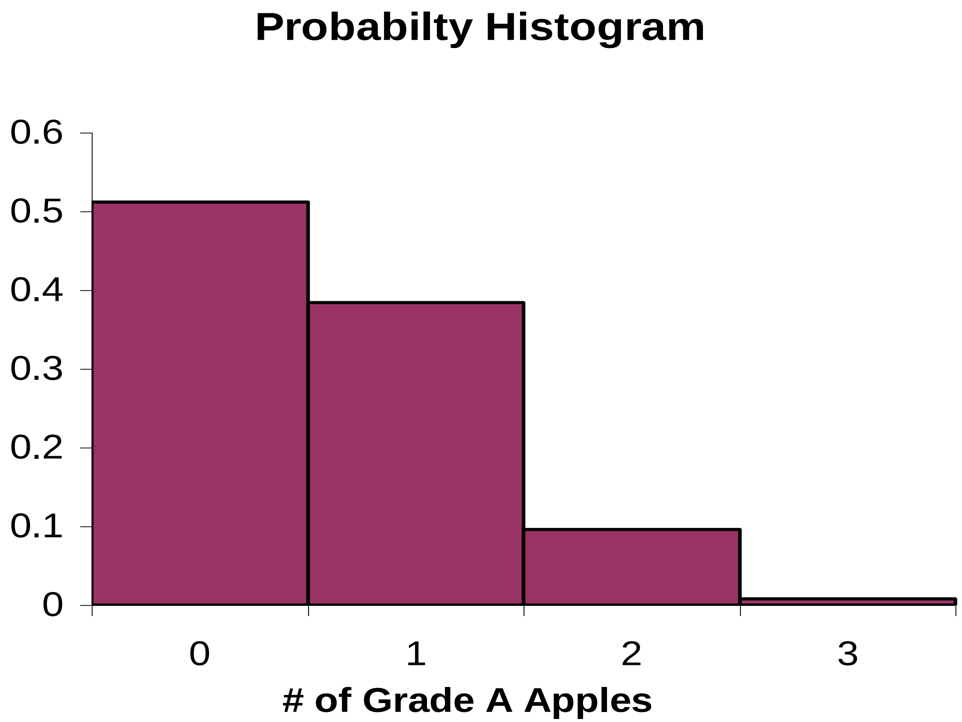
The Results in Table Form

x	p(x)
0	$1(.8)^3$
1	$3(.8)^2(.2)^1$
2	$3(.8)^1(.2)^2$
3	$1(.2)^3$

or

x	p(x)
0	0.512
1	0.384
2	0.096
3	0.008

Results in Graphical Form (Probability Histogram)



For a probability histogram, the area of a bar is the probability of obtaining that value associated with that bar.

Properties of Discrete Probability Distributions

The probabilities p_i must satisfy

1. $0 \le p_i \le 1$ for each i

2. $p_1 + p_2 + ... + p_k = 1$

The probability P(X in A) of any event is found by summing the p_i for the outcomes x_i making up A.

Nov 2021
Chapter 7 – Day 2

Warm-UP: Review Normal distributions

1) What is a *z-score*?

2) Given a normal distribution, what is the proportion of observations that fall between *z-scores* of $-1.6 < x < 0.7$?

3) What is the z-score associated with the Standard normal probability of .892?

4) What is an Apgar score? (don’t know…look at handout or look it up!)

WARM-UP

1) What is a z score?

A *z* score gives the relationship between an observation (x_i) and the mean ($\bar{x}$) of a distribution in terms of some number of standard deviations. It is positive when it's greater than the mean, and negative when it's less than the mean.

2) Given a normal distribution, what is the proportion of observations that fall between z-scores of -1.6 < x < 0.7 ?

Approx. 0.703 or 70.3% of the observations

Warm-UP

3) What is the z-score associated with the Standard normal probability of .892?

***z* score = 1.237**

4) What is an Apgar score?

