MATH 145 **Winter, 2006**

Normal Probability Density Functions

The **Standard Normal Density function** is $f(z) = \dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}$

This is a probability density function, with mean $\mu = 0$ and standard deviation $\sigma = 1$. The corresponding distribution function is important enough to be given a special notation,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2}\, dx$$

Thus, $\Phi(z)$ is an antiderivative of $f(z)$. The function $\Phi(z)$ cannot be expressed in terms of the usual elementary functions. Some calculators have a button to evaluate $\Phi(z)$. Tables of values for $\Phi(z)$ can be found in most statistics books. Many texts provide a table which lists the values of $\Phi(z)$ for $0 \le z \le 3$. Some tables give values of the function $N(z) = \Phi(z) - 0.5$

NOTE $\Phi(-a) = 1 - \Phi(a)$. (By the symmetry of the graph.)

Thus, if X is a continuous random variable whose probability distribution function is a normal curve with mean 0 and standard deviation 1, then

$$P(a < X < b) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}x^2}\, dx = \Phi(b) - \Phi(a)$$

Example (1) The area under the standard normal curve, between the limits $x = -.1$ and $x = .3$, is equal to

$$\Phi(.3) - \Phi(-.1) = .61791 - .46017 = .15774$$

The table gives the value of $\Phi(.3)$; also $\Phi(-.1) = 1 - \Phi(.1)$.

The function $\Phi(z)$ can be used to find areas under a normal curve with mean μ and standard deviation σ. This means finding the area under the curve

$$\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$$

A change of variables, letting $x = (z-\mu)/\sigma$, shows that

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\, dz = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}x^2}\, dx$$

thus,

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\, dz = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

If a quantity Z has a normal probability density function, with mean μ and standard deviation σ, then

$$P(a < Z < b) = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

Together with tables for values of $\Phi(z)$ (or $N(z)$), this can be used to estimate what fraction of a normally distributed quantity lies in a given range.

Example (2) The heights of corn plants in a certain field fit a normal curve, with mean 2.1 meters, and standard deviation .55 meters. Find the fraction of corn plants whose heights are in the range 1.6 to 2.0 meters.

Solution (2) In this problem, the quantity Z is probability density function which is a normal curve with $\mu = 2.1$ and $\sigma = .55$. We need to find $P(1.6 < Z < 2.0)$.

$$\begin{aligned} P(1.6 < Z < 2.0) &= \Phi\Big(\frac{2.0 - 2.1}{.55}\Big) - \Phi\Big(\frac{1.6 - 2.1}{.55}\Big) = \Phi(-.18) - \Phi(-.91) \\ &= .4286 - .1814 = .2472 \end{aligned}$$

Example (3) The body masses M of a population of fish in a lake are normally distributed about a mean of 225 grams with standard deviation 32 grams. What fraction of the fish have masses in the range 200 to 250 grams?

Solution. By the same steps as the previous example,

$$\begin{aligned} P(200 < M < 250) &= \Phi\Big(\frac{250 - 225}{32}\Big) - \Phi\Big(\frac{200 - 225}{32}\Big) = \Phi(.78) - \Phi(-.78) \\ &= .7823 - .2177 = .5646 \end{aligned}$$

Normal Approximations to Binomial Probabilities

Binomial probabilities can be modelled after flipping a biased coin a certain number of times. Precisely, suppose that we have a biased coin (meaning that the probabilities of heads may not be the same as the probability of tails), and toss the coin n times. We use the following notation

$$p = \text{Prob}(H)\,, \qquad q = 1 - p = \text{Prob}(T)\,, \qquad n = \text{number of coin tosses}$$

The probability of getting exactly k heads (and thus $n - k$ tails) is given by

$$C(n,k)\, p^k\, (1-p)^{n-k}\,, \qquad \text{where} \quad C(n,k) = \frac{n!}{k!\,(n-k)!}$$

For example, if the coin has $P(H) = .72$, and thus $P(T) = .28$, and we toss the coin 25 times, then the probability of getting exactly 20 heads equals

$$C(25,20)\,(.72)^{20}\,(.28)^5 \approx .128$$

If this were plotted it would look very much like a normal curve. We can find the formula for the normal curve that approximates the above points using the following formulas for the mean and standard deviation for a binomial experiment,

$$\mu = n\,p\,, \qquad \sigma = \sqrt{n\,p\,(1-p)}$$

For the above coin toss, $p = .72$, and $n = 25$, and thus

$$\mu = 25(.72) = 18\,, \qquad \sigma = \sqrt{25(.72)(.28)} \approx 2.24$$

We plot below the normal curve with $\mu = 18$ and $\sigma = 2.24$, that is, the function

$$f(z) = \frac{1}{2.24\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{z-18}{2.24}\right)^2}$$

and we see that the curves are quite close.

