Figure 5.9. CONTINUOUS RANDOM VARIABLES: Selected Characteristics

The normal distribution is a useful model (or idealization) for the shape of data sets for which the distribution (*e.g.*, as shown by an appropriate stemplot or histogram) has a central peak with a roughly symmetrical falling away on either side. Before we learn about continuous probability models for *other* distribution shapes, we discuss six characteristics of *all* such models:

- ∗ Shape (p.d.f.)
- ∗ Normalization
- ∗ Mean
- ∗ Standard deviation
- ∗ Probability
- ∗ Shape (c.d.f.).

The *c*.d.f. (*cumulative distribution function*), like the p.d.f. (*probability density function*), describes the *shape* of a distribution but is discussed *last* in this Figure because it is the most difficult of the six characteristics to manipulate; Figures 5.11 and 5.12 also deal with these same characteristics but, in these two Figures, the c.d.f. follows immediately after the p.d.f.

1. Shape: (P.d.f.) This characteristic is described by the *probability density function* (*p.d.f.*) of the random variable; a p.d.f. can take only non-negative (*i.e.*, zero or positive) values. For example, the *normal* p.d.f. is as given in equation (5.9.1) at the right above; its graph is the familiar 'bell-shaped' curve representing non-negative values of $f(y)$ over the *whole* real axis.

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{y-\mu}{\sigma}\right]^2} \quad ; \quad -\infty < y < \infty \qquad \text{-----(5.9.1)}$$

2. Normalization: The area of a histogram is 1 (or 100%) and we use an appropriate p.d.f. to model the shape of a histogram of data for a continuous variate (*e.g.*, measurements); hence, the area under a p.d.f. must also be 1 (or 100%). Because area under a curve is found by integration, the expression (5.9.2) at the right above must hold; it is called the *normalizing condition* [for the p.d.f. $f(y)$].

$$\int_{-\infty}^{\infty} f(y)dy = 1 \qquad \text{-----(5.9.2)}$$

Sometimes, a p.d.f. is written as a function which involves a constant (k, say); our task may then be to use the normalizing condition to find the value of k.

Example 5.9.1: If we write the normal p.d.f. as: $f(y) = \text{k}e^{-½[(y-\mu)/\sigma]^2}$; $-\infty < y < \infty$,

then setting: $\int_{-\infty}^{\infty} f(y)dy = \text{k}\int_{-\infty}^{\infty} e^{-½[(y-\mu)/\sigma]^2}dy = 1$ yields: $\text{k} = 1/\sqrt{2\pi}\sigma$.

Example 5.9.2: If the random variable V has the p.d.f. shown at the right, we can use the normalizing condition (5.9.2) to find k = 2.

$$f(v) = \begin{cases} \text{k}v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

3. Mean: This characteristic, a measure of the location of a distribution, is defined in equation (5.9.3) at the right. We see from equation (5.9.3) [and (5.9.2)] that the mean is the *centroid* of the distribution and so represents its point of balance; for a *symmerical* distribution, the mean is the centre of symmetry (or mid-point).

$$\mu \equiv E(Y) = \int_{-\infty}^{\infty} y \bullet f(y)dy \qquad \text{-----(5.9.3)}$$

The usual notation for the mean is μ (lower case Greek mu) or $E(Y)$, the latter reflecting the words *expectation* (or *expected value*) of the random variable Y, which are synonyms for 'mean.'

Example 5.9.3: For the normal distribution, $N(\mu, \sigma)$, the mean is:

$$E(Y) = \int_{-\infty}^{\infty} y \bullet f(y)dy = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} y \bullet e^{-½[(y-\mu)/\sigma]^2}dy = \mu;$$

i.e., the mean of the normal distribution is its parameter μ.

NOTE: 1. We need to distinguish two uses of the symbol μ:

- a parameter of the normal distribution, $N(\mu, \sigma)$, which happens to be its mean;
- a symbol for the mean of *any* distribution – if there is more than one random variable in a particular context, an added subscript avoids ambiguity (*e.g.*, μ_Y is the mean of Y).