Warm-Up

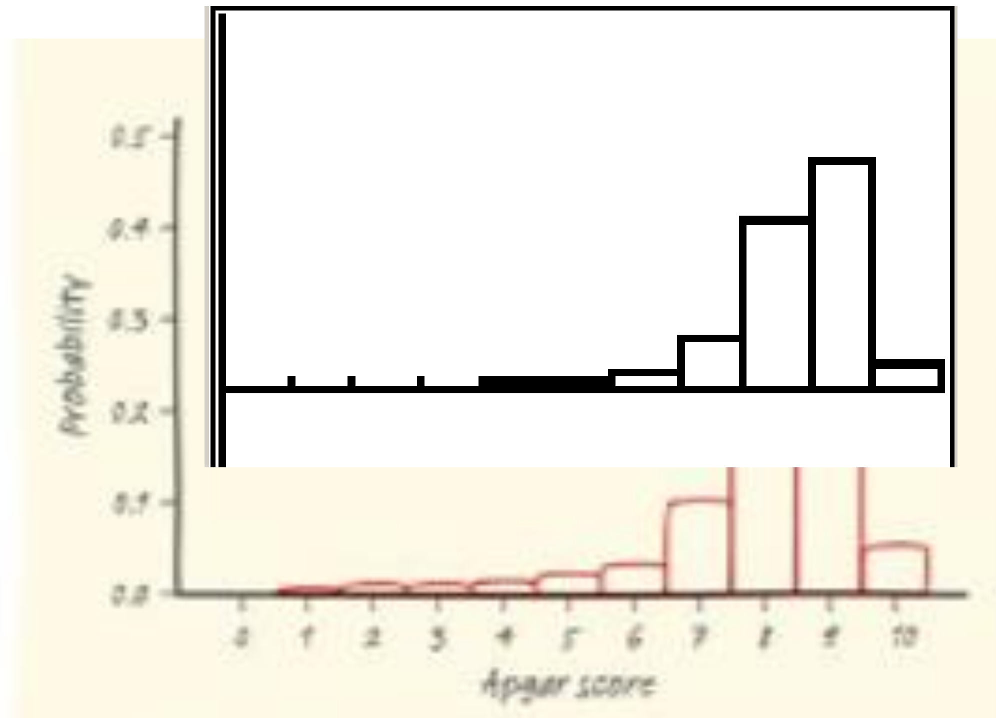
What is an Apgar score? What are the possible values for the random variable?

Virginia Apgar invented the **Apgar score** in 1952 as a method to quickly summarize the health of newborn children. Apgar was an anesthesiologist who developed the score in order to ascertain the effects of *obstetric anesthesia* on babies. The Apgar scale is determined by evaluating the newborn baby on five simple criteria on a scale from zero to two, then summing up the five values thus obtained. The resulting Apgar score ranges from zero to 10.

Histogram of Probability Distribution

L1	L2
0	.001
1	.006
2	.007
3	.008
4	.012
5	.02
6	.038

L2(1)=.001



Can you make this histogram on your calculator?

Apgar Scores

Example: Babies' Health at Birth (Apgar Scores)

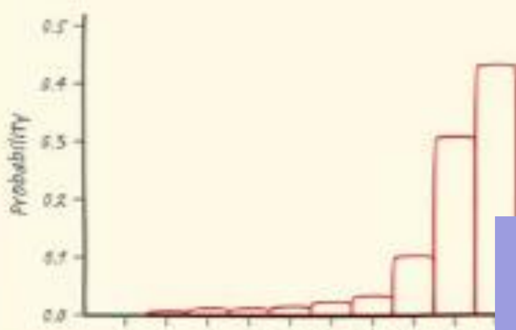
(a) Show that the probability distribution for X is legitimate.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?

Value:	0	1	2	3	4	5	6	7	8	9	10
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

(a)



(c) $P(X \geq 7) = .908$
We'd have a 91 % chance of randomly choosing a healthy

(b)

Notice the difference b/t notation > and ≥
With DRV, these are different, but with CRV, they mean the same!

