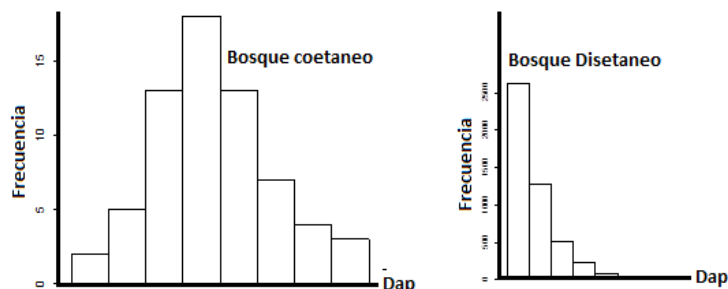


3.4 Teoría sobre Distribuciones diamétricas

En el sentido estricto de la palabra una distribución diamétrica es una relación entre los diámetros y sus frecuencias respectivas, referida a un área determinada en un bosque o rodal. Generalmente se acude al agrupamiento por clases diamétricas en el ajuste por momentos y a los datos crudos cuando se trata de máxima verosimilitud.

En forma simple, una distribución diamétrica es un histograma con los diversos niveles encontrados para los diámetros de un rodal con escogencias subjetivas para el ancho de las clases y los puntos de partida. Por ejemplo, para los datos de las Tablas 3.1 y 3.2 podrían ser unos como estos:



Las distribuciones diamétricas se conciben como herramientas importantes para el manejo y la silvicultura, por ejemplo: cuando, cuanto y, que aclarar; con que probabilidades hacer los cálculos para el crecimiento y, desde el punto de vista dasométrico su mayor importancia sería en el campo de las predicciones tanto en tiempo como en espacio.

Pero no bastan estas gráficas solamente, un aspecto importante son los parámetros que las gobiernan, forma, escala y localización pues como parte de la estructura horizontal de la organización del bosque, estudiada en la silvicultura, se emplean diversas ecuaciones para modelar las frecuencias como función de los diámetros de los árboles o especies que parecen agregarse.

Los enunciados al respecto llevan más de siglo y medio de propuestas teóricas (Loetsch et al, 1962; UNESCO, 1980 y, Uribe, 1984), entre otros, reportan trabajos de autores como Liocourt, en 1899, para comparar las frecuencias de árboles en clases diamétricas sucesivas con lo cual pretendió encontrar una relación constante que, después, se conoció como el cociente de Liocourt.

(Huffel, 1919, 1926, Schaeffer et al, 1930, Goodman, 1930 y Hough, 1932), fueron los primeros en ajustar una función exponencial negativa a las distribuciones. (Meyer, 1933-1952) introduce el concepto del bosque balanceado y le define un modelo exponencial negativo que se extendió debido a su facilidad a otros estudios como el cálculo de incrementos en volumen (Meyer, 1952), regulación y control de cortas (Chagas et al., 1983, Francois, 1938 citados por (Prodan, 1949).

(Leak, 1964), al notar que el cociente de Liocourt no era constante, adopta el criterio de "bosque no balanceado" y ajusta una curva que relaciona dicho cociente con el diámetro. Así mismo se reportan trabajos en este sentido por (Le Cacheux, 1965, Dawkins, 1958,

Pierlot, 1966) citados por (Loetsch et al, 1967). (Zohrer, 1970) introduce la función Beta, (Hervart, 1972) propone modelos hiperbólicos y (Bailey y Dell, 1973) implementan el uso de funciones tipo Weibull, (Moser Jr, 1976) propone estructuras de densidad con base en la distribución J invertida; (Goff y West, 1975) proponen modelos polinomiales etc.). En Colombia los primeros trabajos se dieron por (Gómez, 1983, Uribe, 1984, Arteaga, 1987, Gómez y Ramírez, 1989) y luego otros.

No existe una satisfacción plena en este tema altamente influenciado por decisiones subjetivas o de logística como la escogencia del diámetro mínimo, tamaño y forma de parcelas, pues cuanto mayor sea el número de datos más sensibles las pruebas estadísticas que pueden conducir a rechazar modelos propuestos; se necesitan más conceptos dinámicos que estáticos, saber más del crecimiento, mortalidad, regeneración y manejo de rodales, pues las distribuciones diamétricas son solo una de las características que reflejan un estado del bosque.

3.4.1 Algunas consideraciones importantes. De acuerdo con un resultado gráfico, aportado por (UNESCO, 1980) en el bosque de Imataca en la Guyana venezolana, se resalta la gran variabilidad del, número de árboles presentes por hectárea en cada clase diamétrica, las formas adoptadas por las distribuciones, de acuerdo con el número de parcelas de 1 Ha muestreadas, el número de árboles/ha y las clases adoptadas.

Las distribuciones van desde formas de L para marcas de clases amplias y bajo número de árboles/ha, hasta formas normales o cercanas a una campana de Gauss. Para la clase 10-19 cm aparecen curvas similares para regiones muy distantes entre sí como Venezuela, Gabón y Kampuchea, que permiten de pronto comparaciones o posibles inferencias. Además de lo anterior se denuncia la falta de estudios sistemáticos que presenten las influencias que tienen la pendiente, la altitud, la formación ecológica, el tipo de bosque, quedando solo conclusiones personales al respecto. Se nota que las diferentes formas de la distribución reflejan distintos tipos de comportamiento entre especies de luz y de sombra, así como irregularidades debidas a la historia del bosque, a la geomorfología y topografía y al comportamiento de algunas especies.

Por ejemplo, en la región amazónica, (Heindsdijk, 1957; citado por (UNESCO, 1980) parece distinguir estructuras diferentes según posición topográfica así: las pendientes parecen ser más ricas en árboles pequeños que las mesetas y más pobres en árboles grandes.

(De Milde y Groot, 1970) en la Guyana parecen mostrar que los bosques de ribera y terrazas lateríticas muestran menor número de árboles/Ha especialmente entre 10 y 29 cm que los bosques de terreno quebrado y colinas elevadas.

En Venezuela, Sierra de Imataca se muestra que laderas con pendientes mayores del 30% presentan menos árboles que las de inclinación suave o moderadas (Rollet, 1969). (Pierlot, citado por UNESCO, 1980) no parece encontrar gradiente altitudinal bien definido respecto a las distribuciones diamétricas; sin embargo, en el sudeste de Nigeria se reportan bosques de poca altura y escaso diámetro solamente en crestas estrechas y elevadas, cuando por el contrario (Waytt-Smith, 1960) en Malasia reporta mayor densidad en crestas que en la parte alta o baja de las pendientes.

(Veillon, 1965) en estudios sistemáticos realizados alrededor de la variación de la distribución por alturas, en bosques inalterados, con muestreos entre los 70 y 2.250 msnm, encontró que para $d \geq 20$ cm, muestra el paso de sus frecuencias de 52 a 138 a los 1500 msnm, de ≈ 123 a 1940 msnm, con un máximo absoluto de ≈ 168 a 2960 msnm que se reduce a 71 árboles/Ha al llegar a los 3250 msnm.

Se pretende que algunas formaciones y tipos de bosques pueden ser fácilmente distinguibles por su estructura diamétrica ya que parece evidenciarse dentro de ellos diferencias respecto a la estructura, por ello algunos autores como (Borges, 1972, citado por UNESCO, 1980) relaciona la historia del bosque con la estructura diamétrica.

Algunos autores conceptúan, como (del Valle, 1981), que los bosques maduros cercanos al clímax tienden a distribuirse como una J invertida cuando se da un flujo adecuado de la regeneración hacia los diámetros mayores. Otros proponen teorías como la hipótesis de (Goff y West, 1975), sobre una curva trifásica de la distribución diamétrica del bosque tropical, además de diferencias significativas entre dos tipos de bosques estudiados. Como se ve es un problema visualizable y analizable desde incontantes escenarios y criterios.

3.4.2 Ajuste de la distribución. Ajustar una distribución se refiere al hallazgo de una función matemática que represente la función de densidad de una variable aleatoria, en este caso el dap. Por ejemplo, sean algunas observaciones (x_1, x_2, \dots, x_n) de una población desconocida y deseamos saber si tienen alguna función de densidad $f(X, \theta)$ en que θ sería un vector de parámetros a estimar con los datos dados.

Para el ajuste se identifican 4 pasos principales:

- 1) Visualización gráfica de los datos
- 2) Elección de la función de ajuste, mediante hipótesis de trabajo acerca de ella.
- 3) Estimación de parámetros;
- 4) Pruebas de bondad de ajuste.

3.4.2.1 Visualización gráfica de los datos. Se usarán los datos de la tabla 3.1, y se verá el histograma y su función de densidad, así como la empírica acumulativa ($F(x)$) calculada en R como `ecdf()`:

```
dae<-read.table("clipboard")#Datos de unos eucaliptos
attach(dae)
names(dae)
[1] "mdap" "Frdap"

n<-length(mdap)#número de clases diamétricas
n
[1] 9

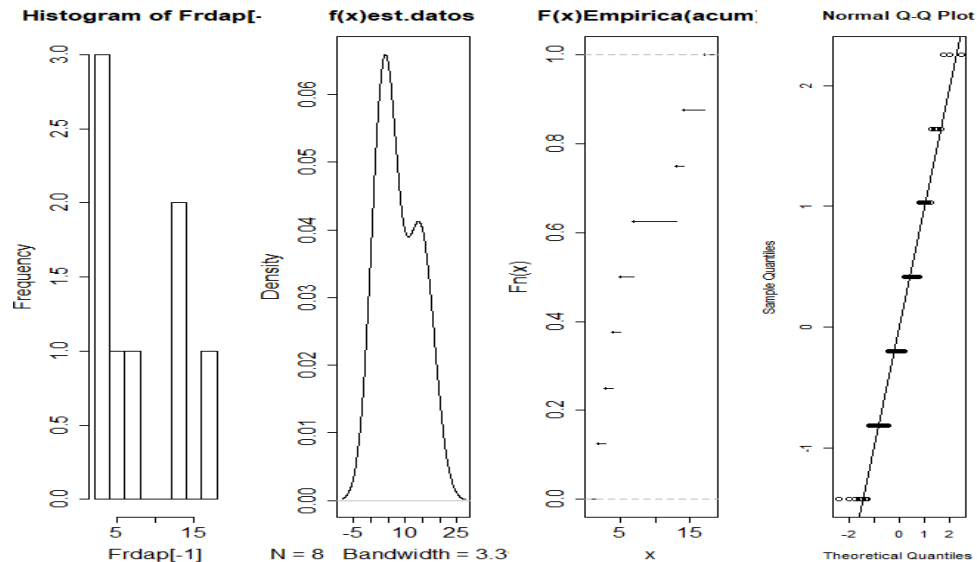
mu<-mdap*%Frdap/sum(Frdap)#Cálculo de la media ponderada
mu
      [,1]
[1,] 48.23077

var<-(Frdap*%(mdap-mu)^2)/(sum(Frdap)-1)#Cálculo de la varianza ponderada
var
      [,1]
[1,] 265.9615

par(mfrow=c(1,4))#Dividimos la ventana en 4 columnas
```

```
hist(Frdap[-1],breaks=6,cex.main=1.7,cex.axis=1.5,cex.lab=1.5)
plot(density(Frdap[-1]),main="f(x)est.datos", cex.main=1.7,cex.axis=1.5,cex.lab=1.5)
plot(ecdf(Frdap[-1]),main="F(x)Empirica (acum.)",cex.main=1.7,cex.axis=1.5,cex.lab=1.5)
```

Si bien el histograma no revela una función conocida, la $f(x)$ podría tratarse de una distribución unimodal (o bimodal), por lo cual usaríamos una normal, en primera instancia.



Debe notarse que una de estas gráficas solo revela detalles aproximados por lo cual es necesario un análisis más profundo de los datos crudos. Acá, por no tener la base original hubo que hacer recomposición de ellos. Para lo cual se crea el archivo como el que tendría los datos crudos para hacer un Q-Q plot:

```
dicru<-data.frame(d=c(rep(25,7),rep(35,14),rep(45,17),rep(55,13),rep(65,7),rep(75,4),rep(85,3))
)
```

El Q-Q plot es un diagrama de dispersión que compara la distribución empírica y la ajustada por medio de los cuantiles empíricos, para determinar si unos datos provienen de una población conocida. En el eje "y" se sitúan los cuantiles empíricos en el x los del modelo teórico. Una línea a 45 grados es la referencia. Si los datos empíricos se parecen a los de la población escogida caerán muy cercanamente a ella. Acá es difícil el análisis por cuanto no se poseen los verdaderos datos crudos, que fueron previamente agrupados en clases.

```
z.dicru<-(dicru$d-mu)/var^0.5
qqnorm(z.dicru) ##QQplot
abline(0,1)
```

3.4.2.2 Elección del número de intervalos. Cuando se adopta el método de momentos, parece haber ventajas estadísticas al agrupar en clases, ya que los datos generalmente son más una situación "familiar" que la probabilidad de representar cabalmente al bosque completo, (Reynolds, 1988), aunque hay opiniones contrarias.

No existen reglas precisas en cuanto al número de intervalos, resultando un asunto más o menos arbitrario, pues con uno demasiado grande no se simplifica el conjunto de datos originales, pudiendo aparecer problemas de ajuste final, y sí es pequeño pueden ocurrir distorsiones de la información. En términos generales se tienen algunas indicaciones: (Snedecor y Cochran, 1971) hablan de formar más de 5; 12 como un número muy aceptable

y 24 como el número máximo, en caso de grandes muestras. *Sturges*, citado por (Díaz y Gutiérrez, 1985) propone que el número de intervalos se obtenga como (3.38 a), otros autores para $n > 200$ proponen 3.38 d.

$$(a) n_i = 1 + 3.32 \log_{10} N; (b) n_i = \sqrt[3]{N}; (c) n_i = \left[\text{entero} \left(\frac{R}{A} \right) \right] + 1; (d) n_i = 4^{0.2} \sqrt{0.75 \sqrt{n-1}} \quad (3.38)$$

en la cual N (n) = número de datos. (Gómez, 1982) habla de la fórmula (b), (Loetsch et al., 1964) proponen la fórmula (c) en que: R = rango, A = amplitud de clase = R/n , redondeado a las cifras decimales del parámetro a evaluar, n_i = número de intervalos. La tabla siguiente, no paramétrica, muestra el comportamiento de la fórmula d), con aspectos tan importantes como que aunque crezca mucho el tamaño de la muestra, el número de clases se va volviendo asintótico.

Número no paramétrico de clases vs tamaño de la muestra

n	k
20	6.8
40	7.86
60	8.54
80	9.05
100	9.47
120	9.82
140	10.13
160	10.41
180	10.66
200	10.89
220	11.1
240	11.29
260	11.47
280	11.65
300	11.81
320	11.96
340	12.11

Para n moderado, k debe restringirse a no sobrepasar $n/5$.

Además de lo anterior, deben escogerse los límites que contengan los valores máximos y mínimos en cuestión, con base en un rango ampliado, cuya diferencia se reparte en ambos extremos.

3.4.3 Error sistemático debido al agrupamiento. (Loetsch et al, 1973) resaltan el efecto de tomar la marca de clase como su diámetro representativo, lo que conduce a un error sistemático causado por no coincidir ni con la media de cada clase, ni con el diámetro del árbol de área basal medio.

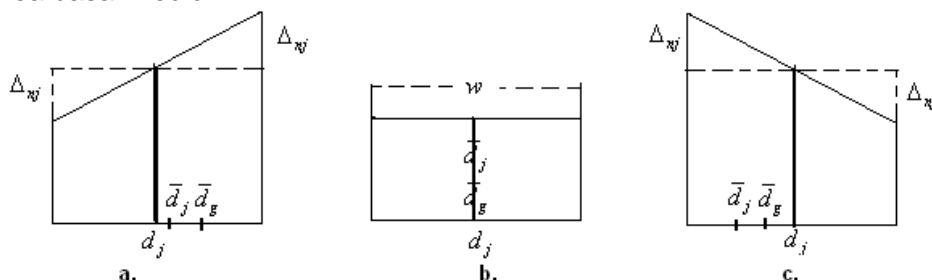


Figura 3.21 Diferencia entre los diferentes elementos diamétricos de clase.

En la Figura 3.21, adaptada del mismo autor, se representan tres intervalos de clase, una distribución creciente (a), una rectangular (b) y otra decreciente (c). Esta última, permite apreciar que el diámetro promedio de la clase, \bar{d}_j , es menor que el diámetro de área basal media \bar{d}_{gj} y ambos a su vez menores que la marca de clase d_j . El área basal de la clase se deriva generalmente con base en el diámetro marca de clase:

$$G'_j = \frac{\pi}{4} n_j d_j^2 \quad (3.39)$$

con n_j : árboles en la la clase j y d_j : marca de clase. Este valor resulta erróneo por cuanto la verdadera área basal de la clase se obtiene por:

$$G_j = \sum_{i=1}^{n_j} g_{ij} = \frac{\pi}{4} \sum_{i=1}^{n_j} d_{ij}^2 = \frac{\pi}{4} n_j \bar{d}_{gj}^2 \quad (3.40)$$

en que g_{ij} : área basal del árbol i de la clase j , d_{ij} : diámetro del árbol i de la clase j , \bar{d}_{gj} : diámetro correspondiente al área basal promedia de la clase j , obtenido como:

$$\bar{d}_{gj} = \sqrt{\frac{\sum_{i=1}^{n_j} d_{ij}^2}{n_j}} \quad (3.41)$$

El error sistemático debido al agrupamiento en clases es entonces:

$$E_{gj} = G'_j - G_j = \frac{\pi}{4} n_j (d_j^2 - \bar{d}_{gj}^2) \quad (3.42)$$

Esta diferencia $d_j - \bar{d}_{gj}$ entre el diámetro del centro de la clase y el diámetro del área basal media puede estimarse en dos fases, dado que generalmente no se cuenta con información individualizada al agrupar:

1) Determinación de la diferencia entre la marca de clase d_j y la media aritmética de la clase:

$$\Delta_j = d_j - \bar{d}_j \quad (3.43)$$

2) Determinación de la diferencia entre la media aritmética de la marca de clase \bar{d}_j y el diámetro del área basal media de clase \bar{d}_{gj} . Estas diferencias se basan en las del cálculo para centros de gravedad, por integración de diámetros cuadrados.

Con respecto a 1), es posible su obtención así:

$$\Delta'_{ji} = \frac{w^* \Delta_{nj}}{6^* n_j} \quad (3.44)$$

siendo el significado de Δ_{n_j} el que se aprecia en Figura 3.21c, el cual es positivo por lo que \bar{d}_j es menor que d_j y Δ_j es positivo, w es el ancho de la clase y n_j el número de árboles de dicha clase. Cuando ambos diámetros sean idénticos entonces $\Delta_{n_j} = 0$, pero si la distribución es creciente Δ_j asume un valor negativo pues Δ_{n_j} lo es y el diámetro medio

aritmético también será mayor que el de la marca de clase.

Si la distribución puede ser matemáticamente definida, también es posible derivar la diferencia Δ_j funcionalmente.

Por ejemplo, para una decreciente lineal se podría obtener con base en un error aceptable, como:

$$\Delta_j = d_j - \bar{d}_j = \frac{\varepsilon w^2}{12} = \Delta = \text{constante} \quad (3.45)$$

Con respecto a la fase 2), el diámetro del área basal media de la clase será:

$$\bar{d}_{g_j} = \sqrt{\bar{d}_j^2 + \frac{w^2}{12}} \quad (3.46)$$

3.4.4 Cálculos para obtener los diámetros corregidos. El Δ_j se halla de la ecuación (3.44) y

$$\bar{d}_j = d_j - \Delta_j \quad (3.47).$$

El Δ_{n_j} , el decremento fraccional del número de árboles, se calcula basado en las frecuencias de la clase anterior y la siguiente, con las expresiones:

$$\Delta_{n_j} = \frac{(n_{j-1} - n_{j+1})}{4} \quad (3.48)$$

para todas las clases, excepto para la primera y la última, donde:

$$\Delta_{n1} = \frac{n_1 - n_2}{2}; \quad y \quad \Delta_{nk} = \frac{n_k}{2} \quad (3.49);$$

al considerar k clases diamétricas (k = 5, 9, 15, ...). Estas relaciones para calcular los diámetros corregidos, se han resumido con las siguientes expresiones:

$$\left. \begin{aligned} d_j^* &= d_j - \frac{w^*(n_{j-1} - n_{j+1})}{24 * n_j}, \text{ para } j = (2, k - 1) \\ d_1^* &= d_1 - \frac{w(n_1 - n_2)}{12 * n_1}, \text{ para la primera clase} \\ d_k^* &= d_k - \frac{w}{12}, \text{ para la última clase} \end{aligned} \right\} \quad (3.50)$$

Obtenidos los diámetros corregidos (d_j^*), las fórmulas para la media y la varianza son:

$$\bar{d} = \frac{\sum_{j=1}^k d_j^* n_j}{\sum_{j=1}^k n_j}; \quad \sigma^2 = \frac{\sum_{j=1}^k n_j d_{gj}^2 - \frac{\left(\sum_{j=1}^k d_j n_j\right)^2}{\sum_{j=1}^k n_j}}{\sum_{j=1}^k n_j} \quad (3.51)$$

para lo cual los diámetros de área basal media se obtienen por (3.46).

3.4.5 Comparación de distribuciones. Las distribuciones diamétricas se vuelven indicadoras de gran utilidad para predecir o estimar aspectos estructurales del mismo bosque, o de varios entre sí, (fundamentalmente su estado actual y su desarrollo) y con buena experiencia, inferencias de manejo y situaciones silviculturales).

A manera de ejemplo se presentan los datos obtenidos en un rodal de eucalipto agrupados como muestran, la Tabla 3.1 y la Figura 3.22. Para efectos de comparación de distribuciones se acostumbra el uso de intervalos de clase, iguales y coincidentes pero dependiendo de los objetivos de la investigación se podría acudir a intervalos diferentes, para lo cual será preciso hacer transformaciones a clases iguales.

