

# 非参数核估计在探讨天然林直径结构规律中的应用\*

王雪峰<sup>1)</sup> 管青军<sup>2)</sup> 武利军<sup>2)</sup> 戴海平<sup>2)</sup>

(1) 中国林业科学研究院资源信息所, 100091, 北京; 2) 内蒙古得耳布尔林业局, 022357, 内蒙古根河;  
第一作者 30 岁, 男, 助理研究员)

**摘要** 应用非参数核密度估计理论和方法, 对从吉林省汪清林业局抽取的 12 块天然林样地进行直径结构模拟, 结果表明: (1) 当 Weibull 及其它分布函数能很好模拟的样地时, 核估计方法也能对该样地进行模拟, 并且要优于分布函数法; (2) 在分布函数不能描述林分时, 核估计方法仍能描述该林分, 并能取得很好的效果; (3) 不论是天然林还是人工林均可用核估计方法对林分直径结构进行模拟, 非参数方法可能成为一种有用的方法。

**关键词** 直径分布; 天然林直径结构; 核密度估计; 非参数统计模型; 林分表  
**分类号** S758.5

直径结构是最基本的林分结构, 不仅因为林分直径便于测定, 而且因为它直接影响树木的树高、干形、材积、材种及树冠等因子的变化, 能为许多森林经营技术及测树制表提供理论依据。在近百年的直径结构研究中, 人们把重点放在对同龄林直径结构的研究上, 但随着研究的深入, 随着目前生态环境的日益恶化, 人们把重点转移到陆地最大的生态系统——森林生态系统中, 探讨天然林直径结构规律, 将是其重要内容之一。天然林多为异龄复层结构, 直径分布曲线类型多样, 变化复杂, 所以应该选择适应性强、灵活性大的分布函数。美国迈耶<sup>[1]</sup>提出用指数方程  $n_i = a \cdot e^{-bd_i}$  来模拟均衡异龄林, 但此种模型仅对反 J 型曲线有效。许多学者都认为, 不论近似正态的直径分布或左偏、右偏乃至反 J 型的递减直径分布, 使用 Weibull 分布都可以取得良好的结果<sup>[2~4]</sup>。通过本研究也表明, Weibull 分布在多数情况下能取得很好的结果, 但对于直径分布曲线波动大的天然林林分, Weibull 曲线变得无能为力。这是因为该林分并不属于某一分布族。本文介绍了以数据为基础的非参数核方法<sup>[5~7]</sup>在模拟天然林直径结构中的应用, 通过与 Weibull、对数正态分布进行比较, 直观地给出了与分布族无关的核方法的巨大优点, 从而为研究天然林结构提供了简单通用的有利工具。

## 1 材料与方法

### 1.1 材料来源及样地概况

研究材料取自吉林省汪清林业局。该局位于 129°56' ~ 131°04' E, 43°05' ~ 43°40' N, 海拔 360 ~ 1 477 m, 全区越琿春河、绥河及嘎呀河流域 3 个水系, 导致全区地形复杂; 年降水量 547

\* 本项目为 1998 ~ 1999 年中国林业科学研究院基金项目“非参数统计在直径结构及林分预测中的应用研究”的部分内容。

mm, 年均气温 3.9, 无霜期 137 d; 土壤以暗棕壤为主, 还有少量沼泽土、草甸土穿插其中。本局属长白山植物区系, 种类繁多, 结构复杂。森林植被主要代表树种深山区以红皮云杉 (*Picea asperata* Koraiensis Nakai) 为主, 有红松 (*Pinus koraiensis* Sieb. et Zucc.)、臭松 (*Abies nephrolepis* (Trautv.) Maxim.) 等。阔叶树多为椴树 (*Tilia mandshurica* var. *ovalis* (Nakai) Liou et Li)、柞树 (*Xylosma japonicum* (Walp.) A. Gray)、枫桦 (*Betula costata* Trautv.)、色树 (*Acremono* Maxim.)、白桦 (*Betula platyphylla* Suk.) 等。林相以针阔混交为主, 林下灌木主要有毛榛子 (*Corylus mandshurica* Maxim.)、忍冬 (*Lonicera* spp.)、胡枝子 (*Lespedeza bicolor* Turcz.)、珍珠梅 (*Sorbaria sorbifolia* (L.) A. Br.)、绣线菊 (*Spiraea salicifolia* L.)、粉枝柳 (*Salix rorida* Laksch.)、悬钩子 (*Rubus idaeus* L.)、黄刺玫 (*Rosa xanthina* Lindl.)、刺五加 (*Acanthopanax senticosus* (Rupr. et Maxim.) Harms) 等。林下草本植物主要有莎草 (*Cyperaceae* spp.)、苔草 (*Cares* spp.)、蕨类 (*Pteridium* spp.)、木贼 (*Equisetum hiemale* L.)、小叶樟 (*Deyeuxia angustifolia* (Kom.) Y. L. Chang) 等。

