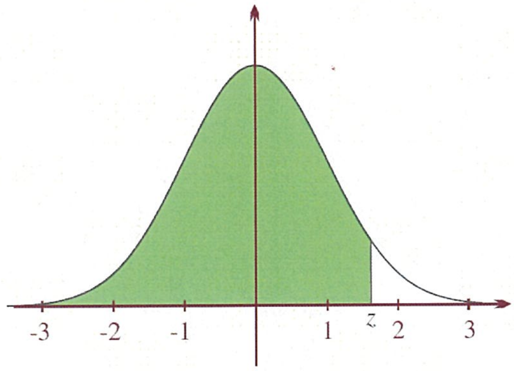
NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION
(from Cambridge textbook)

The standard normal probability distribution

The shaded area in the graph to the right represents a value of the *cumulative standard normal distribution function*

$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \Phi(t)\, dt.$$



The table below gives some further values of the probabilities $P(Z \le z) = \Phi(z)$, allowing two decimal places for z — after that, use interpolation. For example,

$$P(Z \le 1.627) = \Phi(1.627) = \int_{-\infty}^{1.627} \Phi(z)\, dz \doteq 0.9474 + 0.7(0.9484 - 0.9474) \doteq 0.9481.$$

	second decimal place									
z	+ .00	+ .01	+ .02	+ .03	+ .04	+ .05	+ .06	+ .07	+ .08	+ .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

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION
(from Cambridge textbook)

	second decimal place									
z	+ .00	+ .01	+ .02	+ .03	+ .04	+ .05	+ .06	+ .07	+ .08	+ .09
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
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION (from Cambridge textbook)

Note: You may find in this exercise and the next that your answers differ slightly from the text, depending whether or not you interpolate, and whether you use the supplied tables or alternatives such as statistical calculators and spreadsheets that provide more accurate values.

1 A certain binomial distribution for the random variable X consists of 20 independent trials, each with probability of success $p = 0.3$.

a Write this binomial distribution in symbolic form.

b Calculate the probability of obtaining 9, 10 or 11 successes.

c Confirm that $np > 5$ and $nq > 5$, suggesting that a normal approximation to the binomial could be used to estimate this probability.

d Calculate the normal approximation by determining $P(8.5 \le X \le 11.5)$, treating X as approximately $N(\mu, \sigma^2)$, and using the normal tables at the start of this exercise.

e Find the percentage error in the normal approximation. Does the approximation seems fairly accurate for this value of n and p?

a) $P(X=k) = {}^{20}C_k \, 0.3^k \times 0.7^{20-k}$

b) $P(X=9) + P(X=10) + P(X=11) = {}^{20}C_9 \, 0.3^9 \times 0.7^{11} + {}^{20}C_{10} \, 0.3^{10} \times 0.7^{10} + {}^{20}C_{11} \, 0.3^{11} \times 0.7^{9}$

$= 0.06536956 + 0.0308171 + 0.0120066 = 0.1082$ so 10.82%

c) $n \times p = 20 \times 0.3 = 6$ which is indeed greater than 5

$n \times (1-p) = 20 \times 0.7 = 14$ ——— 5

$\therefore$ a normal approximation could be used to estimate this probability

d) For $x = 8.5$ $\quad z = \dfrac{8.5-\mu}{\sigma} = \dfrac{8.5-6}{2.05} = 1.22$

as $\mu = np = 6$ and $\sigma^2 = npq = 20 \times 0.3 \times 0.7 = 4.2$ so $\sigma = 2.05$

For $x = 11.5$ $\quad z = \dfrac{11.5-6}{\sigma} = \dfrac{11.5-6}{2.05} = 2.68$

From the table:

$P(8.5 \leq X \leq 11.5) = P(X \leq 11.5) - P(X \leq 8.5)$

$= P(z \leq 2.68) - P(z \leq 1.22)$

$= 0.9963 - 0.8869$

$= 10.94\%$

e) So the binomial distribution calculation gives 10.82% compared to the calculation with the normal distribution 10.94%

% error is about 1% - it's very accurate.

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION
(from Cambridge textbook)

2 Repeat the steps of Question **1** for these cases.

a $n = 50$, $p = 0.5$, find $P(18 \leq X \leq 20)$, **b** $n = 20$, $p = 0.4$, find $P(8 \leq X \leq 9)$,

2a) Binomial: $P(18 \leqslant X \leqslant 20) = {}^{50}C_{18}\, 0.5^{50} + {}^{50}C_{19}\, 0.5^{50} + {}^{50}C_{20}\, 0.5^{50}$

$= 0.0849$ approx

$np = 50 \times 0.5 = 25 > 5$

$nq = 50 \times 0.5 = 25 > 5$

So estimation with Normal distribution is possible.