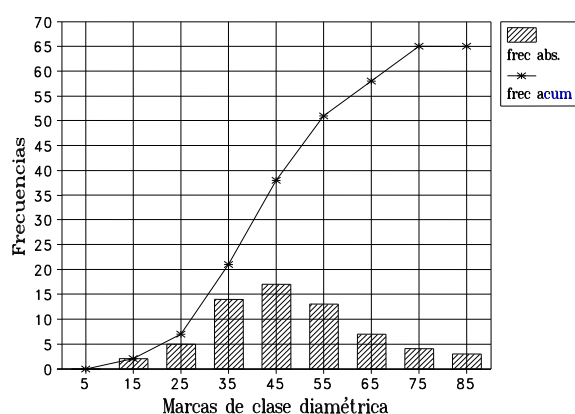


Figura 3.22. Distribuciones de frecuencia absoluta y acumulada de la tabla 3.1

El cálculo de ajuste de distribuciones no censuradas tiene: variable: DIAM, 65 valores agrupados entre 15.0 a 85.0, media = 48.24, desviación estándar = 16.3083. En R, el contraste de Shapiro Wilks, no permite rechazar H_0 :

```
shapiro.test(Frdap[-1])#Eliminamos la primera clase por tener frecuencia cero

      Shapiro-Wilk normality test

data:  Frdap[-1]
W = 0.8865, p-value = 0.217

par(mfrow=c(1,2))
plot(mdap[-1],Frdap[-1],"b",pch=20,main="Frecuencia      vs      dap",    cex.lab=1.5,cex.axis=1.5,
cex.main= 1.8,col="black")
qqnorm(Frdap[-1],main="QQ norm",cex.lab=1.5,cex.axis=1.5,cex.main=1.8)
qqline(Frdap[-1],col=2)# Gráficas pag siguiente.
```

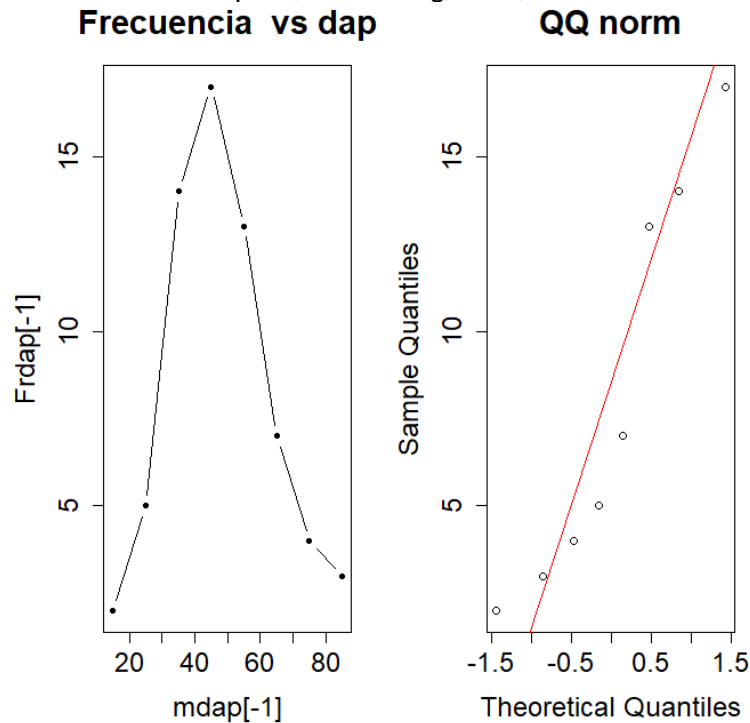
Prueba de bondad de ajuste de Kolmogorov-Smirnov, entre observados y estimados:

```
ks.test(Frdap[2:9], faes)

      Two-sample Kolmogorov-Smirnov test

data:  Frdap[2:9] and faes
D = 0.625, p-value = 0.08702
alternative hypothesis: two-sided
```


en ambas el $p\text{-value} > 0.05$, no permiten rechazar que d provenga de una Distribución Normal con un 95% de confianza pero, ks, más rigurosa, avala la anterior.



Otras pruebas: Sesgo y curtosis de dos colas para normalidad. Un código como el siguiente: que, si da valores mayores que 1 debe usar el complemento, no rechaza la normalidad para valores entre +2 y -2.

```
sesgo <- function (x) {
  m3 <- sum((x-mean(x))^3)/length(x)
  s3 <- sqrt(var(x))^3
  skew <- m3/s3
  ses <- sqrt(6/length(x))
  totest <- skew/ses
  pt(totest, (length(x)-1))
  pval <- 2* pt(totest, (length(x)-1))
  pval
}
sesgo(Frdap[-1])
[1] 1.158126
> 1-sesgo(Frdap[-1])
[1] -0.158126
```

Para curtosis:

```
curtosis<- function (x) {
  m4 <- sum((x-mean(x))^4)/length(x)
  s4 <- var(x)^2
  kurt <- (m4/s4) - 3
  sek <- sqrt(24/length(x))
  totest <- kurt/sek
  pvalue <- 2*pt(totest, (length(x)-1))
  pvalue
}
curtosis(Frdap[-1])
[1] 0.3599187
```

3.4.6. Tipos de distribuciones. Existen, desde las puras matemáticas hasta las gráficas, simuladas, crecientes, decrecientes, bi o polinomiales, y muchas más. Para el ejemplo

mencionado, los datos de la Tabla 3.1 son los dap/ha encontrados en una plantación de eucaliptos. Se presenta en primer lugar el procedimiento realizado con Excell, para una

prueba de bondad de ajuste con chi cuadrada $\chi_c^2 = \frac{\sum_{i=1}^k (f_o - \hat{f})^2}{\hat{f}}$.

TABLA 3.1. Ejemplo de datos y ajustes de Eucalipto para distribuciones diamétricas

mdap	Frdap	yxfr	fr*(y-m)^2	z con Ls	prob ac	probcl	freces	chi2
5	0	0	0					
15	2	30	2208.57	-1.7311	0.0417	0.0417	2.7118	0.1868
25	5	125	2698.34	-1.1179	0.1318	0.0901	5.8558	0.1251
35	14	490	2450.75	-0.5047	0.3069	0.1751	11.3800	0.6032
45	17	765	177.44	0.1085	0.5432	0.2363	15.3601	0.1751
55	13	715	595.69	0.7217	0.7648	0.2216	14.4012	0.1363
65	7	455	1968.45	1.3349	0.9090	0.1443	9.3786	0.6033
75	4	300	2866.37	1.9480	0.9743	0.0653	4.2417	0.0138
85	3	255	4055.93	2.5612	0.9948	0.0205	1.3319	2.0894

sumas 65.000 3135.000 17021.540

64.661

media(m) 48.231
var 265.962
desv 16.308

chi2ob 3.933
chi2t 14.067

La prueba χ^2 de bondad de ajuste no permite rechazar la H_0 , de que es una normal, ya que la calculada no supera la tabulada. Se resalta en este y todos los cálculos similares que se debe trabajar con los límites superiores de clase y con la distribución acumulada. En R:

```
euc<-read.table("clipboard")
attach(euc)
names(euc)
[1] "mdap" "Frdap"

euc
  mdap Frdap
1    5     0
2   15     2
3   25     5
4   35    14
5   45    17
6   55    13
7   65     7
8   75     4
9   85     3

su<-sum(Frdap)
su
[1] 65

medp<-mdap*Frdap/su #Cálculo de la media ponderada
medp
      [,1]
[1,] 48.23077

varp<-sum(Frdap*(mdap-medp)^2/(su-1)) #Cálculo de la varianza ponderada
varp
[1] 265.9615

desp<-sqrt(varp) # Desviación estándar ponderada
desp
```

```
[1] 16.30833

zsu<-(mdap+5-medp)/desp #Conversión a Z, debe hacerse con el límite superior de clase

zsu
[1] -2.3442483 -1.7310647 -1.1178810 -0.5046973  0.1084863  0.7216700  1.3348537 1.9480374
2.5612210

zsu[-1]#El primer valor era para una frecuencia cero.
[1] -1.7310647 -1.1178810 -0.5046973  0.1084863  0.7216700  1.3348537  1.9480374 2.5612210

probzu<-pnorm(zsu[-1])#Función R para el cálculo de las probabilidades normales acumuladas
probzu
[1] 0.04172012 0.13180891 0.30688572 0.54319504 0.76475131 0.90903788 0.97429475 0.99478475

faes<-probzu*su #Frecuencia acumulada estimada si los datos fueran normales
faes
[1]  2.711807  8.567579 19.947572 35.307678 49.708835 59.087462 63.329159 64.661009
```

3.4.6.1 Frecuencias por clase. Se construyó la siguiente función en R para el cálculo de las frecuencias por clase, que se reiterará en los cálculos de las demás distribuciones estudiadas:

```
faesx<-function(x){#cálculos de las frecuencias por cada clase
  n<- length(x)
  fres<- numeric(n)
  for (i in 1:n){
    fres[i]<- x[i+1]-x[i]
  }
  result <- c(x[1],fres)
  return(result)
}

fre<-faesx(faes)#Frecuencia esperada por clase diamétrica
fre
[1]  2.711807  5.855772 11.379993 15.360106 14.401157  9.378627  4.241697 1.331850 NA
```

Se calculan los elementos para la χ^2 observada

```
pchi<-(Frdap[-1]-fre)^2/fre
pchi
[1] 0.186838 0.125063 0.603202 0.175080 0.136325 0.603272 0.013772 2.089368 NA

chi2c<-sum(pchi[-1],na.rm=T)#Suma de los elementos anteriores
chi2c
[1] 3.746085
```

Se calcula la tabulada:

```
qchisq(0.95, 7, ncp = 0, lower.tail = T, log.p = FALSE)
[1] 14.06714
```

Resultado similar al obtenido en Excell.

3.4.6.2 Cálculo de parámetros y frecuencias estimadas Máximo Verosímiles (MLE).

Este método se hizo para estimar los parámetros de la distribución de probabilidad de una variable aleatoria X a partir de una muestra aleatoria simple de ella misma. Si θ es el vector de parámetros de la distribución y $f(x|\theta)$ su función de densidad, el MLE, dada la muestra $\{x_1, x_2, \dots, x_n\}$, es el valor que maximiza el log de verosimilitud (y minimiza AIC):

$$\ell(\theta) = \sum_{i=1}^n \log f(x_i | \theta).$$

R posee procesos de optimización para funciones, e implementa varias de ellas, en la librería **MASS** con lo cual para la MLE bastaría con hacer log-L a partir de la función de densidad, los datos, y optimizarla. Por ejemplo, si **X** sigue una distribución dada con parámetros **a** y **b**, se puede correr una función como la siguiente para calcular la verosimilitud logarítmica como:

```
logver=function(params,x) {
  a=params[1]
  b=params[2]
  lv=sum(log(dfunction(x, a, b)))
  return(lv)
}
```

La anterior puede maximizarse mediante la función `optim()`; con unos valores de partida de los parámetros; (calculados por momentos o, algunas veces a partir de la aproximación (a, b)=(1,1) arbitraria, como se mostrará al estudiar la función Weibull. Se debe notar que el proceso MV *requiere de los datos crudos o sea sin agrupamiento*. Para el ejemplo del eucalipto los construiremos en un *dataframe* que llamaremos *dicru* (diámetros crudos).

La función es *fitdistr* la cual viene implementada para distribuciones: *beta, cauchy, chi-cuadrada, exponencial, f, gamma, geometrica, log-normal, logística, binomial negativ, normal, Poisson, t y Weibull*, en que, para las siguientes distribuciones: *Normal, log-Normal, geométrica, exponencial y Poisson* los cálculos se basan en los errores estándar exactos, mientras para las demás deben aportarse valores de partida.

```
library(MASS)
dicru<-data.frame(d=c(rep(25,7),rep(35,14),rep(45,17),rep(55,13),rep(65,7),rep(75,4),rep(85,3))
)#Eliminamos la clase de frecuencia cero

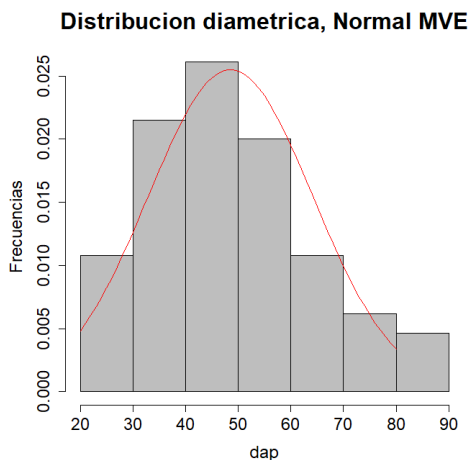
str(dicru)
'data.frame': 65 obs. of 1 variable:
 $ d: num 25 25 25 25 25 25 25 35 35 35 ...

euc.normv<-fitdistr(dicru$d,"normal")#La función fitdistr en MASS (revisarla), da los
parámetros de la función requerida, anotada entre comillas.
euc.normv
      mean      sd
48.538462 15.633373
( 1.939081) ( 1.371138)
```

3.4.6.3 Valores observados y estimados. Para ver el comportamiento de ambas frecuencias construimos el histograma de los datos y la curva ajustada por el modelo usado:

```
hist(dicru$d,breaks=6,freq=F,main="Distribucion diametrica, Normal MVE", xlab="dap", ylab =
"Frecuencias",col="gray",cex.lab=1.5,cex.axis=1.5,cex.main=1.5)
lines(20:80, dnorm(20:80,48.538462,15.633373),col="red")
```

La gráfica obtenida nos ayuda, aún más, a ver el comportamiento de los datos con respecto a la normal. No parece un mal ajuste. Pero se insiste, es solo una visualización gráfica que requiere otros análisis.



3.4.6.4 Distribuciones unimodales. Son características de poblaciones evaluadas bajo un cierto parámetro homogéneo, especialmente la edad. Un rodal coetáneo tiende a la unimodalidad. En términos generales, esta condición es señal de una regeneración en ciclos y no en forma continua, aunque se dan situaciones cambiantes de la distribución en la vida del rodal hasta presentarse condiciones de equilibrio como sucede en etapas de cultivo en las cuales los diámetros se agrupan más aleatoriamente alrededor de la media, pero pronto factores estimulantes positivos o negativos del crecimiento como suelo, micrositio, enfermedades y plagas, disturban el balance y se dan renuevos de mejores diámetros con ventajas aleatorias para alcanzar un incremento sistemático y producir una curva sesgada a la izquierda. Luego aparece la competencia y se acentúa el sesgo hasta el grado máximo de asimetría. Figura 3.24.

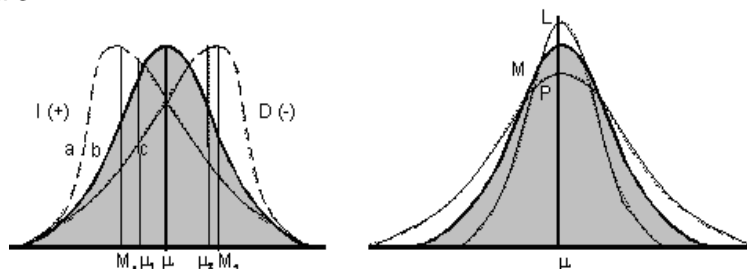


Figura 3.24. Comparación de distribuciones unimodales con la normal, asimetría y sesgos.

En rodales viejos, o por aclareos, el número de individuos desaparece por mortalidad hasta otro máximo, al hacerlo también los suprimidos, dejando diámetros mayores, hasta que las posibilidades de crecimiento tienden a igualarse y se reduce la asimetría. La forma de distribuciones cambia entonces con el tipo e intensidad de los manejos, especialmente las cortas, lo que da lugar a diversas configuraciones unimodales (Loetsch et al., 1964) (Figura 3.24).

3.4.6.4.1 Escogencia del modelo. Este es el primer paso para el ajuste de las distribuciones: escoger el modelo matemático o la función que represente unos datos lo mejor posible, con base en algunas hipótesis acerca de la naturaleza de ellos, para lo cual, como se ha reiterado varias veces, se acude a diversas técnicas gráficas como los histogramas y curvas teóricas y ajustadas, etc. que, pueden resultar bastante subjetivas, por lo cual, existen expresiones analíticas como el criterio K de Pearson y otros que al resolver

una ecuación diferencial particular permite acercarse a varias familias de distribuciones empíricas para un ajuste adecuado, que dependen de la media, variabilidad, sesgo y curtosis.

Sí se *estandarizan* los datos el tipo de curva dependerá únicamente del sesgo y la curtosis

de acuerdo con la siguiente formula: $K = \frac{\gamma_1^2 (\gamma_2 + 6)^2}{4(4\gamma_2 - 3\gamma_1^2 + 12)(2\gamma_2 - \gamma_1^2)}$, en la cual

$r_1 = \frac{\sum_{i=1}^n (x_i - \mu)^3}{n\sigma^3}$ es conocido como coeficiente estandarizado de sesgo de Pearson y

$r_2 = \frac{\sum_{i=1}^n (x_i - \mu)^4}{n\sigma^4} - 3$ coeficiente estandarizado de kurtosis de Pearson.

3.4.6.4.2 Caracterización de las distribuciones unimodales. Como patrón de comparación de las distribuciones unimodales se acude a la normal ya que esta queda inequívocamente definida por la media, y la varianza. Media, mediana y moda coinciden en ella. Al romperse esta coincidencia aparecen la asimetría y la kurtosis.

3.4.6.4.2.1 Asimetría de una distribución. La asimetría refleja la forma geométrica de la función de densidad y está ligada a la posición de la mediana con respecto a la media y la moda. Por convención, cuando la mediana (M) se localiza a la izquierda de la media (μ) se dice que la distribución es sesgada a la izquierda, o que es asimétrica positiva, Figura 3.24. Asimetría negativa o sesgada a la derecha cuando M se coloca a la derecha de μ , pero depende de los autores (algunos lo expresan al contrario) y es importante especificar el sentido. Como a veces no resulta fácil apreciarlo se acude a la estadística, con base en los momentos con respecto a la media, de orden superior, de la distribución definidos por m_k

$$m_k = \frac{\sum_{i=1}^N (X_i - \mu)^k}{N} \quad (3.53)$$

m denota el momento central y k el orden, que dieron lugar al γ_1 y γ_2 mencionados en el numeral anterior. También se acostumbra el Coeficiente de Pearson definido como:

$$CP = \frac{3(\bar{X} - Me)}{s} \quad (3.54)$$

en que Me y s son la mediana y la desviación estándar. El CP, tiene en cuenta que la mediana es insensible a los cambios de los valores de los extremos, cuando la media es arrastrada fácilmente en la dirección de la asimetría. Este coeficiente se interpreta como r_1 el cual se interpreta así: Si $r_1 = 0$ la distribución es simétrica; si $r_1 < 0$ negativa asimétrica y, si $r_1 > 0$ positiva asimétrica. r_1 varía así: $-1 \leq r_1 \leq +1$. Para el ejemplo de la Tabla 3.1: $\bar{X} = 48,231$, $s = 16,308$, y $Me = 42,2$, $r_1 = 0,314$ lo que no conduce a grandes desviaciones con respecto a la normal.

Con R obtuvimos:

```
mom3<-sum(Frdap*(mdap-mu)^3)/65
```

```

mom3
[1] 1201.784

r1<-mom3/(sd^3)
> r1
      sd
0.3145346

cp<-3*(mu-median(dicru$d))/sd
> cp
[1] 0.6737772

```

Los anteriores se encuentran también en la:
library(moments)

```

skewness(dicru$d, na.rm = FALSE)
[1] 0.5133097

```

Para efectos prácticos una curva con r_1 entre $\pm 0,37$ se puede considerar simétrica, el CP es mucho más estricto y oscila así: $-1 \leq \text{simetría} \leq +1$.

3.4.6.4.2.2 Kurtosis. Refleja el grado de empinamiento de la curva y evalúa el exceso o defecto por encima de la media teórica. Si la curva presenta un defecto por aplanamiento, con respecto a un patrón normal llamado mesocúrtico, se llama platicúrtica (plati = plano). Si la curva resulta más puntiaguda que la normal, se llama leptocúrtica (*leptos* = aguja) la cual tiene como patrón un valor estandarizado aproximado de 3. La kurtosis se evalúa con el coeficiente estandarizado de kurtosis de Pearson, ya mencionado:

$$r_2 = \frac{m_4}{\sigma^4} - 3 \quad (3.55)$$

De donde salió el r_2 del numeral anterior. Tiene exceso positivo si $r_2 > 0$ y negativo si $r_2 < 0$ y varía así: $-1 \leq r_2 \leq +1$. Para el ejemplo $r_2 = -0,21$ (platicúrtica). Si las clases se escogen muy amplias y los datos se someten a fuerte redondeo, los índices de asimetría y kurtosis pueden dar valores falsos por signo y magnitud, en cuyo caso es bueno trabajar entre 15 y 20 clases (Loetsch et al, 1973). Con datos de R:

```

mom4<-sum(Frdap*(mdap-mu)^4)/65
> mom4
[1] 189823.4

r2<-mom4/sd^4
r2
      sd
3.17789
kurtosis(dicru$d, na.rm = FALSE)
[1] 2.684276

```

Las diferencias se deben a que unos cálculos se hicieron con los datos crudos y otros agrupados. Sin mucho alejamiento de la estructura normal. Para probarlo se hace la prueba de bondad de ajuste de la χ^2

```

su<-sum(Frdap)
su
[1] 65

pno<-pnorm(d(lsc1),mu,sd)#probabilidades acumuladas

pno<-pnorm(mdap[-1]+5,mu,sd)#probabilidades acumuladas límite superior de clase

```

```

pno
[1] 0.03396407 0.11784550 0.29247494 0.53724220 0.76826517 0.91509317 0.97791426 0.99600056

faes<-su*pno #Frecuencias acumuladas esperadas o estimadas bajo normalidad
faes
[1] 2.207664 7.659958 19.010871 34.920743 49.937236 59.481056 63.564427
[8] 64.740036

fres<-faesx(faes)#se usa la función ya conocida
fres
[1] 2.207664 5.452293 11.350914 15.909872 15.016493 9.543819 4.083371 1.175609 NA

pchi<-sum((Frdap[-1]-fres[-9])^2/fres[-9]) #Prueba chi2 calculada valores de campo y estimados

pchi
[1] 4.531728

qchisq(0.95, 7, ncp = 0, lower.tail = T, log.p = FALSE)
[1] 14.06714

```

Efectivamente también se comportan los datos crudos como una normal.