7.3: Probability Distribution for a Continuous Random Variable

A **probability distribution for a continuous random variable x** is specified by a mathematical function denoted by *f(x)* which is called the **density function**. The graph of a density function is a smooth curve (the **density curve**).

The following requirements must be met:

1. $f(x) \geq 0$
2. The total area under the density curve is equal to 1.

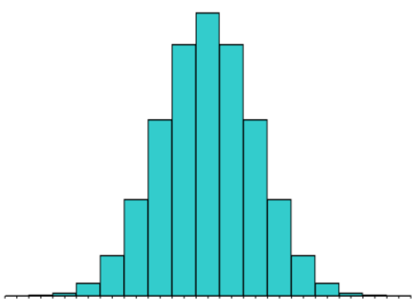
The probability that x falls in any particular interval is the area under the density curve that lies above the interval.

Probability Density Function

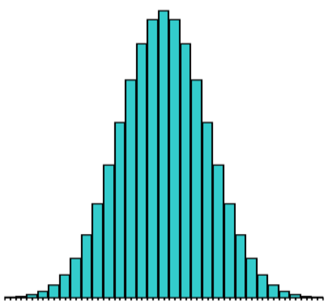
The probability that a discrete random variable X takes on a particular value x, that is, $P(X = x)$, is frequently denoted $f(x)$. The function $f(x)$ is typically called the **probability mass function**, although some authors also refer to it as the **probability function**, the **frequency function**, or **probability density function**. Most college level statistics courses will use the common terminology — the probability mass function — and its common abbreviation — the **p.m.f.**

Continuous Probability Distributions

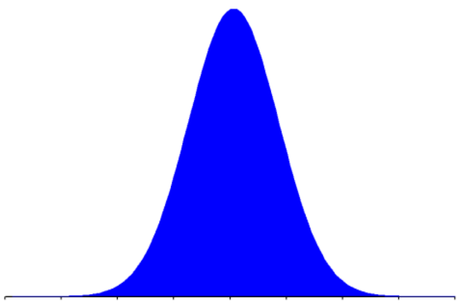
If one looks at the distribution of the actual amount of water (in ounces) in “one gallon” bottles of spring water they might see something such as



Amount measured to nearest hundredths of an ounce.