$\mu = np = 25 \qquad \sigma^2 = npq = 50 \times 0.5 \times 0.5 = 12.5$

$z_1 = \dfrac{17.5 - 25}{\sqrt{12.5}} \simeq -2.12 \qquad z_2 = \dfrac{20.5 - 25}{\sqrt{12.5}} \simeq -1.27$

Hence $P(17.5 < X < 20.5) = P(-2.12 < z < -1.27)$

$= P(1.27 < z < 2.12)$

$= P(z < 2.12) - P(z < 1.27)$

$= 0.9830 - 0.8980 = 0.085$

compared to 0.0849 ☺

2b) Binomial

$P(8 \leqslant X \leqslant 9) = {}^{20}C_8 \times 0.4^8 \times 0.6^{12} + {}^{20}C_9 \times 0.4^9 \times 0.6^{11}$

$\simeq 0.17970579 + 0.1597385 \simeq 0.33944$

$np = 20 \times 0.4 = 8$

$nq = 20 \times 0.6 = 12$

so estimation with Normal distribution is possible.

$\mu = np = 8 \qquad \sigma^2 = npq = 8 \times 0.6 = 4.8$

$z_1 = \dfrac{7.5 - 8}{\sqrt{4.8}} \simeq -0.228 \qquad z_2 = \dfrac{9.5 - 8}{\sqrt{4.8}} \simeq 0.68$

Hence $P(7.5 < X < 9.5) = P(-0.228 < z < 0.68)$

$= P(z < 0.68) - \underbrace{P(z < -0.228)}$

$= P(z > 0.228)$

$= 1 - P(z < 0.228)$

Hence $P(7.5 < X < 9.5) = 0.7517 - 1 + 0.5910 \simeq 0.343$

compared to 0.339

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION (from Cambridge textbook)

3 A barrel contains 600 pink and 400 blue counters. At each stage of an experiment, the barrel is stirred well, then a counter is removed, its colour is noted, and it is returned to the barrel. This experiment is repeated 20 times.

a Explain why each stage of the experiment is a Bernoulli trial.

b Explain why the full experiment is binomial.

c Is it necessary to return the counter after each draw?

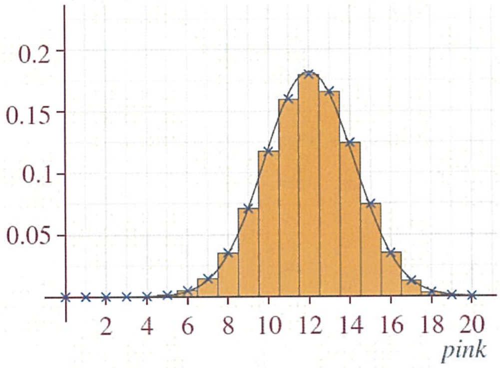
d Write down the probability of drawing a pink counter, and find the mean and standard deviation for this binomial distribution.

e Find the probability of drawing exactly 14 pink counters, correct to 3 decimal places.

f A student constructs a histogram for this distribution and overlays the normal distribution with the same mean and standard deviation. He notes the strong agreement, and to test this he uses his standard normal tables to calculate the area corresponding to drawing 14 pink counters.

i Explain why the required area for the normal distribution with random variable X is $P(13.5 < X < 14.5)$.

ii Hence find this area. Is it in strong agreement with your answer to part **e**?



a) There are only two possible outcomes, pink or blue.

b) There are n stages, each independent, with the same probability of success.

c) Yes, otherwise the stages will not be independent.

d) $p = 0.6$ $\quad \mu = np = 20 \times 0.6 = 12$ $\quad \sigma^2 = npq = 20 \times 0.6 \times 0.4 = 4.8$

so $\sigma = \sqrt{4.8}$

e) $P(X=14) = {}^{20}C_{14} \times 0.6^{14} \times 0.4^{6} = 0.124$

f) ii) $P(13.5 < X < 14.5) = P\left(\frac{13.5-12}{\sqrt{4.8}} < z < \frac{14.5-12}{\sqrt{4.8}}\right)$

