Universidad Austral de Chile
Faculty of Forestry Sciences

Notes on FOREST MENSURATION I. Statics

Oscar García

September 1995

Institute of Forest Management
Chair of Forest Mensuration

Translation from Spanish, January 2004

I decided to translate these notes in order to use them as a framework for Directed Studies courses in Forest Mensuration at the University of Northern British Columbia. They are used to organize the sequence of topics to be discussed, being complemented with readings from the standard textbooks. I resisted the temptation of tampering with the original text; a revision could easily have got out of hand in terms of time and effort. Some of the content is outdated, and in parts reflects local practices and needs. "Dynamics" (growth & yield) is not covered. The translation has, however, served its purpose reasonably well, and perhaps others might find something of interest in it. My apologies for the far from polished English. Adrian Batho suffered with good humor the inagural experiment, and contributed to blunt some of the roughest linguistic edges.

O. García
garcia@unbc.ca

Contents

1 Introduction **4**

2 Logs and forest products **7**
2.1 Length 7
2.2 Diameters 9
2.3 Cubature of logs 13
2.4 Stacked wood 19
2.5 Weight measure 22
2.6 Sawn timber 23
2.7 Change of units 27

3 Trees **28**
3.1 Diameters 28
3.2 Heights 29
3.3 Cubature of trees 31
3.4 Stem analysis 32
3.5 Volume functions (tables) 35
3.6 Form factors and quotients, etc. 37
3.7 Taper functions (curves) 38
3.8 Bark functions 41

4 Stands **42**
4.1 Diameter, basal area 42
4.2 Heights 43
4.2.1 Height-diameter curves 43
4.2.2 Dominant height 44
4.3 Cubature of stands/plots 46
4.4 Volume functions 48
4.5 Stand tables, distributions 48

A Errors **54**
A.1 Error bounds 54
A.2 Significant figures 56

B Regression **62**
B.1 Matrices 62
B.2 The least-squares method 66
B.3 Statistical considerations 71

Chapter 1

Introduction

Forest Mensuration, Dasometrics or Dendrometrics, deals with the quantification of forests, trees, and forest products. We can distinguish in it techniques for direct or indirect measurement, estimation procedures using statistical relationships, and methods of prediction where the variable time takes part. The main topics include:

	MEASUREMENT (direct, indirect)	ESTIMATION (statistical)	PREDICTION (over time)
LOGS (products)	Length, diameter Cubication (volume) Log rules Defects, quality	Volume functions Sawn timber conv. Weight-volume conv. Stacked wood	
TREES	DBH, height, bark Cubication Stem analysis	Volume functions Taper functions Bark functions Product assortment	
STANDS	Stand tables	Height – DBH DBH distributions Volume functions Inventories	Site quality Growth Mortality

We will cover these topics approximately in the order indicated, except that Inventory will be treated in a separate course. In the time available it will only be possible to introduce the concepts and main methods in a rather superficial form. Some subjects of mathematics and statistics will be reviewed and/or explained to the extent and at the time at which it becomes necessary. We will emphasize principles, foundations and generally applicable methods, indicating information sources to refer to when facing specific problems.

An important objective is to develop abilities of quantitative reasoning and of application of mathematical and statistical knowledge to real

problems and new situations. It is expected that many practical details on specific methods and measurements will be learned through reading the literature, exercises, and subsequent courses. The laboratories and subjects treated in them will be an important and integral part of the course.

Topics less important or somewhat peripheral to the central subject will be indicated with smaller font and with a ♡. Items marked with ♡♡ are included mostly for curiosity value.

General references:

INSTITUTO DE MANEJO FORESTAL, Cátedra de Manejo Forestal. Previous years notes. (Much relevant and complementary material, in Spanish).

HUSCH,B. ,MILLER,C.I., and BEERS,T.W. Forest Mensuration. 3rd Edition. Wiley. 1982. (Good elementary text, emphasizing measurement principles).

AVERY, T.E. and BURKHART, H.E. Forest Measurements. Fourth edition. McGraw-Hill. 1994. (Another elementary text, brief explanations and emphasis on North American practices).

AVERY, T.E. Natural Resources Measurements. Second edition. McGraw-Hill. 1975. (Earlier edition, available in the Library).

LOETSCH, F., ZOEHRER, F. and HALLER, K. Forest Inventory, Vol.II. BLV Verlagsgesellschaft. 1973. (Fairly comprehensive, good reference text).

PARDE, J. Dendrométrie. Editions de l'Ecole Nationale des Eauxs et Forêts – Nancy. 1961. (Mensuration text in French. There is an updated edition by Pardé and Bouchon, but not in the Library).

PRODAN, M. Holzmeßlehre. J.D. Sauërlander's Verlag, Frankfurt. 1965. (Classic, in German).

CAB VAN LAAR, A. y AKQA, ? ?? (To appear, but by the authors it should be excellent).

SPURR, S.H. Forest Inventory. Ronald. 1952. (A classic, still interesting in its treatment of volume tables, growth fundamentals, etc.).

CAILLIEZ, F. Estimación del Volumen Forestal y Predicción del Rendimiento. Vol.1 - Estimación del Volumen. Estudio FAO: Montes 22/1. 1980. (Volumes, Spanish translation).

ALDER, D. Estimación del Volumen Forestal y Predicción del Rendimiento. Vol.2 - Predicción del Rendimiento. Estudio FAO: Montes 22/2. 1980. (One of the best on growth models. Spanish translation).

BRUCE, D. and SCHUMACHER, F.X. Medición Forestal. Editorial Herrero. 1965. (Translation of a fairly old text, although still useful in its instrument and measurements part).

CLUTTER, J.L., FORTSON, J.C., PIENAAR, L.V., BRISTER, G.H. and BAILEY, R.L. Timber Management: A Quantitative Approach. Wiley. 1983. (Forest Management text, also good coverage of the most common methods in growth modelling).

ASSMAN, E. The Principles of Forest Yield Study. Pergamon Press. 1970. (Good reference, lots on European growth studies. Suffers of lack of structure and analysis).

CARRON,L.T. An Outline of Forest Mensuration with Special Reference to Australia. Australian National University Press. 1968. (Brief and without much detail, but interesting for mensuration in plantations).

PETERS N.,R., JOBET J.,J. y AGUIRRE A.,S. Compendio de Tablas Auxiliares Para el Manejo de Plantaciones de Pino Insigne. Instituto Forestal, Manual No.14. 1985. (Practices, equations, etc., used in Chile).

HAMILTON, G.J. Forest Mensuration Handbook. Forestry Commission Booklet No.39. Her Majesty's Stationery Office, London. 1975. (Interesting manual for its detailed description of standards and procedures. Discusses appropriate degrees of refinement in measurements according to costs and benefits).

Chapter 2

Logs and forest products

2.1 Length

Lengths are usually measured with tabes or graduated rods, and they do not present major difficulties. It must be remembered nevertheless that the exact definition of a length is subject to conventions that can vary from a situation to another. For example, often it is common to round down log lengths to the nearest foot or decimeter. Depending on the application or established conventions, the length of a curved log could be measured in a straight line, or following the curvature.

♡ Although measuring lengths may seem simple, some aspects are not completely obvious. How long is the coast of Chile? How to measure it? It could be measured on a map by carefully translating a ruler, walking a compass over it, with a curvometer, or tracing it with a thread. In any case, the measurement on a small scale student map surely will give a value smaller than if it is measured in charts at scale 1:50000. It is possible to imagine that if the process is repeated on aerial photographs every time on larger scale, greater lengths will be obtained. Which is the "true" length? Is there such a thing?

Questions, exercises

1. Measure the length of the coast on a map, or of the irregular edge of some leaf, with a compass opened to 3 cm. Repeat with ever decreasing openings.

2. Graph the logarithm of the lengths obtained over the logarithm of the resolution (opening of the compass). See any trend?

3. Repeat the same with a semicircle. Similarities and differences?

4. The limit of the slope in the relationship between the length logarithms vs. resolution is called the *fractal dimension* of a curve. If this dimension is not a whole number, the curve is a *fractal*. A fashionable topic is the use of fractals (curves and generalizations to surfaces and volumes) as models for natural objects and natural phenomena. Numerous articles on this have been appearing in forest research journals. Comment on the following statements:

   (a) Fractals are distinguished in that when measuring their length, this increases with the resolution used.

   (b) Opposite to classic geometry, in nature all curves and surfaces are fractals.

   (c) Consequently, talking about the length or surface of real objects does not make sense.

5. A frequent example is the length of the Mississippi river. When measuring it in segments of 400 km (with a compass, for instance) its length is 1600 km. If segments of 100 km are used a length of 1800 km is obtained. Thus, greater lengths are obtained every time as the resolution is increased. Comment.

6. Outside Valdivia there is a signboard that indicates 187 km to Temuco.

   (a) How do you think that that value was obtained? Suggest alternative methods.

   (b) Do you find any usefulness in this information?

♡ Some references about fractals:

Scientific American, Vol.238, No.4, pág.14. April 1978.

STRAND, L. Crown density and fractal dimension. Medd.Nor.inst.skogforsk. 43(6):1-11. 1990.

ZEIDE, B. and PFEIFER, P. A method for estimation of fractal dimension of tree crowns. For.Sci. 37:1253-1265. 1991.

MANDELBROT, B.B. The Fractal Geometry of Nature. W.N.Freeman. 1983.

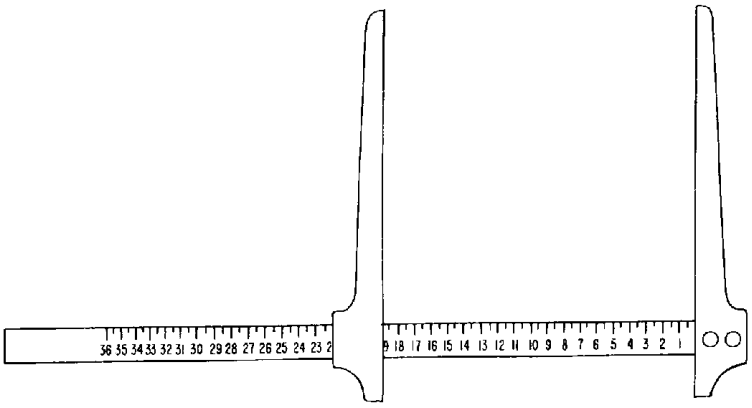
de GUZMAN, M., MARTIN, M.A., MORAN, M. y REYES, M. Estructuras Fractales y Aplicaciones. Una Introducción. Editorial Labor. 1991.

PRUSINKIEWICZ, P. and HANAN, J. Lindemeyer Systems, Fractals and Plants. Lecture Notes in Biomathematics, Vol.79. 1989.

In Physics, philosophical problems about the nature or existence of "real" values have been obviated by taking the *operational* point of view. In this a quantity is defined through the procedure (operations) used to measure it.

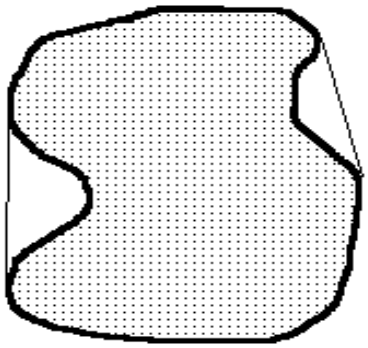
2.2 Diameters

The instruments commonly used are the tape (measuring the circumference) and the calipers. Also diameters at the ends of logs can be measured directly. We will not talk more here of this situation, but of measurements "from outside", as at points far from the log ends, or on standing trees.



CALIPERS

What is of interest generally is not the diameter, but the area of the cross-section with the purpose of estimating volumes. In the first place, it is clear that both the tape and the caliper ignore possible concavities in the section, dealing with the convex closure(*convex hull*):



The difference (positive) between the area of the convex closure and the area of the stem cross-section is called the **convex deficit**.

The tape actually measures the perimeter P of the convex closure. What is called "diameter" is the diameter that a circle with that perimeter would have, that is to say, $D = P/\pi$. In fact, the *diameter tape* used in mensuration is graduated in units of π, so that it gives that diameter directly. But the area of the convex closure will always be less than or equal to the area $\pi D^2/4$ of that circle, since the circle is the figure with the largest area for a given perimeter. The difference is called **isoperimetric deficit**. We have identified then two sources of overestimation when calculating the area of the section by $S = \pi D^2/4 = P^2/(4\pi)$.

In general, the diameter measured with calipers varies with the direction from which it is measured. By a theorem of Cauchy (1841), it is possible to prove that the expected value of a caliper diameter measured in a random direction (or what is the same, the diameter averaged over all the possible directions) is equal to the value obtained with tape (we are ignoring possible errors of measurement). That is to say, $E[D_c] = D_t$.

♡ More precisely, what Cauchy said is that for any convex figure with perimeter P, the expected value of its projection on a random direction is P/π. Clearly, the measurement D_c given by calipers corresponds to a projection, and the "diameter" given by the tape is $D_t = P/\pi$.

Nevertheless, the expected value of the area of the cross-section obtained by a caliper measurement does not agree with that obtained with the tape. The caliper diameters have a certain variance

$$\sigma^2 = V[D_c] = E[D_c^2] - E[D_c]^2 = E[D_c^2] - D_t^2 \ .$$

The areas calculated for the cross-section are $S_c = \pi D_c^2/4$, with expected value

$$E[S_c] = \frac{\pi}{4} E[D_c^2] = \frac{\pi}{4}(\sigma^2 + D_t^2) \ ,$$

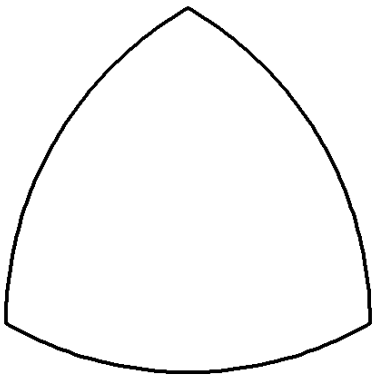
from where

$$E[S_c] = S_t + \frac{\pi}{4}\sigma^2 \ .$$

This value is greater than or equal to S_t, since σ^2 is non-negative.

Loetsch *et al.* in their text, and B. Matérn ("On the geometry of the cross-section of a stem", *Meddelanden från, Statens Skogsforskingsinstitut, Band*46,Nr.11, 1956) analyze the results of averaging pairs of caliper measurements made in several ways.

♡ Could we determine the form of the cross-section from several caliper measurements, and thus to devise some type of correction for the isoperimetric deficit? The following figure has the same projection in all directions, like a circle:



Therefore, the form of a section cannot be determined by measurements "from outside".

♡♡ The previous figure is called the Reuleaux triangle. Matérn gives other examples of figures with this property, called *orbiforms*. Another interesting article is: Gardner, M. "Curves of constant width, one of which makes it possible to drill square holes". *Scientific American*, February of 1963.

Questions, exercises

1. You need to cut a 1.3 m rod. You only have a diameter tape graduated in centimeters and millimeters (of diameter). What reading of the tape corresponds to the length of the rod?

2. One of the latest breakthroughs of the genetic improvement program has been the production of trees of square section, with the consequent advantages in sawing conversion and costs of transport. For a section of 30 cm by side:

   (a) Calculate the diameter indicated by the tape.

   (b) Calculate the isoperimetric deficit.

   (c) Indicate the expected value of the caliper diameter for a random direction.

3. Note that the Reuleaux triangle shown above is formed by three segments of circle, centered at the vertices. Calculate, as percentages of the real area, the areas obtained from measurements with tape and calipers.

4. On a stem cookie, calculate as precisely as possible the cross-sectional area using, for example, a planimeter or a point grid. Calculate, as percentage of the area, the deficit of convexity and the isoperimetric deficit.

5. The following are some properties of the ellipse with greater diameter $2a$ and minor diameter $2b$:

$$\text{equation: } \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

$$\text{area: } \pi ab$$

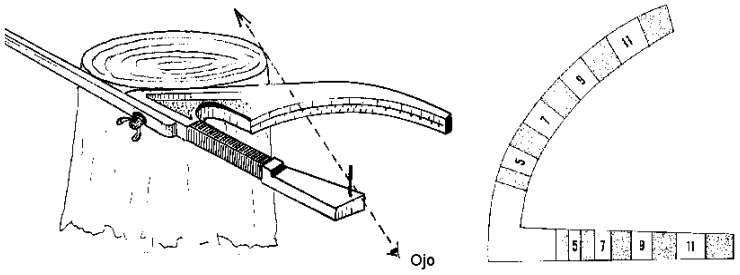
$$\text{excentricity: } e = \sqrt{1 - b^2/a^2}$$

$$\text{perimeter: } \pi[\frac{3}{2}(a+b) - \sqrt{ab}] \text{ (aprox.)}$$

(this is an approximation for the perimeter valid for moderate excentricities, sufficient for this example; the exact value is $a \int_0^{2\pi} \sqrt{1 - e^2 \sin^2\theta}\, d\theta$, an elliptical integral of the second kind whose values can be found in tables).
Take some reasonable excentricity values for trees with elliptical cross-section (for example thinking about a greater diameter of 30 cm and several reasonable small diameters). For these excentricities:

   (a) Calculate the isoperimetric deficit, as a percentage (show that it only depends on the excentricity).
   (b) Calculate the maximum percent error when calculating the area of the section with a caliper measurement.
   (c) Calculate the percent error in the area when using the average of the smallest and largest diameters.