3.4.7 Distribuciones multimodales. Ocurren tanto para rodales simples como para grandes áreas, pudiendo indicar presencia de poblaciones mezcladas, aunque estas no sean necesariamente multimodales. Entre las causas reportadas por (Loetsch et al, 1974) están: la mezcla de especies de diferentes ritmos de crecimiento, diferencias de edad, competencia entre individuos que permiten formar doseles: primario y secundario, excesivas cortas en ciertas clases diamétricas para satisfacer demandas específicas y condiciones de sitio extremadamente diferentes. No hay mucha claridad en el criterio de la multimodalidad, pero el manejo de las aparentes subpoblaciones puede eliminarse al separarlas en poblaciones normales con su propia estructura, historia silvicultural y dinamismo.

3.5 Funciones de distribución.

Tan importante como calcular los parámetros que las caracterizan, con respecto a la normal, es conocer la función matemática que describa la distribución, pues es posible llegar a definir, por ecuaciones, el desarrollo de las distribuciones diamétricas con el incremento de la edad por medio de los modelos reales o simulados (Loetsch et al., 1964). Veremos algunas de las más importantes.

3.5.1 Distribución Gamma $\Gamma(\alpha)$. La distribución gamma toma su nombre de la función gamma de Euler, de forma:

$$\Gamma(\alpha) = \int_0^{\infty} x^{(\alpha-1)} e^{(-x)} dx, \text{ para } \alpha > 0; x \geq 0 \quad (3.56)$$

al integrarla por partes con

$$\mu = x^{(\alpha-1)}; \quad dv = e^{-x} dx \quad (3.57)$$

se llega al resultado que permite evaluarla:

$$\Gamma(\alpha) = (\alpha-1) \int_0^{\infty} x^{(\alpha-2)} e^{(-x)} dx = (\alpha-1) \Gamma(\alpha-1) \quad (3.58)$$

de tal forma que:

$$\Gamma(n) = (n-1)(n-2) \cdots \Gamma(1) \approx (n-1)! \quad (3.59),$$

para n = entero positivo. Además,

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (3.60)$$

La distribución gamma se considera especial para describir variables aleatorias limitadas en un extremo o que varían en un intervalo dado. Su función de densidad está dada en dos versiones a y b, por varios autores, como:

$$f(X; \alpha, \beta) = \begin{cases} a) \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\beta x}, & x \geq 0; \alpha > 0, \beta > 0; \\ b) \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{(\alpha-1)} e^{-\frac{x}{\beta}}, & x \geq 0; \alpha > 0, \beta > 0; \\ 0 & \text{en otros casos.} \end{cases} \quad (3.61)$$

Las gráficas de la Figura (3.24) muestran los cambios sufridos por la función al cambiar los parámetros de forma y/o de escala, dejando constante el otro.

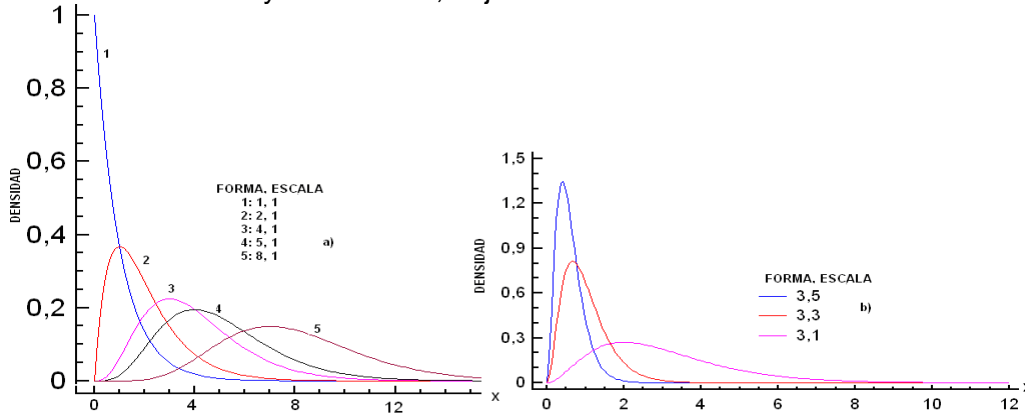


Figura 3.24 Diversas funciones gamma con cambios de forma a) y escala b).

Se obtiene una gran variedad de formas: en general si $\alpha \leq 1$ la curva es singular, decreciente, y cuando $\alpha > 1$ presenta un punto de máxima en $x = (n-1)/\beta$. Al variar β no cambia la forma de la distribución, pero si su escala, por eso se conoce a α como parámetro de forma y a β como de escala. Esta distribución es apropiada para modelar el tiempo requerido para que un total de exactamente α eventos independientes tengan lugar si ellos ocurren a una tasa constante β .

Lo anterior sugiere numerosas aplicaciones, *por ejemplo* si un conjunto de elementos se ordenan en lotes de tamaño α y, una demanda para las partes individuales surge independientemente a una tasa constante β por semana, el tiempo entre los agotamientos de los lotes se comporta como una variada gamma.

Igual podría adaptarse al comportamiento diamétrico de árboles ordenados en grupos de tamaño α , en que se cambia su frecuencia tan pronto se alcancen ciertos valores del DAP y esto ocurriera a una tasa constante β . Muchos fenómenos como el anterior no pueden justificarse teóricamente como variadas gamma, pero si adaptarse o sustentarse empíricamente por este modelo, si se validan con pruebas de bondad de ajuste, manteniendo constante la forma.

La gamma se usa frecuentemente en análisis bayesiano, como un modelo previo para

descubrir la incertidumbre inicial concerniente a la rata de ocurrencia de algún proceso, como el tránsito de un árbol de una a otra categoría diamétrica. Pero es la distribución acumulada la que permite el cálculo de probabilidades.

3.5.1.1 Función de distribución acumulada. Esta se da, para la versión a) por la función:

$$F(X \leq x) = F(X; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x t^{(\alpha-1)} e^{-\beta t} dt; \quad x \geq 0 \quad (3.62)$$

conocida como función gamma incompleta, tabulada por ejemplo en (Harter, 1964) y (Pearson, 1957), las cuales requieren, para su uso, los cálculos de los valores:

$$\mu = \frac{\beta x}{\sqrt{\alpha}} \quad y \quad p = (\alpha - 1) \quad (3.63)$$

en los cuales μ es el transformado de x , que se introduce en $I(\mu, p)$ de las tablas o en algún programa estadístico, como se verá luego, y arrojan la probabilidad de que un valor aleatorio de una Γ con parámetros α y β tome un valor de $X \leq x$, o sea que constituye la probabilidad acumulada $F(X; \alpha, \beta)$.

Ejemplo. Una lancha maderera cruza un río tan pronto es cargada con 9 rastras de madera (lote de tamaño 9). Durante cierto tiempo, el promedio de rastras, transportado por mulas, llegado por hora al embarcadero es de 6. Calcular la probabilidad de que el tiempo entre dos embarques consecutivos sea menor a una hora. De acuerdo con la definición dada, entonces, el tiempo entre las salidas de la lancha es una variada gamma, con $\alpha = 9$ y $\beta = 6$, entonces, con $\mu = 6 \times 1 / \sqrt{9} = 6/3 = 2.0$, y $p = \alpha - 1 = 8$, se trata de encontrar (3.62), o $F(X \leq 1)$, que coincide con $I(2.0, 8) = F(1, 9, 6) = 0.153$, una baja probabilidad de que el tiempo entre dos embarques consecutivos sea menor a una hora.

Esto mismo se puede facilitar en paquetes estadísticos, por ejemplo en Excel (usa versión b) de la ecuación (3.61 o 3.64) se acude a la $F(1, 9, 1/6)$: Con la distribución gamma acumulativa del EXCEL en función gamma se llega al mismo resultado, pero tener la precaución, de trabajar $scale = 1/B$ constante, ya que se trabaja con el modelo a) (3.46), que lo sigue llamando beta, en este caso $escala = 1/6$. Al introducir los datos este valor, va en la casilla que requiere beta.

X	alfa	beta	escala	proac
0	9	6	0.167	0
1	9	6	0.167	0.152
2	9	6	0.167	0.843

En R, se usa la pgamma de la librería MASS:

```
pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE, log.p = FALSE) #si se da rate,
no se da scale y viceversa
```

```
probz<-pnorm(z)
[1] 0.1527625
```

3.5.1.2 Parámetros y aproximación a la normal de la distribución gamma. Desde la formulación de la distribución gamma como un modelo para el tiempo requerido para que exactamente α eventos sucedan asumiendo una rata constante β , se puede usar el

teorema del límite central para ver la aproximación de la gamma a medida que α se incrementa. Con conceptos de esperanza se tienen para ambas versiones a) y b):

$$a) E(X) = \frac{\alpha}{\beta}; \quad \sigma^2(X) = \frac{\alpha}{\beta^2}; \quad b) E(X) = \bar{X} = \alpha\beta, \quad \sigma^2(X) = \alpha\beta^2 \quad \left\} (3.64)\right.$$

X será una variable estandarizada normal cuando α sea grande, por ello

$$F\left(1; \mu = \frac{\alpha}{\beta}; \sigma^2 = \frac{\alpha}{\beta^2}\right) = \int_{-\infty}^1 \frac{\beta}{\sqrt{2\pi\alpha}} \exp\left[-\frac{\beta^2\left(z - \frac{\alpha}{\beta}\right)^2}{2\alpha}\right] dz \quad (3.65)$$

La estandarizada para grandes valores de α , en el modelo a) es

$$Z = \frac{(x - \alpha / \beta)}{(\sqrt{\alpha} / \beta)}; \text{ en el ejemplo: } \rightarrow z = \frac{1 - 9/6}{\sqrt{9/6}} = -1 \rightarrow \text{Pr} = 0.159.$$

En R, se usa la pgamma de la librería MASS:

```
pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE, log.p = FALSE)#si se da rate,
no se da scale y viceversa
```

```
probbz<-pnorm(z)
[1] 0.1527625
```

3.5.1.3 Estimación de los parámetros. Frecuentemente uno o ambos parámetros deben ser estimados para datos observados. La mejor estimación es la de máxima verosimilitud, pero para muchos de los problemas, como los abordados en la dasometría, es suficiente el método de los momentos, especialmente cuando unos datos originales se difundieron en clases:

$$\hat{\beta} = \frac{\bar{x} (n - 1)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\bar{x}}{s^2} \text{ o también } \hat{\beta} = (n - 1) \left\{ \frac{\sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \right\} \quad (3.66)$$

El estimado de:

$$\alpha = \hat{\alpha} = \frac{\bar{x}^2 (n-1)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\bar{x}^2}{s^2} \therefore \hat{\alpha} = \beta \bar{x} \quad (3.67)$$

En muchas aplicaciones α es conocido por experiencia y solo se precisa estimar β . Para la opción b) trabajar con (364).

A menudo los valores exactos de todos los datos, más allá de un límite, son desconocidos, porque no se midieron o fueron removidos de la muestra, por lo cual se conocen como censuradas y tienen formas especiales de estimación (Wilk, et. al, 1962), que, no deben confundirse con las funciones truncadas, las cuales terminaron en el punto de truncamiento, y no existen muestras más allá de este. En las censuradas existen, pero se desconocen. En todos estos procesos de momentos, es importante la clase anterior con frecuencia cero. En el ejemplo de eucalipto, ajustar con EXCEL la distribución gamma a sus datos

mdap	Frdap	yxfr	fr*(y-m)^2	pgrammac	probcl	freces	chi2
5	0	0	0	0.0002	0.0002	0.0115	0.0115
15	2	30	2208.57	0.0157	0.0155	1.0098	0.9709
25	5	125	2698.34	0.1181	0.1024	6.6578	0.4128
35	14	490	2450.75	0.3365	0.2184	14.1947	0.0027
45	17	765	177.44	0.5870	0.2505	16.2817	0.0317
55	13	715	595.69	0.7826	0.1956	12.7138	0.0064
65	7	455	1968.45	0.9004	0.1178	7.6563	0.0563
75	4	300	2866.37	0.9592	0.0589	3.8255	0.0080
85	3	255	4055.93	0.9848	0.0256	1.6611	1.0791

sumas	65.000	3135.000	17021.540			64.012	
						chi2ob	2.579
						chi2t	14.067
media(m)	48.231						
var	265.962			beta=	5.514		
desv	16.308			alfa=	8.746		

Por la prueba de bondad de ajuste, no se rechaza H_0

3.5.1.4 Forma generalizada de la distribución gamma. Una variada gamma se asume que toma valores entre 0 e ∞ . Sin embargo es posible definirla también a partir de un valor particular, por ejemplo entre a e ∞ , lo que posibilita la gamma triparamétrica:

$$f(X; \alpha, \beta, a) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} (x-a)^{(\alpha-1)} e^{-\beta(x-a)}, & x \geq a; \quad a \geq 0; \alpha > 0; \beta > 0; \\ 0 & \text{en otros casos.} \end{cases} \quad (3.68)$$

cuya función de distribución acumulativa puede obtenerse haciendo $x' = (x-a)$ y proceder como si fuera la gamma biparamétrica.

3.5.1.5 Casos especiales de la función gamma. En la definición y fórmula (3.56), $\alpha > 0$, y además por la definición del tiempo requerido para que α eventos ocurran, implica que esta α debiera ser entero positivo. Cuando esto se asume, la función gamma se conoce como la distribución de ERLANG, muy útil en la teoría de colas, en aprovechamientos forestales.

La función χ^2 es un caso especial de la gamma, cuando $\beta = \frac{1}{2}$ y α es múltiplo de 1/2 quedando entonces como único parámetro el entero $\gamma = 2\alpha$, referido como sus grados de libertad.

La aplicación de esta distribución surge del escenario en que la suma de cuadrados de γ observaciones de una distribución normal estandarizada tiene una distribución χ^2 con γ grados de libertad, o sea coincide con $\chi^2 = Z_1^2 + Z_2^2 + \dots + Z_r^2$.

Otro ejemplo. Se va a usar como archivo de prueba la siguiente información salida de un bosque natural, recompuesto para evitar menos de 5 individuos por clase.

:

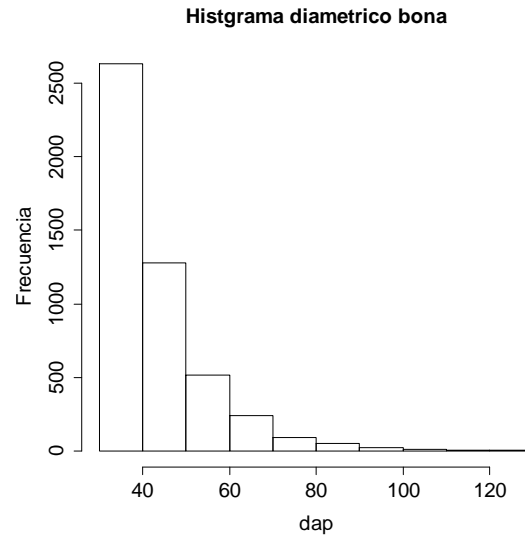
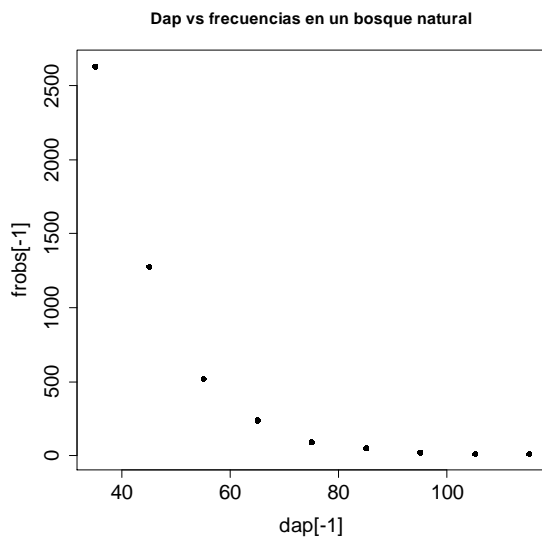
cladia	frobs
35	2634
45	1281
55	518
65	240
75	93
85	50
95	25
105	12
115	7
125	4

cladia	frobs
35	2634
45	1281
55	518
65	240
75	93
85	50
95	25
105	12
115	11

Inicialmente es deseable una representación gráfica simultánea de la distribución de frecuencias observadas en que se ve se trata de una jota invertida

```
plot(frobs[-1]~dap[-1], data=bona, main="Dap vs frecuencias en un bosque natural", col = "black",
pch=20,cex.axis=1.5,cex.axis=1.5,cex.lab=1.5)
```

```
hist(bn.d$d,breaks=11, freq=TRUE, main="Histgrama diametrico bona", xlab="dap", ylab=
"Frecuencia",cex.axis=1.5,cex.lab=1.5,cex.main=1.5)
```



Evaluada en Excell Tabla 3.2:

cladia	frobs		cladia	frobs	a*b	b*(d-u)^2	pgaac	pgacl	fres	chi2
35	2634	1	35	2634	92190	176446	0.42878	0.429	2086	144.2
45	1281	2	45	1281	57645	4221.66	0.73662	0.308	1497	31.26
55	518	3	55	518	28490	72314.4	0.91056	0.174	846	127.2
65	240	4	65	240	15600	114219	0.97626	0.066	319.6	19.82
75	93	5	75	93	6975	94136.3	0.99483	0.019	90.28	0.082
85	50	6	85	50	4250	87426.3	0.99904	0.004	20.49	42.52
95	25	7	95	25	2375	67120.8	0.99984	8E-04	3.919	113.4
105	12	8	105	12	1260	45853.7	0.99998	1E-04	0.654	196.9
115	7	9	115	11	1265	36102.1	1	2E-05	0.098	488.2
sumas			suma	4864	210050	697840		1	chi2c=	1164
media	43.18		var	143.5						
beta=	3.323		alfa	12.996				chist(k-p-1)		12.59

Para nuestro caso, según el método de los momentos (3.67)

$$\sigma^2(X) = \frac{\alpha}{\beta^2} \therefore \beta = \frac{\bar{x}}{\sigma^2(X)} = 0.2923; \quad \bar{x} = \frac{\alpha}{\beta} \therefore \alpha = 43.184 * 0.2923 = 12.63; \text{scale} = \frac{1}{\beta}$$

3.5.1.6 Cálculo de parámetros y frecuencias estimadas. Aparte de los ya mostrados en Excell, en R, hay una implementación de la gamma que varía con respecto a la del Excell, pues en ella los parámetros se relacionan con los estadísticos así: para $x \geq 0$, $a > 0$ (a o α) y $s > 0$, (s: escala), la $E(X) = a*s$ y la $Var(X) = a*s^2$, de donde $s = Var(X)/E(x)$ y $\alpha = E(x)/s$, o sea la versión 1 de (3.61). Una vez recompuesto los datos y el uso de scale, tabla 3.2, se ajusta por momentos así:

```
bona<-read.table("clipboard")
attach(bona)
names(bona)
[1] "cladia" "frobs"

su<-sum(frobs)#
su
[1] 4864

medp<-cladia%*%frobs/su#media ponderada del dap
medp
      [,1]
[1,] 43.18462

varp<-sum(frobs*(cladia-medp)^2/(su-1))#Varianza ponderada
varp
[1] 147.7422

desp<-sqrt(varp)
desp
[1] 12.15492

escala<-varp/medp
escala
      [,1]
[1,] 3.421176

alfa<-medp/escala
alfa
      [,1]
[1,] 12.62274

probg<-pgamma(cladia+5, alfa,scale=escala, lower.tail = TRUE, log.p = FALSE)#Proporciones
gamma acumuladas
probg
[1] 0.4308711 0.7345025 0.9078092 0.9747670 0.9942869 0.9988895 0.9998094
[8] 0.9999705 0.9999958

faes<-probg*su#Frecuencia acumulada esperada gamma
faes
[1] 2095.757 3572.620 4415.584 4741.267 4836.211 4858.599 4863.073 4863.857
[9] 4863.980

fes<-faesx(faes)
fes
      [10] 2095.7569695 1476.8629475 842.9641120 325.6825293 94.9448861
      [6] 22.3873102 4.4743713 0.7834194 0.1230968 NA

pchi<-sum((frobs-fes[-10])^2/fes[-10]) #Prueba chi2 calculada valores de campo
> pchi
[1] 1561.964

chi2t<-qchisq(0.95, 6, ncp = 0, lower.tail = T, log.p = FALSE)
```

```
> chi2t
[1] 12.59159
```

Nuevamente, como se ve en la tabla 3.1 el ajuste no sería bueno, según la prueba de χ^2 .

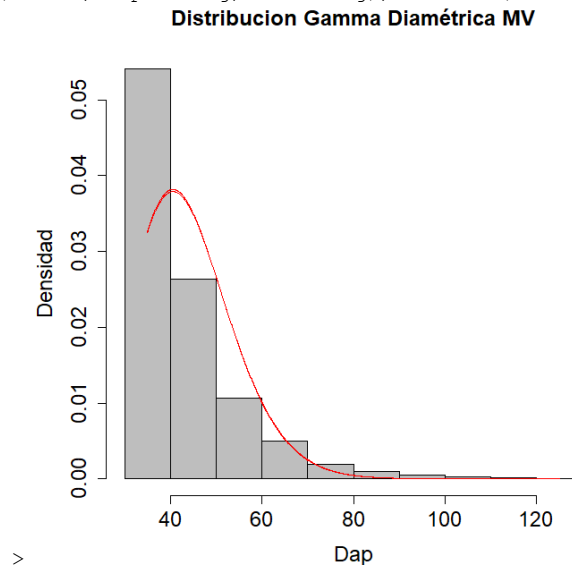
Estimados MLE para la función gamma. De nuevo acudimos a R

```
d<-cladia
f<-frobs
bn.d<-data.frame(d=c(rep(35,2634),rep(45,1281),rep(55,518),rep(65,240),rep(75,93), rep(85,50),
rep(95,25), rep(105,12), rep(115,7),rep(125,4)))

gammv<-fitdistr(bn.d$d,"gamma")
gammv
      shape      rate
16.207792700 0.375242689
( 0.325304002) ( 0.007649045)

alfag<-gammv$estimate[1]#parametro alfa
ratag<-gammv$estimate[2]#Parametro rata

hist(bn.d$d, breaks=11,freq=FALSE,main="Distribucion Gamma Diamétrica MV", xlab="Dap", ylab=
"Densidad",col="gray")
lines (35:125, dgamma(35:125,shape=alfag,rate=ratag),col="red")#350:1250
```



Prueba de bondad de ajuste de los estimados MLE del bona

```
probgmv<-pgamma(d+5, shape=alfag,rate=ratag, lower.tail = TRUE, log.p = FALSE)#Proporciones
gamma acumuladas

probgmv
[1] 0.4122246 0.7535796 0.9306096 0.9858443 0.9977651 0.9997127 0.9999687
[8] 0.9999970 0.9999998

su<-length(bn.d$d)
su
[1] 4864

faes<-su*probgmv
faes
[1] 2005.061 3665.411 4526.485 4795.147 4853.129 4862.602 4863.848 4863.986
[9] 4863.999
```

```

fre<-faesx(faes)
fre
[1] 2005.0605713 1660.3507906 861.0735435 268.6619539 57.9825686
[6] 9.4730000 1.2455423 0.1376245 0.0131896 NA

chi2c<-sum((frops-fre[-10])^2/fre[-10])
chi2c
[1] 11245.64

chi2t<-qchisq(0.95, 6, ncp = 0, lower.tail = T, log.p = FALSE)
chi2t
[1] 12.59159

```

Entonces coincide con el método de los momentos, se rechaza que los datos se comporten como gamma.