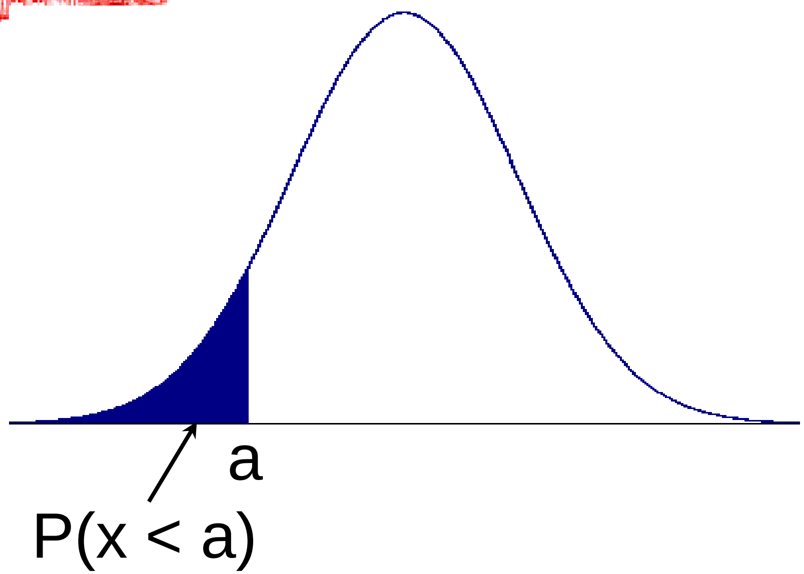


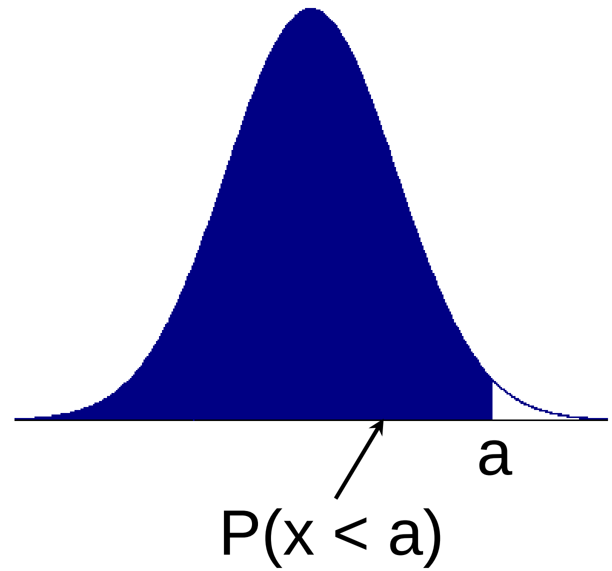
Amount measured to nearest ten thousands of an ounce.



Limiting curve as the accuracy increases

Some Illustrations





Notice that for a continuous random variable x, P(x = a) = 0 for any specific value a because the "area above a point" under the curve is a line segment and hence has o area.

Specifically this means $P(x < a) = P(x \leq a)$.

Continuous Random Variables

Definition. The **probability density function** ("**p.d.f.**") of a continuous random variable X with support S is an integratible function $f(x)$ satisfying the following:

(1) $f(x)$ is positive everywhere in the support S, that is, $f(x) > 0$, for all x in S

(2) The area under the curve $f(x)$ in the support S is 1, that is:

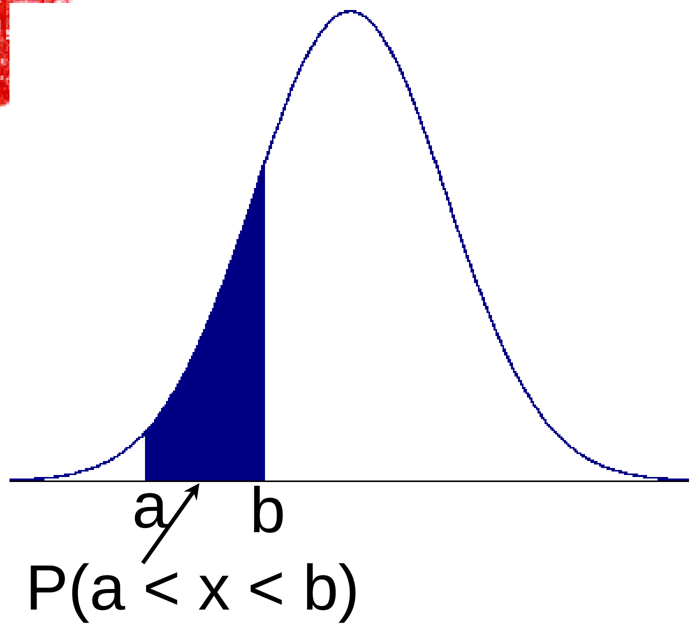
$$\int_S f(x)dx=1$$

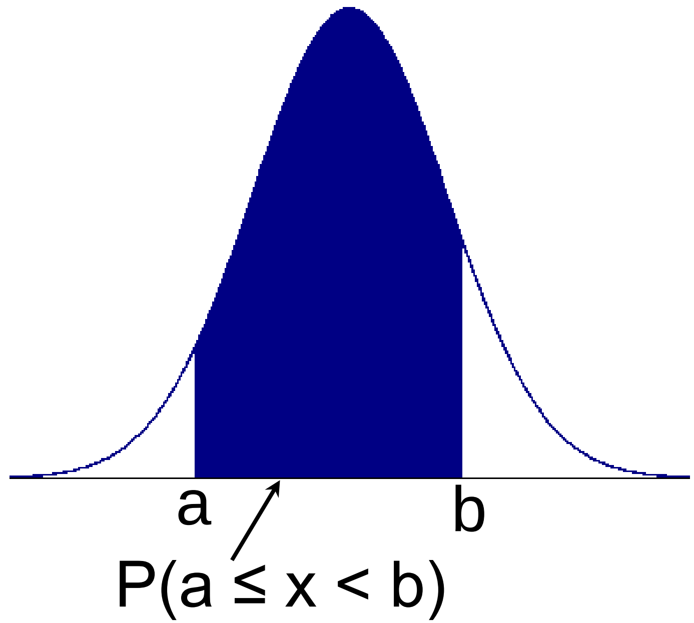
(3) If $f(x)$ is the p.d.f. of x, then the probability that x belongs to A, where A is some interval, is given by the integral of $f(x)$ over that interval, that is:

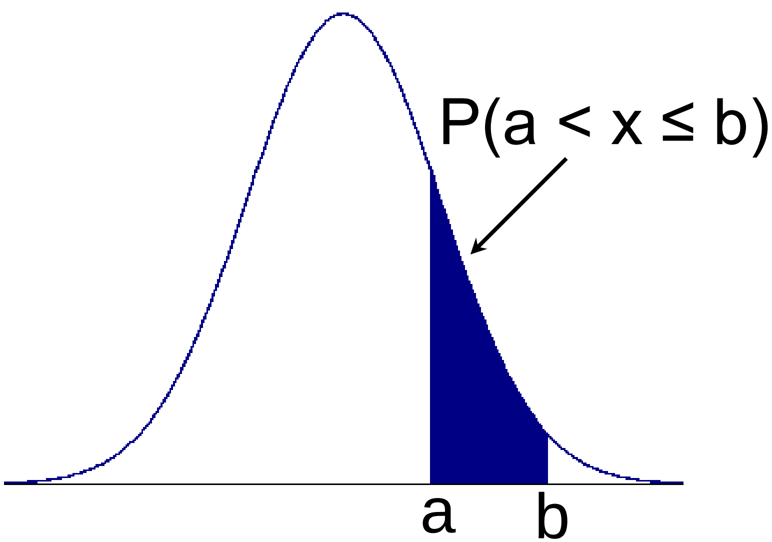
$$P(X\in A)=\int_A f(x)dx$$

** As you can see, the definition for the p.d.f. of a continuous random variable differs from the definition for the p.m.f. of a discrete random variable by simply changing the summations that appeared in the discrete case to integrals in the continuous case.

Illustrations of Finding Areas







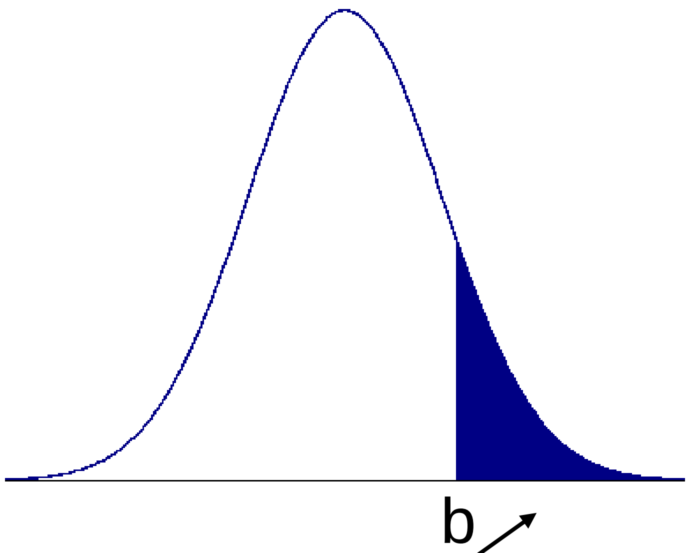
Note: for CRV

P(a < x < b) = P(a ≤ x < b)