汪清局内采用系统抽样在全区建有 1 500 块固定样地, 面积 0.06 hm<sup>2</sup>。本研究从中抽取 99、111、187、555、561、618、622、630、643、1 231、1 235、1 403、1 426 号样地共 12 块, 取 1987 年直径数据用于分析天然林直径结构。

## 1.2 分析方法

当  $X_1, X_2, \dots, X_n$  是概率密度为  $f(x)$  的总体样本, 则任一区间  $[a, b]$  内的频数  $m$  可由下式得到:

$$m = n \int_a^b f(x) dx \quad (1)$$

此式为直径结构模拟中计算频数的基本公式。取  $f(x)$  的一个非参数核密度估计(核函数的定义见参考文献[7])  $f_n(x)$ :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \quad (2)$$

则

$$\begin{aligned} m &= n \int_a^b f_n(x) dx = n \int_a^b \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right) dx \\ &= \frac{1}{h_n} \sum_{j=1}^n \int_a^b K\left(\frac{x-X_j}{h_n}\right) dx = \frac{1}{h_n} \sum_{j=1}^n \int_a^b K\left(\frac{x-X_j}{h_n}\right) dx \end{aligned} \quad (3)$$

其中  $h_n$  为窗宽, 如果样本是分组数据  $d_i, n_i (i=1, 2, \dots, k)$ ,  $d_i$  为  $i$  径阶中值,  $n_i$  为  $i$  径阶株数即实验频数,  $k$  是径阶数。此时核密度修改如下:

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^k n_i K\left(\frac{x-d_i}{h_n}\right) \quad (4)$$

式中  $n = \sum_{i=1}^k n_i$ 。  $[a, b]$  内频数相应变为:

$$\begin{aligned} m &= n \int_a^b f_n(x) dx = n \int_a^b \frac{1}{nh_n} \sum_{j=1}^k n_j K\left(\frac{x-d_j}{h_n}\right) dx \\ &= \frac{1}{h_n} \sum_{j=1}^k n_j \int_a^b K\left(\frac{x-d_j}{h_n}\right) dx = \frac{1}{h_n} \sum_{j=1}^k n_j \int_a^b K\left(\frac{x-d_j}{h_n}\right) dx \end{aligned} \quad (5)$$

韦布尔分布、对数正态分布的概率密度函数分别为 (6)、(7) 式。

$$f(x) = \begin{cases} 0 & x \leq a \\ \frac{c}{b} \left(\frac{x-a}{b}\right)^{c-1} \exp\left[-\left(\frac{x-a}{b}\right)^c\right] & x > a \end{cases} \quad (6)$$

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{x\sigma} \frac{1}{2\pi} \exp\left[-\frac{(\ln x - a)^2}{2\sigma^2}\right] & x > 0 \end{cases} \quad (7)$$

将 (6)、(7) 式分别代入 (1) 式, 可以得到韦布尔分布、对数正态分布在某一区间  $[x_1, x_2]$  内的理论频数  $p_i, i = 1, 2, \dots, k$ , 如 (8)、(9) 式。

$$m = n \left\{ \exp\left[-\left(\frac{x_1 - a}{b}\right)^c\right] - \exp\left[-\left(\frac{x_2 - a}{b}\right)^c\right] \right\} \quad (8)$$

$$m = \frac{n}{\sigma} \frac{1}{2\pi} \int_{x_1}^{x_2} \frac{1}{x} \exp\left[-\frac{(\ln x - a)^2}{2\sigma^2}\right] dx \quad (9)$$