6. Explain the principle of the instruments for measurement of diameters shown here (*Visiermeszwinkel* or Bitterlich diameter sector, and a type of parabolic caliper):



2.3 Cubature of logs

When calculating volumes of logs and trees we normally assume that the sections are circular, or at least that diameters are such that the area of the section is $\pi D^2/4$. This is a simple example of the use of *models*, idealizations that are taken as true in later calculations and decisions.

♡ "In general, a model is a simplified representation of some aspect of reality (not to be confused with the normative meaning of the word, something worthy of being imitated.) We continuously use models, in some form or other. There are mental models, i.e., imagined cause-effect relationships between components of some system through which we try to explain and anticipate its behavior. Models may be stated in verbal form, for example the description in words of the workings of some machine. Material models, such as scale models of buildings and airplanes, are well-known.

A mathematical model is like a verbal model, but using mathematical language. Mathematical language differs from natural language in that it is more concise and less ambiguous. This, together with the availability of prescriptions that can be used mechanically, allows us to reason in more complex situations, with less effort, and with less risk of confusion.

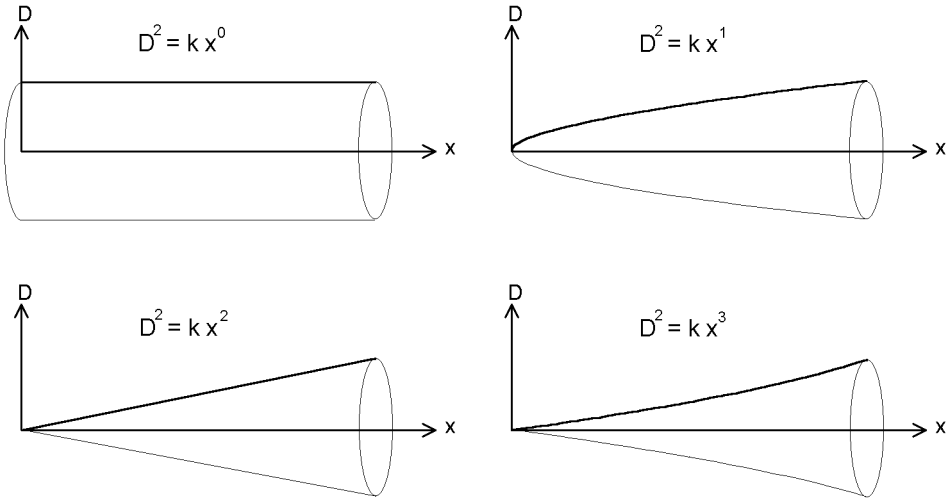
With the progress in computing, it becomes easier to handle increasingly complex models, and in fact, the use of engineering scale models has been decreasing, replaced by mathematical models that are cheaper and more flexible. Computers have become indispensable as tools in the development and use of many models. But note an analogy between *computer modelling* and *typewriter poetry*. Also, realism is not necessarily a virtue in a model, it is better to abstract just those aspects that are most relevant in each instance. In a model airplane for wind tunnel testing, its color or the name of the pilot may not be important. The manual for a video recorder does not say much about its inner workings, but it can be used to predict the effects of pressing different buttons.

It is useful to distinguish between *models for prediction*, and *models for understanding* (Bunell 1989). Models for understanding (e.g. physiological, process models) are mostly useful in research, to help comprehension, to collect and link pre-

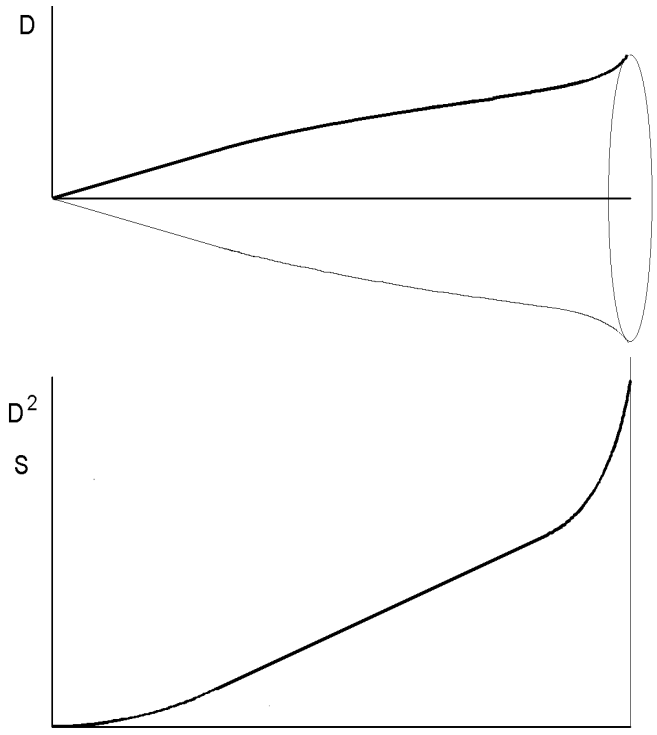
viously isolated bits of knowledge, and to identify gaps where more work is needed. The benefits come from the development of the model, and not so much from its later use, if any. I shall focus on models for prediction, intended for management planning. Typical applications are in forecasting for forest planning purposes, in the comparison and evaluation of silvicultural (pruning, thinning) regimes, and in the updating of stand description databases.

For the most part the models are presented as deterministic. In general, decision-makers use model results as representing a most likely course of events, and introducing "randomness" in the predictions has not been found very helpful in practice. A stochastic (random, probabilistic) component is however necessary for developing rational parameter estimation procedures, and appropriate stochastic structures are discussed later in that context."
(García, O. "The state-space approach in growth modelling". *Canadian Journal of Forest Research 24*, 1894–1903, 1994).

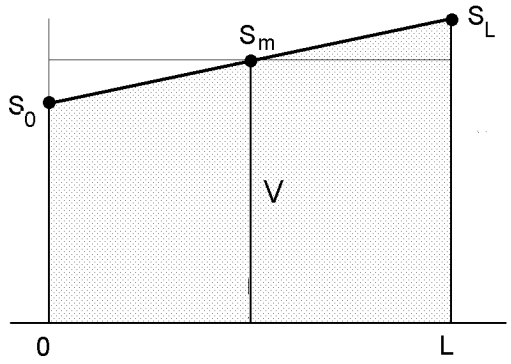
It is customary in forest mensuration to take the shape of logs and trees as similar to certain solids of revolution, the cylinder, paraboloid, cone, or neiloid. The general formula for the variation of the diameter D with the length x is $D^2 = ax^n$ with $n = 0, 1, 2, 3$. In terms of the area of the section, $S = bx^n$.



More generally, different parts from the tree resemble portions of these solids. The crown part, in conifers, tends to the cone form. The stem central part approaches a paraboloid. The base of the tree expands in a form similar to the neiloid, although generally values of n greater than 3 come closer.



The volume of a stem section, for example a log, clearly corresponds to the area underneath the curve of S over length. For a log in the central, paraboloidal, part of the tree, the section changes to a linear form, and the volume is the area of the following trapeze.



The volume (area of the trapeze) can in this case be calculated given the

end sections and the length:

$$V = \frac{S_0 + S_L}{2} L = \frac{\pi}{8}(D_0^2 + D_L^2)L \ .$$

In Mensuration this is known as **Smalian's formula**. The volume/area can also be calculated based on the midpoint diameter:

$$V = S_m L = \frac{\pi}{4} D_m^2 L \ ,$$

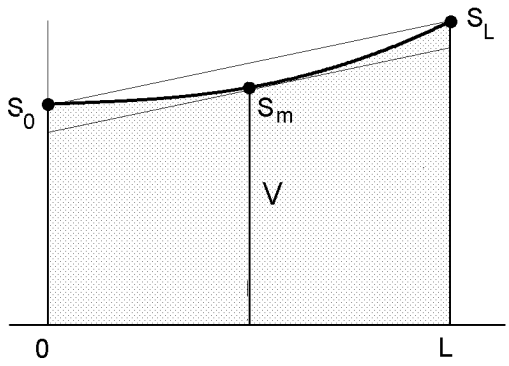
Huber's formula. We assume here that we are using consistent units (e. g., D and L in metres, V in cubic metres).

Questions, exercises

1. Give Smalian's formula for D in centimetres, L in metres and V in cubic metres.

2. The same for D in inches, L in feet and V in cubic feet.

Assume now that the curve of S is convex, as is likely to happen in the first log and in the top end of the tree. (Note: convexity or concavity of a curve or surface is taken as seen from the origin).



It can be seen that Smalian gives the area of the upper trapeze, overestimating the real volume. Huber gives the area of the lower trapeze, producing an underestimate. Comparing the areas between the dotted lines on each side of the curve, one sees that Huber gets closer to the real value.

Huber's formula is generally more exact, and requires measuring one diameter instead of two. In many instances, however, the centre of the log is

not easily accessible, as when logs are stacked up. In addition, if the volume without bark is needed, it is easier to measure the diameters under bark in the extremities of the log. Because of this Smalian's formula, although producing larger errors, tends to be used most frequently.

If we had three diameters, at the ends and centre, a weighted average of Huber and Smalian would reduce the errors. It is shown that the following formula, that can be seen as such a weighted average, gives exact results for polynomials of up to third degree. That is to say, it is exact for all the solids of revolution considered here.

$$V = \frac{S_0 + 4S_m + S_L}{6} L \ .$$

In Mensuration this one is traditionally known as **Newton's formula**, although in mathematics this is the basis of *Simpson's rule*. It is little used in practice, except in the cubature of complete trees as we will see later.

♡ D. Bruce ("Butt log volume estimators", *Forest Science 28* , 489–503, 1982) proposes to estimate the volume of the first log with $\frac{\pi}{16}(3d_0^2+D_L^2)L$ (explain why this might be better than Smalian). He also recommends to correct Huber multiplying by a factor of 1.04 or 1.08 depending on species, and analyzes the errors of several methods in real cases.

Questions, exercises

1. A log of 6 m of length has small, large and centre diameters of 22, 27 and 25 cm, respectively. Compute the volume by (a) Smalian, (b) Huber, and (c) Newton.

2. A log of 3.2 m of length has a small diameter of 30 cm, large of 34 cm, and in the centre the diameter is 32 cm.

   (a) Use the small and large diameters to calculate the central diameter:

      i. Assuming that the log is a truncated cone.

      ii. Assuming that the log is a truncated paraboloid.

   (b) Calculate the volume in cubic metres by the formulas of

      i. Smalian.

      ii. Huber.

3. By geometry and integration it is found that the area under $S = bx^n$ between two points with $S = S_0$ and $S = S_L$, separated by a length L, is

$$V = \frac{L}{n+1} \frac{S_L^{1+1/n} - S_0^{1+1/n}}{S_L^{1/n} - S_0^{1/n}} .$$

Obtain the alternative forms

$$V = \frac{LS_0}{n+1} \frac{r^{1+1/n} - 1}{r^{1/n} - 1} = \frac{LS_0}{n+1} \sum_{k=0}^{n} r^{k/n} ,$$

where $r = S_L/S_0$.

4. Take a reasonable range of values for r (think of approximately paraboloid trees with various diameters and heights). Calculate, for $n = 1, 2, 3, 8$, the percentage of error in the formulas of Smalian and Huber. (Hint: notice that $S^{1/n}$ as a function of length is a straight line).

♡**Numerical integration** Calculating the area under a curve $f(x)$ given some points on it is an important problem in Numerical Analysis: numerical integration or quadrature. The usual approach is to obtain formulas of the form $L \sum w_i f(x_i)$ (the length multiplied by a weighted average of the heights) that are exact for polynomials of the highest degree possible, subject to various constraints on the number of points and on the values of the x_i and/or the w_i. In the case where the x_i are given uniformly spaced one has the *Newton-Cotes* formulas, *closed* if the x_i include the ends of the integration interval, *open* otherwise. The formulas of Smalian and Newton used in mensuration correspond to the closed Newton-Cotes for 2 and 3 points, respectively. They are known as the trapeze and Simpson formulas, and are exact for polynomials of 1st and 3rd degree (the error is a function of the second and third derivatives). The one of Huber could be seen as the open formula with one point, exact for polynomials of 1st degree.

If the x_i can be chosen freely, the *Gauss* formulas are obtained. For example, with three points, $L[5f(L(1-\sqrt{3/5})/2)+8f(L/2)+f(L(1+\sqrt{3/5})/2)]/18$ is exact for polynomials of up to the fifth degree. In the formulas of *Chebyshev* it is required all the w_i to be equal, which reduces the effect of errors in $f(x)$, and facilitates the calculations since it is sufficient to take the mean of the $f(x_i)$. For two points Gauss and Chebyshev coincide, and for one we would get Huber again.

Note that, depending on the irregularity and errors in $f(x)$, it may be preferable to integrate subintervals separately than to use formulas of high degree. With uncertainty in $f(x)$ it may also be better to use a polynomial approximating its values instead of one of higher degree that interpolates them exactly. For more details see Numerical Analysis texts.

2.4 Stacked wood

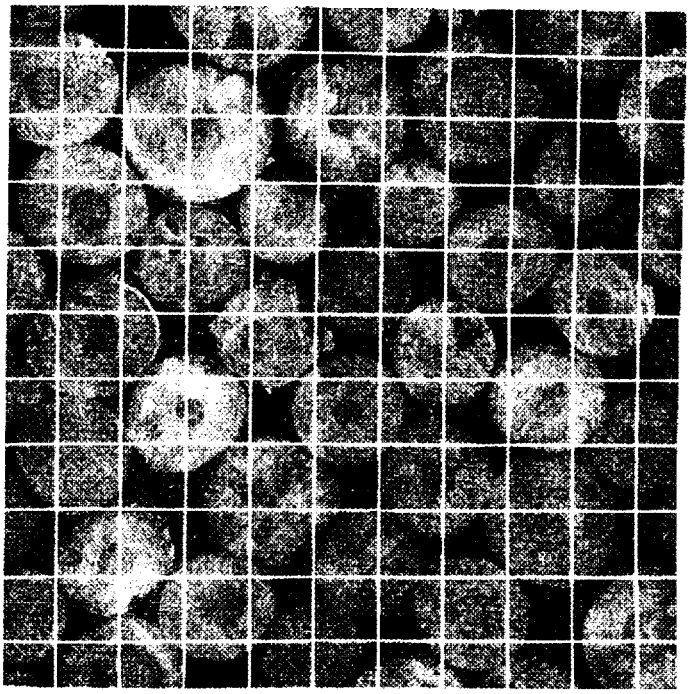
Products such as firewood and pulp logs are frequently commercialized according to their volume in piles or stacks. A **stere metre** is the volume of a stack of $1 \times 1 \times 1$ metres (a cubic metre stacked), and it is used for firewood. For pulp the unit most used in Chile is the **metro ruma**, a section of 1×1 metres in a stack of logs 2.44 m long. It may be with bark or without bark. A unit common in North America is the *cord*, that corresponds to a stack of 4×8 feet with logs 4 feet long.

Due to edge effects the wood content can vary slightly with the stack dimensions, and much with the stacking method, so buyers and sellers usually establish specific norms on dimensions and stacking methods.

Other important factors in the solid content are the irregularity of the logs, the variability of the diameters, and the bark thickness. Movement during transport can also introduce important changes.

It is of interest to have conversion factors, for example solid cubic metres per metro ruma. These may also be expressed as a proportion of solid volume in the stowed volume. Peters *et al* give some conversion factors for radiata pine in INFOR's Manual No. 14. Bruce and Schumacher mention a formula by the second author for the proportion of solid volume over bark: $0.84 - 0.04N$, where N is the mean number of logs by square foot on the face of the stack.

It is possible to obtain more exact conversion factors for specific situations through sampling. Stacks can be measured, and then taken apart to determine the volume of the logs using Smalian's formula, for instance. An often-mentioned alternative consists of taking photographs and measuring in them the proportion of wood with a points grid or some other method. Similar grids have been suggested for direct use in the field. More practical for field use would be the method proposed by Avery: marking a rod or tape at regular intervals, and determining the proportion of marks that fall in wood when placing it on the face of the stack. In any case, sufficient sampling intensity is needed to obtain the desired precision.