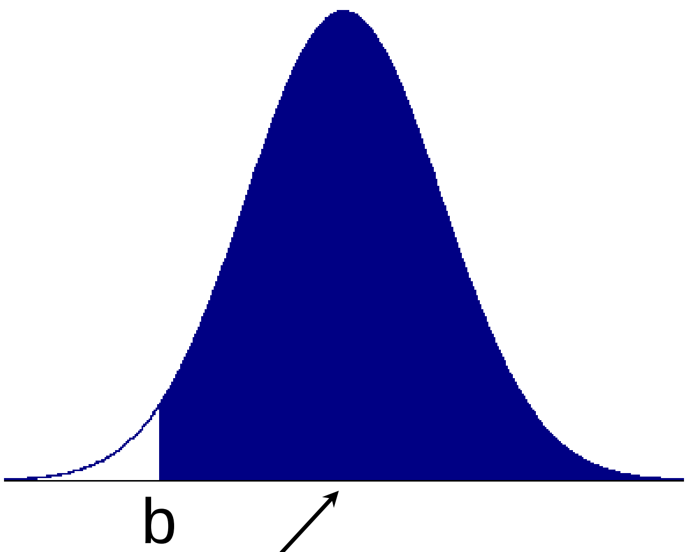
or = P(a < x ≤ b)

or = P(a ≤ x ≤ b)

Equivalent Notations for Finding areas



$P(x > b) = P(b < x)$



$P(x \geq b) = P(b < x)$

Uniformly *Weird* Example

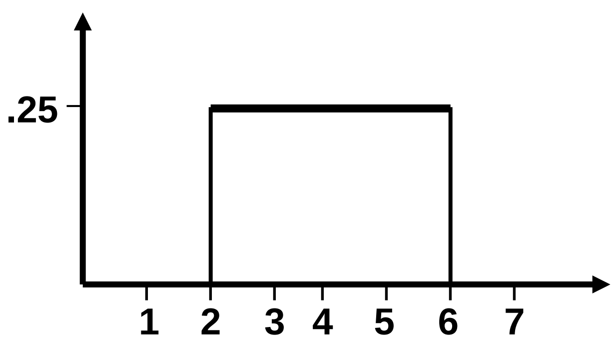
Define a continuous random variable x by

x = the weight of the crumbs in ounces left on the floor of a restaurant during a one-hour period.

Suppose that x has a probability distribution with density function

$$f(x) = \begin{cases} .25 & 2 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

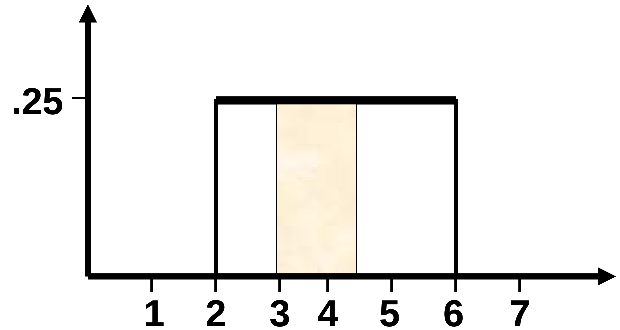
The graph looks like



Example

Find the probability that during a given 1 hour period between 3 and 4.5 ounces of crumbs are left on the restaurant floor.

Note: the graph is a uniform distribution



The probability is represented by the shaded area in the graph. Since that shaded area is a rectangle, area = (base)(height)=(1.5)(.25) = .375