进一步在假设  $H_0$ : “样本是由分布为  $p$  的总体中抽出的” 成立时, 统计量

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - p_i)^2}{p_i} \quad (10)$$

服从自由度  $f = k - r - l$ 。其中  $r$  为分布中的参数个数,  $n_i$  为实验频数。对于非参数方法,  $r = 0$ 。

## 2 结果与分析

分别样地把原始数据按 1 cm 整化, 然后按 (5)、(8)、(9) 式计算各径阶株数即理论频数。再由 (10) 式计算卡方统计量, 将计算结果整理成表 1。

表 1 卡方检验

样地号	株数	卡方统计量			临界值		结 论		
		韦布乐 W	对数正态 L	核估计 K	参数估计	核估计	W	L	K
99	83	28.055 8	39.389 2	16.566 4	31.410 4	33.924 5	√	×	√
630	96	27.008 3	42.358 3	12.750 9	41.337 1	43.773 0	√	×	√
622	90	41.787 9	54.097 2	17.575 1	37.652 5	40.113 3	×	×	√
618	101	33.022 9	50.457 6	19.075 1	47.399 9	49.801 8	√	×	√
561	39	15.720 6	21.851 9	8.036 5	27.587 1	30.143 5	√	√	√
555	54	28.683 2	40.314 8	16.201 8	32.670 6	35.172 5	√	×	√
187	120	89.318 0	173.022 4	21.294 9	65.170 8	67.504 8	×	×	√
1 426	88	23.792 6	38.933 5	10.078 9	31.410 4	33.924 5	√	×	√
1 403	56	32.490 7	43.876 1	16.159 4	32.670 6	35.172 5	√	×	√
1 235	198	50.200 7	41.865 1	15.795 5	33.924 4	36.415 0	×	×	√
1 231	94	43.756 7	66.826 5	19.780 9	43.773 0	46.194 2	√	×	√
111	108	55.724 5	91.302 0	22.371 1	48.602 4	50.998 5	×	×	√

注: 表中核函数计算所选用的核及窗宽分别为 K6, 0.41; K1, 0.68; K4, 0.70; K2, 1.28; K3, 0.82; K5, 0.59; K3, 0.77; K2, 1.19; K5, 0.85; K1, 0.85; K6, 0.34; K4, 0.92 (见参考文献[3])。 “√” 表示通过检验, “×” 表示未通过检验。临界值由卡方分布表得到。

从表 1 可见,在抽取的 12 块样地中,用 Weibull 模拟的 8 块样地通过卡方检验,LogNormal 分布只有 561 号一块样地通过检验。与此相反,采用非参数核密度方法进行模拟,12 块样地全部通过检验。值得注意的是,对于用分布函数及核方法模拟均通过检验的样地,核方法模拟精度更高。比如 99 号样地,Weibull 分布卡方统计量为 28.055 8,Kernel 估计为 16.566 4;1403 号样地 Weibull 分布、Kernel 方法均通过卡方检验,其卡方统计量分别为 32.490 7、16.159 4,可以看到,Kernel 方法卡方统计量最小,其它几块样地也同样得到这种结论。可以看到,Kernel 卡方统计量变动幅度基本在 10~20 之间,而与样地及样本数关系不是很大。

