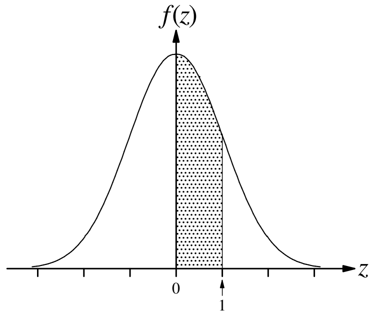
Figure 5.2. PROBABILITY MODELS: Finding Gaussian Probabilities

The following examples illustrate use of the standard Gaussian probabilities given on page Ap. 1 in Appendix B of the Course Materials, to find values for probabilities involving the Gaussian family of distributions. A convenient way to reason is to use sketches which show the required probability as the appropriate *area under a Gaussian p.d.f.* [$f(z)$ or $f(y)$]. Throughout the discussion, Z and Y are random variables such that $Z \sim G(0,1)$ and $Y \sim G(\mu, \sigma)$.

1. Table Look-up of $G(0,1)$ Probabilities

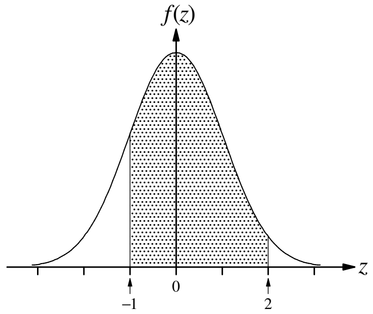
$\Pr(0 < Z \leq 1)$ = area under $G(0,1)$ p.d.f. from 0 to 1
= area under $G(0,1)$ p.d.f. from $-\infty$ to 1
 − area under $G(0,1)$ p.d.f. from $-\infty$ to 0
= 0.8413 − 0.5000 [from the $G(0,1)$ Table]
= 0.3413

[By symmetry, $\Pr(-1 < Z \leq 0)$ is also 0.3413]



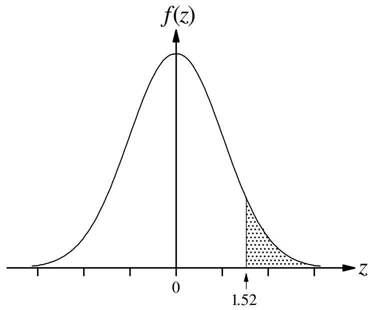
$\Pr(-1 < Z \leq 2)$ = area under $G(0,1)$ p.d.f. from −1 to 2
= area under $G(0,1)$ p.d.f. from −1 to 0
 + area under $G(0,1)$ p.d.f. from 0 to 2
= area under $G(0,1)$ p.d.f. from 0 to 1 [symmetry]
 + area under $G(0,1)$ p.d.f. from 0 to 2
= 0.3413 + 0.4772 [from the $G(0,1)$ Table]
= 0.8185

[By symmetry, $\Pr(-2 < Z \leq 1)$ is also 0.8185]



$\Pr(Z > 1.52)$ = area under $G(0,1)$ p.d.f. from 1.52 to ∞
= 1 − area under $G(0,1)$ p.d.f. from $-\infty$ to 1.52
= 1 − 0.9357 [from the $G(0,1)$ Table]
= 0.0643

[By symmetry, $\Pr(Z \leq -1.52)$ is also 0.0643]



NOTE: 1. Unless context dictates otherwise, we write $\Pr(a < Y \leq b)$, the probability the random variable Y (say) takes values between a and b, so as to *ex*clude the lower end-point of the interval (a, b] and *in*clude the (finite) upper end-point.
- This convention maintains consistency when we find a probability using our definition of the cumulative distribution function $F(y) = \Pr(Y \leq y)$ to find a probability (*e.g.*, see Table 5.3.1 on page 5.8 of Figure 5.3).