♡**Point grids and systematic sampling** It is easy to obtain unbiased estimators of known precision using random sampling. Systematic sampling, of which the point grid is a special case, is generally much more efficient. Its use has been resisted, however, due to the difficulty in estimating its precision, and to the possibility of extreme errors in unfavourable cases. These extreme errors, associated to coincidences of periodicity between the population and the sample, are probably rare in practice, and can usually be avoided by taking suitable precautions. The theory behind its precision is complex, and exact results depend on non-observable population characteristics. However, it has been possible to obtain satisfactory approximations. Thus, in some disciplines there has been a resurgence of interest in systematic sampling procedures, specially in *Stereology*. This deals with the determination of quantitative characteristics mainly in organs and animal tissues, and there is a highly developed mathematical theory. See for instance the articles by Matérn, Kellerer, Mattfeldt and by Cruz-Orive in volume 153, number 3 of the *Journal of Microscopy*, 1989.

The variance in a measurement of area with a square point grid is found to be approximately $0.0728Pa^3$, where P is the perimeter of the area to be measured and a is the spacing of the grid. This formula was obtained by Matheron in 1965, and presented in the forestry literature by J. Bouchon (*Ann.Sci.For. 32*, 131-134, 1975) and R.B. Chevrou (*Resource Inventory Notes 20*, 3-6, 1979). The more exact value 0.0728 instead of 0.0724 is given by Matérn in the reference already mentioned,

and it is used in the equation below. Gundersen and Jensen (*J.Microscopy 147*, 229-263, 1987) use this to give the coefficient of variation of the area A based on the counted number of points N and of the form indicator $P/\sqrt{A}$:

$$\sigma/A = 0.27(P/\sqrt{A})^{1/2}/N^{3/4} \ .$$

The ratio $P/\sqrt{A}$ has a minimum of $2\sqrt{\pi}$ for circles (note the relation with the isoperimetric deficit). Gundersen and Jensen give examples with ratios of up to 33 for very complicated figures, which gives for the coefficient of variation an approximate range of $0.5N^{3/4}$ to $1.5N^{3/4}$.

Questions, exercises

1. How many metros ruma are there in one cord?
2. We have a stack of pulp logs, stowed along a strong slope. Its length, along the ground, is of 12 m. The mean height measured vertically is 1.5 m, and measured perpendicular to the ground is 1.2 m. The logs are 2.44 m long. What is the content in metros ruma?
3. Calculate the solid proportion in a stack of cylinders of uniform diameter stacked in rectangular pattern. The same for the more compact arrangement (triangular).

4. In the previous photograph, estimate the precision obtained with the grid shown. Make a table of precision versus number of points.
5. Loetsch, Zöhrer and Haller, page 34, show coefficient of variation obtained with point grids. Compare with the predictions of the Gundersen and Jensen formula. What do you think of the logarithmic regression coefficient in page 35?
6. Estimate the precision of the method suggested by Avery.
7. An alternative to counting points along a line is to measure the accumulated length of the intersections. Devise a practical method based on this principle.

2.5 Weight measure

A common wood measurement method for pulp logs, and in sometimes for sawlogs, is through weight. Loaded trucks or railroad cars are weighed, and the known weight of the empty vehicle is subtracted. For smaller scale operations and field measurement, a device has been developed in New Zealand that calculates the weight from the pressure in the hydraulic system of a front loader.

It is of interest to know the equivalent in volume and/or dry weight. Factors that affect these conversions are the variations in moisture content and in wood density. In turn, these vary with the time elapsed since harvest, tree size, site, weather, and locality of origin. Variables such as the amount of mud adhered to the logs and the amount of fuel in the trucks also can be important.

The most important source of variation is generally the moisture content. It should be mentioned that often paying by weight is advantageous for the buyer, being an incentive for delivering fresh wood.

Perhaps the most usual is using an average conversion factor. Sometimes corrections based on moisture content (estimated with electrical instruments) are made. If P is green weight, p is dry weight, and V is volume, the required conversion factor to volume V/P depends on the *basic density* p/V, characteristic of the species, and the *moisture content* (%) $100(P-p)/p$. The green volume V does not change appreciably while the moisture content does not go below the fibre saturation point, around 30%.

Regressions are also used that include the mean log size or the number of logs per load. Another predictor that has been used is the date (month), since the variations through the year usually are important. An analysis of the factors that affect conversion factors, specifically for radiata pine, is found in: Ellis, J.C. "Weight/volume conversion factors for logs", New Zealand Logging Industry Research Association, *Technical Release 6*(3), 1984.

Questions, exercises

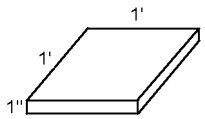
1. Given the moisture content h and the basic density d, obtain formulas for the conversion factors to volume and to dry weight.

2. The price of a metro ruma of debarked *E. anonimus* is \$20000. Calculate an equivalent price to pay for the ton. It is known that: the proportion of solid wood is 0.72; basic density is 0.74 g/cm^3; moisture content is 80%.

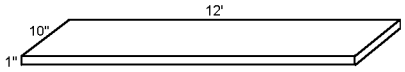
2.6 Sawn timber

In Chile it is customary to express the volume of sawn wood in *pulgadas madereras* ("lumber inches"), a unit related to the North American *board foot*. Although the metric system gains ground, in sawmilling practice measures in feet and inches are still in current use. One inch (abbreviation 1") is 2.54 cm. A foot (abbreviation 1') is 12 inches, or 30.48 cm.

The **board foot** is the volume of a square piece of 1 foot by side and one inch of thickness. A **pulgada maderera** is the volume of a board 10" wide by 12' of length and 1" of thickness.



Clearly, one pulgada maderera equals 10 board feet. This is used commonly for native timbers. In plantations, more common is the **pulgada corta** or **pulgada pinera**, which is defined with a length of 10.5' instead of 12'. These are *nominal* measures, that is to say, they may include tolerances and/or planning and shrinking losses.



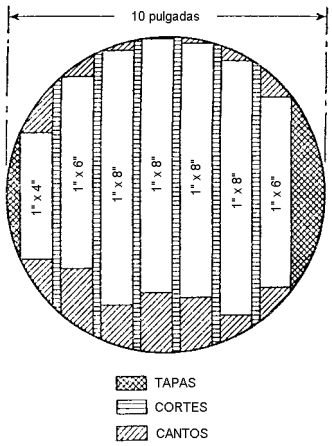
Questions, exercises

1. Compute the volume in pulgadas madereras of pieces with the following dimensions:
   (a) 2" × 4" × 16'
   (b) 3/4" × 3 3/4" × 8'
   (c) 4" × 4" × 10'
2. Calculate the conversion factor between pulgadas madereras and pulgadas cortas.
3. How many cubic metres (nominal) are in one pulgada maderera?

In order to estimate the amount of sawn timber obtainable from a log, standard log rules are used, or empirical relations are developed for a specific sawmill.

Log rules are tables or formulas that give sawn volume based on the small diameter and length of a log. They have been obtained through diagrams, or as formulas derived from geometric reasoning.

In the diagram method, circles of various sizes were drawn at scale, representing the small end of logs. In these, boards that would be obtained in sawing were indicated, and their volume calculated.



The best known log rule of this kind is the one of **Scribner** (1846). A slight variant, the Scribner decimal C, is one of the most used nowadays in the U.S.A. There is an approximation formula:

$$V = 0.79D^2 - 2D - 4$$

(logs of 16' of length, V in board feet, D in inches truncated to the lower whole number).

Among the formula rules, one of most exact is the **International** (Clark, 1906). The reasoning is as follows (see the figure above). First, it is considered that to each board of 1"of thickness corresponds a cut where 1/8" is lost in sawdust, and 1/16" in shrinkage tolerance. That is, 1" in 19/16" is utilized, so that the area $\pi D^2/4$ of the small end is reduced multiplying it by 16/19. Then the area that would be lost in slabs and edges is subtracted. It is seen that this one has an irregular shape, but approximates an outer ring, with a thickness independent of the diameter. Alternatively, it may be assumed that from each board cut, approximately a same amount is lost from the edges. At any rate, this term to be subtracted is proportional to D, with a constant estimated experimentally. For a length of 4' these two

terms are multiplied by the length to obtain the recoverable volume, which in board feet and for diameter in inches is

$$V = 0.22D^2 - 0.71D \ .$$

This formula can be modified for sawkerfs different from 1/8", and the most used is the one of 1/4". With this sawkerf the area of the circle should have been reduced by a factor of 16/21 instead of 16/19, so that the formula above is adjusted simply multiplying it by 19/21, obtaining

$$V = 0.199D^2 - 0.642D \ .$$

Unlike Scribner's rule, the International takes into account taper, assuming that boards of 4' of length are usable, and that the average taper is 1/2" in 4'. The previous formula is therefore applied starting at the small of each 4' section, incrementing the diameter by 1/2" for the following section. Thus, for logs of 16' the 1/4" rule would give the following formula:

$$\begin{aligned} V &= 0.199D^2 - 0.642D \\ &+ 0.199(D+0.5)^2 - 0.642(D+0.5) \\ &+ 0.199(D+1.0)^2 - 0.642(D+1.0) \\ &+ 0.199(D+1.5)^2 - 0.642(D+1.5) \\ &= 0.769D^2 - 1.374D - 1.23 \ . \end{aligned}$$

In their commercial application these log rules are accompanied by detailed standards for measuring, rounding, and defect allowances. Specially in over-mature natural forests, the defects (rot, cracks, sweep) are usually important. The general idea is to enclose the defect in a parallelepiped, thinking about the way in which the cuts will be made, and deducting its volume before applying the rule. Practices used in the U.S.A. are described by Bell, J.F. and Dilworth, J.R. "Log Scaling and Timber Cruising", O.S.U. Book Inc. Stores, Corvallis, Oregon, 1988, 1993.

The traditional log rules can only give rough estimates, since the conversion varies much with the technology of the sawmill, species, product dimensions, etc. For a given situation it is possible to develop an empirical log rule or conversion function through sawing studies. A certain number of logs covering the desired range of diameters is measured, sawn, and the lumber obtained is measured. With these data it is then possible to fit a regression equation giving the sawn volume as a function of small diameter, and possibly of length if this varies. From the log rule formulas already seen, it is seen that a reasonable function might take the form (for a given length)

$$V = b_0 + b_1 D + b_2 D^2 \ .$$

♡ F.X. Schumacher and W.C. Jones (*Journal of Forestry 38*, 889–896, 1940) proposed an interesting method to obtain empirical log rules without counting on detailed individual log data. The basic idea is that the previous equation can be added over all logs processed in a day:

$$\sum V = b_0 N + b_1 \sum D + b_2 \sum D^2 \ .$$

It is then possible to estimate the coefficients from daily total production $\sum V$, number of logs N, and sums of small diameters and of their squares. Formulas for variable-length logs are handled in an similar way.

The method can be useful also for weight/volume conversion factors, and in other applications. Clearly, to obtain reliable results, long daily production series with important day to day variations in the characteristics and number of logs are needed.

In many instances it may be more convenient to express the sawn yield in relative terms, as a conversion factor of pulgadas or cubic metres sawn per cubic metre of logs. The cubic volume can be estimated from the small diameter assuming some value for taper, for instance the 1/2" per 4' of length (1:96) of the International rule.

Questions, exercises

1. A log 3.2 m long has small and large diameters (under bark) of 30 and 34 cm, respectively.

   (a) Obtain the sawn volume in pulgadas madereras using the International 1/8" rule.

   (b) Based on Smalian, give the conversion factor (% of recovery).

2. L.R. Grosenbaugh (U.S.Forest Service, Southern Exp.Sta. Ocassional Paper 126, 1952) gave the following simple approximation to the International 1/4" rule for 16' logs: $V = 0.8(D-1)^2$. Evaluate its accuracy.

3. Express the International 1/4" rule for sawn volume in cubic metres, diameter in centimetres, and length in metres.

4. Under the International rule taper assumption (1/2" in 4', or 1:96), obtain formulas for cubic volume as a function of small diameter using: a) Smalian, b) Huber, c) volume of a truncated cone.

5. Obtain a formula for the conversion factor to sawn timber, as % in function of the small diameter, using the 1/4" International log rule and its taper assumption. Hint: consider a 4' section.

6. Assume a standard log length of 12'. We have daily values of number of processed logs, production in board feet according to the International rule, and total log volume (calculated from the small diameter, Smalian, and taper of 1%). Can you obtain an empirical log rule by the method of Schumacher and Jones? How to calculate $\sum D$ and $\sum D^2$?

7. Avery (cited by Loetsch *et al*) wrote: "There is little justification for log rules aside from its use in the past and the resistance to change...". When converting to the metric system New Zealand abandoned the use of all kind of log rules, presumably for being considered incompatible with the new system. Comment.

2.7 Change of units

One has a formula, and it is required to express it in other units. For instance, the Smalian's formula with length and diameters in metres, $\frac{\pi}{8}(d^2 + D^2)L$, we want to convert it to length in metres and diameters in centimetres. For this, simply replace the variables by their corresponding values in the new units. For example, D metres $= D$ centimetres$/100$. We have

$$\frac{\pi}{8}([d/100]^2 + [D/100]^2)L \ .$$

Simplifying we get

$$\frac{\pi}{80000}(d^2 + D^2)L \ .$$

As another example, Scribner's formula

$$V = (0.79D^2 - 2D - 4)(L/16)$$

with V in board feet, D in inches, and L in feet, can be converted to pulgadas madereras, centimetres and metres in the following way:

$$[10V] = (0.79[D/2.54]^2 - 2[D/2.54] - 4)([L/0.3048]/16)$$

$$V = (0.0025D^2 - 0.016D - 0.083)L \ .$$

Chapter 3

Trees

3.1 Diameters

The most commonly used diameter is the *diameter at breast height* (DBH). It is defined as the diameter, over bark unless stated otherwise, at a height above the ground that in most of the countries that use the metric system is 1.3 meters. Some exceptions are New Zealand (1.4 m), and Japan (1.25 m). In the US, 4.5' is used. These heights are convenient for measurement with callipers, and are somewhat distant from the influence of the butt-swell at the base of the tree (although perhaps a greater height might have been preferable).

When measuring DBH it is desirable to rely on more precise specifications, which unfortunately are not standardized. For example, on a slope it is customary to measure the DBH height either from the mean ground level at the tree base, or from ground level on the upper slope side. In case of stem deformation at breast height, the measurement may be displaced upward, downward, or the average of two measurements may be taken.

Questions, exercises

1. Discuss the advantages and disadvantages of the several criteria of measurement of the DBH just mentioned.

2. How important can the difference be between measuring the DBH at 1.3 or at 1.4 m? A typical stem taper in the first log is on the order of 1:100.

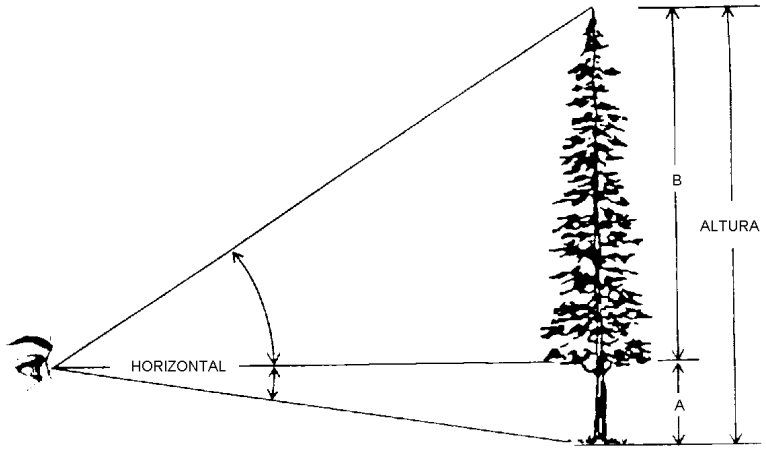
The most common instruments used for measuring DBH are the caliper and the diameter tape (graduated in units of π). Upper diameters are

measured by climbing, with instruments mounted on rods, or with various kinds of optical dendrometers. In all instances, the considerations on variability and error sources, and the relations with cross-sectional area already discussed in the section on log measurement are applicable.

In order to turn diameters over bark to under bark, the bark thickness can be measured, or estimated from pre-established relationships. The Swedish bark gauge is most commonly used. Its proper use requires periodic practice and calibration, and biases can be important. Two readings, on opposite sides of the stem, are commonly taken, adding them up to obtain the "double bark thickness", considered as the difference between the diameters.

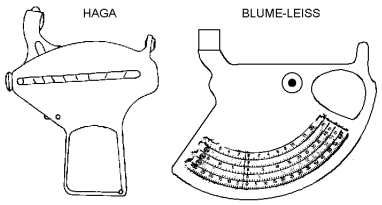
3.2 Heights

Heights of up to 10–15 m are preferably measured with telescopic poles. For greater heights, clinometers (instruments that measure vertical angles) or hypsometers (specialized instruments that indicate height) are used. Some hypsometers (Christen, Merrit) use similarity of triangles, but at the present time most are based on trigonometrical principles.