Example 5.9.4: If the random variable V has the p.d.f. shown at the right,

the mean of V is: $E(V) = \int_{-\infty}^{\infty} v \bullet f(v)dv = \int_{0}^{1} v \bullet 2v dv = \frac{2}{3}$.

$$f(v) = \begin{cases} 2v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

As described overleaf in Section 4, an *extension* of the idea of expectation is involved in both calculating and defining the standard deviation of a distribution – see also Note 7 overleaf on page 5.22.

Another measure of the location (or 'centre') of a distribution is its *median*; finding the median is discussed on the third and fourth sides (pages 5.23 and 5.24) of this Figure in Examples 5.9.9 and 5.9.10 of the topic **Probability**.

(*continued overleaf*)

4. S.d.: This characteristic, a measure of *variation* for a distribution, is found as shown at the right in equations (5.9.4) and (5.9.5). [It may be an aid to memory to think of the upper expression as: *the square root of the mean square minus the squared mean*; the type of operation represented by these terms is reminiscent of the *calculation* formulae for the standard deviation of *data* in Section 5 of Figure 4.8 of the Course Materials.]

$$\sigma \equiv s.d.(Y) = \sqrt{E(Y^2) - [E(Y)]^2} \quad \text{-----(5.9.4)}$$

$$\text{where:} \quad E(Y^2) = \int_{-\infty}^{\infty} y^2 \cdot f(y)\,dy \quad \text{-----(5.9.5)}$$

The usual notation for (probabilistic) standard deviation is σ (lower case Greek sigma) or *s.d.* (or one of its equivalents such as s.d. or SD) followed by the random variable in brackets.

Example 5.9.5: For the normal distribution, $N(\mu, \sigma)$, we can find that:

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 \cdot f(y)\,dy = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} y^2 \cdot e^{-\frac{1}{2}[(y-\mu)/\sigma]^2}dy = \sigma^2 + \mu^2;$$

hence: $s.d.(Y) = \sqrt{E(Y^2) - [E(Y)]^2} = \sqrt{(\sigma^2 + \mu^2) - [\mu]^2} = \sigma;$

i.e., the standard devation of the normal distribution is its parameter σ.

NOTE: 2. As with the mean in Note 1 overleaf, we distinguish two uses of the symbol σ:

- a parameter of the normal distribution, $N(\mu, \sigma)$, which happens to be its standard deviation;
- a symbol for the standard devation of *any* distribution – if there is more than one random variable in a particular context, an added subscript avoids ambiguity (*e.g.*, σ_Y is the standard deviation of Y).

Example 5.9.6: If the random variable V has the p.d.f. shown at the right:

$$f(v) = \begin{cases} 2v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

$$E(V^2) = \int_{-\infty}^{\infty} v^2 \cdot f(v)\,dv = \int_{0}^{1} v^2 \cdot 2v\,dv = \tfrac{1}{2}.$$

Hence: $s.d.(V) = \sqrt{E(V^2) - [E(V)]^2} = \sqrt{\tfrac{1}{2} - [\tfrac{2}{3}]^2} = \sqrt{\tfrac{1}{18}} = \tfrac{1}{3\sqrt{2}} \simeq 0.2357.$

NOTES: 3. It is *strongly* recommended that the calculation of standard deviation be done in three *distinct* steps:

(1) find $E(Y)$; (2) find $E(Y^2)$; (3) find $s.d.(Y)$.

Doing more than one step simultaneously makes the process *much* more prone to mistakes.

4. As a measure of variation, the standard deviation is most useful for *symmetrical* distributions; for *un*symmetrical (or *skewed*) distributions, a measure of variation like the interquartile range (or IQR) is usually more appropriate.

5. For the *normal* distribution, one 'interpretation' of the standard deviation is that the two *points of inflection* of the p.d.f. lie one standard deviation either side of the mean; *i.e.*, they lie at $\mu \pm \sigma$. However, in general, there is no simple or 'natural' interpretation of the value of a standard deviation, such as the value of $1/3\sqrt{2}$ for the p.d.f. in Example 5.9.6 above, although when we *compare* the standard deviations of *two* distributions, we can say that the distribution with the *larger* standard deviation has greater variation – its pd.f. will (usually) appear more dispersed (or spread out).