$\boxed{1}$ How would the values of the probabilities calculated above be *altered* if the inequalities involved a strictly less than sign (*i.e.*, <) in place of each ≤ sign? Justify your answer by reference to appropriate properties of the definite integral.
- Answer the same question if the ≤ or ≥ signs were replaced by < or > signs.
  – What do you conclude from your previous answers about use of the Gaussian distribution for modelling continuous quantities? Explain briefly.

$\boxed{2}$ Draw a sketch to illustrate the probability found by symmetry in the last line of each of the three examples given above.

$\boxed{3}$ With*out* further look-up of a table of standard Gaussian probabilities, use the information in the three examples given above to evaluate the five probabilities which follow, and draw a sketch to illustrate each one:
- $\Pr(Z > 1)$
- $\Pr(Z > 2)$; $\Pr(Z \leq -1)$
- $\Pr(Z \leq 1.52)$; $\Pr(Z > -1.52)$.

$\boxed{4}$ Discuss briefly the advantage(s) and disadvantage(s) of a *two*-panel table of $G(0,1)$ probabilities [for example, as in Appendix B of the course Notes in Table 2 (the overleaf side)], compared with the *one*-panel version in Table 1 on the front of this page.

2. Table Look-up of $G(\mu, \sigma)$ Probabilities

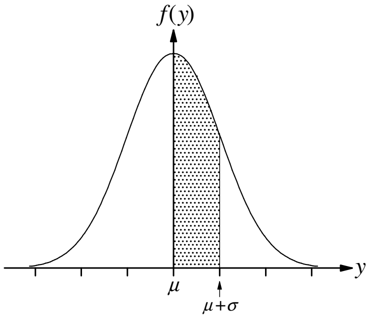
$$\frac{Y-\mu}{\sigma} \sim G(0, 1) \qquad \text{-----(5.4.1)}$$

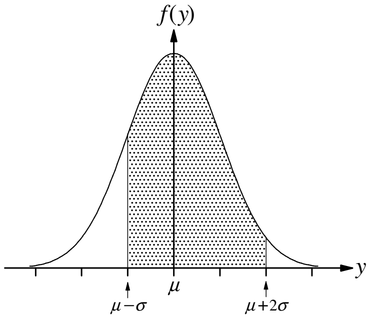
To find probabilities for Gaussian distributions *other than* the standard [or $G(0,1)$] case, we use the operation of *standardization* (*i.e.*, subtract the mean and divide the difference by the standard deviation) to obtain the *equivalent* probability for the $G(0,1)$ distribution; the basis of this operation is the result given as equation (5.4.1) at the right above. We illustrate this matter using probabilities equivalent to those in the three examples overleaf on page 5.3; recall that $Y \sim G(\mu, \sigma)$ and $Z \sim G(0,1)$.

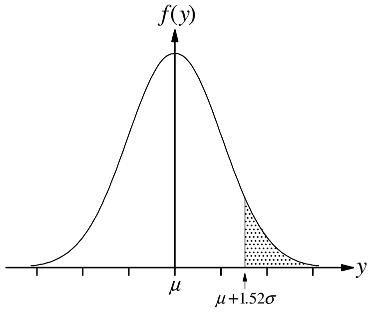
$$\begin{aligned} &\Pr(\mu < Y \le \mu+\sigma) \\ &= \Pr\left(\frac{\mu-\mu}{\sigma} < \frac{Y-\mu}{\sigma} \le \frac{\mu+\sigma-\mu}{\sigma}\right) \\ &= \Pr[0 < G(0,1) \le 1] \\ &= \Pr(0 < Z \le 1) \\ &= 0.3413 \end{aligned}$$