Thus the area under the normal curve approximates the probability values for the binomial experiment. The approximation is based on the following idea: the probability of getting exactly k heads is approximately equal to the area under the normal curve between the values $x = k - \frac{1}{2}$ and $x = k + \frac{1}{2}$. (This means x varies over an interval of length 1 centered at k.)

In the example of the coin toss experiment, to approximate the probability of getting exactly 20 heads, we would take the normal curve with $\mu = 18$ and $\sigma = 2.24$, and find the area under this curve between $x = 19.5$ and $x = 20.5$

By the previous section, this is

$$\begin{aligned} P(19.5 < X < 20.5) &= \Phi\Big(\frac{20.5-18}{2.24}\Big) - \Phi\Big(\frac{19.5-18}{2.24}\Big) = \Phi(1.17) - \Phi(.67) \\ &= .8790 - ..7486 = .1304 \end{aligned}$$

This is pretty close to the correct answer of .128. We have used the **histogram correction**.

The normal approximation to binomial probabilities becomes even more time saving in calculating the probability that the number of heads lies in some range. For example, the probability that the number of heads in the above experiment is at least 15 and at most 20, is approximated by the area under the curve between 14.5 and 20.5 (using the histogram correction). This calculation is

$$\begin{aligned} P(14.5 < X < 20.5) &= \Phi\Big(\frac{20.5-18}{2.24}\Big) - \Phi\Big(\frac{14.5-18}{2.24}\Big) = \Phi(1.17) - \Phi(-1.56) \\ &= 0.8790 - 0.0594 = 8196 \end{aligned}$$

To calculate this using the binomial formula we would have to calculate

$$\begin{aligned} C(25,15)(.72)^{15}(.28)^{10} &+ C(25,16)(.72)^{16}(.28)^{9} + C(25,17)(.72)^{17}(.28)^{8} \\ +C(25,18)(.72)^{18}(.28)^{7} &+ C(25,19)(.72)^{19}(.28)^{6} + C(25,20)(.72)^{20}(.28)^{5} \end{aligned}$$

This works out to be 0.806, which is pretty close to the true answer.

The normal approximation becomes even better as n gets bigger; however, it is important that the mean μ be not too close to the endpoints. In statistics, the rule of thumb is:

The normal approximation to binomial experiments can be used provided that $n\,p \geq 5$ and $n\,(1-p) \geq 5$

This means that the mean μ is between 5 and $n-5$, which forces n to be at least 10.

Example (4) In the human population, the probability of a live birth producing a male child is .52, and the probability that it produces a female child is .48. Use the normal approximation to answer the following: If 100 children are randomly chosen, what is the probability that the number of female children in the collection is at least 45 and at most 55?

Solution (4) This is a binomial experiment, since the experiment can be modelled by recording the gender of 100 children by flipping a biased coin 100 times. In this case, let $p = P(F) = .48$, and $1-p = P(M) = .52$. Then the mean and standard deviation are given by

$$\mu = np = 100(.48) = 48\,, \qquad \sigma = \sqrt{100(.48)(.52)} \approx 5.00$$

Note that the mean is between 5 and 95, so the rule of thumb says we can apply the normal approximation. The normal approximation says that the probability of at least 45 females and at most 55 females is approximately equal to

$$\begin{aligned} \Phi\Big(\frac{55.5-48}{5}\Big) - \Phi\Big(\frac{44.5-48}{5}\Big) &= \Phi(1.5) - \Phi(-.7) \\ &= .9332 - .2420 = .6912 \end{aligned}$$

Example (5) Suppose that you toss a fair coin 1000 times. Use the normal approximation to estimate the probability that the number of heads is between 490 and 510.

Solution (5) This is binomial, with $n = 1000$, $p = 1-p = .5$. Thus,

$$\mu = 1000(.5) = 500\,, \qquad \sigma = \sqrt{1000(.5)(.5)} \approx 15.8$$

The normal approximation says that the probability that the number of heads is at least 490 and at most 510 is approximately

$$\begin{aligned} \Phi\Big(\frac{510.5-500}{15.8}\Big) - \Phi\Big(\frac{489.5-500}{15.8}\Big) &= \Phi(.66) - \Phi(-.66) \\ &= .7454 - .2546 = .4908 \end{aligned}$$