6. Although standard deviations are usually calculated from the expressions (5.9.4) and (5.9.5) above (as illustrated in Examples 5.9.5 and 5.9.6 above), the analytically equivalent (see Exercise 5 on the fifth side of this Figure) *definition* (5.9.6) of the standard deviation of a random variable (Y, say) is *the square root of the expected squared deviation (or difference) of Y from its mean*; symbolic versions of this definition are given at the right, where $E(Y)$ could be written as μ or μ_Y. Two expressions [*viz*. a definition (5.9.6) and its calculational equivalent (5.9.4)] for the standard deviation of a *random variable* is the probabilisitic analogue of the situation for *data* in Section 5 of Figure 4.8.

$$\sigma \equiv s.d.(Y) = \sqrt{E[Y - E(Y)]^2} \quad \text{-----(5.9.6)}$$

$$= \sqrt{\int_{-\infty}^{\infty} [y - E(Y)]^2 \cdot f(y)\,dy} \quad \text{-----(5.9.7)}$$

- The discussion in this Section 4 is about 'probabilistic' (*not* 'data') standard deviation – recall Note 1 near the middle of page 5.4, the second side of Figure 5.1.

7. The definition of standard deviation in Note 6 involves a special case of the result (5.9.8) at the right for the expectation (or mean) of a *function* [$h(Y)$, say] of a random variable (Y). Equation (5.9.8) enables us to establish two useful expressions for the mean and standard deviation of a *linear function* (a + bY, say, where a and b are constants) of the random variable:

$$E[h(Y)] = \int_{-\infty}^{\infty} h(y) \cdot f(y)\,dy \quad \text{-----(5.9.8)}$$

$$E(\text{a} + \text{b}Y) = \text{a} + \text{b}E(Y) \quad ; \quad s.d.(\text{a} + \text{b}Y) = |\text{b}|s.d.(Y). \quad \text{-----(5.9.9)}$$

8. The square of the standard deviation is the *variance*; *i.e.*, $var(Y) = [s.d.(Y)]^2$ -----(5.9.10)

Figure 5.9. CONTINUOUS RANDOM VARIABLES Selected Characteristics: (continued 1)

5. Probability: In a histogram, proportion (or relative frequency) is represented by *area* and we use an appropriate p.d.f. to model the shape of a histogram of continuous data (*e.g.*, measurements); because probabilities are estimated by observed proportions, it is *area* under a p.d.f. that represents probability. Thus, as shown at the right in equation (5.9.11), probability for a continuous random variable (Y, say) is found by *integrating the p.d.f. of Y between appropriate limits* (a and b, say, where a < b).

$$\Pr(a < Y \le b) = \int_a^b f(y)dy \qquad \text{-----(5.9.11)}$$

Example 5.9.7: If the random variable $Z \sim N(0, 1)$, so that: $f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$; $-\infty < z < \infty$,

then: $\Pr(a < Z \le b) = \int_a^b f(z)dz = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-½z^2}dz.$

This integral can only be evaluated *numerically* (*e.g.*, using Simpson's rule) for given values of a and b; thus, we now understand:

- why we obtain probabilities for the standard normal distribution *in practice* by table look-up rather than by integration;
- how the values are obtained in a table of standard normal probabilities (like Figure 5.4 of the Course Materials).

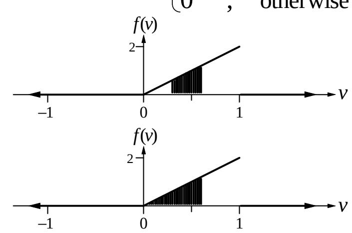
Example 5.9.8: If the random variable V has the p.d.f. shown at the right:

$$f(v) = \begin{cases} 2v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

$$\Pr(0.3 < V \le 0.6) = \int_{0.3}^{0.6} f(v)dv = \int_{0.3}^{0.6} 2v\,dv = 0.27;$$

$$\Pr(-1 < V \le 0.6) = \int_{-1}^{0.6} f(v)dv = \int_{0}^{0.6} 2v\,dv = 0.36;$$

the diagrams at the right show the p.d.f. with shaded areas representing the two probabilities.