Example 2: Uniform Random Variable

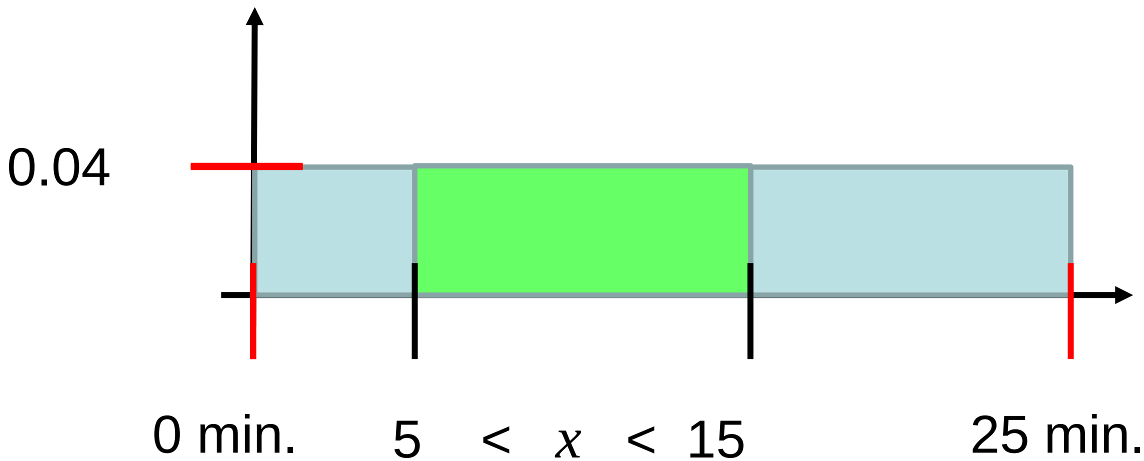
Let x be the amount of time that commuter waits for a TARC bus at a local stop. Suppose that a bus comes every 25 minutes, but the commuter is not sure of the next arrival.

If you assume a uniform distribution, what is the probability that the bus will arrive within the next ***5 to 15 minutes****?*

Sketch a graph of the density function, with labels, and then determine the probability of the interval.

Uniform Distribution: Bus arrival

Sketch a graph of the density function, with labels, and then determine the probability of the interval.



Probability of arriving between 5 and 15 minutes:

$P(5 < x < 15) =$ *is the same as* $P(5 \le x \le 15)$

Probability of arriving between 5 and 15 minutes:

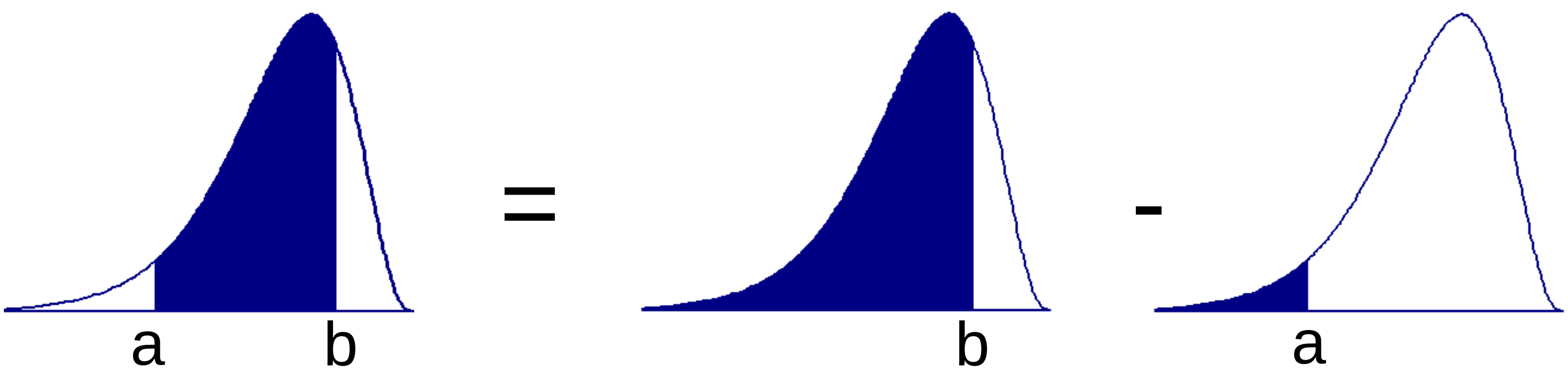
$$P(5 < x < 15) = 0.04(10) = 0.4$$

Method of Probability Calculation

The probability that a continuous random variable x lies between a lower limit a and an upper limit b is