3.5.2 Distribución Beta, (β). Otra distribución importante es la β de Euler que, por su simpleza y la cantidad de formas que puede asumir, resulta muy adaptable a la mayoría de las distribuciones diamétricas así como para representar variables físicas en intervalos de longitud finita (Canavos, 1987).

Se dice que una variable aleatoria X posee una distribución beta si su función de densidad de probabilidades se da por la ecuación:

$$f(x, \alpha, \gamma) = \begin{cases} \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)\Gamma(\gamma)} x^{(\alpha-1)} (1-x)^{(\gamma-1)}, \\ 0 \leq x \leq 1; \alpha, \gamma > 0; \\ 0, \text{ para cualquier otro valor de } x \end{cases} \quad (3.69)$$

Los valores α y γ son parámetros de forma y sus cambios producen diversos perfiles:

- si $\alpha < 1$ y $\gamma < 1$ la función tiene forma de U,
- si $\alpha < 1$ y $\gamma \geq 1$ se comporta como una J inversa,
- si $\alpha \geq 1$ y $\gamma < 1$ el perfil es una J,
- cuando α y $\gamma > 1$, ambos, la función es unimodal con un punto de máxima en

$$x = \frac{\alpha - 1}{\alpha + \gamma - 2} \quad (3.70).$$

- cuando $\alpha = \gamma$ la distribución es simétrica. El nombre proviene de su asociación con la función beta definida como:

$$f(\alpha, \gamma) = \int_0^1 x^{(\alpha-1)} (1-x)^{(\gamma-1)} dx \quad (3.71)$$

relacionada con la gamma por la expresión:

$$\beta(\alpha, \gamma) = \frac{\Gamma(\alpha)\Gamma(\gamma)}{\Gamma(\alpha + \gamma)} \quad (3.72).$$

3.5.2.1 Función de distribución acumulativa. Se define por la ecuación(3.73), Conocida como función beta incompleta, tabulada entre otros por (Pearson, 1948) quien le define criterios para su uso (está en muchos paquetes estadísticos), para condiciones similares, resaltando siempre que x estará entre 0 y 1, para lo cual para nuestros datos convencionales se acude a (3.79)

:

$$P(X \leq x) = F(x, \alpha, \gamma) = \begin{cases} 0 & \text{para } x \leq 0 \\ \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \int_0^x t^{(\alpha-1)} (1-t)^{(\gamma-1)} dt; & 0 < x < 1; \\ 1 & \text{para } x \geq 1 \end{cases} \quad (3.73)$$

Cuando la suma de α y γ es relativamente grande, la ecuación (3.73) se puede aproximar de manera adecuada por la diferencia de dos funciones de distribución normalizadas con correcciones para continuidad, así:

$$F(x, \alpha, \gamma) \approx F_n(Z_u; 0, 1) - F_n(Z_\ell; 0, 1) \quad (3.74)$$

en que:

$$Z_u = \frac{[\gamma] - 0.5 - (\alpha + \gamma - 1)(1 - x)}{\sqrt{(\alpha + \gamma - 1)(x)(1 - x)}}; \quad Z_\ell = \frac{(\alpha + \gamma - 1)(1 - x) + 0.5}{\sqrt{(\alpha + \gamma - 1)(x)(1 - x)}} \quad (3.75)$$

y $[\gamma]$ denota el entero más grande que no exceda a γ , descritas por la función beta.

3.5.2.2 Cálculos para la estimación de los parámetros. Los estimadores máximo verosímiles de la beta son los ideales pero, para distribuciones diamétricas, a veces, basta el método de los momentos con una pequeña pérdida de precisión, basado en las siguientes ecuaciones:

$$\hat{\gamma} = \frac{(1 - \bar{x})}{s^2} [\bar{x}(1 - \bar{x}) - s^2]; \quad \hat{\alpha} = \frac{\bar{x}\hat{\gamma}}{1 - \bar{x}} \quad (3.76)$$

y, para mejorar su estimación se recomienda agrupar los diámetros en clases y reajustar los límites a y b , primero y último de la distribución ajustada así:

$$a = x_1 - \frac{w}{2}; \quad b = x_k + \frac{w}{2} \quad (3.77),$$

x_1 = marca de clase de la primera clase diamétrica; x_k = marca de clase de la última clase diamétrica, w = amplitud de clase. Como generalmente las últimas clases resultan subestimadas por las distribuciones β , se recomienda una expansión del límite superior para b , así: $b = b + (b - a)/4$ como máximo.

3.5.2.2.1 Comparación del modelo estimado con los datos observados. Los métodos de momentos se basan en la tabla de frecuencias, los MV, en todas las observaciones crudas. Usaremos el archivo de bosque natural ya propuesto:

```
bona<-read.table("clipboard")
attach (bona)
names(bona)
[1] "d"    "fro"

su<-sum(fro)
su
[1] 4864

mu<-fro**d/su
mu
      [,1]
[1,] 43.18462

var<-sum(fro**d*(d-mu)^2/(su-1))
var
[1] 147.7422
```

```

a<-d[1]-5
a
[1] 30

b<-d[length(d)]+5
b
[1] 120

xpri<-(d-a)/(b-a)
xpri
[1] 0.05555556 0.16666667 0.27777778 0.38888889 0.50000000 0.61111111 0.72222222
[8] 0.83333333 0.94444444

mur<-fro%%xpri/su #Media relativizada
mur
      [,1]
[1,] 0.1464958

varr<-sum(fro%%(xpri-mur)^2/(su-1))#varianza relativizada
varr
[1] 0.01823978

gar<-((1-mur)/varr)*(mur*(1-mur)-varr)#parametro gamma relativizado
gar
      [,1]
[1,] 4.99732

alr<-mur*gar/(1-mur) #parametro alfa relativizado
alr
      [,1]
[1,] 0.8577419

```

Datos crudos

```

bnocr<-data.frame(dp=c(rep(xpri[1],2634),rep(xpri[2],1281),rep(xpri[3],518),rep(xpri[4],240),
rep(xpri[5],93),rep(xpri[6],50),rep(xpri[7],25), rep(xpri[8],12), rep(xpri[9],11)))
str(bnocr)
'data.frame': 4864 obs. of 1 variable:
 $ dp: num 0.0556 0.0556 0.0556 0.0556 0.0556 ...

library(MASS)

betamv<-fitdistr(bnocr$dp,dbeta, start=list(shape1=0.9, shape2=5))#Requiere valores de
partida

betamv
      shape1      shape2
1.30424033 7.16543359
(0.02385519) (0.15289870)

lscd<-function(x,nc){#Limite superior de clase diametrica
  lsu<- numeric(nc)
  for (i in 1:nc){
    lsu[i]<- (x[i]+x[i+1])/2
  }
  lisu <- c(lsu)
  return(lisu)
}

lscd(xpri,length(d))
[1] 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667 0.7777778 0.8888889 NA

lsuc[9]<-0.9999999
lsuc
[1] 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667 0.7777778 0.8888889 0.9999999

probbeta<-pbeta(lsuc, 1.30424,7.165434, lower.tail = TRUE, log.p = FALSE)#Prob acumuladas

```

```

probbeta
[1] 0.4443549 0.7587067 0.9129378 0.9748220 0.9946303 0.9992847 0.9999593 0.9999997
1.0000000

faes<-su*probbeta #frecuencias estimadas acumuladas
faes
[1] 2161.342 3690.350 4440.529 4741.534 4837.882 4860.521 4863.802 4863.999 4864.000

fresc<-faesx(faes)

fresc
[1] 2.161342e+03 1.529007e+03 7.501798e+02 3.010048e+02 9.634756e+01 2.263922e+01
3.280914e+00 1.966403e-01 1.428712e-03 NA

fresc<-fresc[-10]
fresc
[1] 2.161342e+03 1.529007e+03 7.501798e+02 3.010048e+02 9.634756e+01 2.263922e+01
3.280914e+00 1.966403e-01 1.428712e-03

pchi<-sum((fro-fresc)^2/fresc)#chi2 calculada
pchi
[1] 85782.93
qchisq(0.975,6)
[1] 14.44938

```

Los datos no se comportan como una beta.

3.5.2.3 Forma generalizada de la distribución beta. Puede generalizarse una forma para cubrir el intervalo (a, b) en vez del (0, 1) que define la ecuación original, así:

$$f(x, \alpha, \gamma, a, b) = \begin{cases} \frac{1}{(a-b)} \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \left(\frac{x-a}{b-a}\right)^{(\alpha-1)} \left(1-\frac{x-a}{b-a}\right)^{(\gamma-1)}, & a \leq x \leq b, \alpha, \gamma > 0 \\ 0, & \text{para cualquier otro valor de } x \end{cases} \quad (3.78)$$

3.5.2.4 Probabilidades acumulativas. Las probabilidades acumulativas para la Beta generalizada pueden obtenerse transformando una variable beta al intervalo (0, 1), usando:

$$x' = \frac{(x-a)}{(b-a)} \quad (3.79)$$

y usando la forma biparamétrica en forma similar a lo visto con la gamma.

En la Tabla 3.3, Tomada de (Uribe, 1983), con base en datos de clases diamétricas y sus respectivas frecuencias, se dan los pasos secuenciales para el ajuste de la beta con base en las ecuaciones (3.79) y (3.76). Se insiste en que para lograr las probabilidades beta acumuladas se trabaja con los límites superiores de clase (ls). Los límites de la distribución quedaron definidos sin correcciones, pues no hay elementos claros para su propuesta.

Tabla 3.3. Elementos para el ajuste de la función beta con ayuda de Excell.

cladia	frobs	x'	fr*x'	fr(x'-pr)^2	ls	probbet	freesac	frecles	chi2
35	2634	0.056	146.33	21.78	0.111	0.51	2497.59	2497.59	7.5
45	1281	0.167	213.50	0.52	0.222	0.76	3712.99	1215.41	3.5
55	518	0.278	143.89	8.93	0.333	0.89	4350.55	637.56	22.4
65	240	0.389	93.33	14.10	0.444	0.96	4663.37	312.82	17.0
75	93	0.500	46.50	11.62	0.556	0.99	4799.74	136.37	13.8
85	50	0.611	30.56	10.79	0.667	1.00	4849.04	49.30	0.0
95	25	0.722	18.06	8.29	0.778	1.00	4862.06	13.02	11.0
105	12	0.833	10.00	5.66	0.889	1.00	4863.94	1.88	54.5
115	11	0.944	10.39	7.00	0.999	1.00	4864.00	0.06	2003.2
	4864	1.000	712.56	88.70				Chi2cal=	2132.9

media	0.146	r=	4.997
var=	0.018	a=	0.858

No se corrió la prueba de falta de ajuste dado el valor de la χ^2 calculada. Se deja constancia que el método de los momentos tiene muchas limitaciones para esta función, por lo cual veremos qué pasa con el ajuste máximo verosímil.

3.5.2.4.1 Ajuste, MV de la función beta R. Empezamos por el archivo de datos crudos:

Datos crudos

```
bncr<-data.frame(dp=c(rep(xpri[1],2634),rep(xpri[2],1281),rep(xpri[3],518),rep(xpri[4],240),
rep(xpri[5],93),rep(xpri[6],50),rep(xpri[7],25), rep(xpri[8],12), rep(xpri[9],11)))
str(bncr)
'data.frame': 4864 obs. of 1 variable:
 $ dp: num 0.0556 0.0556 0.0556 0.0556 0.0556 ...

library(MASS)

betamv<-fitdistr(bncr$dp,dbeta, start=list(shape1=0.9, shape2=5))#Requiere valores de
partida

betamv
      shape1      shape2
1.30424033  7.16543359
(0.02385519) (0.15289870)

lscd<-function(x,nc){#Limite superior de clase diametrica
  lsu<- numeric(nc)
  for (i in 1:nc){
    lsu[i]<- (x[i]+x[i+1])/2
  }
  lisu <- c(lsu)
  return(lisu)
}

lscd(xpri,length(d))
[1] 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667 0.7777778 0.8888889 NA

lsuc[9]<-0.9999999
lsuc
[1] 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667 0.7777778 0.8888889 0.9999999

probbeta<-pbeta(lsuc, 1.30424,7.165434, lower.tail = TRUE, log.p = FALSE)#Prob acumuladas
probbeta
[1] 0.4443549 0.7587067 0.9129378 0.9748220 0.9946303 0.9992847 0.9999593 0.9999997
1.0000000

faes<-su*probeta #frecuencias estimadas acumuladas
faes
[1] 2161.342 3690.350 4440.529 4741.534 4837.882 4860.521 4863.802 4863.999 4864.000

fresc<-faesx(faes)

fresc
[1] 2.161342e+03 1.529007e+03 7.501798e+02 3.010048e+02 9.634756e+01 2.263922e+01
3.280914e+00 1.966403e-01 1.428712e-03 NA

fresc<-fresc[-10]
fresc
[1] 2.161342e+03 1.529007e+03 7.501798e+02 3.010048e+02 9.634756e+01 2.263922e+01
3.280914e+00 1.966403e-01 1.428712e-03

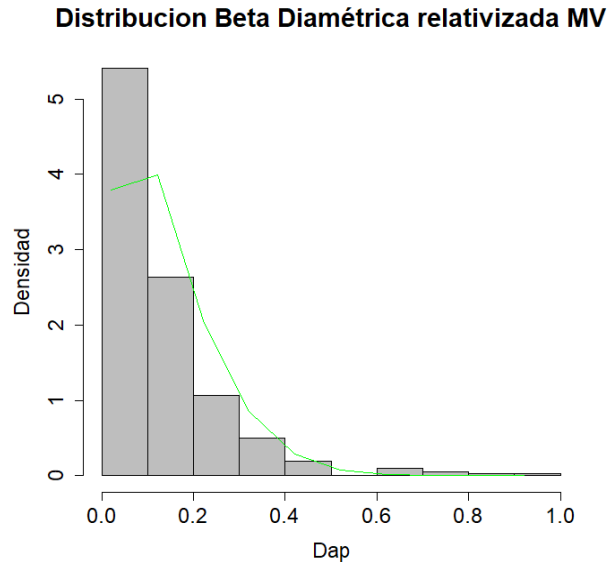
pchi<-sum((fro-fresc)^2/fresc)#chi2 calculada
pchi
```

```
[1] 85782.93
qchisq(0.975,6)
[1] 14.44938
```

Los datos no se comportan como una beta.

Construcción del histograma.

```
hist(dr, breaks=10,freq=FALSE,main="Distribucion Beta Diamétrica relativizada MV",
xlab="Dap", ylab="Densidad",col="gray",cex.lab=1.4,cex.axis=1.5,cex.main=1.8)
lines(0.21:9.90/10,dbeta(0.1:9.9/10,1.355057,8.51005473),col="green")
```



3.5.2.5 Casos especiales de la distribución beta. Por su versatilidad, se presentan algunos casos especiales para la distribución beta. Figura 3.26.

3.5.2.5.1 La distribución uniforme o rectangular. Surge cuando en la beta se hacen $\alpha = \gamma = 1$, entonces:

$$f(x; 1, 1) = \begin{cases} 1; & 0 \leq x \leq 1; \\ 0, & \text{en otros casos} \end{cases} \quad (3.81)$$

y al generalizarla en el intervalo (a, b):

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & 1; \quad a \leq x \leq b; \\ 0, & \text{en otros casos} \end{cases} \quad (3.82)$$

La probabilidad que una variada uniformemente distribuida tome un valor en un intervalo (x_1 , x_2) tal que $x_1 \geq a$; $x_2 \leq b$, es directamente proporcional a la longitud relativa del intervalo, o sea:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x, a, b) dx = \int_{x_1}^{x_2} \frac{1}{b-a} dx = \frac{x_2 - x_1}{b-a} \quad (3.83)$$

que, se constituye en un modelo apropiado para evaluar tiempos de ocurrencia de eventos igualmente probables en cualquier instante durante un intervalo.

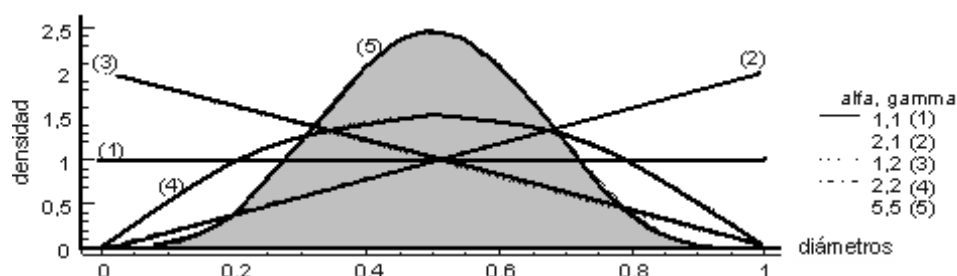


Figura 3.26. Algunos casos especiales de la distribución beta.

Por ejemplo, en labores de aprovechamiento, al pie de un camino maderero es igualmente probable que la madera sea recibida en cualquier momento x durante un período de 6 horas (p. ej. entre las 10 am y las 4 pm) o sea entre las 10 y las 16 horas. Cuál será la probabilidad de que llegue a las 11 am o durante la primera hora? Entonces:

$$f(x; 10, 16) = \frac{1}{16-10} = \frac{1}{6}; \quad 10 \leq x \leq 16 \approx f(x; 0, 6) = \frac{1}{6-0} = \frac{1}{6}$$

$$\therefore F(1; 0, 6) = \int_0^1 \frac{1}{6} dx = 0.166$$

3.5.2.5.2 La distribución triangular recta y la distribución parabólica.

a) La distribución triangular β recta se obtiene cuando $\alpha=2$ y $\gamma=1$, Figura 3.26 (2) cuya

$$f(x; 2, 1) = \begin{cases} 2x; & 0 \leq x \leq 1 \\ 0 & \text{forma triangular} \end{cases} \quad (3.84)$$

o haciendo $\alpha=1$ y $\gamma=2$, Figura 3.26 (3) cuya:

$$f(x; 1, 2) = \begin{cases} 2(1-x); & 0 \leq x \leq 1 \\ 0 & \text{forma semitriangular} \end{cases} \quad (3.85)$$

b) La distribución parabólica beta recta se obtiene cuando $\alpha=2$ y $\gamma=2$, Figura 3.26 (4) cuya:

$$f(x; 2, 2) = \begin{cases} 6x(1-x); & 0 \leq x \leq 1 \\ 0 & \text{forma parabólica} \end{cases} \quad (3.86)$$

que puede usarse como una aproximación simple a la normal. En la medida en que conjuntamente $\alpha = \gamma \geq 5$ se asemeja cada vez más a la normal, Figura 3.26 (5).

3.5.3 Distribución binomial negativa. Surge de un escenario binomial cuando en una secuencia de ensayos independientes con probabilidad de éxito constante p , en lugar de fijar el número de ensayos n en $b(x; n, p)$ y encontrar el número de éxitos $x = k$, se continúan los ensayos hasta lograr los k éxitos, es decir, se busca la variable aleatoria que representa un número de ensayos n para observar k éxitos. El número de maneras distintas en las que puedan observarse $k-1$ éxitos en $n-1$ ensayos es:

$$\binom{n-1}{k-1} \quad (3.87);$$

por lo que la probabilidad de tener k éxitos en n ensayos cuando el último sea un éxito es:

$$p(n, k, p) = \binom{n-1}{k-1} p^k (1-p)^{(n-k)} \quad (3.88) ;$$

conocida también como la distribución de Pascal, (Canavos, 1987), en la cual al sustituir $n = X + K(R)$ en (3.88), se llega a la binomial negativa, en que x es el valor de una variable aleatoria que representa el número de fracasos hasta que se llega de manera exacta a k éxitos, con R reemplazos para lograrlo.

Es una distribución discreta. Se llama binomial negativa debido a que las probabilidades al hacer el reemplazo (R) anotado en (3.88) corresponde a los términos sucesivos de la expansión binomial de:

$$[q + (-p)]^{-k} \quad (3.89)$$

Cuando k no es entero la función de probabilidad se puede escribir como:

$$P(X; k, p) = \frac{\Gamma(k+x)}{x! \Gamma(k)} p^k (1-p)^x \begin{cases} x = 0, 1, 2, \dots; \\ k > 0; 0 \leq p \leq 1 \end{cases} \quad (3.90)$$

La aplicación primaria de esta distribución no es más que una alternativa para el modelo de Poisson cuando la ocurrencia no es constante sobre el tiempo o el espacio, o se espera que la frecuencia entre grupos no sea la misma.

Por ejemplo, en bosques disetáneos, es más probable una alta frecuencia de diámetros menores que mayores, por lo cual no se toma la binomial negativa en términos del número de ensayos necesarios para alcanzar un determinado número de éxitos sino como el número de frecuencias por clase diamétrica cuando las ocurrencias no son constantes. La literatura la presenta como un modelo matemático para el estudio de poblaciones de organismos regidos por leyes de contagio, en las cuales las probabilidades de encontrar un individuo aumenta a medida que de antemano se haya encontrado otro de ellos en un espacio dado. Puede expresarse de muy diversas maneras, como la dada por (Robles y Carrillo, 1971) citado por (Gómez, 1984) para interpretaciones biológicas, así:

$$P(X=x) = \frac{\Gamma(k+x)}{\Gamma(x+1) \Gamma(k)} \left[\frac{m}{(k+m)} \right]^x \left[\frac{k}{(k+m)} \right]^k \quad (3.91)$$

en la cual $P(X=x)$ es la posibilidad de obtener x organismos en una unidad de muestreo y $\Gamma(k)$ la función gamma; m la media de los individuos por unidad de muestreo; k índice de agregación que a medida que el contagio disminuye su valor aumenta llegando a una Poisson. Además si $k > 8$ se puede pensar en distribuciones al azar. La media y la varianza de esta distribución están dadas por:

$$E(X) = m; \quad \sigma^2(X) = \frac{(m+m^2)}{k} > E(X) \quad (3.92)$$

y su función generadora de momentos por:

$$M_x(t) = b^k (1 - ae^t)^{-k} \quad (3.93),$$

en la cual $a = \frac{m}{k+m}$ y $b = \frac{k}{k+m}$.