Heights can be total (up to the apex of the crown) or merchantable (up to a diameter limit, or up to the the point where the stem branches). One measures the distance to the tree, and angles from the horizontal to the base and to the upper end. The distance is measured with tape or with a fixed-distance rangefinder built-into the instrument. The distance is generally sloped, which must be taken into account in the calculations. The most recommendable hypsometers are the Suunto, Haga and Blume-Leiss,

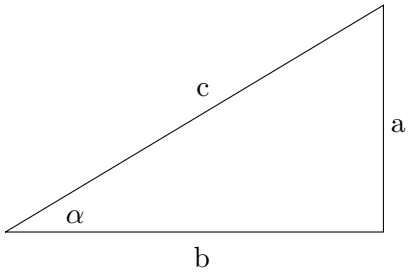
all based on the use of a pendulum or counterweight to establish a vertical reference. Clinometers such as the Abney level, that uses an air bubble as reference, may also give good results, although its use is more cumbersome. For high precision it is necessary to resort to theodolites or tachymeters.



Hypsometers are subject to large errors if they are not used with care. The height pole is preferable when practical. Several readings must always be taken; the median of three would be recommendable. The calibration of the instrument must also be verified periodically. In particular, the accuracy of the rangefinder must be checked, since substantial factory variations have been found.

Questions, exercises

1. Based on the above figure, derive the principles and formulas for the calculation of heights. The general approach in this type of problem consists of establishing or identifying right-triangles and applying some of the following formulas



$$\frac{a}{c} = \sin\alpha$$

$$\frac{b}{c} = \cos\alpha$$

$$\frac{a}{b} = \tan\alpha$$

$$a^2 + b^2 = c^2$$

2. What happens if the level of the line of vision is below the base of the tree?

3. You are at a horizontal distance of 20 m from a tree. The following readings have been taken with a clinometer. Reading to the apex: 48°, Reading to the base: -8°.

(a) What is the height of the tree?

(b) The angles have an error of $\pm 1°$. By substitution, calculate limits of error for the height.

3.3 Cubature of trees

Several types of tree volume are distinguished. The *cubic* volume is the volume of wood contained in a portion of the stem. It can be *total* volume, for the whole tree from base to apex, or *merchantable*, from stump height to the height where a certain *diameter limit* is reached. Sometimes by total volume it is meant all the usable merchantable volume, for example in the tables of volume for pine of INFOR, which give volumes from the stump to a diameter limit of 10 cm (at present smaller diameters are accepted, of 8 cm or less). Cubic volumes between several diameter limits may be considered, for example *sawn volume* between the stump and a limiting diameter of 25 cm, and *pulp volume* between diameter limits of 25 and 10 cm. Often by sawn volume it is understood not the cubic volume, but the volume of sawn wood estimated in "pulgadas" or in board feet (*sawed volume*). Similarly, pulp volume could be specified in "metros ruma". Finally, volumes can be *over bark* or *under bark*.

As already seen for logs, the cubic volume of the whole tree or of a portion of it is given by the area under the cross-section curve as a function of length or height. Sections are usually calculated as circular areas for a sequence of diameter measurements at various heights. Diameters are measured on the fallen tree, by climbing, or at a distance with optical instruments (dendrometers).

The calculation of cubic volume is then essentially an integration problem. We need to calculate the area under the section–height relationship, knowing the sections at several heights. Some possibilities are the following:

- Graphical method. Plot on graph paper, for each measurement height, the cross-sectional area or the diameter squared. Draw a curve passing through this data. Calculate the area under the curve with a planimeter, dot grid, or or in some other way. Apply the appropriate proportionality factor to obtain the volume. Nowadays this method is little used.

- By sections. For pairs of consecutive measurements, the volume between those heights is calculated with Smalian's formula or the formula for the cone. These volumes, for the required portion of the stem, are

added. For uniform height intervals, Huber or Newton can also be used. If they are included, usually the stump section is taken as a cylinder, and the top section as a cone. This is, perhaps, the most common procedure. If the sections are sufficiently short, errors and differences between the various formulas are negligible.

- More sophisticated numerical integration methods. For uniform intervals it is possible to use Simpson's method, or methods similar to it. There are methods for irregular intervals that could be used, such as the one of Gill and Miller (*The Computer Journal 15*,80–83, 1972). J. Bouchon ("Les Tarifs de Cubage", Ecole Nationale du Génie Rural des Eaux et Forêts, 1974), and later others, have used integration with *splines*. The integration of taper curves (see below) could be considered as another example of this.

- Grosenbaugh proposed a method based on measurements spaced at constant diameter intervals. It is suited to the use of optical dendrometers. The principle is similar to interchanging variables and integrating under the height *vs* section curve, instead of under the curve of section *vs* height.

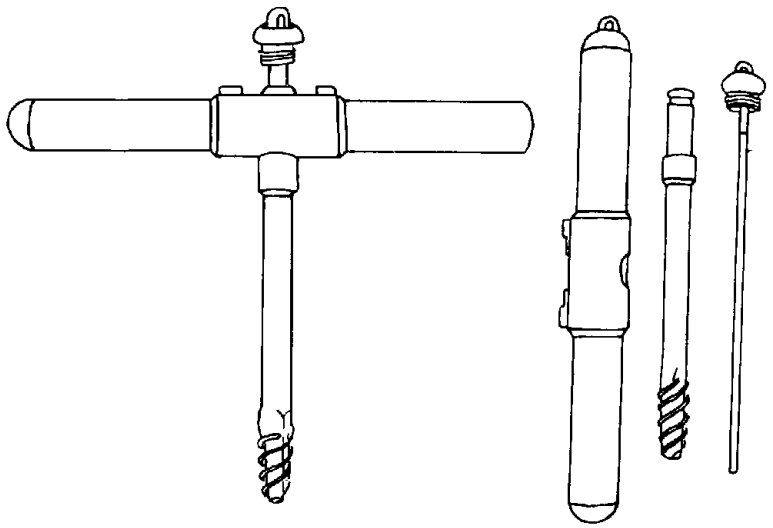
- Sampling. There has been interest recently in methods based on a careful selection of one or more points of measurement. The objective is to estimate, in inventory, total volumes per hectare without the use of volume tables. Strategies of systematic, restricted, and/or variable probability sampling are used. See, for instance, H.T. Valentine *et al., Forest Science 38*,160–172, 1992, and H.V. Wiant *et al., Forest Science 38*,187–191, 1992.

In research work, volumes have been measured by water displacement in devices known as *xylometers*. These are basically water tanks in which the logs are submerged. The volume is determined by the overflow or water level change, or through the change in the weight of the log when submerged.

3.4 Stem analysis

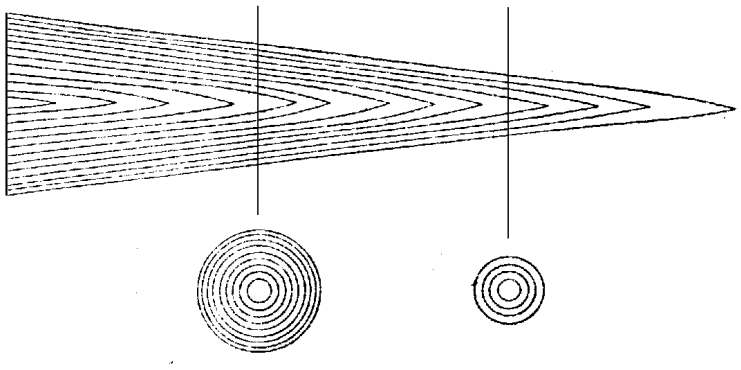
In many species of trees the wood produced at the beginning of the growing season differs in its anatomy from that produced at the end. In a stem cross-section, *growth rings* or annual rings can then be distinguished, with a visible discontinuity marking the limit between the growth of successive years. The boundaries between rings can be observed in cross-sections, or

in cylindrical wooden cores extracted at right angles to the surface of the stem with an *increment borer*:



The visibility of the rings may be improved with colorants and other treatments. X-ray or gamma-ray densitometers are also used to detect the changes of density associated with the rings. There is a number of possible sources of error that need to be kept in mind. The presence of *false rings*, produced by abrupt climatic variations or other factors, causes difficulties. Eccentricity of the rings, and inclination of the borer, can produce serious errors. There may be compression of the outer rings in the wooden core, especially if the borer is not sharp enough. Loss of moisture below the fiber saturation point (approximately 30%) produces contraction of the wood. A good manual on increment boring is: Jozsa, L. "Increment core sampling techniques for high quality cores", Forintek, Spec. Pub. No. SP-30, 1988.

The number of annual rings between the pith and the cambium indicates the year in which the tree reached the corresponding height. This can be used to estimate the age of the tree. The time elapsed between two heights can also be estimated, and from there the height growth rate. Measurement of the rings provide estimates of growth in diameter and in basal area. For values over bark, it is necessary to estimate the bark thickness indirectly, usually with bark–dbh relationships.



Three-dimensionally, what we have is wood layers that form annually on top of each other. Observing the intersections of these layers with cross sections of the stem at various heights (the rings), the past dimensions of the stem can be reconstructed. Thus, data can be obtained for volume tables, taper curves, and site indices (growth in height). The principles of this reconstruction are more or less obvious, except possibly for height estimation.

The problem with height is that the end of each height increment can happen at any height between the levels of two successive cuts, the exact height being unknown. Sometimes it is possible to guide oneself by external indicators (whorls, scars left by the bracts of the apical bud, arrangement of the foliar primordia), and to make cuts that coincide with the end of each annual increment. Otherwise, it is necessary to do some kind of interpolation. Dyer and Bailey (*Forest Science 33*,3–13, 1987) found that a simple method proposed by Carmean gives good results.

Carmean's method is based on assuming a constant increase in height for every year between two successive levels, with the cuts occurring at the middle of an increment. The distance between two successive cuts is divided by the difference in ring numbers, obtaining a mean increment k. The heights estimated above the level of the lower cut are then $k/2, k/2 + k, k/2 + 2k, \ldots$ (see the figure).

Questions, exercises

1. In a stem analysis, a tree with 67 rings in the stump is cut into 5 m logs. The numbers of rings in the upper ends of the logs are 53, 37, and 19. Consider the time intervals in which the tree passed from a cut level to the next. For each one of these intervals, estimate the average rate of growth in height.

2. In a stem analysis (in 1995), we have obtained diameters for rings 5 and 10, counted from the outside of each section. With these diameters the following areas were calculated (m^2):

Height (m)	Ring 5	Ring 10
0.3	0.1080	0.0740
3.3	0.0829	0.0581
6.3	0.0558	0.0412
9.3	0.0384	0.0283
12.3	0.0137	0.0093
14.1	–	0
15.0	0	–

Calculate the growth rate in m^3/year between 1984 and 1989. Use Smalian, taking the stump (height 30 cm) as a cylinder.

3.5 Volume functions (tables)

The measurements necessary to cube a tree and calculate its volume are expensive and slow. It is therefore of interest to be able to estimate the volume indirectly through easier to measure variables, such as DBH and height. The relationships that allow achieving this are the tree volume functions, also called volume tables for historical reasons.

Volume functions are obtained by regression, using a sample of trees in which the volume and the predictor variables are measured. The volume can be total, merchantable, sawn, etc. Once having the function, the volume of other trees can be estimated knowing only the value of the predictors.

Volume functions whose only predictor is the DBH are called *local volume tables*. They are only applicable to the stand for which they were constructed, or at most for very similar stands, since the volume also depends to a large degree of the height, and the relationship between this and DBH vary with the stand density, its age, etc. Typical local volume functions may take the forms $V = a + bD^2$, $V = aD^b$, or $V = aD + bD^2$. The parameters a and b may be estimated by simple linear regression, using in the last two cases the transformations $\log V = \log a + b \log D$, and $V/D = a + bD$.

Volume functions for general use (*general volume tables*) include as predictors, in addition to the DBH D, the height H (total or merchantable), and in some instances also some upper diameter or form indicator. A commonly used form indicator is *Girard's form quotient*, defined as the ratio

between the diameter under bark at 5.19 m of height and the DBH over bark. 5.19 m corresponds to the end of a first 16 feet log.

Some common forms for volume functions are

$$V = a + bD^2 + H + dD^2H \ ,$$

and the variants obtained with various combinations of a, b and c set equal to zero, and

$$\log V = a + b \log D + c \log H \ .$$

We are dealing with typical linear regression problems, without major complications. Nevertheless, three particular aspects may be mentioned.

It often happens that the dispersion of the regression residuals for V tends to increase with the values of the variable, an instance of heterocedasticity. The logarithmic transformation, when it is used, can eliminate or reduce this effect, because if σ is proportional to V then the deviations of $\log V$ have a variance approximately constant.

Another way of facing heterocedasticity is to use *weighed least-squares*, applied to volume tables by Cunia in 1964. The assumption is that the variance of ε_i is $\sigma^2 w_i$, where the w_i are known. This is a special case of generalized least-squares, with W diagonal (only its diagonal elements $w_{ii} \equiv w_i$ are different from zero). The parameters are then estimated by minimizing the weighed sum of squares $\sum e_i^2/w_i$. A program for ordinary linear regression can also be used, noticing that the model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ with variance $\sigma^2 w_i$ reduces to one with variance σ^2 if we divide both sides by $\sqrt{w_i}$.

A typical example is the volume equation $V = a + bD^2H$. It is often the case that the residuals suggest a σ proportional to the independent variable D^2H. Better estimates for the parameters a and b are therefore obtained by fitting the equation $\frac{V}{D^2H} = a\frac{1}{D^2H} + b$.

Another topic that is often mentioned is the fact that when using logarithmic transformations of the dependent variable, such as $\log V$, biased estimators are obtained for the original variable, in this case V. Corrections to the results of the linear regression have been proposed in order to reduce the bias. It is not clear, however, if this is really justified, because generally the bias is reduced at the cost of increasing the MSE and reducing the likelihood.

It is also possible to avoid the bias by using nonlinear regression, for example fitting by least-squares the equation $V = kD^bH^c$ instead of the logarithmic equation shown above. This has become fashionable with the advances in computing. Considering the already mentioned stabilization of

the variance that is generally obtained with the logarithms, it seems probable that in most instances this remedy is worse than the disease.

The third topic is the comparison of models that use different dependent variable transformations. Obviously in this case it does not make sense to compare the regression SE, MSE or R^2, since these refer to different variables. Probably the best thing to do is a graphical analysis of residuals, since the quality of the fitting may be different for different values of the variables. Another possibility is to calculate the SE or MSE for the untransformed variables, or with a same transformation, preferably separately for various predictor ranges. In a different approach, Furnival (*Forest Science* 7,337–341, 1961) proposed an index that is frequently used for this purpose. It is essentially an approximation to a transformation of the likelihood function. It must be kept in mind, therefore, that it measures as much the plausibility of the regression function, as that of the error distribution implicit in it.

3.6 Form factors and quotients, etc.

Various combinations of diameters are used to describe form and to estimate tree volumes, especially in Central Europe. Girard's form quotient, widely used in North America, has already been mentioned.

Form factors are ratios between the volume of a tree and the volume of a certain cylinder. One speaks of the *artificial* form factor, based on the under-bark volume and the volume of a cylinder of equal diameter to the DBH (over bark), and of *natural* form factors based, for instance, on the cylinder with diameter equal to that at the middle of the tree, or at some other proportion of the total height. Hohenadl's method, used in Germany, is based on the diameter at 1/10 of the height.

Thus, the artificial form factor is

$$f = \frac{V}{\frac{\pi}{4} D^2 H} \ ,$$

using consistent units. Regressions are used to estimate f as a function of predictors such as D and H. This does not differ substantially from the volume functions already discussed, especially when using transformations such as Cunia's. Therefore we will not study it in detail.

A useful application is for quick-and-dirty volume estimates, assuming a constant form factor. This is clearly equivalent to using a volume function of the form $V = bD^2H$.

Questions, exercises

1. What percentage difference in the calculated volume would exist when measuring the DBH at 1.4 instead of 1.3 m above ground? Assume a constant form factor, and the same taper used in the International Log Rule (0.5" in 4', that is, 1:96). Compute for DBH's of 20 and 40 cm.

2. In a tree we have measured a DBH (over bark) of 25 cm, height of 20 m, and double bark thickness at breast height of 30 mm. Assume that the form of the stem above breast height is somewhere between a cone and a paraboloid. Below breast height, approximate it by a cylinder.

   (a) Calculate a minimum and a maximum for Girard's form quotient.

   (b) Calculate the total cubic volume under bark in both instances.

3. By integration, prove that the volume of the classic solids of revolution with equation $d^2 = kh^n$ is also given by $V = bD^2H$, as functions of the height H and of the diameter at the base D.