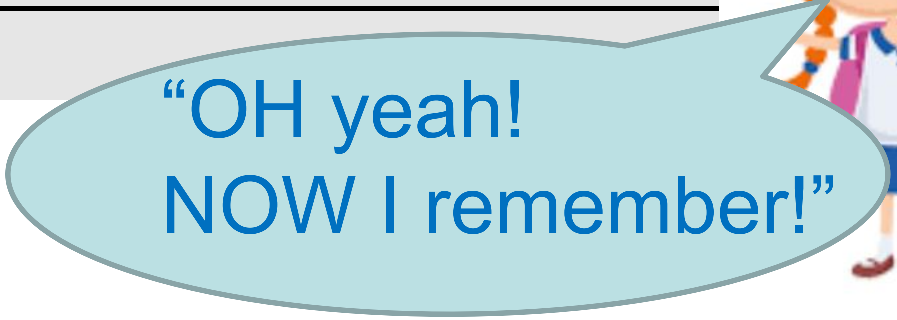
P(a < x < b) = (cumulative area to the left of b) – (cumulative area to the left of a)

= P(x < b) – P(x < a)



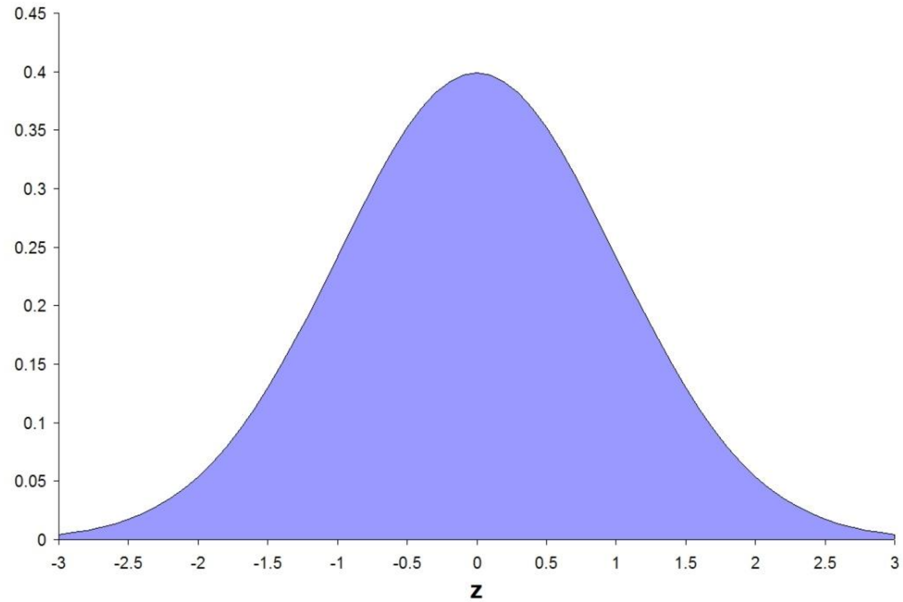
Method of Probability is *review*, not NEW!

Reminder: We use this process of subtraction between two boundaries for finding the *area under the curve* or the interval probability or the percentage of observations for the Standard Normal distribution



Standard Normal Distribution

A normal distribution with mean 0 and standard deviation 1, is called the standard (or standardized) normal distribution.



Normal Tables

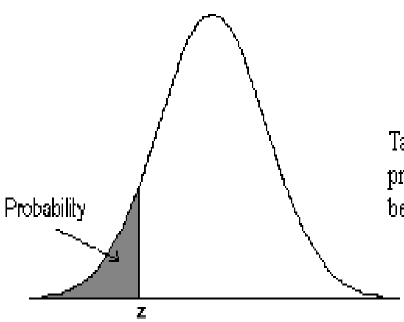


Table entry is probability at or below z.

z*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0010	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0019	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0046	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0061	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Normal Tables

z*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

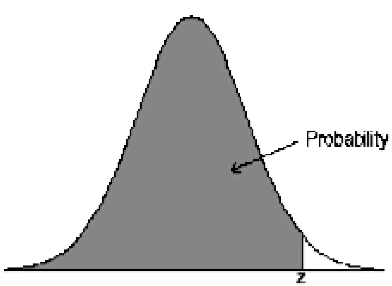


Table entry is probability at or below z.