$= P(0.68 < z < 1.14)$

$= P(z < 1.14) - P(z < 0.68)$

$= 0.8729 - 0.7517$

$= 0.1212$

So in close agreement with the estimate using a Binomial Distribution

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION (from Cambridge textbook)

4 In a college of 3000 pupils, 1320 are girls. A teacher selects a group of 15 pupils at random from the college rolls, without revealing the result, and asks her students to determine the probability that the group has more than 8 girls in it.

a Write down the probability p of selecting a girl from the population of 3000.

b Write down the mean and standard deviation of the binomial distribution obtained by selecting 15 pupils from the population.

c Use the exact binomial distribution to determine the probability of obtaining 9 or 10 girls.

d Is the sample size big enough to use the normal approximation to the binomial? Use the criterion $np > 5$ and $n(1 - p) > 5$.

e Use the normal approximation to the binomial to estimate the probability in part **c**.

f Find the percentage error in the estimation.

a) $p = \frac{1320}{3000} = 0.44$

b) $\mu = np = 15 \times 0.44 = 6.6$ $\sigma^2 = 15 \times 0.44 \times 0.56 = 3.696$
so $\sigma = \sqrt{3.696}$

c) $P(X=9) + P(X=10) = {}^{15}C_9 \times 0.44^9 \times 0.56^6 + {}^{15}C_{10} \times 0.44^{10} \times 0.56^5$
$= 0.09541273 + 0.044980288 = 0.1404$

d) $np = 15 \times 0.44 = 6.6$
$nq = 15 \times 0.56 = 8.4$
so the sample size is big enough to approximate using a Normal Distribution.

e) $P(8.5 < X < 10.5) = P\left(\frac{8.5 - 6.6}{\sqrt{3.696}} < z < \frac{10.5 - 6.6}{\sqrt{3.696}}\right)$
$= P(0.9883 < z < 2.0286)$
$= P(z < 2.0286) - P(z < 0.9883)$
$= 0.97875 - 0.8385$ using calculator
$= 0.1402$

f) $\frac{0.1402}{0.1404} = 0.9985$ so % error $\simeq 0.1\%$

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION
(from Cambridge textbook)

5 A commonly used rule of thumb states that the normal approximation to a binomial distribution will be reasonable if $np > 5$ and $n(1 - p) > 5$. This means that if p is further away from 0.5, the sample size needs to be bigger to get a reasonable approximation.
According to this rule, how big does the sample need to be if:

a $p = 0.5$, **b** $p = 0.25$, **c** $p = 0.125$, **d** $p = 0.01$,
e $p = 0.75$, **f** $p = 0.875$, **g** $p = 0.9$, **h** $p = 0.55$?

a) $np > 5 \Rightarrow n \times 0.5 > 5 \Rightarrow n > 10$

b) $np > 5 \Rightarrow n \times 0.25 > 5 \Rightarrow n > 20$

c) $np > 5 \Rightarrow n \times 0.125 > 5 \Rightarrow n > 40$

d) $np > 5 \Rightarrow n \times 0.01 > 5 \Rightarrow n > 500$

e) $np > 5 \Rightarrow n \times 0.75 > 5 \Rightarrow n > 6.7$
$n(1-p) > 5 \Rightarrow n \times 0.25 > 5 \Rightarrow n > 20$
so $n > 20$

f) $np > 5 \Rightarrow n \times 0.875 > 5 \Rightarrow n > 5.7$
$n(1-p) > 5 \Rightarrow n \times 0.125 > 5 \Rightarrow n > 40$
so $n > 40$

g) $n(1-p) > 5 \Rightarrow n \times 0.1 > 5 \Rightarrow n > 50$

h) $n(1-p) > 5 \Rightarrow n \times 0.45 > 5 \Rightarrow n > 11$

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION (from Cambridge textbook)

6 According to some estimates, eight per cent of males in the world are colour-blind. A representative random sample includes 854 people, and a statistician wishes to determine the probability that between 7% and 9% of the people in the sample are colour-blind.

a Give a reason why the researcher might want to use a normal approximation to the binomial to calculate this probability.

b Comment on the assumption that the sample is representative and random.

c Calculate the required probability, using a normal approximation without a continuity correction.

d As a measure of the accuracy of ignoring any continuity correction, calculate the probability $P(76 < X < 76.5)$ associated with the boundary of 76 successes.