4. Verify the values of b equal to 0.20 for the neiloid, 0.26 for the cone, 0.39 for the paraboloid, and 0.79 for the cylinder.

We saw that volumes can be specified up to and between various diameter limits. It is possible to develop volume functions for variable diameter limits, including the diameter limit as another independent variable in the regression. Taper functions are another route to estimating these volumes.

3.7 Taper functions (curves)

Taper curves or functions describe the expected diameter, under or over bark, at different heights up the stem. The functions contain also DBH and total height, and occasionally other variables. Among other things, with this information cubic volumes, sawn, etc., can be calculated for any portion of the stem, for instance between two given diameter limits. It is even possible, given the specifications (minimum and maximum diameters and lengths, position in the tree) for products such as pulp logs, sawlogs or veneer logs of various grades, to simulate tree bucking, or to determine the best way of doing it.

One of the simplest taper functions, recommended by Kozak, Munro and Smith in Canada, is a regression of the form

$$\frac{d^2}{D^2} = b_1 + b_2\frac{h}{H} + b_3\frac{h^2}{H^2} \ ,$$

where d is the diameter at height h, and D and H are DBH and total height, respectively.

Note in the first place that it might be desirable to force the function so that when $h = H$ (at the top) the diameter is zero. Clearly, one must have $b_1 + b_2 + b_3 = 0$. Substituting b_1 in terms of b_2 and b_3, it is seen that this can be achieved with a regression

$$\frac{d^2}{D^2} = b_2(\frac{h}{H} - 1) + b_3(\frac{h^2}{H^2} - 1) \ ,$$

or

$$\frac{d^2/D^2}{h/H - 1} = b_2 + b_3(\frac{h^2}{H^2} + 1) \ ,$$

for example. If d is a diameter over bark, one could also make $d = D$ for $h = 1.3$ m, leaving a single free parameter.

A second observation is that a second degree polynomial in h will not well represent the form of the stem near the base. This could be improved adding a term in h^3/H^3 (recall the neiloid $d^2 = kh^3$). In practice it has been found that to better represent the butt-swell, a term with a quite high power h, such as h^8 or h^{40}, is generally needed.

Another characteristic of this kind of equation is that it implies a "shape" that does not change with tree size. Graphing d *vs* h for trees with different D and H, it is clear that the curves can be matched over all their length by choosing appropriate scale factors for the axes d and h (the curves of d/D *vs* h/H are equal). Often better results are obtained if form is allowed to vary with size. For this, in equations like the previous one, functions of D and/or H are substituted for some of the b_i. Note that if these functions are linear in their parameters, the regression is still linear after the substitution. A way of finding appropriate forms for these functions is to fit an initial regression separately for each tree, and then to graph the b_i over D and H. This it is an example of problems sometimes called, for historical reasons, "harmonization of curves", and that appear frequently in Forest Mensuration.

A great variety of models and estimation methods has been used to obtain taper functions. We will only examine another two examples.

P. L. Real and J. A. Moore (pages 1037–1044 in *Forest Growth Modelling and Prediction*, USDA Forest Service, General Technical Report NC-120, 1988), used the following initial model for Douglas-fir, fitting it independently to the data of each tree:

$$y = b_1(x^3 - x^2) + b_2(x^8 - x^2) + b_3(x^{40} - x^2) \ ,$$

where y is $d^2/D^2 - x^2$ and x is $(H-h)/(H-4.5)$. Note the use of high powers of h, and the conditioning to ensure a zero diameter at the apex, and $d = D$ at breast height (d is over bark, heights are in feet, and 4.5 is the breast height). Then, the b_1, b_2 and b_3 for each tree were fitted to three regressions containing (not necessarily in linear form) H, D, and crown length.

This model is atypical in including another predicting variable in addition to D and H, the crown length. This seems a good idea, recall the differences in stem form within the crown (approx. conical), and below it (approx. parabolic). In addition, crown length reflects stand characteristics, being associated with its density, and it is known that two trees with a same DBH and height, one dominant in a dense stand, and the other suppressed in a more open stand, would have different forms. On the other hand, it would probably have been advisable to re-estimate the parameters directly with the full data set after substituting the expressions for the b_i.

The other example is from A. Gordon (*New Zealand Journal of Forestry Science 13*,146–155, 1983), for radiata pine. The function is

$$d^2 = \frac{4V}{\pi H}(b_1 z + b_2 z^2 + b_5 z^5 + b_{16} z^{16}) \ ,$$

with the constraint $\sum b_i/(i+1) = 1$ enforced through regression with a transformed function. The variable z is $1 - h/H$, and V is the volume calculated with a cubic volume function obtained from the same data. This is what is called a *compatible* taper function, a concept developed by Demaerschalk in Canada. It has the property that integrating $\pi d^2/4$ with respect to h between 0 and H produces exactly the same V estimated by the volume function. Although the compatibility between taper and volume functions is appealing, it is not very clear if the possible sacrifices of flexibility and precision are worthwhile.

Functions based on polynomials seem to be most common, but other forms have also been used, such as rational and trigonometric functions. The use of *splines* has also become popular. These are combinations of several functions, usually cubic polynomials, each one valid over certain range of the independent variable, and with continuity and smoothness constraints at the joining points.

From a statistical point of view, we may comment that in taper functions the assumptions necessary for the optimality of least-squares are far from being fulfilled. The homoscedasticity assumption is unrealistic, considering that there are perfectly known points (the apex and possibly the DBH), near which errors should be smaller. Similarly, the independence assumption is untenable since, on a given tree, diameters taken close together tend to deviate from the mean in the same direction.

Questions, exercises

1. We have a taper function (under bark):

$$(d/D)^2 = 1.0 - 1.5h/H + 0.5(h/H)^2 \ .$$

   For a tree of 25 m of height and 30 cm DBH calculate:

   (a) Girard's form quotient.

   (b) The total cubic volume.

   (c) The number of 4 m sawlogs with small diameter of 20 cm or more (assume a 30 cm stump).

3.8 Bark functions

In many instances it is necessary to estimate bark thickness, its volume, or the diameter under bark from diameter over bark or vice versa. As a first approach one can obtain regressions for bark thickness, or for the ratio between diameters under and over bark, as a function of the diameter at that point. Frequently this relationship between bark and diameter varies along the stem, and more precise estimates are obtained by including the height of measurement as an additional predictor.

It must be kept in mind that measurements with the bark gauge are imprecise and can have considerable biases, so that in many instances it is preferable to use an estimate instead of the direct measurement.

Chapter 4

Stands

We refer here to aggregate characteristics, applicable jointly to all the trees in a piece of land. This may be a stand, a hectare, or a sample plot.

4.1 Diameter, basal area

We have already considered before the measurement of diameters on logs and on individual trees, in particular the DBH and its definition. We saw that generally there was more interest in the cross-sectional area, derived from the diameter measurement, that in the diameter in itself. The same happens in a stand. At this level the basal area and the mean DBH associated with it are important.

The *basal area* is the sum of the cross-sections at breast height, per unit of land area. Normally, the section for each tree is calculated from the DBH, assuming a circular section, and the basal area is expressed in square meters by hectare. In the USA and some other countries, square feet by acre are used. Sometimes basal area estimates are obtained without measuring DBH, for instance with Bitterlich's point sampling method, or with regressions based on measurements on aerial photographs, as it will be seen in the Forest Inventory course. The area of a circle with diameter equal to the DBH of a given tree is also called its basal area (tree basal area).

In Forest Mensuration *mean DBH*, or mean diameter, usually implies the diameter of the tree of mean basal area. That is, unless specified otherwise, the mean DBH is the quadratic mean diameter $\sqrt{\sum d_i^2/n}$. Given the basal area B and the number of trees per hectare N, the mean DBH (meters) is $D = \sqrt{\frac{4}{\pi} B/N}$.

The arithmetic mean diameter is less used. Occasionally dominant diameters or top diameters are defined, based on largest trees. These are related to the top heights described below.

Note that the variance of a sample of diameters is

$$s^2 = \overline{d^2} - \overline{d}^2$$

(the bar over an expression denotes an average). It is seen then that the (quadratic) mean diameter is always larger than the arithmetic mean diameter (unless the variance is zero).

♡ In 1959 a committee of the International Union of Forestry Research Organizations (IUFRO) prepared recommendations to standardize the notation in Mensuration. Small-case letters would be used for tree-level variables, and upper-case for stands. The letters were chosen considering the most common initial for the words in English, French and German: H for height, D for diameter, G for basal area. Diverse variants would be indicated by subscripts (IUFRO, 1959, "The standardization of symbols in Forest Mensuration", Reprinted in: *Technical Bulletin 15*, U. of Maine, Orono, 1965).

The attempt at standardization has not been entirely successful. Both G and B are common for denoting basal area. Although the recommendation to use small-case for trees and upper-case for stands is applied with certain frequency, often lower-case letters would be confused with parameters.

4.2 Heights

4.2.1 Height-diameter curves

For time and cost reasons, often the heights of all the trees in a stand or sample plot are not measured. Heights measured on a sub-sample are used in a regression to estimate heights based on DBH. Simple linear regressions with transformations of the variables are generally used.

Comparisons of height–DBH curves have been done, for example, by Curtis (*Forest Science 13*,365–375, 1967), A.R. Ek (p. 67–80 *Statistics in Forest Research*, Proc. of meeting of IUFRO Subject Group S6.02, Vancouver, 1973), García (INFOR, Nota Técnica No. 19, 1973), and Arabatzis and Burkhart (*Forest Science 38,*192–198, 1992). Some equations that have given good results are $H = b_1 + b_2 \log D$, $H = b_1 + b_2/(D + 10)$, and $\log H = b_1 + b_2/D$.

It is a typical regression problem, without greater complications. It is convenient, nevertheless, to make sure that the resulting curve is not

decreasing, something that frequently happens with some functions, with small samples, and/or with large measurement errors.

Questions, exercises

1. Consider general and local volume functions of the form $V = a+bD^2H$ and $V = c + dD^2$ (a, b, c and d are parameters). If these relationships were exact, what form would the height–diameter equation have?

4.2.2 Dominant height

In addition to the (arithmetic) mean height, several definitions of *dominant height*, also called top height, are used. The objective is to have a variable not appreciably affected by stand density (trees per hectare), nor by silvicultural treatments, specially thinnings. Among other things, such a variable is useful to assess land productivity, as we shall see later. Height does not vary as much with density as diameter or basal area. Nevertheless, when doing a selective thinning in which trees smaller than average size are extracted, the mean height of the residual stand increases. A dominant height, based on largest trees, changes much less.

We can distinguish three general approaches to the problem of defining a dominant height measure. The first one is based on the visual classification of trees in crown classes, typically as dominant, codominant, intermediate, and suppressed. One takes then the average height of the dominants, or of dominants and codominants. This height is little affected by density and thinnings (excluding thinnings from above, which are not very common), and has been frequently used. An objection relates to its subjectivity; it depends on the crown classification, and different observers probably would obtain different values.

In order to avoid this problem, diverse measures based on some objectively determined proportion of the largest trees have been proposed. This is the second type of dominant height measures, developed mainly in Germany. Some correspond to the mean height of a fixed proportion of the largest trees, such as the 20% of the trees of largest diameter (proposed by Weise in 1880). Other proposals take the trees that exceed the average by more than one or two standard deviations. It is clear that this approach is somewhat deficient in its sensitivity to low thinnings.

The third type of dominant height uses a fixed number of trees by hectare instead of a fixed proportion. For instance, the 100 or 200 largest trees per hectare. This is done in sample plots, where the proportion corresponding to

the plot area is taken: for the 100 largest per hectare, the 10 largest in a 1000 m^2 plot are selected, or the 5 largest in a 500 m^2 plot. There is a number of variants of this approach. The m tallest, or the m trees of largest diameter may be chosen. The second way is often easier. When the heights are estimated from a height–DBH curve, both methods are obviously equivalent. Once the m trees are selected, the arithmetic mean of the heights, either measured or estimated, can be calculated. Another alternative is to calculate the quadratic mean of the DBH for the m trees (called sometimes dominant DBH or top DBH), and to use the height given by the height–DBH curve for that mean. The height of the highest or fattest tree in 1/100 ha plots or sub-plots is also used. In all cases, malformed trees (broken, forked, etc.) are excluded.

♡ This type of heights defines an objective calculation procedure, and at the same time it is not appreciably affected by low thinnings or natural mortality. There is nevertheless a definition problem that is generally ignored, but that can be important in some situations. The expected mean height of the 100 largest trees in one hectare is not the same as that for the 10 largest ones in 1/10 ha, or that of the largest tree in 1/100, etc. This effect of plot size is easy to understand thinking about an area of 1/10 ha subdivided into 10 parts of 1/100. If the largest tree in each one of the 10 parts is selected, it is clear that the mean height will be in general lower than if the 10 largest ones are chosen, independently of their location.

The problem was identified by J. Fries (Sveriges Skogsvårdsförbunds Tidskift:72, 559–563, 1974) and B. Matérn (Sv.Skogs.Tids.:74, 51–53. 1976). On the basis of observations and theoretical approximations they found, for example, that the expected value of the mean height for the 10 largest trees in 0.1 ha corresponds to the height of the one largest tree in 0.015 to 0.018 ha (instead of 0.01 ha). Therefore, in Sweden the dominant height is defined more precisely as the mean of the 10 largest in 0.1 ha. K. Rennolls (*Comm.For.Rev. 57*:215–219, 1978) also treated the subject and proposed as standard the largest in 1/100 ha.

♡♡ Sometimes Lorey's height is mentioned. This is the mean height weighed by the tree basal areas, $\sum b_i h_i / B$. The idea is that multiplying by the basal area gives a quantity more closely related to the volume per hectare than using the arithmetic mean. This because the volume of tree i is approximately a linear function of $b_i h_i$.

Questions, exercises

1. The class of 1994 obtained the following data for a 500 m^2 *Eucalyptus nitens* plot:

DBHs: 30 7 33 22 23 24 26 13 30 25 6 23 19 14 19 29 28 25 26 42 20 21 35 42 14 45 24 14 41 23 13 28 42 26 20 22 25 40 27 21 26 27 38 23 46 21 20 16 22 25 31 29 21 21 22 25 25 22 21 17

Sample trees:
DBH: 30 22 26 30 28 42 20 21 35 42 14 45 27 26 27 31 29 25
HT: 22 24 26 30 23 28 20 20 25 27 14 32 27 25 27 27 28 24

Calculate DBH and dominant height according to the several alternatives described.

2. Separate the DBH data in four groups of approximately equal size. Find the two largest trees in each group. Compare to the eight largest trees in the plot.

4.3 Cubature of stands/plots

In inventory the volume of a stand is estimated from the volumes for some number of sample plots. In order to calculate a plot volume one could calculate the volume of each tree, either by direct cubature or with a general volume function, and add. However, in most instances cubing all the trees, or even measuring all the heights in order to apply a volume function, is too expensive. One then resorts to an indirect estimation using the DBH of all the trees, and the heights or volumes in a part of them. The plot trees that are measured in detail are sometimes called "sample trees".

There are two commonly used methods. The first one is used with general volume functions (tree volume as a function of DBH and height). We assume that a suitable volume function is available. We begin by using the sample trees, in which height and DBH have been measured, to fit to a height–diameter curve. With this, the height for the trees in which only DBH has been measured is estimated. Having now DBH and height for all the trees, the volumes are calculated with the given volume function, and added. The volume by hectare can be calculated dividing by the corresponding plot area. This method can also be seen as a substitution of the height given by the height– DBH function $H = f(D)$ into the general volume function $V = g(D, H)$, obtaining a local volume function $V = g[D, f(D)] \equiv h(D)$.

In the second method, the volume of each sample tree is determined first. This can be done by direct cubature (from diameters measured at several heights), or with a suitable volume function. With the sample tree volumes and DBH, a regression of volume as function of DBH (a local volume function) is fitted. With this, the volumes for all the trees are calculated,

and added. Often the data are well fitted by a simple linear regression of volume on DBH squared (or on tree basal area), the so-called *volume line* or volume–basal area line[1].

The first method is most used, and may be somewhat less time consuming when the height–diameter curve is needed anyway, for instance for calculating dominant height. The local volume function may turn out to be somewhat biased because of the indirect way in which it is obtained. The second method is more direct and more general (it is not limited to the use of volume functions). The volume line normally presents a much tighter relationship between the variables that the height–diameter curve. Usually the second method gives somewhat better results, although the differences are not great.

It is customary not to include as sample trees those that display malformations such as forks, twists, or broken leaders. This provides more consistent measures at the stand level, both for heights and for volumes. On the other hand, some overestimation takes place. It is important also to remember that what is calculated is a standing volume that, due to logging losses and waste, differs from the volume extracted.