NOTES: 9. The second probability calculation in Example 5.9.8 reminds us that only those part(s) of a p.d.f. where the function takes *positive* values contribute to a probability; thus, for the p.d.f. in Example 5.9.8, the probability for the interval from –1 to 0.6 comes *only* from the sub-interval *0* to 0.6, where *this* p.d.f. is non-zero.

10. We know, from knowledge of integration, that as the limits of integration (a and b) approach each other, the area under a given p.d.f. gets *smaller*; in the limit, when a = b, the area is *zero*. Hence, for *any* continuous random variable, the probability it takes on any *particular* value is *zero*. [Thinking of this another way, a continuous random variable can take on an *infinite* number of values so the probability of any *one* of these values is zero.]

Three comments about this property of continuous random variables are:

- We are reminded that mathematical models of the real world are only *approximations* – values of many quantities we model by continuous random variables (such as measurements) have *non*-zero probabilities of occurrence.
- Because there is *zero* probability associated with either end-point of an interval, when we find the probability a continuous random variable lies in an interval from a to b (with a < b), for *any* continuous random variable we have:

$$\Pr(a < Y \le b) = \Pr(a \le Y < b) = \Pr(a < Y < b) = \Pr(a \le Y \le b). \qquad \text{-----(5.9.12)}$$

- The zero probability of any *particular* value of any continuous random variable is why the word *density* is part of the name (*viz*. probability *density* function) of the function that describes the shape of a continuous distribution; similarly, the scale on the vertical axis of a histogram is a *density* scale, *proportion per unit* or *percent per unit*, the 'unit' being that of the variate on the horizontal axis.

Example 5.9.9: The *median* of a distribution is the value of the random variable that divides the area under the p.d.f. in *half*. Because the normal distribution is *symmetrical*, its median is *equal to* its mean – recall the central diagram in Section 8 of Figure 4.1 of the Course Materials.

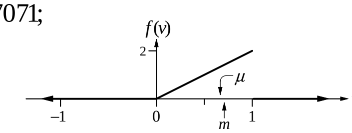
Example 5.9.10: If the random variable V has the p.d.f. shown at the right, the median (m, say) of V is given by:

$$f(v) = \begin{cases} 2v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

(*continued overleaf*)

Example 5.9.10: (continued) $0.5 = \int_{-\infty}^{m} f(v)dv = \int_{0}^{1} 2v\,dv$, which yields: $m = \frac{1}{\sqrt{2}} \simeq 0.7071$;

we see that, for this distribution, the median is a little larger than the mean of $\frac{2}{3} = 0.\dot{6} \simeq 0.6667$ (recall Example 5.9.4 on the first side of this Figure).



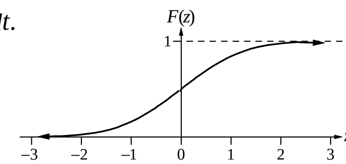
6. Shape: (C.d.f.) Instead of describing a continuous probability distribution by its probability density function, an equivalent alternative is to give its *cumulative (distribution) function* [the word *distribution* is optional in the name], abbreviated *c.f.* or *c.d.f.*; the definition of this function (for the random variable Y) is given as equation (5.9.13) at the right. [This definition uses a (dummy) variable of integration (t) to distinguish it from y, the argument of the c.d.f. and the upper limit of integration.]

$$F(y) = \Pr(Y \le y) = \int_{-\infty}^{y} f(t)dt \qquad \text{-----(5.9.13)}$$

Example 5.9.11: If the random variable $Z \sim N(0, 1)$, so that: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$; $-\infty < z < \infty$,

the c.d.f. of Z is:: $F(z) = \Pr(Z \le z) = \int_{-\infty}^{z} f(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-½t^2} dt.$

This integral can only be evaluated *numerically* (*e.g.*, using Simpson's rule) for a given value of z; hence:



- this integral expression is the symbolic form of the standard normal cumulative distribution function;
- values of 0.5 and greater of this function are given in the lower right-hand section of the first side of Figure 5.4 of the Course Materials [only the values for the *upper half* of the function are given in this table because $F(-z) = 1 - F(z)$]; its graph is shown above at the right – the function only *reaches* 0 on the left at $-\infty$ and 1 on the right at $+\infty$.