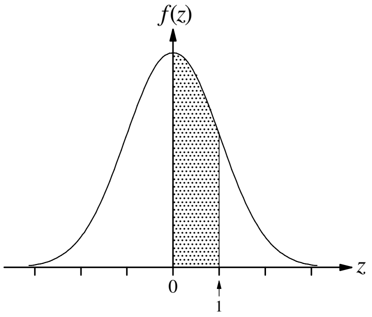
$$\begin{aligned} &\Pr(\mu-\sigma < Y \le \mu+2\sigma) \\ &= \Pr\left(\frac{\mu-\sigma-\mu}{\sigma} < \frac{Y-\mu}{\sigma} \le \frac{\mu+2\sigma-\mu}{\sigma}\right) \\ &= \Pr[-1 < G(0,1) \le 2] \\ &= \Pr(-1 < Z \le 2) \\ &= 0.8185 \end{aligned}$$

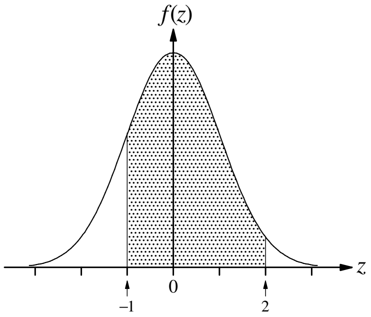
$$\begin{aligned} &\Pr(Y > \mu+1.52\sigma) \\ &= \Pr\left(\frac{Y-\mu}{\sigma} > \frac{\mu+1.52\sigma-\mu}{\sigma}\right) \\ &= \Pr[G(0,1) > 1.52] \\ &= \Pr(Z > 1.52) \\ &= 0.0643 \end{aligned}$$

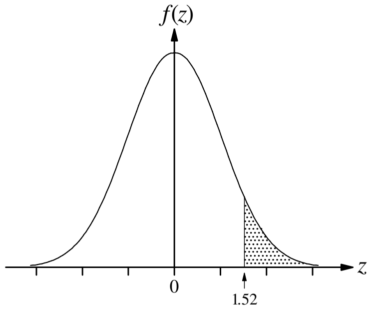












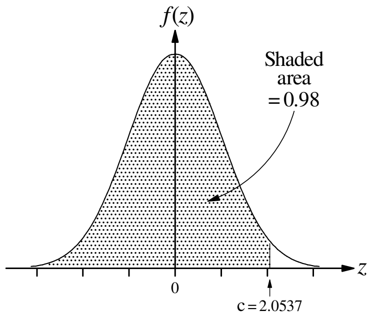
3. 'Inverse' Table Look-up of Gaussian Percentiles

The probability calculations overleaf and above for the Gaussian random variables Z and Y involve 'direct' use of a table of standard Gaussian probabiliies; *i.e.*, we have a value (z) which we locate in the left and top *margins* of the Table and we then read off the value of the corresponding *table entry*. However, in some situations we need the 'inverse' process of table look-up; *i.e.*, we have a *probability* (or *area*) which is, or is close to, the value of a table entry and we have to find the corresponding *z-value* (a Gaussian quantile or percentile) in the table margins.

As an illustration, suppose we want the value of c such that $\Pr(Z \le c) = 0.98$. We look in the *body* of the table for 0.98000 and find 0.97982 for $z = 2.05$ and 0.98030 for $z = 2.06$; hence, to 2 decimal places, c = 2.05. To obtain a more accurate value for c, we can use *linear interpolation* between the z-values of 2.05 and 2.06; we have:

$$c = 2.05 + (2.06 - 2.05) \times \frac{0.98000 - 0.97982}{0.98030 - 0.97982} = 2.05375.$$

Although it is important to *understand* this interpolation procedure for obtaining additional decimal places for z-values, we can often avoid having to actually *do* it by using the separate lower right section of Table 1 in Appendix B; this Table gives z-values [to 4 decimal places] with*out* interpolation for selected probabilities [denoted $F(z)$]. We see that the c value for our example is actually 2.0537 (or 2.053 749); thus, linear interpolation gives us, in this instance, the correct value to five decimal places. A diagrammatic representation of this example is given at the right.



5 With*out* further table look-up, use the information in the example to find the value of d such that $\Pr(Z \le d) = 0.02$; draw a sketch to illustrate your answer.