As a general comment, it could be said that perhaps the importance of calculating cubic volumes tends to be overrated. In practice, mean conversion factors of doubtful accuracy are applied to the cubic volume for estimating sawn volumes, dollar values, etc. Its direct utility is mostly as a conventional unit traditionally used for comparative purposes.

Questions, exercises

1. With the data from the *E. nitens* plot of Section 4.2.2, estimate the volume by hectare with both methods. Graph the data and relationships.

2. Compare the MSE of the local volume functions obtained.

3. Compare the fit to the data for this height–DBH function with that for the other functions previously indicated, graphically and through the MSE.

4. What form of local volume function would imply the best height–DBH function found in the previous point? Compare with the volume line.

[1] Unrelated to the similarly named Gray's "volume line", which is the line of squared diameter over height on the stem for the parabolic part of a tree.

5. In a 500 m^2 plot, the following DBH (cm) were measured, sorted in increasing order:

   6.3 7.4 12.9 13.1 13.6 13.7 13.9 16.2 16.7 19.0 19.1 19.9 20.1 20.5 20.6 21.1 21.2 21.2 21.2 21.5 21.7 21.7 21.8 22.0 22.5 22.7 22.7 22.9 23.4 24.3 24.4 24.6 24.9 25.1 25.2 25.4 25.4 25.7 26.0 26.3 26.4 26.9 27.2 28.1 28.1 28.7 29.4 29.8 30.0 31.0 33.0

   ($n = 51$, $\sum D = 1136.5$, $\sum D^2 = 26935.99$). With sample trees, the height–DBH curve $H = -15.1 + 12.0 \ln D$ and the volume line $V = -0.115 + 0.00082D^2$ were obtained.

   (a) Calculate the volume per hectare.

   (b) Calculate a dominant height based on the 100 largest trees per hectare.

   (c) Same as (b), but with another method/definition giving a different result.

4.4 Volume functions

Using data obtained by cubature of a number of plots, one can obtain regressions of volume as a function of stand variables such as basal area and dominant height. These are stand volume functions (or "tables").

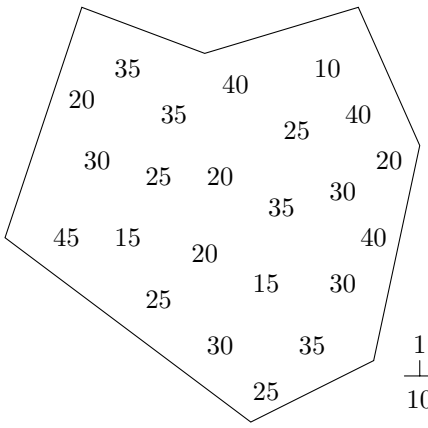
It is easier to use the volume function than cubing a plot directly, so that a function might be used for convenience. More frequently, stand volume functions are useful when information on the individual trees is not available. An example might be when determining basal area indirectly through Bitterlich's method. Another one is when basal area and height come from a growth model projection.

It is a typical regression problem, without greater complications. Functions similar to those of volume by tree are used, with basal area in place of the DBH squared. For example, $V = b_1 + b_2 BH$. A function that has given good results in coniferous plantations is $V/B = b_1 + b_2 H$, where the division by B helps with heteroscedasticity.

4.5 Stand tables, distributions

The individual DBH measured in a plot or stand can be used as a more detailed description than the summary provided by the mean DBH or the basal area. The set of observed diameters (a "tree list") may be used directly,

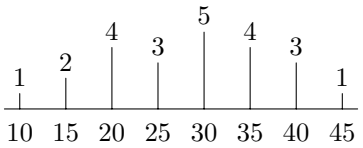
or these may be presented in the form of a histogram or of a frequency table (a "stand table").



Tree list:

20,30,35,45,15,15,30,30,35,35 ,30,20,
20,25,40,40,35,10,25,25,20,25 ,35

Stand table:



In the stand table, traditionally the DBH are grouped in *DBH classes*, and the number of trees in each class is shown in terms of its per hectare equivalent. It is customary to include also the volume by hectare calculated for each DBH class, what is called a "stock table", and sometimes also the estimated heights and other variables. The grouping into classes is done *a posteriori*, or it may result from recording only the diameter class at the time of measurement.

Stand tables were often used to facilitate the calculation of aggregate variables such as the basal area, mean DBH, and volume per hectare, adding class values weighted by frequency. With the advances in computing this application has lost its importance. Nowadays it is not recommendable to group data in this way for calculation purposes, since the loss of precision is not justified. Stand and stock tables may still be useful as a simple and convenient summary of stand characteristics.

An alternative to the stand table, which is essentially a histogram of the diameter distribution, is the approximation of the distribution by continuous functions. Probability distribution functions are used, making an analogy between the proportion of trees with diameters in a certain interval, and the probability that a random variable should take values within the interval.

The *distribution function* $F(x)$ for a random variable X is the probability of this having a value less than or equal to x:

$$F(x) = \Pr\{X \leq x\} \ .$$

If the variable is continuous and the derivative exists, $f(x) = dF(x)/dx$ is the *density* function. Given a distribution or density function, the probability

of X having a value between a and b, or analogously the proportion of trees with DBH between a and b, can be calculated then as

$$\Pr\{a < X \leq b\} = F(b) - F(a) = \int_a^b f(x)\, dx \ .$$

In general, one talks of a probability *distribution*, defined by its distribution function (or cumulative distribution), or by its density.

For uneven-aged stands, *Liocourt's law* (1898) says that the frequencies by diameter classes would diminish with diameter following a geometric progression. This corresponds to an exponential distribution with density $f(x) \propto e^{-kx}$. In reality this does not always agree with the observations.

In even-aged stands, unimodal densities are normally used, that is to say, densities with a unique well-defined maximum. Among the distributions that have been used or proposed there are the normal, log-normal, beta, gamma, Johnson S_B, and Weibull. In general, no large differences are found between the representation of data by any of them. This is not surprising, since to reliably determine the form of a distribution one would need extraordinarily large samples, something that has not been duly appreciated by the majority of researchers and users of these distributions.

Currently, the most used it is the Weibull distribution, since it gives acceptable results, and is mathematically convenient for having explicit expressions for both the density and the distribution function:

$$F(x) = 1 - \exp[-(\frac{x-a}{b})^c], \quad f(x) = \frac{c}{b}(\frac{x-a}{b})^{c-1} \exp[-(\frac{x-a}{b})^c], \quad a \leq x < \infty \ .$$

This is the Weibull with three parameters, a, b and c. The assumption that it is impossible to find a DBH smaller than a is conceptually somewhat questionable, and it produces statistical difficulties in estimation. It is probably preferable to use the two-parameter Weibull, where $a = 0$. Notice that in this case if the distribution of DBH is Weibull, the distribution of DBH squared (or of tree basal areas) is also Weibull.

Growth models are common that use a diameter distribution as description (state) of the stand, and that predict how the parameters of this distribution change over time. Other models project tree lists or stand tables. The distributions, lists and stand tables are also used to estimate different specification product mixes that could be obtained from a stand, being in many instances the basis of sophisticated computer systems that include bucking simulators and optimizers. See, for example, D.M. Hyink and J.W. Moser (*Forest Science 29*, 85–95, 1983), D.J. Depta ("Integrated Forest planning systems at Weyerhaeuser Company", in Nagumo H. et al

(eds) *Proc. IUFRO Symp. on Forest Management Planning and Managerial Economics*, U. of Tokyo, 1984), B.E. Borders and W.D. Patterson (*Forest Science 36*, 413–424, 1990).

Although undoubtedly distributions (in all their forms, including tree lists and stand tables) are useful and necessary for estimating product mixes and sizes, their reliability tends to be overestimated. Aside from the sampling variability already mentioned, the analogy with probability distributions has been taken too far. The use of sample plots produces an overrepresentation of pairs of trees separated by short distances, and the DBH are not distributed at random on the ground. Competition induces negative correlations in the DBH of nearby trees, whereas microsite variations induce positive correlations. Consequently, the distributions derived from plot data, those that are usually obtained, can be considerably different from the distributions for a whole stand, those are usually required in the applications (O. García, " What is a Diameter Distribution?", in Minowa,M. and Tsuyuki,S. (eds) *Proc. Symp. Integrated Forest Management Systems*, Japan Soc. of Forest Planning Press, 1992). These models must be used with caution, appreciating their limitations.

Diverse methods have been used for estimating the parameters of DBH distributions, the main ones being the maximum likelihood method (ML) and the method of moments. This last one consists of making the two or three first moments of the distribution (depending on if 2 or 3 parameters need to be estimated) to agree with the respective moments of the sample. That is, with two parameters one takes the values for which the mean and variance in the sample and in the theoretical distribution are the same. Although statistically not as efficient as ML, there are certain advantages in the consistency of the observed quadratic mean DBH with the one calculated from the distribution (clearly, this happens if either the distribution of the diameters or that of the basal areas is considered).

♡ The mean and variance for the Weibull (with $a = 0$) are

$$\mu = b\Gamma(1 + 1/c) \,,$$

$$\sigma^2 = b^2[\Gamma(1 + 2/c) - \Gamma^2(1 + 1/c)] \,.$$

(Γ is the so-called *gamma function*, defined by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$, whose values are tabulated and for which there are various approximations). The moment estimators are obtained solving these two equations for b and c as functions of μ and σ^2. This solution cannot be obtained algebraically, being necessary to resort to successive approximation procedures.

More convenient is to use an approximation formula that gives c as a function of the coefficient of variation $z = \sigma/\mu$, sufficiently accurate for practical purposes (O. García, *New Zealand Journal of Forestry Science 11*, 304-306, 1981):

$$1/c = z[1 + (1 - z)^2(-0.22102 + z(0.010061 + z(0.11736 - 0.050999z)))] \ .$$

Then one obtains $b = \mu/\Gamma(1 + 1/c)$. For tables and approximations to the gamma function see for example Abramowitz, M. and Stegun, I. A., "Handbook of Mathematical Functions", or the article just cited (in APL the function ! calculates $\Gamma(z + 1)$, which for z integer equals the factorial of z). The approximation given is not valid for $c \leq 1$, where the Weibull takes the form of an inverted J instead of being unimodal.

Questions, exercises

1. Calculate the distribution function and the proportionality factor for the exponential $f(x >) \propto e^{-kx}$. Notice that the total area under $f(x)$, which is the same as $F(\infty)$, must be 1.

2. Fit a two-parameter Weibull to the DBH of *E. nitens* of Section 4.2.2. Use the method of moments.

3. Graph the obtained density and cumulative distribution. Compare with a histogram of the data.

4. What is the estimated number of trees per hectare between 10 and 20 cm?

5. The diameter distribution in an uneven-aged stand is exponential, with density $f(D) = 0.03e^{-0.03D}$. What percentage of the trees has more than 40 cm of DBH?

6. Given n observations, the empirical distribution is the one that assigns probability $1/n$ to each observed value. In a 500 m^2 plot the following DBH were measured (cm), sorted in increasing order:

   6.3 7.4 12.9 13.1 13.6 13.7 13.9 16.2 16.7 19.0 19.1 19.9 20.1 20.5 20.6 21.1 21.2 21.2 21.2 21.5 21.7 21.7 21.8 22.0 22.5 22.7 22.7 22.9 23.4 24.3 24.4 24.6 24.9 25.1 25.2 25.4 25.4 25.7 26.0 26.3 26.4 26.9 27.2 28.1 28.1 28.7 29.4 29.8 30.0 31.0 33.0

   Using the empirical distribution, estimate the number of trees per hectare with DBH between 20 and 25 cm.

7. The mean for the exponential with distribution function $F(x) = 1 - e^{-kx}$ is $1/k$. Assume that in a stand with 800 trees per hectare and a basal area of 40 m^2/ha, the squared diameters (or tree basal areas) follow an exponential distribution.

   Calculate the number of trees smaller than 25 cm, estimating k by the method of moments.

Appendix A

Errors

All measurements are subject to error and uncertainty. Error sources are varied, and could be classified in many ways. For instance, there are what we might call "mistakes", due to wrong readings on an instrument scale, transcription errors, etc. There are instrumental errors, due to defects or bad use of an instrument Personal errors, caused by deficiencies in the observer senses, or by subconscious influence of his interests or preferences. Very important and often ignored are errors due to the model; for instance, in most calculations with tree diameters and cross-sections it is assumed that the cross-section is circular. Systematic errors are those that always act in the same direction.

In relation to an instrument or method that generates a (real or hypothetical) series of measurements, it is useful to distinguish between *accuracy* and *precision*. Accuracy refers to the closeness between measurements and the true value. Precision has to do with consistency, closeness of the measurements among themselves. Measurements can be precise but inaccurate. Some authors understand accuracy as the absence of systematic errors ("bias"), closeness of the measurements mean to the true value.

A.1 Error bounds

In engineering calculations it is common to work with uncertainties or estimated errors assumed to span the true value. That is, a value is given as $x \pm \Delta x$, where x is the estimated value and Δx is a maximum error bounding the true value (it is taken as a positive number, the error absolute value). In other words, by *error* here we understand an error bound.

In particular instances the error in the result of calculations with quan-

tities subject to error can be determined by substituting all possible combinations of negative and positive errors, and taking the extreme results (the combinations to be tried can be reduced if it is clear which are the most unfavorable situations). It is a good idea to do this in important instances. The methods described below are more convenient, and can provide useful relationships between errors and variables.

It is clear that in a sum or difference errors add up, because they are assumed independent and the direction of their action is unknown (for a bound, the most unfavorable situation must be taken):

$$(x \pm \Delta x) + (y \pm \Delta y) = (x + y) \pm (\Delta x + \Delta y)$$

$$(x \pm \Delta x) - (y \pm \Delta y) = (x - y) \pm (\Delta x + \Delta y) \ .$$

Multiplication and division is somewhat more complicated:

$$(x \pm \Delta x)(y \pm \Delta y) = xy \pm x\Delta y \pm y\Delta x \pm \Delta x \Delta y \ .$$

The last term is small relative to the others, and omitting it we can write (assuming that x and y are positive)

$$(x \pm \Delta x)(y \pm \Delta y) = xy \pm xy(\Delta x/x + \Delta y/y) \ .$$

$\Delta x/x$ is the *relative error* for x (while Δx is the *absolute error*). It is seen, then, that the relative error for the product is approximately the sum of the relative errors for the factors. The same happens with division.

More generally, the error for a function of x and y may be approximated by the initial terms of its Taylor series:

$$g(x + \Delta x, y + \Delta y) = g(x, y) + \frac{\partial g(x, y)}{\partial x}\Delta x + \frac{\partial g(x, y)}{\partial y}\Delta y + \dots \ .$$

The omitted terms contain products of errors and, as in the multiplication, can be neglected if the errors are not too large. Considering the uncertainty in the error signs, we find then that in the worst case the error in g is approximately

$$\Delta g = \left|\frac{\partial g(x, y)}{\partial x}\right| \Delta x + \left|\frac{\partial g(x, y)}{\partial y}\right| \Delta y \ .$$

The generalization to any number of variables is obvious.

Let us see two simple examples.

(i) Let $z = g(x, y) = xy$. Then

$$\Delta z = |y|\Delta x + |x|\Delta y \ ,$$

which agrees with the results above.

(ii) The error in the one-variable function $g(x) = \ln x$ is

$$\Delta \ln x = \left|\frac{1}{x}\right| \Delta x = \frac{\Delta x}{x}$$

(x must be positive), so that *the relative error in x is approximately equal to the absolute error in* $\ln x$.

Questions, exercises

1. Use the relationship $\ln xy = \ln x + \ln y$ and the result from example (ii) to obtain the relationship between the relative errors of xy, x and y. Obtain also the relative error of x/y.

2. Calculate the error (bound) for a tree height given the errors in the distance measurement and in the top and base angle measurements.

3. Assume that the height error is dominated by the error in the angle α between the top and the horizontal, and that this error is independent of α (other errors are negligible). Show that the error is a minimum when $\alpha = 45°$.

A.2 Significant figures

Using significant figures is an alternative to expressing an error as $x \pm \Delta x$. Significant figures are the digits, excluding zeros used only for establishing the position of the decimal point. For instance, the numbers 1302, 0.8206, 0.0002135, 60.60 and 1.490×10^3, all have 4 significant figures. Without further information, it is nor known if 1490 has 3 or 4 significant figures.

The indication of errors through significant figures is not fully standardized. Usually, uncertainty in the last given figure is assumed, with that digit giving an idea of the most likely value (the figures "signify something"). Some authors (e.g. Husch) use an stricter criterion, that the error must not exceed one unit in the last figure. Others accept some uncertainty in the before-last figure. In general, it is considered that it does not make sense to specify more than one or two figures in Δx, and that x should be given up to the digit corresponding to the last figure in the error. 15.04 ± 0.15 is correct, not 15.036 ± 0.15. More figures would suggest false accuracy, less would result in unnecessary accuracy loss.