El método de máxima verosimilitud para el valor de k es, según (Gómez, 1984):

$$N [\ln(k + \bar{x}) - \ln k] = \left[\frac{n_1 + n_2 + \dots + n_r}{k} + \frac{n_2 + n_3 + \dots + n_r}{k+1} + \dots + \frac{n_r}{k+r-1} \right] \quad (3.94)$$

con n_i , frecuencia observada con x_i individuos ($i = 1 \dots r$) y r = número máximo de individuos encontrados en una unidad de muestreo, y:

$$n = \sum_{i=0}^r n_i; \quad \bar{X} = \frac{1}{n} \sum_{i=0}^r x_i n_i \quad (3.95),$$

n = número total de unidades observadas, \bar{X} = media de individuos por unidad de muestreo. El proceso, puede iniciarse con algún valor de k, obtenido antes, y se hacen ajustes de comparación aumentándolo o disminuyéndolo, hasta que se satisfaga la ecuación anterior.

Ejemplo: Para estudiar la distribución diamétrica de un rodal, luego de discretizar sus clases se encontró el número de individuos por clase y por ha., primeras dos columnas Tabla 3.4 Encontrar el valor de k.

Tabla 3.4 Número de árboles por clase de un rodal y estimados

Xrel	Árboles/ha	fiXri	fi(Xri-Xrp)^2	prob	Frest	prob	FresMV	para chi	para chi
0	70	0	92039	0,45	67,49	0,496	74366	0,093	0,256
1	38	38	0,817	0,26	39,03	0,252	37782	0,027	0,001
2	17	34	12379	0,14	20,96	0,128	19196	0,748	0,251
3	10	30	34348	0,07	10,97	0,065	9753	0,085	0,006
4	9	36	73274	0,04	5662	0,033	4955	1968	3302
5	3	15	44545	0,02	2,9	0,017	2517	0,003	0,093
6	2	12	47,11	0,01	1477	0,009	1279	0,185	0,406
7	1	7	34262	0,01	0,75	0,004	0,65	0,084	0,189
8	0	0	0	0	0,379	0,002	0,33	0,379	0,33
suma	150	172	338773			chi² est	3572	4835	
promedio	1147	k MV	1025						
varianza	2274	q	1983			chi² tabla	14.44		
k	1167	p	0,983						

Con ellos se obtuvo: $\sum_{i=0}^r f_i = \sum_{i=0}^8 f_i = 150$; $\bar{X} = 1.147$; $s^2 = 2.273$, con los cuales se calculó con un valor de partida para $k = 1.15$, con lo cual:

$$150 [\ln(1.15 + 1.147) - \ln 1.15] = \frac{38 + 17 + L + 0}{1.15} + \frac{17 + 10L + 0}{1.15 + 1} + L \frac{1}{1.15 + 8 - 2} + \frac{0}{1.15 + 8 - 1}$$

que no equilibra ambos lados. El valor que lo hace es aproximadamente 1.025.

(Prodan, 1968) presenta la binomial negativa como la expansión de (3.89), la cual puede ser evaluada de acuerdo con el método de los momentos así:

$$s^2 = k.p.q = \bar{X} * q = \bar{X} * \left(1 + \frac{\bar{X}}{k} \right) \quad (3.96),$$

con lo cual:

$$k = \frac{\bar{X}^2}{s^2 - \bar{X}} ; \quad q = \frac{s^2}{\bar{X}} \quad (3.97)$$

y el cálculo de las frecuencias surge la siguiente relación general:

$$h(x) = \binom{k+x-1}{x} p^x q^{-(x+k)} = \left[\frac{(k+x-1)!}{x! (k-1)! q^k} \right] \left[\frac{\bar{X}^x}{(k+\bar{X})^x} \right] \quad (3.98)$$

Había sido ensayada con éxito, propuesta por el autor, en condiciones de bosque heterogéneo con buen ajuste a (Giraldo y Ramírez), citados por (Duque y Builes, 1986), y por ser una función discreta que solo responde a valores enteros, 0 a N, con transformaciones de clases mediante:

$$X_{r_i} = \frac{M_{C_i} - M_{C_{mic}}}{\text{amplitud}} \quad (3.99),$$

en la cual M_{C_i} = marca de clase i; $M_{C_{mic}}$ = primera marca de clase, X_{r_i} = nueva marca de clase relativa (i = 0...n), con los cuales se procede a la búsqueda de \bar{X}_r , s_r^2 y k_r que son los valores equivalentes en la nueva escala, en la cual:

$$si \begin{cases} X_r = 0; h(0) = q^{-k} \\ X_r = 1; h(1) = \frac{k \cdot p \cdot h(0)}{q} \\ \vdots \\ X_r = x; h(x) = \frac{(k+x-1)}{x} \frac{p \cdot h(x-1)}{q} \end{cases} \quad (3.100)$$

Para el ejemplo analizado se encontraron los valores de los relativizados así: $\bar{X}_r = 0.8192$, $q = 1.8187$, $p = 0.81873$, $s_r^2 = 1.49$ y $k_r = 1.0007$, y con ellos las probabilidades de las X_r obtenidas de la modificada propuesta por (Prodan, 1968). Para los datos de la Tabla 3.4 se estimó la probabilidad con los valores de k dados por (3.97) y la k máximo verosímil. Parece lograrse un buen ajuste, y además se resalta que se trata de una distribución muy fácil de evaluar.

Estimación máximo verosímil en R. Abrir la librería MASS.

```
bona<-read.table("clipboard")
attach(bona)
names(bona)
[1] "d" "f"

#datos crudos

bnbn<-data.frame(d=c(rep(35,2634), rep(45,1281), rep(55,518), rep(65,240), rep(75,93),
rep(85,50), rep(95,25), rep(105,12), rep(115,11)))
binonegm<-fitdistr(bnbn$d, "negative binomial")

binonegm
```

```

      size      mu
23.2827913 43.1839996
( 0.6978840) ( 0.1592005)

```

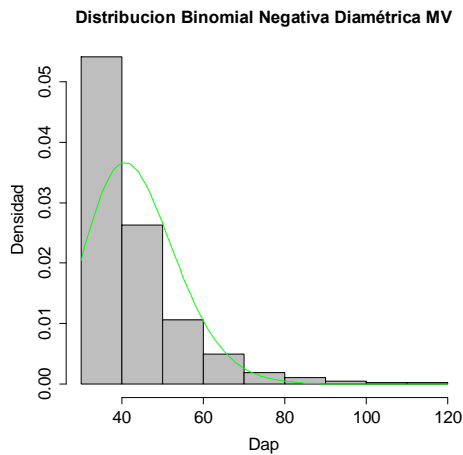
El parámetro μ es la media, R exige un parámetro p

$$= \mu = a \frac{(1-p)}{p}; \quad \therefore \quad p = \frac{\hat{a}}{\hat{a} + \hat{\mu}}$$

```

p<-a/(a+mu)
p
      size
0.3487214
hist(bn$d, breaks=11,freq=FALSE, main="Distribucion Binomial Negativa Diamétrica MV",
xlab="Dap", ylab="Densidad",col="gray",cex.lab=1.5,cex.axis=1.5,cex.main=1.5)
lines(30:120,dnbinom(30:120,a,p),col="green")

```



Prueba de bondad de ajuste. Existe un paquete (*library, vcd*) que tiene pruebas de bondad de ajuste para algunas distribuciones, entre ellas la binomial negativa:

```

library(vcd)

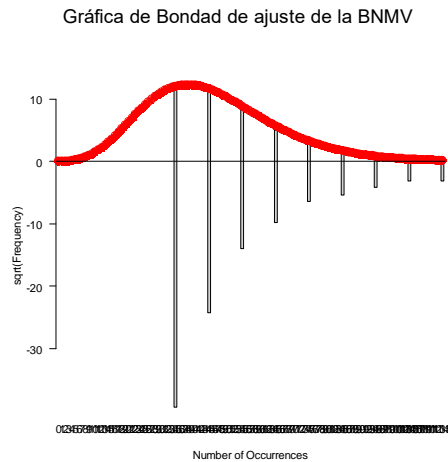
bdajbinon <- goodfit(bn$d, type = "nbinomial",method="MinChisq")#bondad de ajuste
summary(bdajbinon)
      Goodness-of-fit test for nbinomial distribution

      X^2  df P(> X^2)
Pearson 66425.86 123      0
Warning message:
In summary.goodfit(bdajbinon) : Chi-squared approximation may be incorrect

bdajbinon$par
$size
[1] 12.40789

$prob
[1] 0.2260371
plot(bdajbinon,main="Gráfica de Bondad de ajuste de la BNMV")

```



3.5.4 Distribución Lognormal. Una variable X , se dice que es una lognormal si sus logaritmos siguen una distribución normal con parámetros, de ellos, μ_l y σ_l^2 . Por tanto, la variable

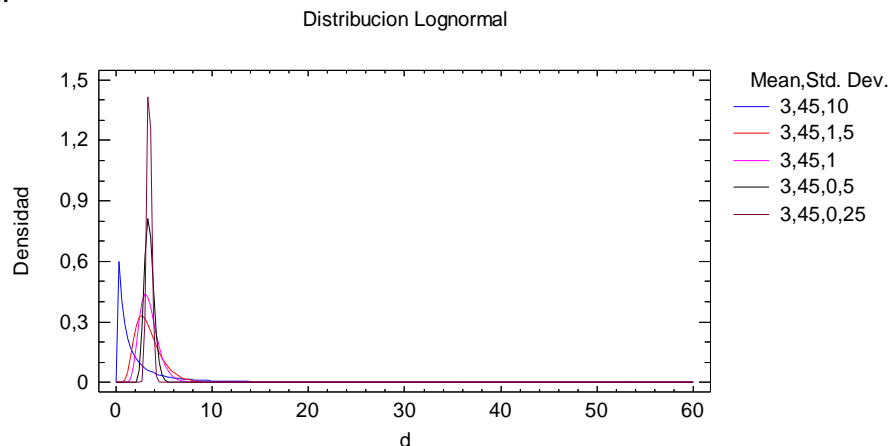
$$\frac{\ln x - \mu_l}{\sigma_l} \quad (3.101)$$

es un variable normal estandarizada de media 0 y desviación estándar 1.

3.5.4.1 Propiedades. Entre otras, la distribución lognormal tiene las siguientes propiedades:

- Para valores de $x < 0$ asigna una probabilidad 0 o sea sólo considera valores positivos.
- Por depender de dos parámetros, es de gran versatilidad y se ajusta bien a un gran número de situaciones empíricas y por tanto útil para parámetros que son a su vez producto de numerosas cantidades aleatorias (como puede ser el caso de distribuciones diamétricas).
- Al poseer su media mayor a su mediana, da más importancia a los valores grandes de X que la distribución normal con los mismos percentiles del 5% y 50% tendiendo, por tanto, a interpretar mejores situaciones de bosques con árboles de gran tamaño, a veces.

Es una función extremadamente versátil, lo que la hace candidata para distribuciones diamétricas.



La figura anterior muestra la versatilidad que ofrecería esta distribución para el caso de regeneración de bosque naturales.

3.5.4.2 Función de densidad. Sí se hace: $Y = \ln(X)$ entonces,

$$g(y) = \begin{cases} \frac{1}{2\pi\sigma_l x} \exp\left[-\frac{1}{2\sigma_l^2}(\ln x - \mu_l)^2\right]; & x > 0; \quad -\infty < \mu_l < \infty; \\ 0 & \text{en otros casos} \end{cases} \quad (3.102)$$

Presenta su función de densidad de probabilidades como:

$$f_1(x) = \frac{\beta}{\sqrt{2\pi}(x-a)} \exp\left\{-\frac{1}{2}\left[\alpha + \beta \ln\left(\frac{x-a}{\gamma}\right)\right]^2\right\}; \quad \begin{cases} x \geq a, \beta > 0, & -\infty < \alpha < \infty \\ \gamma > 0 \end{cases} \quad (3.103)$$

Esta distribución adopta muchas formas para variables no negativas, es sesgada a la derecha y el sesgo crece al hacerlo σ . Se puede resaltar que el parámetro de localización de la distribución normal original se ha convertido ahora en un parámetro de escala coincidente con la media. La dispersión α es también un parámetro de forma, o sea que en esta distribución μ y σ son ambos parámetros de forma. Acudiendo a nuestro archivo bona, transformado, ya trabajado, se llegó a:

```
bn<-data.frame(d=c(rep(25,0),rep(35,2634), rep(45,1281), rep(55,518), rep(65,240), rep(75,93),
rep(85,50), rep(95,25), rep(105,12), rep(115,11)) )

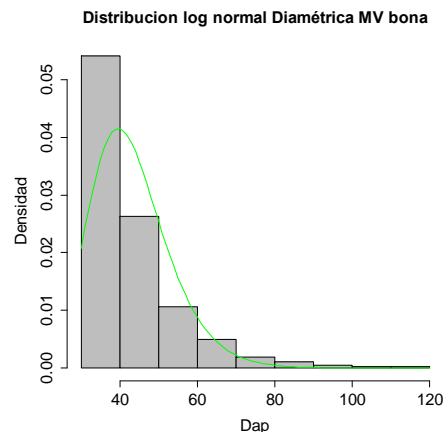
library(MASS)

ldi<-log(bn$d)#logaritmos
head(ldi)
[1] 3.555348 3.555348 3.555348 3.555348 3.555348 3.555348

mu<-mean(ldi)
mu
[1] 3.734441

var<-var(ldi)
var
[1] 0.05573749

ajlogn<-fitdistr(bn$d,"lognormal")
ajlogn
      meanlog      sdlog
 3.734440920  0.236063621
(0.003384794) (0.002393411)
hist(bn$d, breaks=11,freq=FALSE, main="Distribucion log normal Diamétrica MV bona",
xlab="Dap", ylab="Densidad",col="gray",cex.axis=1.5,cex.lab=1.5,cex.main=1.5)
lines(30:120,dlnorm(30:120,mu,var^.5),col="green")
```



Para valores medios y altos de σ , la distribución lognormal es significativamente asimétrica, pero a medida que “a” decrece la distribución es más simétrica. Si σ se acerca a la unidad, la distribución lognormal es equivalente aproximadamente a la distribución exponencial negativa. También se puede observar que para valores de $\sigma < 0.2$ la distribución lognormal se aproxima a la distribución normal.

Para cubrir un rango diferente al que se inicia se acude a un parámetro de localización, el a mostrado en (3.106), con lo cual haciendo

$$\alpha^* = \alpha - \beta \ln \gamma \quad (3.104)$$

se llega a

$$f_1(x) = \frac{\beta}{\sqrt{2\pi}(x-a)} \exp \left\{ -\frac{1}{2} \left[\frac{\alpha}{\beta} + \ln(x-a) \right]^2 \right\}; \quad \begin{cases} x \geq a, \beta > 0, & -\infty < \alpha^* < \infty \\ \gamma > 0 \end{cases} \quad (3.105)$$

que no es más que la log normal triparamétrica:

$$\left\{ \begin{aligned} f_1(x; \mu, \sigma, a) &= \frac{1}{\sqrt{2\pi}\sigma(x-a)} \exp \left\{ -\frac{1}{2\sigma^2} (\ln(x-a) - \mu)^2 \right\}; \\ \begin{cases} x \geq a, \beta > 0, & -\infty < \mu < \infty \\ \sigma > 0, & -\infty < \mu < \infty \end{cases} \end{aligned} \right\} \quad (3.106)$$

en que: $\beta = \frac{1}{\sigma}$ y $\alpha^* = -\frac{\mu}{\sigma}$; conocida como la distribución familia S_L de Johnson.

Ejemplo. Los datos de la Tabla 3.5, de un muestreo de regeneración natural en un bosque natural se obtuvieron los valores:

$\bar{x} = \hat{\mu} = 0.285$; $s = \hat{\sigma}^* = 0.00826$, con sus respectivos $m_3 = 0.00045018$ y $m_4 = 0.0002285$.

Tabla 3.5. Datos de dap y frecuencias para Pinus patula

	di	ds	d	f
1	0.0750	0.1000	0.0875	1
2	0.1000	0.1250	0.1125	2
3	0.1250	0.1500	0.1375	10
4	0.1500	0.1750	0.1625	23
5	0.1750	0.2000	0.1875	36
6	0.2000	0.2250	0.2125	42
7	0.2250	0.2500	0.2375	43
8	0.2500	0.2750	0.2625	55
9	0.2750	0.3000	0.2875	40
10	0.3000	0.3250	0.3125	43
11	0.3250	0.3500	0.3375	31
12	0.3500	0.3750	0.3625	22
13	0.3750	0.4000	0.3875	27
14	0.4000	0.4250	0.4125	18
15	0.4250	0.4500	0.4375	11
16	0.4500	0.4750	0.4625	3
17	0.4750	0.5000	0.4875	2
18	0.5000	0.5250	0.5125	4
19	0.5250	0.5500	0.5375	3
20	0.5500	0.5750	0.5625	0

21	0.5750	0.6000	0.5875	2
22	0.6000	0.6250	0.6125	1

```
dicru<- data.frame(d=c(rep(0.0875,1), rep(0.1125,2), rep(0.1375,10), rep(0.1625,23),
rep(0.1875,36), rep(0.2125,42), rep(0.2375,43), rep(0.2625,55), rep(0.2875,40), rep(0.3125,43),
rep(0.3375,31), rep(0.3625,22), rep(0.3875,27), rep(0.4125,18), rep(0.4375,11), rep(0.4625,3),
rep(0.4875,2), rep(0.5125,4), rep(0.5375,3), rep(0.5625,0), rep(0.5875,2), rep(0.6125,1)))
mean(dicru$d)
[1] 0.285531
> sd(dicru$d)
[1] 0.08880521
```

Al estimarla de acuerdo con (3.105) y (3.106), los ajustes no fueron satisfactorios pues se supuso un $\alpha = 0.012$, por ser el valor inferior de la primera clase. Su $\hat{\beta} = 2.6724$ y su $\hat{\alpha} = 4.288217$. Al evaluarla recomponiendo clases efectivamente el mejor ajuste se logró por medio de la lognormal como lo muestra la Tabla 3.6 y la Figura 3.27, usando como base de datos las marcas de clase repetidas tantas veces como su frecuencia.

Estimación máximo verosímil.

Para que sea corrida como lognormal, se multiplican los valores por 100 (dicr100), para no tener log negativos, en cuyo caso no corre la distribución.

library(MASS)

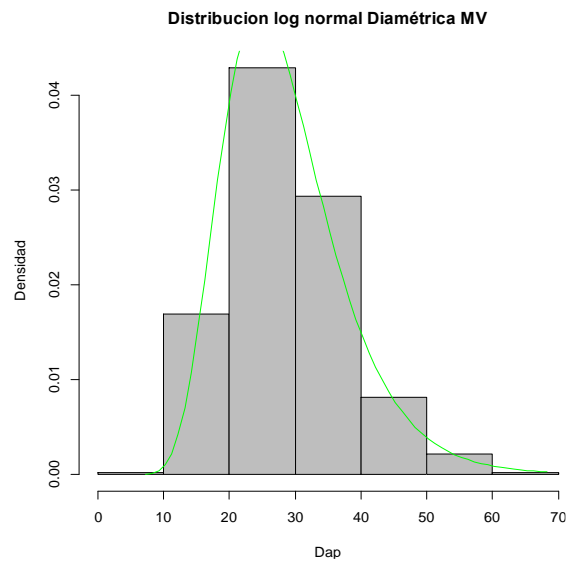
```
dicr100<-data.frame(d=c(rep(0.0875*100,1), rep(0.1125*100,2), rep(0.1375*100,10),
rep(0.1625*100,23), rep(0.1875*100,36), rep(0.2125*100,42), rep(0.2375*100,43), rep(0.2625*100,55),
rep(0.2875*100,40), rep(0.3125*100,43), rep(0.3375*100,31), rep(0.3625*100,22), rep(0.3875*100,27),
rep(0.4125*100,18), rep(0.4375*100,11), rep(0.4625*100,3), rep(0.4875*100,2), rep(0.5125*100,4), rep(
0.5375*100,3), rep(0.5625*100,0), rep(0.5875*100,2), rep(0.6125*100,1)))
> str(dicr100)
'data.frame': 419 obs. of 1 variable:
 $ d: num 8.75 11.25 11.25 13.75 13.75 ...
ajlogn<-fitdistr(dicr100$d,"lognormal")
ajlogn
      meanlog      sdlog
3.30331994 0.31498128
(0.01538784) (0.01088085)
mu<-ajlogn$estimate[1]
sdl<-ajlogn$estimate[2]
sdl
      sdlog
0.3149813
lndicr<-log(dicr100)
frec <- cut(lndicr$d, breaks = 7, dig.lab = 4, right = FALSE)
tab.fr <- table(frec)
tab.fr
frec
[2.167,2.447) [2.447,2.725) [2.725,3.003) [3.003,3.281) [3.281,3.559)
      3          10          59          140          114
[3.559,3.837) [3.837,4.117)
      81          12
tadefr<-as.data.frame(tab.fr)
> tadefr
      frec Freq
1 [2.167,2.447) 3
2 [2.447,2.725) 10
3 [2.725,3.003) 59
4 [3.003,3.281) 140
5 [3.281,3.559) 114
6 [3.559,3.837) 81
7 [3.837,4.117) 12
```

```

ls<-c(2.447, 2.725, 3.003,3.281, 3.559, 3.837,4.117)#limites superiors de clase
> ls
[1] 2.447 2.725 3.003 3.281 3.559 3.837 4.117
exls<-exp(ls)
exls
[1] 11.55363 15.25641 20.14588 26.60236 35.12805 46.38611 61.37484

probln<-plnorm(exls,mu,sdl, lower.tail = TRUE, log.p = F)#probabilidades acumuladas lognormales
probln
[1] 0.00327757 0.03317547 0.17018023 0.47175413 0.79152699 0.95489805 0.99510649
hist(dicr100$d, breaks=7,freq=FALSE, main="Distribucion log normal Diamétrica MV", xlab="Dap",
ylab="Densidad",col="gray")
lines(7.25:68.75,dlnorm(7.25:68.75,mu,sdl),col="green")

```



Prueba de bondad de ajuste.

```

faes<-probln*su
faes
[1] 1.373302 13.900522 71.305515 197.664980 331.649809 400.102282 416.949620

faesx<-function(x){
  n<- length(x)
  fres<- numeric(n)
  for (i in 1:n){
    fres[i]<- x[i+1]-x[i]
  }
  result <- c(x[1],fres)
  return(result)
}
fres<-faesx(faes)
> fres
[1] 1.373302 12.527220 57.404993 126.359465 133.984829 68.452473 16.847338
[8] NA
tadefr$Freq
[1] 3 10 59 140 114 81 12
frob<- tadefr$Freq
frob
[1] 3 10 59 140 114 81 12
pchi<-(frob-fres)^2/fres#numerador de la prueba chi2

chi2c<-sum(pchi,na.rm=T)
chi2c
[1] 10.62907

```

```
chi2t=qchisq(0.975,7-2)
> chi2t
[1] 12.8325
```

El resultado coincide con lo mostrado en la figura, los datos se comportan como una lognormal.