a) Very large numbers are involved. A normal calculator cannot handle numbers such that ${}^{854}C_{76}$. Other calculating devices may not be accurate dealing with the large numbers involved. Also, there are a large number of cases to consider, from 60 to 76 successes. (i.e. 7% and 9% of "successes")

b) It is hard to get a representative sample of the whole world, because different ethnic groups will have different tendencies to colour blindness.

c) $p = 0.08 \qquad \mu = np = 854 \times 0.08 = 68.32$

$\sigma^2 = npq = 854 \times 0.08 \times 0.92 = 62.8544$ so $\sigma \simeq 7.93$

$P(60 < X < 76) = P\left(\frac{60 - 68.32}{7.93} < z < \frac{76 - 68.32}{7.93}\right)$

$= P(-1.0491 < z < 0.9685)$

$= 0.8336 - P(z < -1.0491)$

$= 0.8336 - P(z > 1.0491) = 0.8336 - [1 - P(z < 1.0491)]$

$= 0.8336 - 1 + 0.85293 = 0.6865$

d) $P(76 < X < 76.5) = P\left(0.9685 < z < \frac{76.5 - 68.32}{7.93}\right)$

$= P(0.9685 < z < 1.0315)$

$= 0.84885 - 0.8336 = 0.0152$ so 1.5%

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION
(from Cambridge textbook)

7 A horticulturalist is attempting to cross two species of flowers to strengthen certain characteristics. One of the plants has red flowers and the other's flowers are white, but the horticulturalist wishes to retain the strong red colour in the offspring. According to Mendel's theory of inheritance, there is a 25% chance that the flowers of the offspring will be red. The horticulturalist crosses 15 pairs of parent plants and notes the colour of their offspring plant. Find the probability that of the offspring:

a none will have red flowers,

b there will be at least one with red flowers,

c at least twenty per cent will have red flowers (use a normal approximation here).

a) $P(\text{no red flowers}) = \left(\frac{3}{4}\right)^{15} \approx 1.3\%$

b) $P(\text{at least one with red flowers}) = 1 - P(\text{no red flowers})$

$= 1 - 0.013$

$= 0.987$ so 98.7%

c) 20% of 15 is 3 flowers.

So we are looking at $P(3 < X < 15)$

$np = 15 \times \frac{1}{4} = 3.75$

$\sigma^2 = npq = 3.75 \times 0.75 = 2.81$ so $\sigma = \sqrt{2.81}$

$P(2.5 < x < 15) = P\left(\frac{2.5 - 3.75}{\sqrt{2.81}} < z < \frac{15 - 3.75}{\sqrt{2.81}}\right)$

$= P(-0.7457 < z < 6.711)$

$= P(z < 6.711) - P(z < -0.7457)$

$= 1 - [P(z > 0.7457)]$

$= 1 - [1 - P(z < 0.7457)]$

$= P(z < 0.7457) \approx 77\%$

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION (from Cambridge textbook)

8 Suppose that it is known that 45% of eighteen-year-olds in a particular city do not have a driver's licence. If a random sample of 20 eighteen-year-olds is taken in the city, what is the probability that more than half of them will not have a driver's licence?

$P(X \geqslant 11) = P\left(z > \frac{10.5 - \mu}{\sigma}\right)$ applying continuity correction

$\mu = 20 \times 0.45 = 9$

$\sigma^2 = npq = 20 \times 0.45 \times 0.55 = 4.95$ so $\sigma = \sqrt{4.95}$

$P(X \geqslant 11) = P\left(z > \frac{10.5 - 9}{\sqrt{4.95}}\right) = P(z > 0.6742)$

—— $= 1 - P(z < 0.6742) = 1 - 0.7486 = 0.2514$

so approx 25%

9 Long-term studies show that 60% of the residents and visitors to Nashville Tennessee prefer Country music to Western. The local council provides Country music for those eating their lunch in the park to listen to. How confident can they be that in a group of thirty, more than twenty of them prefer Country? Comment on the assumption of independence in this question.

$P(X > 20) = P(20.5 < X < 30)$ applying continuity correction

$\mu = 30 \times 0.6 = 18$

$\sigma^2 = npq = 18 \times 0.4 = 7.2$ so $\sigma = \sqrt{7.2}$

$P(X > 20) = P\left(\frac{20.5 - 18}{\sqrt{7.2}} < z < \frac{30 - 18}{\sqrt{7.2}}\right)$

—— $= P(0.93 < z < 4.47)$

—— $= P(z < 4.47) - P(z < 0.93)$

—— $= 1 - 0.8238 = 0.176$

so approx 17.6%