Anyhow, the number of significant figures reflects the relative error, while a number of decimal places reflects absolute error. The precision indicated

by significant figures, or the relative error, are independent of the measurement units: 3.24 m and 324 mm carry the same precision.

These relationships between figures and errors allow us to establish certain rules about the significant figures to be used in results from arithmetical operations. The error in a sum or difference is dominated by the largest absolute error in their components (as seen above, maximum errors add up; other error measures combine with less weight on the smaller errors, as will be seen below). Therefore, a rule is adopted to give the result with a *number of decimal places* equal to the least number of decimal places among the terms added or subtracted:

```
  123
   32.3
+   0.276
  -------
  156
```

In multiplication and division the same happens with the relative errors, so that in the result the least *number of significant figures* among the factors is used:

```
754.1 x 0.052 = 39
```

In the intermediate steps of a calculation sequence it is advisable to retain additional figures, and round the final result.

It is important to take into account that in some operations important losses of significant figures (precision) can occur. This is the case of "catastrophic cancellation" when subtracting large nearly equal numbers.

Questions, exercises

1. Indicate the number of significant figures in: (a) 1.00025 (b) 0.002710 (c) 10.003 (d) 100000

2. In the examples of sum and multiplication just given, assume errors of ± 2 units in the last significant figure. Compute the error limits by extreme value substitution. Compare to the significant figures.

3. In an evaluation of silvicultural regimes, incomes of \$3,274,531 and costs of \$3,256,890 are obtained. Compute the expected profit.

(a) Assume now an error of about 1%. Repeat the profit calculation using the appropriate number of significant figures. What can you say about the profitability?

(b) With the 1% errors, obtain error limits by substituting the most optimistic and most pessimistic values.

4. A sample variance can be computed as $\frac{1}{n}\sum(x_i - \overline{x})^2$, where $\overline{x}$ is the mean $\frac{1}{n}\sum x_i$. It is often suggested to simplify calculations by using the formula $\frac{1}{n}\sum x_i^2 - \overline{x}^2$.

   (a) Show that both formulas are mathematically equivalent.

   (b) Compute with both formulas the variance for the three numbers $x_1 = 100001$, $x_2 = 100002$ and $x_3 = 100003$. What happens?

♡ The statistical approach

In calculating error limits we took the most unfavorable situation, with signs for the various errors such that the error in the result is the largest possible. For instance, when adding x to y it is assumed that Δx and Δy act in the same direction, positive or negative, compounding their effects. This is useful because it provides an upper error bound. However, specially with several variables, these limits may be too wide to be useful, and it may seem unrealistic for all errors conspiring to produce the worst possible result. Instead of error limits, it is therefore possible to work with a statistical or probabilistic model of measurement uncertainty.

Statistics deals with the use of information in situations of uncertainty. It uses *Probability Theory*, which deals with the mathematical properties of some uncertainty models.

An uncertain quantity can take any value within a set of possible values. Some values are more plausible than others, so that we give them different weights. These weights might represent relative frequency under repeated observation, a subjective degree of credibility for the various values, etc. In the model we represent the uncertain quantity by a *random variable*, and the weights by a probability. As always, the theory and mathematical manipulation of the model are independent of its interpretation, but obviously this is important when assessing the applicability of the results.

For now, we consider quantities that take on numerical values, so that the weights can be represented by a probability density function defined on the real numbers. The probability for the random variable X to be between a and b is $\int_a^b f(x)\,dx$. Obviously, $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Sometimes it is convenient to distinguish between the random variable X and the observed values x.

PRACTICAL SITUATION		PROBABILISTIC MODEL
Uncertainty in x	$\rightsquigarrow$	X is a random variable
Weighting of possible values	$\rightsquigarrow$	density $f(x)$
Weighted mean of $g(x)$	$\rightsquigarrow$	expected value $E[g(X)]$

The expected value or expectation of a function $g(X)$ is the weighted mean

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx \ .$$

Important special cases:

$$\text{Mean: } E[X] = \mu$$

$$\text{Variance: } E[(X-\mu)^2] = E[X^2] - (E[X])^2 = \sigma^2 = V[X] \ .$$

The meal is a measure of location, the center of gravity around which uncertainty is distributed. The standard deviation $\sigma = \sqrt{\sigma^2}$ is an important measure of spread.

Back to errors, let us represent a measurement or observed or calculated value by a random variable X, and denote the true value as x_0. The error (another random variable) is $\varepsilon = X - x_0$, that is, $X = x_0 + \varepsilon$. Then, $E[\varepsilon] = \mu_\varepsilon$ is the *bias*. A measure of precision is the *standard error* $\sqrt{V[\varepsilon]} = \sigma_\varepsilon$ (it is common to call standard error to the standard deviation of an estimator). Another measure that combines accuracy and precision is the *mean squared error*: $MSE = \sqrt{E[\varepsilon^2]}$. Note that

$$MSE^2 = \sigma_\varepsilon^2 + \mu_\varepsilon^2 = \text{variance} + \text{bias}^2 \ .$$

The error bound or absolute maximum that we used previously would be (if it exists): $\Delta x = \max|\varepsilon|$, and the relative one, $\Delta x / x_0$ (or $\Delta x/(x_0+\varepsilon)$ which is almost the same if the error is small).

To study the propagation of errors when computing with variables subject to error (random variables), we need some simple properties of expectations and variances. From its definition as integral it is easily found that expectation is a linear operator:

$$E[aX + bY] = aE[X] + bE[Y] \ .$$

Let us find the variance of a linear combination.

$$\begin{aligned} V[aX+bY] &= E[(aX+bY-E[aX+bY])^2] = E[\{a(X-E[X]) + b(Y-E[Y])\}^2] \\ &= E[a^2(X-E[X])^2 + 2ab(X-E[X])(Y-E[Y]) + b^2(Y-E[Y])^2] \\ &= a^2V[X] + b^2V[Y] + 2abE[(X-E[X])(Y-E[Y])] \ . \end{aligned}$$

The expectation in the last term is the *covariance* between X and Y, $Cov[X,Y]$. Therefore we have

$$V[aX+bY] = a^2V[X] + b^2V[Y] + 2abCov[X,Y] \ .$$

The covariance is related to the *correlation coefficient*

$$\rho = \frac{Cov[X,Y]}{\sqrt{V[X]V[Y]}} \ ,$$

which is zero if X and Y are independent (more precisely, uncorrelated), and can reach 1 if X and Y tend to vary jointly or -1 if the vary in opposite ways. Finally, note that if a is not random,

$$V[X + a] = V[X] \; .$$

♡♡ The density $f(x)$ that defined the probability for intervals on the x line generalizes to higher-dimensional spaces. For instance, the *joint density* $f(x, y)$ applied to the plane of points specified by coordinate pairs (x, y). (These pairs and their analogs in more dimensions can be seen as lists of numbers, or *vectors*). It is said that the random variables X and Y are independent if their joint density is of the form $f(x, y) = f_1(x)f_2(y)$. A consequence that derives from the definition of expectation as a multiple integral is that if the variables are independent, then $E[XY] = E[X]E[Y]$. It is easily verified that this implies $Cov[X, Y] = 0$. It may be mentioned that zero covariance (uncorrelated variables) does not necessarily imply independence, except in the important case of the Normal distribution.

We are ready now to examine error propagation. Let us see first the addition case.

$$E[\varepsilon_{x+y}] = E[(X + Y) - (x_0 + y_0)] = E[\varepsilon_x + \varepsilon_y] = E[\varepsilon_x] + E[\varepsilon_y] \; ,$$

so that biases add up.

$$V[\varepsilon_{x+y}] = V[\varepsilon_x + \varepsilon_y] = V[\varepsilon_x] + V[\varepsilon_y] + 2Cov[\varepsilon_x, \varepsilon_y] \; .$$

If errors act independently, it is seen that the standard error for the sum is

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2} \; .$$

Measured this way, the error grows more slowly than the maximum error Δ.

For the general case we use, as before, the Taylor series:

$$\begin{aligned} \varepsilon_g &= g(X, Y) - g(x_0, y_0) = g(x_0 + \varepsilon_x, y_0 + \varepsilon_y) - g(x_0, y_0) \\ &\approx \frac{\partial g(x_0, y_0)}{\partial x_0}\varepsilon_0 + \frac{\partial g(x_0, y_0)}{\partial y_0}\varepsilon_y \; , \end{aligned}$$

Assuming independent errors, we have then approximately

$$\sigma_g^2 = \left(\frac{\partial g(x_0, y_0)}{\partial x_0}\right)^2 \sigma_x^2 + \left(\frac{\partial g(x_0, y_0)}{\partial y_0}\right)^2 \sigma_y^2 \; .$$

In the derivatives we could have used the means or the observed values, instead of the actual values x_0 e y_0. The approximations would still be valid, provided that the errors are not too large.

Let us use this to calculate the standard error for a logarithm:

$$\sigma^2_{\ln x} = (1/x_0)^2 \sigma^2_x \ .$$

Using the mean instead of x_0,

$$\sigma_{\ln x} = \sigma_x / \mu_x \ .$$

The expression in the right-hand-side is the coefficient of variation (CV) for x.

Questions, exercises

1. Obtain an expression for the coefficient of variation of the product of two independent variables X and Y as a function of the coefficients of variation of the factors.
2. For the previous problem, graph $\mathrm{CV}(XY)/\mathrm{CV}(X)$ over $\mathrm{CV}(Y)/\mathrm{CV}(X)$ for $\mathrm{CV}(X) > \mathrm{CV}(Y)$. What effect have the smaller and larger errors on the error of the result? Implications for model building?

Appendix B

Regression

Regression methods are fundamental in Forest Mensuration. For a more concise and general presentation, we shall first review some matrix concepts.

B.1 Matrices

An order $n \times m$ *matrix* is simply a table of numbers with n rows and m columns:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}] \ .$$

The a_{ij} are the matrix *elements*. Instead of the square brackets, round parenthesis or double vertical lines are also used: $\|a_{ij}\|$.

A *vector* is a list of numbers. In matrix algebra they are taken as one-row matrices (row vector) or one-column matrices (column vector). Unless stated otherwise, we shall assume columns. They are usually represented by lower-case letters, often underlined or in bold-face:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_i] \ .$$

The *transpose* matrix is the matrix obtained exchanging rows and columns. The transpose of A is denoted as A' or A^T.

The *sum* of two matrices is the matrix of sums of their elements:

$$\mathrm{A} + \mathrm{B} = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] \; .$$

Obviously, A and B must be of the same order.

A single number, to distinguish it from a vector o matrix, is called a *scalar*. The product of a scalar and a matrix is obtained by multiplying the scalar and each of the elements of the matrix:

$$k\mathrm{A} = k[a_{ij}] = [ka_{ij}] \; .$$

From this, the subtraction or difference of matrices is

$$\mathrm{A} - \mathrm{B} = \mathrm{A} + (-1)\mathrm{B} = [a_{ij} - b_{ij}] \; .$$

The *matrix product* $\mathrm{AB} = \mathrm{C}$ is obtained in the following way:

$$[c_{ij}] = [\sum_{k} a_{ik} b_{kj}] \; .$$

That is, the element ij in the product is the sum of products of the elements from row i of A and those from column j of B. Clearly, for the product to be defined the number of columns in the first matrix must equal the number of rows in the second one.

Defining the product in this way is useful, for example, in handling systems of linear equations. The system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written simply as

$$\mathrm{A}\mathbf{x} = \mathbf{b} \; .$$

A sum of squares is

$$e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} \; ,$$

where

$$\mathbf{e} = [e_1 e_2 \cdots e_n]' \; .$$

Even if two matrices have the proper dimensions for calculating the products AB and BA, in general the results are different (the matrix product is not commutative). Other than this, and that the operations are not always possible (certain relationships between numbers of rows and columns must be satisfied), the sum, difference, and product of matrices behave as the corresponding operations on scalars. For instance,

$$\begin{aligned} \mathrm{A}(\mathrm{B}+\mathrm{C}) &= [a_{ij}][b_{ij}+c_{ij}] = [\sum_k a_{ik}(b_{kj}+c_{kj})] = [\sum_k a_{ik}b_{kj} + \sum_k a_{ik}c_{kj}] \\ &= \mathrm{AB}+\mathrm{AC}\ . \end{aligned}$$

Questions, exercises

1. Show that the sum is commutative, $\mathrm{A}+\mathrm{B} = \mathrm{B}+\mathrm{A}$, and associative, $(\mathrm{A}+\mathrm{B})+\mathrm{C} = \mathrm{A}+(\mathrm{B}+\mathrm{C})$.

2. Show that $(\mathrm{AB})' = \mathrm{B}'\mathrm{A}'$.

3. Compute AB and BA, where

$$\mathrm{A} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\ ,\ \ \mathrm{B} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}\ .$$

4. Compute $\mathbf{x}'\mathbf{y}$ and $\mathbf{y}'\mathbf{x}$, where

$$\mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}\ ,\ \ \mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}\ .$$

   Note: Often a matrix with just one element is considered as a scalar.

5. Show that $p(\mathrm{A}+\mathrm{B}) = p\mathrm{A}+p\mathrm{B}$ y $(p+q)\mathrm{A} = p\mathrm{A}+q\mathrm{A}$.

6. Show that the product is associative: $(\mathrm{AB})\mathrm{C} = \mathrm{A}(\mathrm{BC})$.

The matrices with ones on the diagonal and zeroes elsewhere,

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}\ ,$$

are known as *identity*. They act as the number 1 among the scalars; multiplying any matrix and the identity (of the proper order) does not change it:

$$\mathrm{IA} = \mathrm{AI} = \mathrm{A} \; .$$

Now for an analogue to scalar division. In the same way as subtraction may be seen as summing a negative, $a - b = a + (-b)$, division may be seen as multiplication with a reciprocal: $a/b = a(1/b) = ab^{-1}$, where $b^{-1}b = 1$. With matrices, the analogue of a reciprocal is the *inverse*,

$$\mathrm{A}^{-1}\mathrm{A} = \mathrm{A}\mathrm{A}^{-1} = \mathrm{I} \; .$$

Note that for this to make sense, A must be square (same number of rows and columns). Even thus, not all square matrices have an inverse. Those that do not have one are called *singular*.

Using the inverse we could write the solution of the equation system $\mathrm{A}\mathbf{x} = \mathbf{b}$ given earlier:

$$\mathbf{x} = \mathrm{A}^{-1}\mathbf{b} \; .$$

For this solution to exist, A must be square ($m = n$, that is, the number of equations must equal the number of unknowns). In addition, for A not to be singular, the equations must be "linearly independent" (there must be no redundant equations). There are various methods for inverting matrices, one of the most common being Gaussian elimination. This may be also used to solve equation systems without computing the full inverse.

It is not difficult to verify the following properties:

$$(\mathrm{AB})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$$

$$(\mathrm{A}')^{-1} = (A^{-1})' \; .$$

Finally, one can define vector and matrix derivatives. The derivative of a matrix with respect to a scalar, and the derivative of a scalar with respect to a matrix, are defined simply as the matrix of derivatives. It is then easy to verify results like these (A and $\mathbf{a}$ contain constants):

$$\frac{d\mathrm{A}\mathbf{x}}{dt} = \mathrm{A}\frac{d\mathbf{x}}{dt}$$

$$\frac{d\mathbf{a}'\mathbf{x}}{d\mathbf{x}} = \mathbf{a}$$

$$\frac{d\mathbf{x}'\mathbf{x}}{d\mathbf{x}} = 2\mathbf{x}$$

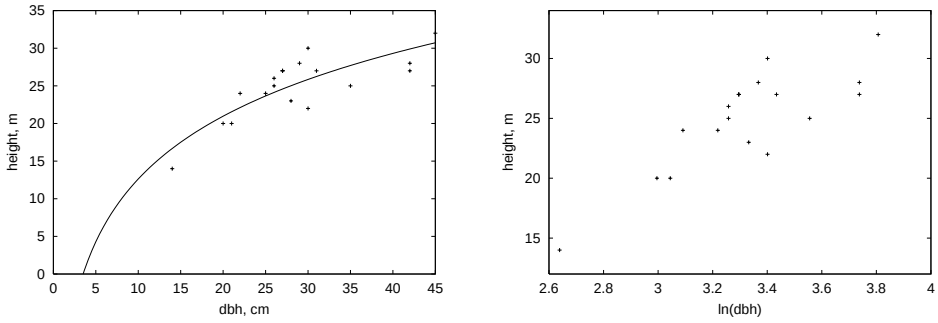
$$\frac{d\mathbf{x}'\mathrm{A}\mathbf{x}}{d\mathbf{x}} = (\mathrm{A} + \mathrm{A}')\mathbf{x}$$

etc. In general, the results are similar to those for scalars, taking into account the no commutativity of products.