La lognormal da lugar a la función de supervivencia que se puede escribir:

$$S(t)=1-F(y) \quad (3.107)$$

siendo $F(y)$ la función de distribución acumulativa de la normal estandarizada. Por lo tanto un modo gráfico de verificar esta distribución es comparar la función de supervivencia dibujada en papel lognormal con una recta, el qqplot nuestro actual.

Para el ajuste siguiente se hizo por medio de la chi cuadrada. Los datos se multiplicaron por 100 para en un cambio de escala para cumplir las condiciones del EXCELL y, las probabilidades ahora si de acuerdo con la normal. La distribución e histograma se presentan en la Figura 3.27. Se muestra solo por procedimiento

Tabla 3.6. Prueba de bondad de ajuste para el dap y prueba de χ^2

	li	ls	d	f	f*Ind	f*desv^2	lnls	prln	n*pr	frest	pchi2		
1	8.697	16.250	12.474	13	32.807	8.872	2.788	0.036	14.925	14.925	0.248		
2	16.250	23.750	20.000	101	302.569	12.657	3.168	0.279	117.006	102.082	0.011		
3	23.750	31.250	27.500	138	457.358	0.174	3.442	0.617	258.328	141.322	0.078		
4	31.250	38.750	35.000	96	341.313	4.058	3.657	0.838	351.229	92.900	0.103		
5	38.750	46.250	42.500	56	209.972	8.949	3.834	0.940	393.901	42.672	4.163		
6	46.250	53.750	50.000	9	35.208	2.845	3.984	0.979	410.301	16.400	3.339		
7	53.750	61.300	57.525	6	24.313	2.961	4.116	0.993	416.093	5.792	0.007		
				419	1403.541	40.518					chi2c	7.951	
media				3.3497	var	0.0969	des	0.311				chi2t	9.488

La lognormal puede derivarse como el modelo para procesos cuyos valores resultan de la multiplicación de muchos pequeños errores (mientras la normal lo hace para la adición de ellos). Es una distribución popular por la facilidad para obtener sus percentiles puesto que los valores acumulativos para $Y = \ln X$ pueden salir de tablas de Z y sus correspondientes valores surgir al tomar antilogaritmos.

3.5.5 Distribución de Weibull. Desarrollada por Fisher en 1928 y Weibull en 1939, con evidencias empíricas, encontró modelos de esfuerzos a que se someten ciertos materiales mediante ella. Se dice que una variable aleatoria X tiene una distribución de Weibull si su densidad de probabilidades es:

$$f(x; c, b) = \begin{cases} \frac{c}{b^c} x^{c-1} \exp \left[-\left(\frac{x}{b} \right)^c \right] = \frac{c}{b} \left(\frac{x}{b} \right)^{c-1} \exp \left[-\left(\frac{x}{b} \right)^c \right]; & x > 0, c, b > 0 \\ 0 & \text{para otros valores} \end{cases} \quad (3.108)$$

llamada forma biparamétrica (Bailey, 1973; Seth et al., 1975; Uribe, 1983; Canavos, 1987). con media y varianza:

$$\left. \begin{aligned} E(X) &= b\Gamma\left(1 + \frac{1}{c}\right); \quad \text{VAR}(X) = \left(\frac{b^2}{c}\right) \left\{ 2\Gamma\left(\frac{2}{c}\right) - \left(\frac{1}{c}\right) \left[\Gamma\left(\frac{1}{c}\right) \right]^2 \right\}; \\ \sigma^2(X) &= b^2 \left\{ \Gamma\left(1 + \frac{2}{c}\right) - \left[\Gamma\left(1 + \frac{1}{c}\right) \right]^2 \right\} \end{aligned} \right\} \quad (3.109)$$

La literatura reporta que esta distribución presenta grandes posibilidades para las estimaciones del porcentaje de los árboles para un diámetro especificado lo mismo que para predicciones implícitas del crecimiento, cuyas constantes pueden ser relacionadas con la forma y fisonomía de ellos y hasta recoger algunas características propias de un rodal (Bailey, 1973; Seth et al., 1975; Uribe, 1983).

Se caracteriza por un parámetro de escala b que permite transformar un área en otro parámetro, por ejemplo el número de árboles; y un parámetro de forma c que depende de los perfiles adoptados por la distribución desde una J invertida hasta una normal. (Para distribuciones decrecientes, por ejemplo $c < 1$).

Además cuando se le introduce un tercer factor adicional al cambiar la variable aleatoria x , por $(x - a)$, que representa el umbral o valor mínimo de la distribución, por ejemplo el diámetro mínimo de un rodal, aparece la forma triparamétrica:

$$f(x) = \begin{cases} \frac{c}{b} \left[\frac{(x-a)}{b} \right]^{c-1} \exp \left[- \left(\frac{x-a}{b} \right)^c \right]; & a \leq x < \infty; \\ 0 & \text{para otros valores} \end{cases} \quad (3.110)$$

El parámetro a es conocido como parámetro de localización, b de escala y c de forma. b y c deben ser siempre positivos, a puede ser 0, positivo y hasta negativo, (Clutter, 1983).

3.5.5.1 Función de distribución acumulada de Weibul $F(x)$. Integrando $f(x)$ por sustitución: $u = (x/b)^c$ aparece una función de forma: $\int e^{-u} du$ útil para determinar el número de árboles situados por debajo de un determinado diámetro X .

$$F(x) = \begin{cases} 1 - \exp \left[- \left(\frac{x}{b} \right)^c \right]; & \text{biparametrica, } 0 \text{ para otros valores} \\ 1 - \exp \left[- \left(\frac{x-a}{b} \right)^c \right]; & a \leq x < \infty; \text{ triparametrica} \end{cases} \quad (3.111)$$

Las constantes en el modelo están íntimamente ligadas con la forma y localización de la distribución, por lo cual las variaciones en ellos son consistentes con las características del rodal.

Obtención de parámetros. Uno de los métodos utilizados en Distribuciones diamétricas es una extensión de la propuesta de (Dubey, 1967) reportado por (Seth et al., 1975), por medio de percentiles. Sean x_1, x_2, \dots, x_n los puntos medios de n intervalos de clases, equidistantes una cantidad h (por comodidad), con sus respectivas frecuencias f_1, f_2, \dots, f_n . La

proporción de observaciones que se encuentran hasta o por debajo de $x_i + \frac{h}{2}$ o sea la acumulada será entonces:

$$\frac{f_1 + f_2 + \dots + f_i}{f_1 + f_2 + \dots + f_n} \quad (3.112)$$

lo cual podría notarse como $F\left(X_i + \frac{h}{2}\right)$, probabilidad acumulada. Para la forma biparamétrica:

$$F\left(X_i + \frac{h}{2}\right) = \left[1 - \exp\left\{-\left(\frac{x_i + h/2}{b}\right)^c\right\} \right] \quad (3.113)$$

es importante porque permite evaluar la distribución acumulada absoluta por debajo de un determinado valor X, en este caso el número de árboles con un diámetro menor que uno fijado d_i , así $N_x = NF(x)$, cuando se conoce el número total de árboles N. Acudiendo al proceso de regresión simple se pueden determinar los parámetros b y c, en que por simplicidad $x_i + h/2$ se reemplaza por x, límite superior de las clases, así:

$$F(x) = 1 - e^{-\left(\frac{x}{b}\right)^c}; \quad \therefore e^{-\left(\frac{x}{b}\right)^c} = 1 - F(x) \quad (3.114)$$

Tomando logaritmos a ambos lados y multiplicando por -1, se tiene:

$$-\left(\frac{x}{b}\right)^c = \ln[1 - F(x)]; \rightarrow \left(\frac{x}{b}\right)^c = -\ln[1 - F(x)] \quad (3.115)$$

Tomando nuevamente logaritmos se llega a:

$$c \ln\left(\frac{x}{b}\right) = \ln[-\ln(1 - F(x))] \therefore c \ln x - c \ln b = \ln[-\ln(1 - F(x))] \quad (3.116)$$

Haciendo:

$$Y = \ln[-\ln(1 - F(x))] = -c \ln b + c \ln x; \text{ como } Y = b_0 + b_1 X \quad (3.117)$$

$$\text{entonces } \ln b = \frac{-b_0}{c} = \frac{-b_0}{b_1}.$$

Como se puede apreciar, el intercepto $b_0 = -c \ln b$, y la pendiente $b_1 = c$.

Para el problema anterior entonces con los datos de la Tabla 3.3, se genera la Tabla 3.7. De ello se obtuvo: $b_0 = -7.9833 \rightarrow b = 42, 2$, $b_1 = 2,133 \rightarrow c = 2,133$, con los que se obtienen los estimados con EXCEL mostrados en la Tabla 3.7.

La forma tripamétrica se puede resolver a partir de la recuperación de distribuciones diamétricas de los atributos de los rodales con base en los dos primeros momentos de su estructura, (Chang, 1987) que se obtienen así:

$$E(x) = \bar{X} = a + b\Gamma\left(1 + \frac{1}{c}\right), \quad E(x^2) = \bar{x}^2 = a^2 + 2ab\Gamma\left(1 + \frac{1}{c}\right) + b^2\Gamma\left(1 + \frac{2}{c}\right) \quad \bar{X}^2 = 12732.4 \frac{G}{N}$$

$$Cv = \frac{s}{\bar{X} - a} = \frac{\left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right)^{\frac{1}{2}} \right]}{\Gamma\left(1 + \frac{1}{c}\right)} \quad (3.118)$$

en las cuales: X = diámetro variable > a (en cm), a = parámetro de localización (> 0), b = parámetro de escala, c = parámetro de forma. Dados unos valores específicos de a tanto como los estimados de \bar{X} y \bar{X}^2 , el valor de c se puede obtener por métodos numéricos, con lo cual:

$$b = \frac{(\bar{X} - a)}{\Gamma(1 + c)} \quad (3.119)$$

TABLA 3.7. Pasos secuenciales en la solución Weibull

i	xi+h/2=xi	Fobs	F ac	F(hi+h/2)	1-F(hi+h/2)	-1*LN(1-F(hi+h/2))	Y	X'=LN(xi+h/2)	F(x)	frec estWe
1	40	2634	2634	0,542	0,458	0,780	-0,249	3,689	0,590	2868,0
2	50	1281	3915	0,805	0,195	1,634	0,491	3,912	0,762	3704,2
3	60	518	4433	0,911	0,089	2,424	0,885	4,094	0,879	4277,2
4	70	240	4673	0,961	0,039	3,237	1,175	4,248	0,947	4606,4
5	80	93	4766	0,980	0,020	3,905	1,362	4,382	0,980	4766,2
6	90	50	4816	0,990	0,010	4,618	1,530	4,500	0,993	4831,9
7	100	25	4841	0,995	0,005	5,354	1,678	4,605	0,998	4855,0
8	110	12	4853	0,998	0,002	6,092	1,807	4,700	1,000	4861,8
9	120	7	4860	0,999	0,001	7,103	1,961	4,787	1,000	4863,5
10	130	4	4864	1,000	0,000	18,421	2,913	4,868	1,000	4863,9

4864

	Coefic	Err est
Intercepción	-79,833	0,927813
X=LN(xi+h/2)	21,328	0,211144

$$\begin{aligned} c &= 2,13 \quad b_1 \\ \ln b &= 3,74 \\ b &= 42,2 \quad b_0 \end{aligned}$$

Una vez obtenidos a, b y c se genera un estimativo del área basal G, de acuerdo con:

$$G = 7,854E - 05 N \sum_i X_i^2 f_i, \text{ en que: } N = \text{número de árboles, } X_i = \text{marca de la } i\text{ésima clase,}$$

$f_i = F(X_i + 0.5) - F(X_i - 0.5)$ = proporción de los árboles en la clase i. La distribución acumulativa de Weibull con parámetros a, b y c, es:

$$F(x) = 1 - \exp\left\{-\left[(x-a)/b\right]^c\right\} \quad (3.120)$$

Si el área basal obtenida a través de los cálculos anteriores no iguala la confrontada del rodal (G*) se debe refinar el valor "a" hasta que se llegue a una aproximación aceptable previamente fijada tal que: $|G - G^*| < \varepsilon$.

Estimación máximo verosímil (MLE)

Maximizando las derivadas parciales con respecto a b y c se llega a los estimados maximoverosímiles de ellos.

$$c = \left[\frac{\sum_{i=1}^n d_i \ln d_i}{\sum_{i=1}^n d_i^2} - \frac{\sum_{i=1}^n \ln d_i}{n} \right]^{-1}; \quad b = \left[\frac{\sum_{i=1}^n d_i^c}{n} \right]^{-\frac{1}{c}}$$

n número de observaciones de la muestra en una distribución Weibull y di la variable (en este caso el diámetro de cada árbol).

En R. Usaremos el ya trabajado archivo con datos bn, ya usado:.

```
library(MASS)
weibmv=fitdistr(bn$d,"weibull")# función del MASS

weibmv
      shape      scale
3.31449527  47.71349017
( 0.03087463) ( 0.21994539)

a<-weibmv$estimate[1]
a
      shape
3.314495

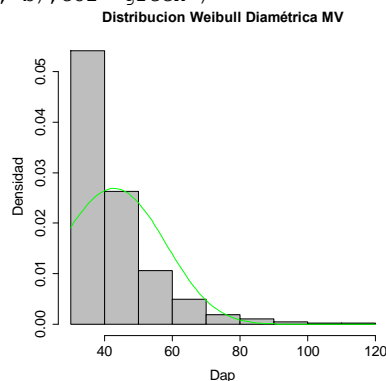
b<-weibmv$estimate[2]
b
      scale
47.71349

names(weibmv)
[1] "estimate" "sd"      "vcov"      "loglik"    "n"

weibmv$vcov
      shape      scale
shape 0.0009532428 0.00234558
scale 0.0023455798 0.04837597
weibmv$loglik
[1] -19308.05

confint(weibmv)#
      2.5 %      97.5 %
shape 3.253982 3.375008
scale 47.282405 48.144575

hist(bn$d,breaks=11,freq=FALSE, main="Distribucion Weibull Diamétrica MV", xlab="Dap",
ylab="Densidad",col="gray",cex.lab=1.5,cex.axis=1.5,cex.main=1.5)
lines(30:120,dweibull(30:120,a, b),col="green")
```



```

confint(weibmv)#límites de confianza para los parámetros
      2.5 %      97.5 %
shape 3.232005 3.351599
scale 47.291851 48.160096

```

3.5.5 Función exponencial negativa Meyer Tipo I.

La función exponencial negativa es un caso especial de los modelos Weibull y Gamma. Sirve para modelar el lapso entre dos eventos consecutivos de Poisson que ocurran de manera independiente y a una frecuencia constante, como es el caso de lo expuesto desde 1889 por el francés *Liocourt d'Lisle*. Si una variable aleatoria X tiene una distribución exponencial, su función de densidad de probabilidad se da, en dos versiones, por:

$$f(x) = \begin{cases} a) \theta \exp[-\theta x]; & x > 0 \quad \theta > 0; \\ b) \frac{1}{\theta} \exp\left[-\left(\frac{x}{\theta}\right)\right]; & x > 0 \quad \theta > 0; \\ 0 & \text{para otros valores} \end{cases} \quad (3.121)$$

el parámetro θ , que representa un lapso de tiempo entre dos eventos Poisson por lo cual recibe el nombre de tiempo promedio entre fallas o aparición del fenómeno, no tiene sin embargo un contexto claro en distribuciones diamétricas, pero se acude a esta función como posible modelador que equipara un fallo a un cambio de categoría diamétrica. La función de distribución acumulada se puede obtener como:

$$F(x) = P(X \leq x) = \begin{cases} a) 1 - \exp[-\theta x]; & x > 0 \quad \theta > 0; \\ b) 1 - \exp\left[-\left(\frac{x}{\theta}\right)\right]; & x > 0 \quad \theta > 0; \\ 0 & \text{para cualquier otro valor} \end{cases} \quad (3.122)$$

Según (Canavos, 1987), y así es asumida por algunos paquetes estadísticos, las expresiones para cuantiles, media y varianza se obtienen así para ambas versiones:

$$\left. \begin{aligned} a) E(X) &= \frac{1}{\theta}; & \sigma^2(x) &= \frac{1}{\theta^2} \\ b) X_q &= \theta \ln\left[\frac{1}{(1-q)}\right]; & E(X) &= \theta; & \sigma^2(x) &= \theta^2 \end{aligned} \right\} \quad (3.123)$$

En las ciencias forestales, fue propuesta por Meyer en 1933, citado por (Loetsch et al., 1974), para dar cuenta matemáticamente de la ley de Liocourt (1898), el primero en expresar para rodales coetáneos que el cociente entre las frecuencias sucesivas de los diámetros en bosques de selección, permanece constante, o dicha de otra forma: el cambio en número de árboles (Y) debido a un cambio en el diámetro (X), se puede considerar proporcional a(Y), o sea:

$$\frac{\Delta Y}{\Delta X} = -bY \quad (3.124)$$

Estableciendo clases pequeñas del diámetro de modo que X se vuelva una escala continua se llega a:

$$\lim_{\Delta X \rightarrow 0} \frac{\Delta Y}{\Delta X} = -bY = \frac{dY}{dx} = -bY \therefore \begin{cases} \ln Y = -bX + c \\ Y = ke^{-bx} = k \exp(-bx) \end{cases} \quad (3.125)$$

Esta ley se utilizó durante mucho tiempo para expresar el concepto del bosque balanceado en la cual Y : número de árboles, X : clase diamétrica, k , b : parámetros constantes.

3.5.7.1 Función de distribución de probabilidades $F(x)$

La función de distribución de probabilidades $F(x)$ variará entonces entre 0 y 1 y se podrá expresar como el área entre 0 y una clase diamétrica X con respecto a toda el área de distribución así:

$$F(x) = \frac{\int_a^x k \exp(-bx) dx}{\int_0^x k \exp(-bx) dx} = 1 - \exp(-bx), \quad x > 0, \quad b > 0 \quad (3.126)$$

La función de densidad de probabilidades de X se define entonces como:

$$f(x) = \frac{dF(x)}{dx} = b \exp(-bx) \quad (3.127)$$

3.5.7.2 Media y varianza de la distribución de Meyer

La media se define como:

$$\mu = \int_0^\infty x f(x) dx = \int_0^\infty x b e^{-bx} dx \therefore \mu = \frac{1}{b} \quad (3.128)$$

La varianza, como

$$\sigma^2(x) = \int_0^\infty (x - \mu)^2 f(x) dx = \int_0^\infty (x - \mu)^2 b \exp(-bx) dx = \frac{1}{b^2} \quad (3.129)$$

o sea $\mu = 1/b = \sigma(x)$. Un estimador insesgado de la media puede obtenerse como:

$$\frac{n\bar{x}}{n-1} = \frac{1}{b} \therefore b = \frac{(n-1)}{n\bar{x}} \therefore \bar{X} = \frac{(n-1)}{nb} \quad (3.130)$$

El ajuste logrado con ella se muestra en la Tabla 3.8

Tabla 3.8 Ajuste por Meyer

i	X	xi+h/2=xi	Fobs	F ac	ProbMe	FacMe
1	35	40	2634	2634	0,555	2700,541
2	45	50	1281	3915	0,647	3147,588
3	55	60	518	4433	0,720	3502,259
4	65	70	240	4673	0,778	3783,643
5	75	80	93	4766	0,824	4006,883
6	85	90	50	4816	0,860	4183,993
7	95	100	25	4841	0,889	4324,507
8	105	110	12	4853	0,912	4435,985
9	115	120	7	4860	0,930	4524,428
10	125	130	4	4864	0,945	4594,596
sumas			4864			

\bar{X}	43,193
b	0,023

Estimación MV de los parámetros. Nuevamente partimos del vector de los datos crudos:

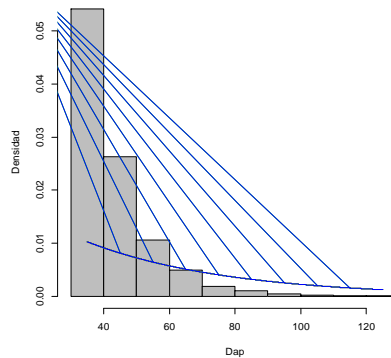
```
bn.d<-data.frame(d=c(rep(35,2634),rep(45,1281),rep(55,518),rep(65,240),rep(75,93), rep(85,50),  
rep(95,25), rep(105,12), rep(115,7),rep(125,4)) )
```

```
library(MASS)  
expdiam<-fitdistr(bn.d$d,"exponential")  
expdiam  
rate  
0.0231519825  
(0.0003319643)
```

Se crea el histograma de los datos crudos

```
hist(bn.d$d, breaks=11,freq=FALSE, main = "Distribucion Diamétrica Exponencial MV",xlab="Dap",  
ylab = "Densidad", col="gray")
```

```
a<-expdiam$estimate[1]  
a  
rate  
0.02315198  
lines(3500:12500/100, dexp(3500:12500/100,a),col="blue")  
Distribucion Diamétrica Exponencial MV
```



Prueba de Bondad de ajuste. (con el archivo (bona))

```
probexp<-pexp(d+5,0.02315198, lower.tail = TRUE, log.p = FALSE)# con el li sup de clase
```

```
probexp  
[1] 0.6038963 0.6857602 0.7507051 0.8022276 0.8431018 0.8755285 0.9012534  
[8] 0.9216616 0.9378521 0.9506964
```

```
faes<-probexp*sum(f)  
> faes  
[1] 2937.352 3335.538 3651.430 3902.035 4100.847 4258.570 4383.696 4482.962  
[9] 4561.712 4624.187
```

```
faesx<-function(x) {  
  n<- length(x)  
  fres<- numeric(n)  
  for (i in 1:n){  
    fres[i]<- x[i+1]-x[i]  
  }  
  result <- c(x[1],fres)  
  return(result)  
}
```

```
fre<-faesx(faes)
```

```
pchi<-(f-fre)^2/fre
```

```
chi2c<-sum(pchi,na.rm=T)
```

```
gchisq(0.975,8)
[1] 17.53455
```

Por tanto, no da un buen ajuste.