Example 5.9.12: If the random variable V has the p.d.f. shown at the right, the c.d.f. of V is:

$$f(v) = \begin{cases} 2v & ; \quad 0 < v \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

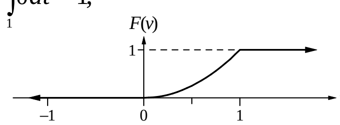
for $-\infty < v \le 0$: $F(v) = \Pr(V \le v) = \int_{-\infty}^{v} f(t)dt = \int_{-\infty}^{v} 0\,dt = 0;$

for $0 < v \le 1$: $F(v) = \Pr(V \le v) = \int_{-\infty}^{v} f(t)dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{v} 2t\,dt = v^2;$

for $1 < z < \infty$: $F(v) = \Pr(V \le v) = \int_{-\infty}^{v} f(t)dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} 2t\,dt + \int_{1}^{v} 0\,dt = 1;$

$$F(v) = \begin{cases} 0 & ; \quad v \le 0 \\ v^2 & ; \quad 0 < v \le 1 \\ 1 & ; \quad v > 1 \end{cases}$$

The diagram at the right shows the c.d.f. of V.



NOTES: 11. A probability can be found as the *difference* between values of a c.d.f.; for example, for the random variable Y and given constants a and b:

$$\Pr(\text{a} < Y \le \text{b}) = \int_{\text{a}}^{\text{b}} f(y)dy = \int_{-\infty}^{\text{b}} f(y)dy - \int_{-\infty}^{\text{a}} f(y)dy = F(\text{b}) - F(\text{a}); \qquad \text{-----(5.9.14)}$$

as an illustration, the c.d.f. in Example 5.9.12 yields: $\Pr(0.3 < V \le 0.6) = F(0.6) - F(0.3) = 0.27$, as also obtained by integrating the corresponding p.d.f. in Example 5.9.8 overleaf on page 5.23.

12. At the *median* (m, say) of a distribution, $F(m) = 0.5$. -----(5.9.15)

13. The following Table 5.9.1 summarizes some properties of p.d.f.s and c.d.f.s:

Table 5.9.1:

Property	P.d.f.	C.d.f.:
Domain and range	$(-\infty, \infty)$; $f(y) \ge 0$	$(-\infty, \infty)$; $0 \le F(y) \le 1$
Values at infinity	$f(-\infty) = 0$; $f(\infty) = 0$	$F(-\infty) = 0$; $F(\infty) = 1$
Relationships	$f(y) = \frac{d}{dy}F(y)$	$F(y) = \int_{-\infty}^{y} f(t)dt = \Pr(Y \le y)$
Probability	Area under $f(y)$: $\Pr(\text{a} < Y \le \text{b}) = \int_{\text{a}}^{\text{b}} f(y)dy$	Difference in values of $F(y)$: $\Pr(\text{a} < Y \le \text{b}) = F(\text{b}) - F(\text{a})$

All c.d.f.s are *non-decreasing* functions; we can justify this statement in two ways:

- because $F(y) = \Pr(Y \le y)$ and probabilities are non-negative;
- because $f(y)$ is the *slope* of the c.d.f. and $f(y)$ is non-negative.