B.2 The least-squares method

Many mensurational methods are based on relationships between a *dependent variable* and one or more *independent variables*. One is interested in describing the relationship between the variables, or in estimating or predicting the value of the dependent variable knowing the value of the predictors. For instance, the relationship between heights and diameters may be used to estimate the height of a tree knowing its dbh, which is more easily measured. Or estimate the volume knowing its dbh and height. Or predict the volume of a stand at a given age.

As an example, take a relationship between two variables. It is useful to make a *scatter diagram*, plotting the available observations with the predictor in the abcissa ("x-axis"), and the dependent variable in the ordinate ("y-axis"). The graph on the left shows observations of height and dbh in a stand of *Eucalyptus nitens* taken by the 1994 class.



A curve like the one shown may be used for estimating the heights of trees in the stand for which only the dbh is known. Clearly, knowing the dbh helps in estimating the height, that is, contributes to reduce the uncertainty about its value. The curve is a "model" that provides height values to be used in place of the unknown ones, or that can serve as a summary description of the observations. At any rate, it is convenient to have an equation for the curve to facilitate its use, and the curve should pass "close" to the observations.

In some instances there are theoretical reasons that suggest a specific kind of equation. In others, as in this example, the equation is purely em-

pirical, chosen with convenience and data-fitting criteria. In general, there will be a class of equations or models $y = f(\mathbf{x}, \mathbf{b})$, where y is the dependent variable, $\mathbf{x}$ is a vector of independent variables, and $\mathbf{b}$ is a vector of parameters whose values will be determined for producing a good fit. With a two-dimensional $\mathbf{x}$ we obtain a surface instead of a curve, and for higher dimensions a hypersurface. To choose the equation form one may use experience with similar problems, trial and error, graphs with transformations producing linear data trends, considerations about the form that the curve should take for the extremes, etc. In the example we have used $H = f(D, b_1, b_2) = b_1 + b_2 \ln D$, seeing in the right-hand-side scatter diagram that the relationship between H and $\ln D$ is roughly linear (note in passing that extrapolation to small diameters outside the range of the data eventually produces negative heights). It would be always possible to choose a curve that passes close to each one of the observations. Although in some sense this would describe perfectly the observed data, in general much less irregular curves, with a small number of parameters, will produce better estimates for future or unobserved values.

Once the form of the equation to be tried is decided, it is necessary to choose parameter values that result in a good fit. It can be assumed that, for a given D, the difference between the unknown H and $f(D, b_1, b_2)$ would tend to be smaller if these differences are small for the observed values. That is, $\mathbf{b}$ should be such that the absolute values of the *deviations*, *residuals* or "errors" $e_i = H_i - f(D_i, b_1, b_2)$ are small for all the observations (D_i, H_i). Obviously, if we try to reduce one e_i beyond some point the other e_i will increase, so that we need some criterion that takes into account the whole set of these. A possible criterion would be to minimize the sum of absolute values $\sum |e_i|$ ("L_1-norm regression"). Another possibility would be to minimize the largest error ($\min \max |e_i|$, the *minimax* criterion). The criterion most commonly used, because of mathematical convenience and of possessing in some instances certain statistical justifications that we will examine later, is that of *least-squares*, which consists of minimizing $\sum e_i^2$.

We have then a model $y = f(\mathbf{x}, \mathbf{b})$, n observations $(y_i, \mathbf{x}_i),\ i = 1, 2, \ldots n$, and we look for a $\mathbf{b}$ such that it minimizes

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - f(\mathbf{x}, \mathbf{b})]^2 \ .$$

Equivalently, we minimize the *root mean square error* (RMSE) $\sqrt{\frac{1}{n} \sum e_i^2}$, which is a useful measure of goodness-of-fit. In general, this optimization problem cannot be solved analytically, and it is necessary to resort to it-

erative numerical optimization methods. An important exception occurs when the model is a linear function of the parameters $\mathbf{b}$. In this *linear regression* situation, it is possible to obtain explicit solutions for the optimal (least-squares) values of the parameters or *coefficients*.

Our example of H *vs* D is an instance of linear regression. It can be written

$$y = b_1 + b_2 x \ ,$$

with $y = H,\ x = \ln D$. This is a straight line, taking here the variable x as predictor. In general, both y and x can be transformations of the original variables. Ideally, the data would satisfy the n equations system

$$\begin{aligned} y_1 &= b_1 + b_2 x_1 \\ y_2 &= b_1 + b_2 x_2 \\ \vdots &\quad \vdots \\ y_n &= b_1 + b_1 x_n \end{aligned}$$

which in matrix notation can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\mathbf{y} = \mathrm{X}\mathbf{b} \ .$$

If we had $n = 2$, we would have a system of two equations in two unknowns (b_1 y b_2), usually with a unique solution. In matrix terms, $\mathbf{y} = \mathrm{X}\mathbf{b}$ with X square and invertible has the solution $\mathbf{b} = \mathrm{X}^{-1}\mathbf{y}$.

With $n > 2$, in general not all the observations are co-linear, and the equation system is incompatible. The objective is to find a $\mathbf{b}$ such that the approximation $\mathbf{y} \approx \mathrm{X}\mathbf{b}$ is the best possible, in the sense of minimizing the length $|\mathbf{e}|$ of the vector $\mathbf{e} = \mathbf{y} - \mathrm{X}\mathbf{b}$ computed from a generalization to n dimensions of Pithagoras Theorem:

$$|\mathbf{e}|^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} \ .$$

There are algorithms, based on matrix factorization, that produce directly the least-squares solution of $\mathbf{y} \approx \mathrm{X}\mathbf{b}$. These are used in the better statistical packages. Sometimes, *pseudoinverses* or *generalized inverses* X^{-}

are used, in terms of which the solution can be written as $\mathbf{b} = \mathrm{X}^{-}\mathbf{y}$. The APL computer language to be used in the laboratories has a generalized inversion and generalized matrix division operator that makes very simple the computation of linear regressions. In APL notation, the matrix product $\mathrm{X}\mathbf{b}$ is `X+.×B` (indicating that we are dealing with sums of products). The coefficients can be obtained with the generalized inverse, `B←(⌹X)+.×Y` or, preferably, with the generalized matrix `B←Y⌹X`.

Before presenting the least-squares solution most commonly used in textbooks and manual calculations, let us examine the more general *multiple* linear regression situation, where in contrast to the previous *simple* linear regression example in which there was just one predictor x there are now p predictors. The model is

$$y = b_1 x_1 + b_2 x_2 + \ldots b_p x_p = \mathbf{b}'\mathbf{x} = \mathbf{x}'\mathbf{b} \; .$$

Simple linear regression is the special case $p = 2,\ \mathbf{b} = (b_1, b_2),\ \mathbf{x} = (1, x)$. The system of equations, including now the deviations e_i, is

$$\begin{aligned} y_1 &= b_1 x_{11} + b_2 x_{12} + \cdots + b_p x_{1p} + e_1 \\ y_2 &= b_1 x_{21} + b_2 x_{22} + \cdots + b_p x_{2p} + e_2 \\ \vdots &\quad\ \vdots \\ y_n &= b_1 x_{n1} + b_2 x_{n2} + \cdots + b_p x_{np} + e_n \end{aligned}$$

that is,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathbf{y} = \mathrm{X}\mathbf{b} + \mathbf{e} \; .$$

The matrix equation is the same as before, again we have to minimize $\mathbf{e}'\mathbf{e}$, and the direct factorization and APL solutions do not change. Almost always a constant is included in the model, and then x_1 and the x_{i1} equal 1.

The most usual explicit solution form is obtained as follows. To minimize the sum of squares $Q = \mathbf{e}'\mathbf{e}$, we make the derivative equal to zero:

$$\begin{aligned} Q &= \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathrm{X}\mathbf{b})'(\mathbf{y} - \mathrm{X}\mathbf{b}) \\ &\quad\ \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathrm{X}\mathbf{b} + \mathbf{b}'\mathrm{X}'\mathrm{X}\mathbf{b} \\ \frac{dQ}{d\mathbf{b}} &= -2\mathrm{X}'\mathbf{y} + 2\mathrm{X}'\mathrm{X}\mathbf{b} = 0 \; , \end{aligned}$$

what gives us the *normal equations*:

$$\mathrm{X}'\mathrm{X}\mathbf{b} = \mathrm{X}'\mathbf{y} \ .$$

The p equations may be solved numerically for the p unknowns $\mathbf{b}$. The solution may also be written explicitly:

$$\mathbf{b} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathbf{y} \ .$$

The goodness of fit can be evaluated through the sum of squares $\mathbf{e}'\mathbf{e}$, through the root mean square error RMSE $= \sqrt{\mathbf{e}'\mathbf{e}/n}$, or the standard error SE $= \sqrt{\mathbf{e}'\mathbf{e}/(n-p)}$. The number of parameters p in the SE penalizes somewhat the complexity of the model when comparing alternatives, and has also a statistical justification explained in the following section.

Although this expression is useful in theoretical derivations, in general it is not the most advisable from the numeric point of view. First, the normal equations system can be solved with less work that that necessary for computing the inverse and the matrix product. Second, important catastrophic cancellation errors can occur, similar to those for the computation of variances demonstrated in the error propagation section. As already mentioned, the most accurate procedures are based on the factorization of X.

When the model includes a constant (column of ones in X), cancellation errors in the normal equations can be much reduced by "centering" the variables, as in the case of the variance, using deviations from the means instead of the original variables. For a model $y = b_0 + \mathbf{x}'\mathbf{b} + e$ it is seen that with the least-square parameters the means satisfy $\overline{y} = b_0 + \overline{\mathbf{x}}'\mathbf{b}$, since the first of the normal equations ensures that the sum of residuals is zero: $0 = \mathrm{X}'(\mathbf{y} - \mathrm{X}\mathbf{b}) = \mathrm{X}'\mathbf{e}$. Subtracting, we have the equivalent model $y - \overline{y} = (\mathbf{x} - \overline{\mathbf{x}})'\mathbf{b} + e$. We estimate $\mathbf{b}$ with this model, and the constant is obtained from $b_0 = \overline{y} - \overline{\mathbf{x}}'\mathbf{b}$.

Questions, exercises

1. Estimate b in the model $y = b + e$. Do you recognize the result?

2. Verify that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \ .$$

   Use this to obtain formulas for the two parameters in simple linear regression (with the original variables).

3. Obtain formulas for simple linear regression using the centered variables (deviations from the means).

4. Estimate by least squares the parameters of the model $y = b_1 + b_2 x + b_3 x^2$. Use centered variables.

5. Research the solution of linear equation systems and matrix inversion by Gaussian elimination.

B.3 Statistical considerations

We have presented least squares as a more or less reasonable and mathematically convenient method for "fitting" functions to observed data. Under certain probabilistic models for the deviations, the least squares criterion can also be justified by statistical arguments.

Assume first that the observations y_i are generated according to a model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \ ,$$

where the ε_i are uncorrelated random variables with mean 0 and unknown variance σ^2. That is,

$$E[\varepsilon_i] = 0 \, , \ \ E[\varepsilon_i^2] = \sigma^2 \, , \ \ E[\varepsilon_i \varepsilon_j] = 0 \text{ si } i \neq j \ ,$$

or, with matrix notation,

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, \ \ V[\boldsymbol{\varepsilon}] = \sigma^2 \mathrm{I} \, ,$$

where $V[\cdot]$ is the covariance matrix. The $\mathbf{x}_i$ are known predictor vectors, and $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated.

We look for an estimator $\hat{\boldsymbol{\beta}} = \mathbf{b}$ unbiased, i. e. $E[\mathbf{b}] = \boldsymbol{\beta}$, and with a variance as small as possible. Let us restrict the search also to estimators that are linear functions on the observations, $\mathbf{b} = \mathrm{A}\mathbf{y}$ for some matrix A. Then, the Gauss-Markov theorem says that for the linear minimum variance unbiased estimator $\mathrm{A} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$. This is the least-squares estimator.

The restriction to estimators that are linear on the observations may seem somewhat arbitrary. If we add the assumption that the deviations follow a normal distribution, the least squares criterion is obtained through a different route. Let the model, not necessarily linear, be

$$y_i = f(\mathbf{x}_i, \beta) + \varepsilon_i$$

with the ε_i normal, with mean 0, variance σ^2, and independent. That is,

$$\mathbf{y} = \mathbf{f}(\mathrm{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \ ,$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathrm{I}) \ .$$

The *likelihood function* is the probability of the model generating data like the observed. The *maximum likelihood* (ML) estimation method consists of estimating the unknown parameters as the values that maximize this function. Besides being intuitively reasonable, the MV estimators have a number of desirable statistical properties, especially in large samples.

Here the likelihood function equals the joint probability density of the y_i, considered as a function of $\boldsymbol{\beta}$ and σ^2. From the independence assumption, the joint density is the product of the (normal) densities of each y_i:

$$L = f_1(y_1) f_2(y_2) \cdots f_n(y_n) \ ,$$

with

$$f_i(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{\{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})\}^2}{2\sigma^2}] \ .$$

The likelihood is then

$$L = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp[-\frac{\sum \varepsilon_i^2}{2\sigma^2}] \ .$$

Clearly, the $\boldsymbol{\beta}$ that maximizes L is that which minimizes the sum $\sum \varepsilon_i^2$. We conclude that, under this model, the ML estimator of $\boldsymbol{\beta}$ is the least-squares estimator.

It is also found, taking the derivative of L with respect to σ^2 and making it equal to zero, that the ML estimator of σ^2 is *mean square error* (MSE) $\sum \hat{\varepsilon}_i^2/n = \sum e_i^2/n$, the square of the RMSE. The expected value of $\sum e_i^2$, for linear models, turns out to be $(n-p)\sigma^2$, so that the MSE is biased. It is customary to use the unbiased estimator SE^2 for the residual variance σ^2, and the standard error SE as estimator for σ.

Another goodness-of-fit indicator often used, incorrectly, is the coefficient of determination $R^2 = 1 - \mathrm{MSE}/S_y^2$, where $S_y^2 = \sum (y_i - \overline{y})^2/n$ is the variance of the observations y_i when the predictors are ignored. For comparing models with the same data, R^2 provides the same information as the MSE or RMSE. With different data sets, however, an R^2 close to one does not imply necessarily a tight relationship or a good model. Among other things, the total variance depends of how the sample has been selected, and unless this can be considered as a random sample from a multivariate distribution, it does not represent a characteristic of the population.

QUESTIONS, EXERCISES

1. Compute a linear regression between y and x with the following data:

x	1	2	3	4	5	6	7	8	9	10
y	1	4	9	16	25	36	49	64	81	100

2. Compute R^2.

3. Plot the data and the regression line.

♡ It is seen that for the linear regression

$$E[\mathbf{b}] = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'E[\mathbf{y}] = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{X}\boldsymbol{\beta} = \boldsymbol{\beta}\ ,$$

so that $\mathbf{b}$ is an unbiased estimator. The same happens with any function linear on the parameters, and, in particular, the prediction expected value $\hat{y}(\mathbf{x}) = \mathbf{x}'\mathbf{b}$ equals $y(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ for any $\mathbf{x}$.

Because the covariance matrix $V[\mathrm{A}\mathbf{z}]$ for a linear transformation is $\mathrm{A}V[\mathbf{z}]A'$, it is found that

$$V[\mathbf{b}] = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}\ .$$

If $\boldsymbol{\varepsilon}$ is normal, this and the fact that any linear transformation of a normal vector is normal allow us to obtain confidence intervals and hypothesis tests for linear functions of b.

Obviously, in real life these statistical models cannot be expected to be fulfilled exactly. But it can be expected that the more we approach the assumptions, the better the estimators will be. For instance, if it is seen that the scatter of the residuals is not quite uniform (*heterocedasticity*), it would be advisable to employ some transformation that changes this situation. Another possible problem is the presence of autocorrelation (correlation among consecutive measurements). In particular, hypothesis tests are subject to the plausibility of the statistical model.

♡**Generalized least squares** Assume that in the linear model the covariance matrix for $\boldsymbol{\varepsilon}$ has the form $\sigma^2\mathrm{W}$, with a known matrix $\mathrm{W} \neq \mathrm{I}$. Maintaining the other assumptions, it is then found that both the minimum variance unbiased and the ML estimator are obtained by minimizing $\mathbf{e}'\mathrm{W}^{-1}\mathbf{e}$. The solution is $\mathbf{b} = (\mathrm{X}'\mathrm{W}^{-1}\mathrm{X})^{-1}\mathrm{X}'\mathrm{W}^{-1}\mathbf{y}$.

A good introduction to statistical inference is found in Chapter 2 of Graybill, for which there is a Spanish translation among the course materials.

A general text with a good treatment of linear regression is Peña Sánchez de Rivera, D. "Estadística, Modelos y Métodos" (2 Vols.), Alianza Editorial, Madrid, 1992.