3.5.7.3 Polinomial de grado r y curva simple exponencial

Una ecuación de la forma:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n \quad (3.131)$$

se presta a menudo, para distribuciones diamétricas dependiendo la bondad de su ajuste, del número de términos, encontrándose en general acertada para rodales coetáneos. Liocourt la presentó como una serie geométrica decreciente de frecuencias n_1, n_2, \dots, n_n tal que $n_{i+1} < n_i$, con una razón $q > 1$. Entonces:

$$n_1; n_2 = n_1 q^{-1}; n_3 = n_2 q^{-1}; \dots; n_{i+1} = n_i q^{-1}; \text{ o } \frac{n_i}{n_{i+1}} = q \quad (3.132)$$

siendo esta la ecuación presentada por Meyer como

$$Y = k e^{-bx} \quad (3.133)$$

Si h define el intervalo de clase, entonces la frecuencia de la clase diamétrica d se obtiene como:

$$\left. \begin{aligned} Y_d &= \int_d^{d+h} k e^{-bx} dx = \frac{k}{b} \left[e^{-bd} - e^{-b(d+h)} \right] = \frac{k}{b} e^{-bd} (1 - e^{-bh}) \\ Y_{d+1} &= \int_{d+h}^{d+2h} k e^{-bx} dx = \frac{k}{b} e^{-b(d+h)} (1 - e^{-bh}) \end{aligned} \right\} \quad (3.134)$$

Si se hace

$$n = \frac{k}{b} e^{-bd} (1 - e^{-bh}) \quad (3.135)$$

se puede derivar la serie:

$$n, n e^{-bh}, n e^{-2bh}, \dots, n e^{-kbh} \quad (3.136)$$

en la cual puede hacerse

$$q = e^{bh} \quad (3.137)$$

que corresponde con la serie geométrica presentada. (Prodan, 1961)

Cociente q . Como en la distribución presentada la frecuencia refleja el grado de probabilidad y viceversa, es posible presentar una expresión para q de la función de distribución acumulada así:

$$q = \frac{F(x-h) - F(x)}{F(x) - F(x+h)} = e^{bh} \quad (3.138)$$

siendo h : intervalo de clase común.

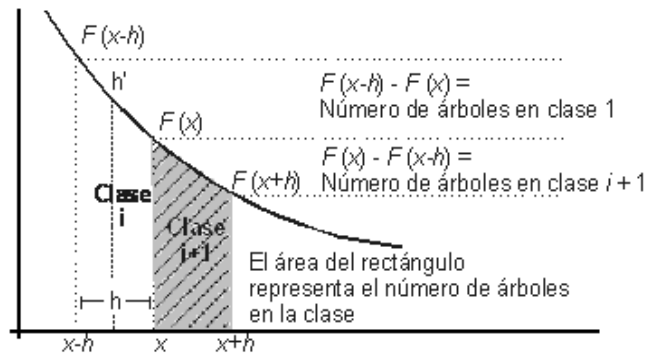


Figura 3.18. Propuesta de modificación de la función de densidad de la polinomial de grado r.

Cuando se hacen intentos para obtener el valor de q , si los q_i , q_{i+1} resultan muy diferentes, se debe cambiar de expresión como lo proponen (Loetsch et. al., 1973), Figura 3.18, modificando la función de densidad, dado que el número de árboles está representado por un rectángulo de ancho h y altura h' , que debe ser corregida pues a medida que se incrementa la amplitud cambia sensiblemente h' , la altura real de dicho rectángulo.

3.5.7.4 Determinación de la función de densidad corregida. Por medio de mínimos cuadrados es posible linealizar la primera ecuación mostrada en (3.146). La función modificada de densidad estará dada por:

$$Y_i = hh'_i = hk_1 \exp(-bd_i) \quad (3.142)$$

Y_i : número de árboles en las clases i , con lo cual se llega a

$$\ln Y_i = \ln h + \ln k_1 + b(-d_i) \quad (3.142)$$

debiendo notarse que la variable independiente se vuelve $-d_i$ con lo cual los parámetros de la ecuación serán:

$$b_0 = (\ln h + \ln k_1); \text{ y } b_1 = +b \quad (3.143)$$

Por medio de la regresión lineal simple se despeja k_1 de las ecuaciones anteriores, previa la regresión lineal simple entre el logaritmo neperiano de las frecuencias contra el negativo de las marcas de clase. Para el ejemplo anterior se hizo la regresión:

$$\ln F_0 = b_0 + b_1(-d); \quad F_0 = \text{frec. obs.} \quad (3.144)$$

de la cual se obtuvieron los siguientes valores: $b_0 = 10.296$, $b_1 = 0.0734$, entonces $\ln k_1 = 10.296156 - \ln 10$, $\rightarrow k_1 = 2961.856$, con los cuales se llegó a la distribución ajustada: $Y_i = 29618.5641e^{-0.07339835909 \cdot d_i}$; y a los resultados de frecuencias estimadas de 2269.27, 1089.23, 522.83, 250.95, etc que complementarían la Tabla 3.8.

3.5.7.5 Distribución de Meyer Tipo II.

Aparte de la modificación reportada a la función de Meyer, se dan otras como el caso propuesto por (Loetsch, 1973; citado por Uribe, 1983), cuando las frecuencias logaritmizadas contra diámetro no están en una línea recta, si no que se encuentran incrementos sucesivos para el valor de q , que generan funciones de progresión geométrica creciente; para lo cual se propone en vez de usar el diámetro simple, introducir el diámetro al cuadrado, con lo cual:

$$N_i = k \exp(-bd_i^2) \quad (3.145)$$

que se resuelve de acuerdo con la propuesta de (Loetsch et al., 1961) similar a la usada para Meyer I. Además de lo anterior, (Vásquez, 1987) propone que en vez del exponente 2 se defina un modelo:

$$N_i = ke^{-ad^m} \quad (3.146)$$

que se denominaría por nomenclatura Meyer III, el cual es no lineal en sus parámetros por lo cual se correría en regresión no lineal.

3.6 Análisis de Goff y West.

Para el caso de los incrementos del diámetro medio, (Goff y West, 1975), en 49 pequeños rodales de Wisconsin, ilustran el desarrollo de la estructura de rodales, y encuentran que las funciones de densidad de distribuciones diamétricas en rodales viejos de estructura uniforme, son de forma sigmoide rotada, más que exponencial negativa, quedando esta para caracterizar áreas con parches localizados de rodales coetáneos.

En pequeños rodales donde las interacciones entre sotobosque y dosel son importantes, la curva típica de la estructura de los rodales muestra una fase en el sotobosque de lento crecimiento y alta mortalidad, una alta relativa vigorosa fase de dosel con baja mortalidad y una vieja fase de copas de crecimientos reducidos y mortalidades crecientes, lo que obliga replantearse modelos que no las consideren por lo menos en circunstancias similares, pues las interacciones del sotobosque pueden ser causas adicionales de variación en crecimiento y mortalidad. La expresión ya analizada (3.146), reescrita como:

$$Ydx = ke^{-bx} dx \quad (3.147)$$

para frecuencia de árboles en pequeños intervalos diamétricos, hacen que k y b se presten para caracterizar diversas estructuras. Meyer mostró que esta función podría utilizarse para medir el desplazamiento del diámetro y predecir la cantidad de cosecha necesaria en cada intervalo para mantener el bosque en función sostenida. La forma negativa de la exponencial de una estructura poblacional implica un porcentaje constante de pérdida en las clases del rodal, lo que ha permitido a varios investigadores hipotetizar sobre tales estructuras.

3.6.1 Modelo potencial de Goff y West. Más que una función de distribución como equívocamente se le conoce es una constante pesquisa con modelos desde el lineal simple, como uno de los casos analizados hasta modelos potenciales de forma:

$$Y = a + bX + cX^2 + \dots + nX^n \quad (3.148)$$

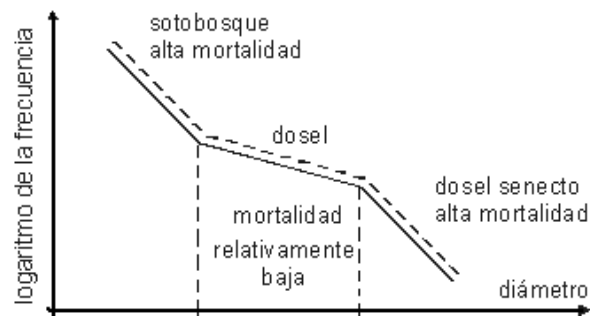


Figura 3.29. Equilibrio hipotético para las estructuras de rodales en tres fases. La curva punteada se comporta como sigmoide rotada.

Como conclusión al trabajo de Goff y West está la evidencia, para ellos, que las curvas formadas al graficar el logaritmo de la densidad contra el diámetro en rodales naturales de mediano a maduros son sigmoides rotadas, más razonables biológicamente como características de estructuras de equilibrio poblacional o de poblaciones de rodales viejos uniformes, más que la recta configurada por una exponencial negativa, cuando relativas áreas pequeñas o áreas estructuralmente uniformes son cubiertas en una muestra y, que por otro lado, la distribución exponencial negativa resulta apropiada para grandes regiones con rodales locales coetáneos de áreas aproximadamente iguales.

3.6.2 Modelo polinomial de Goff y West. De la visión de estos autores parece sugerirse el modelo

$$\log \text{frec}_i = b_0 + b_1 d_i + b_2 d_i^2 + \dots + b_m d_i^m \quad (3.149)$$

debido a que las razones de crecimiento y mortalidad, en poblaciones naturales de estructura uniforme, no resultan uniformemente distribuidas para los números de árboles por clases diamétricas.

Para algunos autores, ente ellos (de Barros et al, 1978), el criterio utilizado para seleccionar el grado apropiado de la ecuación es la correlación simple de las variables independientes d , d^2 , d^3 , ..., etc. con la variable dependiente $\log \text{frec}$, para lo cual se esperarían valores de $R \geq 0.95$, validos en su época, ahora se proponen procesos paso a paso como lo muestran los valores obtenidos para el ejemplo de la Tabla 3.10.

Tabla 3.9 Análisis de regresión de Goff y West para los datos de la Tabla 3.10.

Variable dependiente LOG (frec)					
Parámetro	Estimado	Error estándar	T	Valor p	
Const	4.11663	0.0593242	69.3921	0.0000	
D2	-0.0007709	6.5763E-05	-11.7224	0.0000	
D3	6.6453E-06	1.077E-06	6.18795	0.0080	
D4	-1.84E-08	4.71E-09	-3.89923	0.0080	
ANAVA					
Fuente	SS..	G.L	MS..	F	Valor p
Modelo	8.42405	3	2.8082	4306.7000	0.0000
Residual	3.91E-03	6	0.000652		
Total (Corr)	8.42796	9			

$R^2 = 99.95$, $R_{aj} = 99.93$. Error estándar de est = 0.0255345. Error promedio absoluto =

0.016409. Estadístico de Durbin-Watson = 3.11446. El valor de χ^2 resulta menor que el tabulado.

Tabla 3.10. Valores estimados por Goff y West.

Diámetro	Frecuencia observada	Frecuencia estimada	Chi cuadrada
35	2634	2694.26	1.35
45	1281	1223.20	2.73
55	518	531.24	0.33
65	240	230.01	0.43
75	93	102.42	0.87

85	50	47.90	0.09
95	25	23.78	0.06
105	12	12.55	0.02
115	7	6.97	0.00
125	4	3.99	0.00
		Suma	5.89

3.7 Distribución diamétrica para bosques no balanceados y disetáneos

Un bosque disetáneo es aquel en el cual se entremezclan árboles de diferentes edades y tamaños bien sea individualmente o en pequeños grupos, y se caracterizan como ya se mencionó por una curva diamétrica de J invertida. Las distribuciones de Meyer se aplicaron principalmente a bosques balanceados disetáneos, aquellos que producen un rendimiento sostenido mientras mantienen una estructura y volúmenes esencialmente constantes o, también asimilados como rodales en los cuales el número de árboles decrece en progresión geométrica constante.

El concepto de balanceado puede aplicarse tanto a bosques disetáneos como a bosques manejados de una clase de edad, cuando estas están especialmente separadas, en contraste a los bosques de selección. El término también se aplica a cualquier área con árboles comerciales de altas y bajas dimensiones comerciales en las cuales es posible extraer el crecimiento anual sin perturbarlas drásticamente.

Para determinar si un bosque es balanceado o no Meyer sugiere gráficos del número de árboles versus d en papel semilogarítmico. Si siguen una buena recta se considera un bosque balanceado (Leak, 1964). Este autor encontró incrementos en los cocientes q (también se dan decrementos de ellos) contra las marcas de clase consecutivas y propuso un método que pudiera caracterizar a los bosques balanceados y desbalanceados por medio de una transformación del cociente q/d en una relación lineal en que su pendiente b y su intercepto a representarán la distribución diamétrica. Su metodología consistió en generar una ecuación entre q_i y las marcas de clase y dependiendo de la perfección de las líneas rectas se podría dar un diagnóstico del balance o no del rodal o bosque, ya que la magnitud de b diferente de $b = 0$ refleja la lejanía de la condición balanceada y a se aproxima al q promedio de la condición de balance. Despejando

$$a = \bar{Y} - b\bar{X} \quad (3.150)$$

en que \bar{Y} iguala el cociente \bar{q} y, \bar{X} es el punto de clase promedia, entonces cuando $b_1 = 0$, $b_0 = \bar{Y}$, o sea que movimientos hacia cero del b_1 van hacia el balance y b_0 se aproxima al q que caracterizará la distribución diamétrica.

En resumen, el autor muestra que esta expresión de distribución diamétrica no especifica la densidad del rodal, pero que: 1) la magnitud de b_1 refleja el alejamiento de la condición balanceada y 2) que si un rodal o bosque se mueve hacia esa condición el intercepto b_0 tenderá hacia el valor promedio del q que caracterizaría la distribución balanceada. Para el ejemplo analizado se presenta la **¡Error! No se encuentra el origen de la referencia.:**

Tabla 3.1. Ajuste por el modelo de Leak

Marca de Clase	N° de árboles	Q	Parámetro entre marcas de clase
35	2634	2.056	40
45	1281	2.473	50
55	518	2.158	60
65	240	2.581	70
75	93	1.860	80
85	50	2.000	90
95	25	2.080	100
105	12	1.714	110
115	7	1.750	120
125	4		

De acuerdo con los datos: $q_i = 2.60442 - 0.007063X_i$.

Para el bosque balanceado: $b_0 = \bar{Y} - b_1 \bar{X}$, cuando $b_1 = 0 \rightarrow b_0 = 2.0765$. En este caso $b_1 \rightarrow 0$, desde este punto de vista parece tenderse hacia un bosque balanceado, con una $q = 2.0765$, Figura 3..

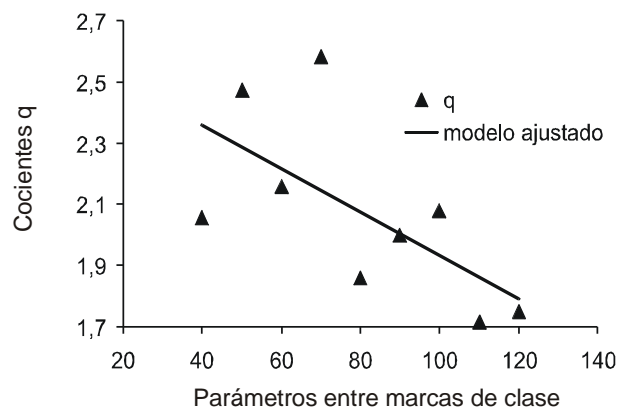


Figura 3.30. Modelo de relación de Leak entre cocientes de Liocourt y marcas de clase

Se encuentran algunas opiniones cuestionadoras acerca de las distribuciones como elementos técnicos de manejo. La mayoría de trabajos que han sustentado la necesidad de distribuciones diamétricas balanceadas para una mejor predicción o manejo del bosque pasaron por alto el efecto económico de tal búsqueda. Usando un bosque disetáneo de maderas duras en Norteamérica se planteó un estudio de la eficiencia de tal inversión o costo adicional en obtener la información con distribuciones de Weibull y concluyen no ser necesarias para una mayor sustentación de la producción sobre un ciclo fijo de corta, y además que los precios resultan demasiado sensitivos por precio de árbol y tamaño, (Bruce y Opalach, 1988). Ello implica investigar este aspecto para las condiciones específicas nuestras.

3.8 Pruebas de bondad de ajuste

3.8.1 Criterios usuales

La mayoría de investigadores del tema han recurrido para la evaluación de la bondad

desajuste a varios criterios entre los cuales son muy populares: la χ^2 (Chi cuadrada), el criterio de razón logarítmica de verosimilitud (Criterio G de Sokal y Rohlf) y la prueba de Kolmogoroff-Smirnoff. Si los etimados son máximo verosimilis el log de verosimilitu, loglik que se maximiza cuando el AIC es menor

3.8.1.1 Criterio de χ^2 La siguiente prueba es comúnmente utilizada como medida de desviación entre una distribución observada y una ajustada por alguna función:

$$\chi^2 = \frac{\sum_{c=1}^k (f - \hat{f})^2}{\hat{f}} \quad (3.151)$$

en que f : frecuencia observada, \hat{f} : frecuencia esperada, k : número de clases diamétricas, $k-1$: grados de libertad, (Zar, 1989), cuando la hipótesis nula es cierta y establece la igualdad de tales distribuciones. Algunos autores se refieren a ella como una prueba de pobreza de ajuste (Zar, 1984), puesto que una prueba es contundente cuando rechaza la hipótesis nula, y no es este el deseo al aplicarla a distribuciones diamétricas, lo que obligaría a calcular pruebas de potencia que a menudo resultan complicadas. A pesar de lo anterior es bastante usada incluso en paquetes sofisticados, pero presenta ciertos inconvenientes como, estar penetrada por el tamaño de la muestra, ser asintótica y que al extrapolar datos a una hectárea, como en nuestro caso, se puede perder confianza. Por ello se debe recurrir a otras propuestas en casos de duda.

3.8.1.2 Criterio G. Razón logarítmica de verosimilitud. El criterio de la razón logarítmica de verosimilitud se expresa como

$$\sum_{c=1}^k f_c \ln \left(\frac{f_c}{\hat{f}_c} \right) \quad (3.152)$$

y puede ser también escrito como

$$\sum_{c=1}^k f_c \ln f_c - \sum_{c=1}^k f_c \ln \hat{f}_c \quad (3.153)$$

con algunas desventajas como el desconocimiento de su distribución teórica, pero con la ventaja de que su dupla conocida como

$$G = 2 \sum_{c=1}^k f_c \ln \left(\frac{f_c}{\hat{f}_c} \right) = 4.60517 \sum_{c=1}^k f_c \log_{10} \left(\frac{f_c}{\hat{f}_c} \right) \quad (3.154)$$

se distribuye como una χ^2 y se maneja con sus mismos criterios, pero preferible a esta cuando $|f_c - \hat{f}_c| < \hat{f}_c$ (0.1)

Al usar (3.150) y (3.151), no se debe trabajar nunca con frecuencias esperadas menores que 5 árboles por clase.

3.8.1.3 Criterio de Kolmogoroff Smirnoff. Fue originalmente calculado para datos continuos, pero tiene un gran uso con datos discretos y nominales. Para cada X_i se registran las frecuencias observadas f_i y se determinan sus frecuencias acumulativas F_i , con las cuales se calculan las frecuencias relativas acumulativas así:

$$relF_i = \frac{F_i}{n} \quad (3.155)$$

que son simplemente la proporción de las muestras medidas $\leq X_i$. Para cada muestra X_i entonces se calcula la frecuencia relativa acumulada esperada, $relF_i$ y la prueba se conduce de acuerdo con las siguientes diferencias:

$$D_i = |relF_i - rel\hat{F}_i| \quad (3.156)$$

y

$$D'_i = |relF_{i-1} - rel\hat{F}_i| \quad (3.157)$$

asumiendo que $F_0 = 0$, con lo cual $D'_1 = rel\hat{F}_1$. La prueba de decisión estadística será:

$$D = \max[(\max D_i), (\max \hat{D}_i)] \quad (3.158)$$

contrastados contra valores críticos de unas tablas generalmente expresadas como $D_{\alpha, n}$, de tal forma que si $D > D_{\alpha, n}$, se rechaza H_0 . Algunos autores lo presentan como

$$K = \max \left| \frac{F}{N} - F(x) \right| \quad (3.159)$$

en el cual F/N : función empírica de distribución de muestra $x_i (i=1, 2, \dots, N)$, $F(x)$: función teórica de la distribución. Existe una modificación a este, cuando se trata de datos continuos, conocido como prueba de Watson para medidas angulares que escapa a este texto. Es una prueba muy rigurosa.

3.8.2 Consideraciones generales

Estudios recientes, con base en simulaciones completas, sugieren que en muchas situaciones el uso de los anteriores criterios puede resultar inapropiado, ya que en la mayoría de los casos el modelo de distribución de probabilidades es más bien una familia de distribuciones indizadas por varios parámetros.