(*continued*)

Figure 5.9. CONTINUOUS RANDOM VARIABLES: Selected Characteristics (continued 2)

7. Appendix: The Gamma Function, $\Gamma(\alpha)$

The *gamma function* of α is defined as: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx.$ -----(5.9.16)

Three properties or values of the gamma function are of interest in the present context:

- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. -----(5.9.17)
- $\Gamma(\alpha) = (\alpha-1)!$ if α is a positive integer; *e.g.*, $\Gamma(6) = \int_0^\infty x^5e^{-x}dx = 5! = 120$. -----(5.9.18)
- $\Gamma(½) = \sqrt{\pi}$; *i.e.*, $\int_0^\infty x^{-½}e^{-x}dx = \sqrt{\pi}$; as a consequence: $\int_{-\infty}^{\infty} e^{-½z^2}dz = \sqrt{2\pi}$. -----(5.9.19)

8. Exercises

1. In Example 5.9.1 on page 5.21 of this Figure, show how the integral arising from using the normalizing condition (5.9.2) yields the value of $1/\sqrt{2\pi}\sigma$ for k.

2. In Example 5.9.2 on page 5.21 of this Figure, show how using the normalizing condition (5.9.2) leads to the value k = 2.

3. In Example 5.9.3 on page 5.21 of this Figure, show how the integration leads to μ for the mean of the normal distribution.

4. In Example 5.9.5 on page 5.22 of this Figure, show how the integration leads to the value of $\mu^2 + \sigma^2$ for $E(Y^2)$.

5. Establish the equivalence of the *definition* of the standard deviation of a random variable [equation (5.9.6) in Note 6 on page 5.22 of this Figure] and the *calculational* expression [equation (5.9.4)] at the top right of the same page.

6. Show why, in Example 5.9.6 on page 5.22 of this Figure, the first integration over $(-\infty, \infty)$ becomes integration over $(0, 1]$ in the second integral.

7. Establish the two results (5.9.9) given in Note 7 near the bottom of page 5.22 of the Figure for the mean and standard deviation of the linear function $(a + bY)$ of the random variable Y.

8. If $E(Y) = \mu$ and $s.d.(Y) = \sigma$, show that the *standardized* variable $Z = (Y - \mu)/\sigma$ has a mean of 0 and a standard deviation of 1.
   - Combined with the fact that if Y is normally distributed, Z also is normal, these results are the basis of the process of *standardizing* for obtaining probabilities for $N(\mu, \sigma)$ random variables from the $N(0, 1)$ table in Figure 5.4.

9. Give a clear and concise statement of the (common) meaning of the two phrases:
   - *the standard deviation of a random variable* and: • *the standard deviation of a distribution*.

10. Justify the statement in Example 5.9.11 on page 5.24 of this Figure that, for the standard normal c.d.f.: $F(-z) = 1 - F(z)$.

11. On the diagram of the c.d.f. in Example 5.9.12 on page 5.24 of this Figure, show the two probabilities calculated for this distribution in Example 5.9.8 on page 5.23 of the Figure.

12. The term 'otherwise' permits more *compact* presentation of the equations of some p.d.f.s, as illustrated in Examples 5.9.2, 5.9.4, 5.9.6, 5.9.8, 5.9.10 and 5.9.12. Write the p.d.f. in these Examples with*out* using 'otherwise'.
    - Explain briefly why 'otherwise' is *not* similarly useful for *c*.d.f.s.

13. Using the result given in Note 12 near the bottom of page 5.24 of this Figure, find the median of the distribution in Example 5.9.12; check that your answer agrees with that obtained by integration of the *p*.d.f. of this distribution in Example 5.9.10 at the top of page 5.24.

14. Using integration by parts, verify the first property (5.9.17) given above for the gamma function.

15. Using mathematical induction, prove the second property (5.9.18) given above for the gamma function.

16. If you have an adequate calculus background, verify as follows the result given above in equation (5.9.19): $\int_{-\infty}^{\infty} e^{-½z^2}dz = \sqrt{2\pi}$.
    - write the integral in terms of x and *again* in terms of y;
    - write the *product* of the two (single) integrals as a *double* integral;
    - change from Cartesian to plane polar coordinates, remembering that $dxdy$ becomes $rdrd\theta$;
    - evaluate the resulting double integral.

    Then show, by the change of variable $v = x^{½}$, that: $\Gamma(½) = \sqrt{\pi}$.

17. Using an appropriate change of variable, find the relationship between: $\int_0^\infty x^n e^{-x}dx$ and: $\int_{-\infty}^{0} x^n e^{x}dx$.

Blank page