Un modelo construido de forma que los parámetros se relacionen con otras variables agregadas del rodal, como edad, densidad, índice de sitio, etc. se referirá más concretamente a cada tipo de rodal que al bosque general, y las pruebas de bondad de ajuste diseñadas para probar la hipótesis de que la muestra pertenece a una distribución específica podrían no ser aptas para resolver el problema. Esto no ocurre tan drásticamente en los primeros pasos de la modelación, donde simplemente se desea ver si se da cierto ajuste con los datos, pero en etapas posteriores, cuando las distribuciones se relacionan con las variables agregadas, al usar información de muchas parcelas surgen grandes dificultades y el problema del ajuste es complicado por las diversas formas de obtención de los parámetros, (Reynolds et al, 1988).

Cuando se considera una variable aleatoria X con una cierta función de distribución acumulativa $F(x)$, se plantea para la bondad de ajuste una hipótesis nula como: $H_0: F(x) = F_0(x)$ en que $F_0(x)$ es completa o parcialmente especificada. En tal caso las citadas pruebas mencionadas pueden ser usadas. Pero si θ_0 : en $F(X, \theta_0)$ involucra parámetros desconocidos, entonces las probabilidades del modelo dependerán de ellos. De acuerdo con ello la prueba de χ^2 de Pearson resultaría buena para datos discretos o continuos que puedan ser naturalmente agrupados en clases, y se puede aplicar usando los

criterios críticos estándar si los parámetros han sido propiamente estimados. La desventaja es que para datos continuos no es tan poderosa como otras pruebas de bondad de ajuste, lo que implicaría usar la corrección sugerida por Yates para continuidad:

$$\chi_c^2 = \sum_{i=1}^n \sum_{j=1}^k \frac{(|f_{ij} - \hat{f}_{ij}| - 0.5)^2}{\hat{f}_{ij}} \quad (3.160)$$

Cuando se trata de evaluar estadísticos derivados de funciones de distribuciones empíricas de muestras como

$$F^*(x) = \frac{\text{Número de } X_i \leq x}{n} \quad (3.161)$$

la prueba de K-S es un buen estadístico porque mide la distancia entre la función teórica y la observada.

3.7.1.5 Otros criterios

Es bueno para complementar el problema, reportar otros criterios más complejos, según el mismo autor, como los basados en el estadístico de Cramér-von Mises:

$$CM = n \int_{-\infty}^{\infty} |F^*(x) - F_0(x)|^2 dF_0(x) \quad (3.162)$$

y en el estadístico de Anderson-Darling (AD):

$$AD = n \int_{-\infty}^{\infty} \frac{|F^*(x) - F_0(x)|^2}{F_0(x)[1 - F_0(x)]} dF_0(x) \quad (3.163)$$

en que la expresión $dF_0(x)$ puede interpretarse como $f_0(x)dx$, si $F_0(x)$ tiene densidad $f_0(x)$. Cuando $F_0(x)$ es completamente especificada, los estadísticos K-S, CM, AD son distribuciones libres bajo la hipótesis nula.

Cuando aparecen las variables agregadas ya no son libres, y una de sus evaluaciones acudiendo al uso de cantiles debe ser estimada para cada familia de distribuciones. Una técnica que no deja de usarse es el ajuste gráfico, juzgado normalmente a ojo, incluso por medio de patrones para su comparación, papeles de probabilidades, etc. De todos modos como problema final se expone el hecho de que a no ser que el tamaño muestral sea lo suficientemente grande, la potencia de las pruebas usuales no será alta a menos que $F(x)$ este muy alejada de $F_0(x)$. Surgen otros problemas como los diámetros correlacionados y otras situaciones que pueden agravar aun más la situación.

3.7.1 Uso de una propuesta de un índice de error

Dadas las circunstancias anotadas, propone (Reynolds et al, 1988), un indicador de la cercanía del modelo a los datos, estimador- p , que consiste en una suma ponderada de diferencias absolutas entre números de diámetros predichos y observados en cada clase, aconsejable de acuerdo con la naturaleza específica del problema tratado, y de las necesidades y criterio del usuario, ya que por ejemplo al usar una distribución es más notorio un desajuste en árboles de 50 cm de d que en los de 5 para ciertos usos o viceversa.

Para desarrollar el índice se dividen los datos en k clases de interés, dígame I_1, I_2, \dots, I_k . Sea $w(x)$ un ponderador asociado con cada diámetro x , podría ser el volumen total, el precio alcanzado de acuerdo con un rango de utilización, el valor alcanzado por una determinada especie en bosques mezclados, etc. Sea $F_e(x)$ la función estimada de distribución acumulada de los diámetros en una parcela predicha por el modelo, y sea $F^*(x)$ la distribución observada. El índice propuesto de error es

$$e = N \sum_{j=1}^k \left| \int_{I_j} w(x) dF_e(x) - \int_{I_j} w(x) dF^*(x) \right| \quad (3.164)$$

en que N : número de árboles por hectárea y la expresión:

$$\int_{I_j} w(x) dF^*(x) \text{ equivalente a } \sum_{x_i \in I_j} \frac{w(x_i)}{n} \quad (3.165)$$

donde x_1, x_2, \dots, x_n son los n diámetros en la parcela. Si por ejemplo $w(x)$ corresponde al volumen, entonces $N \int_{I_j} w(x) dF_e(x)$, es el volumen predicho/ha para los árboles en la j -ésima clase. Se debe notar que e es una suma de errores absolutos tal que una sobre o subpredicción en una clase no cancela lo contrario en otra, lo que significa que el índice solo será pequeño cuando el modelo prediga bien para todas las clases diamétricas, por lo cual será de gran ayuda examinar las diferencias componentes individuales para tratar de identificar los problemas del modelo. Otra posibilidad es la conversión de este índice a porcentaje o a una base proporcional dividiendo por

$$N \int w(x) dF^*(x) \quad (3.166)$$

en el cual el usuario contempla el porcentaje de error dispuesto a correr, sin que ello tenga necesariamente algún significado usual.

El índice de error puede ser aplicado a cualquier distribución diamétrica, basadas o no en probabilidades y que el modelador o usuario pueda seleccionar la función de ponderación y las clases que le convengan para un determinado propósito. Existen situaciones aun mas complejas, como conocer número de árboles y área basal, en cuyo caso puede resultar difícil un modelo que prediga bien ambas características, por ello se acude a las distribuciones calibradas diamétricas con información adicional (Kangas y Maltano, 2000) que escapan al nivel de este texto.

3.8 Medidas de densidad promedia en rodales

Las medidas de densidad propician herramientas importantes para la toma de decisiones en los manejos y aprovechamientos forestales. Se agrupan en dos categorías básicas: densidades promedias por grupos o, medidas puntuales, de acuerdo con la expresión del grado de ocupación de un área determinada, o del nivel de competencia que se de entre árboles aislados. De estas las más importantes se refieren a rodales, como número de árboles por hectárea, distribuciones diamétricas, área basal, etc., ya que las individuales son más específicas. Para que tengan validez plena, se debe buscar que sean fácil y objetivamente medibles e interpretables biológicamente.

3.8.1 Número de árboles y área basal por unidad de área. En general para la mayoría de situaciones silviculturales acompañadas de otras medidas de caracterización, el número de árboles por hectárea resulta una buena medida de densidad, pero especialmente en rodales coetáneos de edad conocida. Esta variable jerarquiza bastante bien en ecuaciones de crecimiento sobre todo de rodales manejados. La misma situación se presenta para el área basal a la altura del pecho.

El área basal por unidad de área, se estudiará en el capítulo siguiente, pero es de las más usadas como medida de densidad.

3.8.2 Índices de densidad por unidad de área o rodal. Algunas medidas de la ocupación promedia de rodales han sido propuestas, con referencia a condiciones limitantes predeterminadas, entre número de árboles/ha y tamaños promedios de estos.

Especialmente conocido es el índice de Reyneke-Frothingham con base en el trabajo de Frothingham (Reyneke, 1933), expuesto por (Clutter et al, 1983). el SDI (*Stand Density Index*) en su sigla inglesa, se puede expresar así: en rodales coetáneos plenamente ocupados, las relaciones limitantes entre el número de árboles/ha N y el diámetro cuadrático medio \bar{D}_q con corteza, se relaciona a menudo linealmente en coordenadas logarítmicas.

Ello quiere decir que para un \bar{D}_q dado, existe un límite para el número esperado de árboles/ha, a pesar de la edad o la calidad del sitio o dicho de otra forma, los rodales al límite se asume que experimentan el mismo grado de ocupación. En coníferas y aun para maderas duras Reyneke observó tal relación en rodales coetáneos de varias especies, con una pendiente de la línea limitante aproximadamente igual a -1.6, para sus datos. La relación limitante se presenta entonces de la forma:

$$N = \alpha \bar{D}_q^\beta \quad (3.177)$$

α y β parámetros de la relación. El mismo autor expresa que el grado de ocupación común en las circunstancias anotadas se alcanza cuando $\bar{D}_q = 10$ pulgadas (25.6 cm), con lo cual:

$$SDI = \alpha 10^\beta; \text{ o } SDI = \alpha 25.6^\beta \quad (3.178)$$

despejando α de (3.177) y llevándola a (3.178), el índice citado queda como:

$$SDI = N \left(\frac{10}{\bar{D}_q} \right)^\beta \quad (3.179)$$

lo que quiere decir que todos los rodales con la misma proporción de número de árboles limitantes/ha presentan el mismo SDI a pesar del diámetro promedio de los rodales. De acuerdo con (3.79) para rodales de un determinado \bar{D}_q el SDI se incrementa proporcionalmente con el N y consecuentemente con el área basal: G . Para el ajuste de los modelos se requiere de rodales plenamente ocupados. El modelo obviamente es de la forma:

$$\ln N = \ln \alpha + \beta \ln \bar{D}_q \quad (3.180)$$

Por ejemplo, para rodales de plantaciones medidos en 40 rodales del norte de Antioquia en *P. patula* en parcelas sin manejo de 17 años se encontró un $\bar{D}_q = 17 \text{ cm}$, y el modelo: $N = 12820.36 \bar{D}_q^{-0.76425}$ con $R^2 = 0.328$, o sea $\alpha = 12820.4$, $\beta = -0.7643$, con lo cual el $SDI = 12820.36 * 17^{-0.76425} = 1471 \text{ arb/ha}$.

Para un rodal dado con $N = 800 \text{ arb/ha}$ y un $D_q = 25 \text{ cm}$, se espera un $SDI = 800 \left(\frac{17}{25} \right)^{-0.76425} = 1074 \text{ arb/ha}$ con una ocupación plena 1095 árboles/ha como limite.

En la actualidad se trabajan otras propuestas como los diagramas de densidad para el manejo forestal que complementan lo expuesto, especialmente las propuestas para *Balsam Fir* de (Penner et. al., 2002) y los de *White Spruce* (Saunders y Puettman, 2000) que se recomienda consultar.

3.8.3 Índices de densidad razón árbol-área (RAA). Otra propuesta, de Chisman y Schumacher, también citados por (Clutter, 1983) para estimar la densidad contempla modelos entre el área ocupada por un árbol, por la proyección horizontal de su copa, área de copa, contra su d :

$$A = \beta_0 + \beta_1 d + \beta_2 d^2 \quad (3.181)$$

y, el área total ocupada por n árboles/ha sería la sumatoria de las áreas ocupadas por ellos:

$$\sum_{j=1}^n A_j = \beta_0 n + \beta_1 \sum_{j=1}^n d_j + \beta_2 \sum_{j=1}^n d_j^2 \quad (3.182)$$

Si se seleccionan N parcelas de interés de una hectárea, o convertidas a ella, aparecerán en cada una de estas, los valores n_i, d_{ij} para $(i = 1, 2, \dots, N)$, $\sum_{j=1}^{n_i} d_{ij}, \sum_{j=1}^{n_i} d_{ij}^2$ con los que será posible obtener estimados de los parámetros de (3.181) por medio del ajuste mínimocuadrático de

$$\sum_{j=1}^n \left(1 - b_0 n_i + b_1 \sum_{j=1}^n d_j + b_2 \sum_{j=1}^n d_j^2 \right)^2 \quad (3.183)$$

que podrán usarse para evaluar la ecuación (3.182) para cualquier rodal dado con n y $\sum_{j=1}^n d_j, \sum_{j=1}^n d_j^2$ conocidos. El RAA resultante será $\left(\sum_{j=1}^n A_j \right)$, una medida de la densidad de rodales relativa a la relación promedia generada en la muestra original y con buen sentido si la muestra original se restringe a rodales plenamente ocupados.

Ejemplo suponga que $b_0 = 0,00006$, $b_1 = 0,00002$ y $b_2 = 0,0000045$. El área ocupada por un árbol de 25 cm de diámetro sería de 0,0033725 has, al usar (3.191) y para un rodal hipotético con $n = 400$ árboles/ha, $\sum_{j=1}^n d_j = 9600$, $\sum_{j=1}^n d_j^2 = 100000$, se llegaría a un RAA = 0.666, usando (3.183) con los estimados.

3.8.4 Índices de espaciamiento o espaciamiento relativo (IER).

Este índice se conoce como índice de Hart, pero evolucionó por el trabajo de varios autores al juzgar ocupaciones deseables de los rodales, estudiar las razones entre distancias de los árboles y las alturas promedio de los árboles dominantes. Se define entonces como

$$IER = \frac{\text{distancia promedio entre árboles}}{\text{altura promedio de dominantes}} \quad (3.184)$$

por ejemplo, para árboles sembrados en cuadro, el índice sería:

$$IER = \frac{100}{\sqrt{N\bar{H}_d^2}} \quad (3.185)$$

en que \bar{H}_d es la altura promedio de árboles dominantes en metros, y N el número promedio de árboles/ha sembrados en cuadro. En general se puede escribir este índice como

$$IER = \frac{K}{\sqrt{N\bar{H}_d^2}} \quad (3.186)$$

K puede lograrse de varias formas, incluidos los métodos PPS que serán abordados en el próximo capítulo.

El hecho más notable de este índice es que parece darse que a pesar de la edad inicial y de la calidad del sitio, las especies parecen aproximarse asintóticamente a un espaciamiento mínimo relativo, a pesar del sitio y de la edad inicial, en ciertas circunstancias, a medida que ellos envejecen, como se aprecia en la Figura 3.31, tomada de (Clutter et al, 1963) de ahí las múltiples utilidades que se le asigna en manejos silviculturales.

Para una curva de crecimiento de alturas de dominantes, dada, este límite mínimo establece un máximo número de árboles por ha esperado a una edad dada, También anota (Parker, 1978, citado por Clutter, 1983) que una combinación de curvas de crecimiento de alturas y una tendencia temporal del IER, implícitamente define una curva de supervivencia.

El IER se ha utilizado para controles de la densidad en rodales intensamente manejados, incluso tablas de producción en las cuales la altura promedio de dominantes reemplaza las variables usuales como edad y calidad de sitio para la predicción del crecimiento.

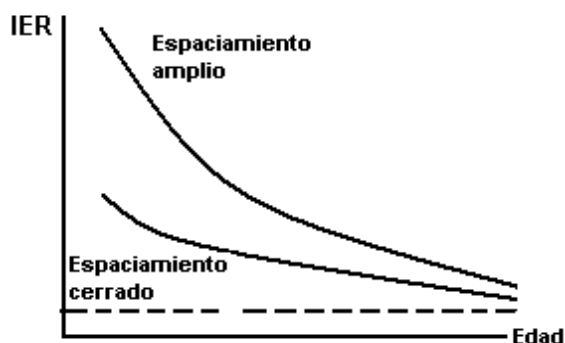


Figura 3.31 Tendencia con el tiempo del espaciamiento relativo en rodales con diferentes densidades

Como un ejemplo del uso del IER, considérese un rodal coetáneo con una altura media de dominantes de 17 m y 600 árboles por hectárea, o sea separados aproximadamente 4.1 m

con lo cual el IER = 0.24. Considere dos cortas hasta dejar un espaciamiento relativo de 0.28 en el rodal remanente. La primera corta será cuando la altura promedio de dominantes alcance 18 m y la segunda cuando alcance 22. El numero de árboles/ha dejados después de la primera corta será, acudiendo a la ecuación (3.185)

$$N = \frac{10000}{(IER * \bar{H}_d)^2} \quad (3.187)$$

Para la primera corta $N_1 = 393$ árboles y la segunda, $N_2 = 264$.

En planes de manejo los regímenes de cortas se diseñan, como acá, para obtener un IER preespecificado, y los intervalos de corta se programan para coincidir con determinados incrementos en las alturas promedias de dominantes.

BIBLIOGRAFIA

- ARTEAGA, FRANCISCO. 1987. Modelos matemáticos para la descripción de la estructura diamétrica de varios tipos de cativales. Medellín, Departamento de Ciencias Forestales, Universidad Nacional de Colombia. 165 p.
- BAYLEY, R. Y DELL, T. 1973. Quantifying diameter distributions with the Weibull function. *For. Sci.*, 19 (2): 97-104.
- BAYLEY, R.L. 1973. Weibull model for *Pinus radiata* diameter distributions. IUFRO. Vancouver. 51-61.
- CANAVOS, G.C. 1987. Probabilidad y Estadística. México. Mc Graw-Hill. 651 p.
- CLUTTER, L. ET AL. 1983. Timber Management: A Cuantitative Aproach. John Wiley and Sons. N.Y. 333 p.
- CONTENTE ET AL. 1979. Comparação de modelos descritivos da distribuição diamétrica en una floresta tropical. *Floresta (Sao Paulo)*. 10 (2): 19-33.
- CHAGAS, C., RIBEIRO, J.C. AND L. COUTO. 1983. Emprego da distribuição diamétrica determinação da intensidade de cote en matas naturais submetidas ao sistema de seleção. *Arvore*, 7 (2): 110-122.
- DE BARROS, ET. AL. 1978. *Revista Floresta*.
- DEL VALLE J.I. Y A. LEMA. 1999. Crecimiento de cohortes de árboles coetáneos en rodales espacialmente dispersos: el caso de *Campnosperma panamensis*. *Revista de la Academia de Ciencias Exactas, Fisicas y Naturales*. Santafé de Bogotá. XXIII (87): 249-250.
- DELL, T.R. 1973. Why not the Weibull for diameter distribution modeling. IUFRO. Vancouver. 51-61.
- DIAZ, A., Y A. GUTIÉRREZ. 1985. Curso: Estadística General. Medellín U. de A. 199 p.
- DUQUE, C. Y S., BUILES. 1986. Metodología para tablas de volumen integral con base en ecuaciones de conicidad para bosques heterogéneos. Medellín. Depto de Ciencias Forestales U.N. Tesis de grado. 139 p.
- EMLEM, J.T. 1967. A rapid method for measuring arboreal canopy cover. *Ecology*, 48 (1): 158-160.
- GADOW, VON, K. 1983. Fitting distributions in *Pinus patula* stands. Stellenbosch. Suid-Afrikaanse Bosboutydskrif. 20-29
- GOFF, F.G. Y D. WEST. 1975. Canopy - understory interaction effects on forest population structure. *For. Sci.* 21 (2): 95-108.
- GOMEZ, HERNÁN. 1982. Notas de clase. s.d.
- HARTER, SD. 1964. Tables of the incomplete Γ function, *Biometrika*, s.d.
- KANGAS A Y M. MALTANO. 2000. Calibrating predicted diameter distribution with additional information *For Sci.*, 46 (3): 390-396.
- LEAK, W. 1965. The J-shaped probability distribution. *For Sci.*, 11 (4): 405-409.
- , 1964. An expression of Diameter Distributions for Unbalanced, uneven-aged stands and forests. *For. Sci.* 10 (1).
- LEMA, A. 1995. Elementos de Geometría Fractal aplicados al cálculo de errores en la cubicación con forcípula de árboles apeados. *Crónica Forestal y del Medio Ambiente*, # 10 : 55-65.
- LOETSCH, F., ZOHRER, F. AND K.E. HALLER. 1973. Forest inventory. BLV Verlagsgesellschaft, Munich. 469 p.
- MOSER, JR. JOHN. 1976. Specification of density for the inverse J-shaped diameter distributions. *For. Sci.*, 22:177-180.
- PEARSON, L. E. 1957. Tables of the incomplete Γ function, *Biometrika*.
- PENNER, M., SWITT, R. GAGNON Y J. KNOX. 2002. A stand density management diagram for Balsam Fir in New Branswick. Canadian Forest Service. 9p.

- REYNOLDS, JR.M., BURK, T. AND W.C.HUANG. 1988. Goodness-of-fit tests and model selection procedures for diameter distributions models. *For. Sci.*, 34: 373-399
- REINEKE, L. H. 1933. Perfecting a stand-density index for even-aged forest. Washington. *Journal of Agricultural Research*: 46 (7): 627-638.
- SANDRASEGARAN, K. 1971. Height-diameter-age multiple regression models for *Rhizophora apiculata* Bl. In: Matang: manglares, tapping, west Malaysia. *Malayan Forester*, 34 (4). 260-275 .
- SETH, S. SATYAMURTHI, K. & S., NEGI. 1985. A comparative study of some distribution functions as applied to diameter at breast height in even- aged stands. *Indian Forester*. 20-32.
- SNEDECOR, G. Y W. COCHRAN. 1971. Métodos estadísticos. México. C.E.C.S.A. 703 p.
- SAUNDERS, M. Y K.J PUETTMAN. 2000. A preliminary White Spruce density management diagram for the Lake States. Minnesota. U. of Minnesota. Staff paper series N° 145: 15 p.
- UNESCO. 1980. Ecosistemas de los Bosques Tropicales. UNESCO, CIFCA. 771 p.
- URIBE, ALBERTO. 1985. Comportamiento de las distribuciones diamétricas de frecuencia de bosques disetáneos. Medellín, Departamento de Ciencias Forestales, Universidad Nacional de Colombia. Seminario, 90 p.
- VANCLAN, J. K. 1994. Modeling forest grow and vield. Applications to mixed tropical forest. CAB/International, Wallingford. UK. 312 p.
- WILK, M.B., R. GANADESIKAN Y J. HUYETT. 1962. Estimation of parameters of the gamma distribution using order statistics. *Biometrika*: 49: 525.
- ZAR, H JERROLD. 1984. Biostatistical analysis. 2ª Ed. Prentice-Hall. New Jersey. 718 p.
- ZARNOCH, S. Y T., DELL. 1985. An evaluation of percentile and maximun likelihood estimators of Weibull parameters. *For. Sci.*, 31 (1): 260-268.
- of Weibull parameters. *For. Sci.*, 31 (1): 260-268.