为了进一步分析天然林的直径结构,我们将 Weibull 分布、LogNormal 分布均不能模拟的 4 块样地数据以直方图形式给出,见图 1~4。

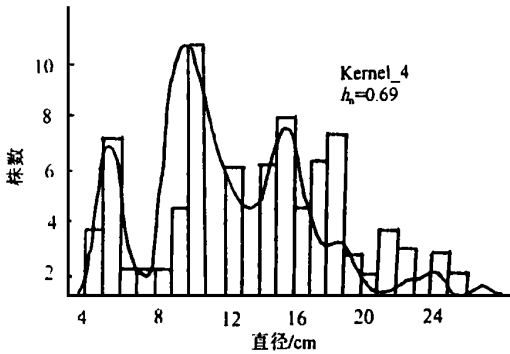


图 1 622 号样地核函数模拟

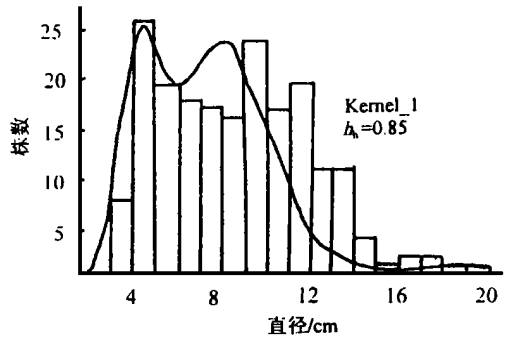


图 2 1235 号样地核函数模拟

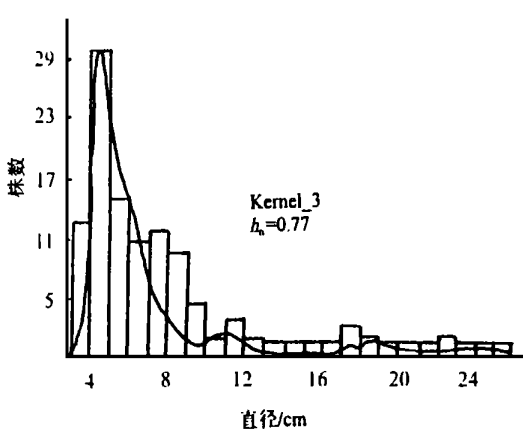


图 3 187 号样地核函数模拟

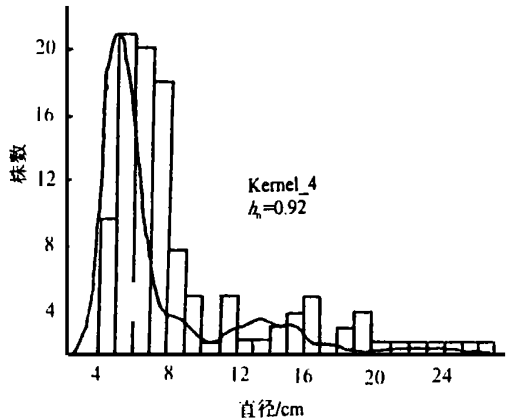


图 4 111 号样地核函数模拟

从图 1~4 清楚可见,这几块样地直径结构相对复杂,呈现很不规则的形状。大体上可认为图 1 是 4 峰结构,图 2 为双峰结构;而图 3 和图 4 相对离散程度较大,树木大多集中在小径阶,但较大径阶仍有树木存在。这些客观存在,很难满足某一分布族所要求的条件,自然用 Weibull 等分布函数来模拟,也不可能达到人们的要求。对于这种复杂林分,我们用非参数核方法来模拟,却得到了满意的结果(见图 1~4 曲线)。

天然林直径结构比人工林要复杂得多,其分布曲线受自身的演替过程、林分年龄结构、树

种特征及组成、自然灾害、采伐方式及强度、立地条件、更新方式及过程等多种因素影响。一般认为保留林及群状同龄林的直径分布呈间断的或波纹状的反J型曲线;具有明显层次的复层异龄林分,直径分布呈双峰山状曲线;林冠层次不整齐的异龄林分,则呈不规则的山状曲线;平均直径较小的林分,曲线尖峭,偏度较大,但随着平均直径的加大,峭度从正到负,偏度逐渐变小。显而易见,天然林直径结构很难属于同一分布族,为描述如此纷繁复杂的天然林分的直径结构,分布曲线法未必是最好的方法。在这种情况下,利用与分布无关的非参数核估计方法,往往能满足我们的要求。

表2给出了Webull分布及LogNormal分布与Kernel方法相比较的符号检验结果。表中所显示的 $P$ 值(右尾概率)为0.0002,对于 $\alpha=0.01$ , $P$ 值也足够小,故有理由说,核方法要优于分布函数法。

表2 符号检验(显著水平0.95)

分布型	原假设	备择假设	$n$	$S_+$	$S_-$	右尾概率	结论
韦布尔	$P_+ = P_-$	$P_+ > P_-$	12	12	0	0.0002	核优于韦布尔
对数正态	$P_+ = P_-$	$P_+ > P_-$	12	12	0	0.0002	核优于对数正态

注:核估计比某分布方法好为成功,概率记 $P_+$ ,相反记 $P_-$ 。 $S_+$ 、 $S_-$ 分别为正负符号数目; $n$ 是符号总数。

### 3 结论与讨论

(1) 由于林分条件的复杂性,往往很难确定林分究竟属于哪一分布族 $f(x, \theta)$ ,  $\theta \in H$ ,此时用参数方法来模拟林分,不可能得到满意结果。而非参数方法假设条件少,并不要求总体分布遵从什么具体形式,故在分布型未知的情况下,不论是天然林还是人工林,均可用非参数核方法来模拟林分的直径结构。

(2) 已定核函数的模拟过程,可以通过调节窗宽使之达到最优平滑曲线。具体操作时,可使实际值与理论值的卡方统计量达到最小的窗宽定为最优窗宽。

(3) 原则上,非参数核估计也能用于林分的树高结构、材积结构、形数结构等,如何应用,效果怎样,正在进一步研究。

### 参 考 文 献

- 1 孟宪宇. 测树学. 北京: 中国林业出版社, 1996. 66 ~ 80.
- 2 孟宪宇. 使用Webull分布对树高分布和直径分布的研究. 北京林业大学学报, 1988, 10(1): 40 ~ 48.
- 3 寇文正. 林分直径分布研究. 南京林产工业学院学报, 1982, (2): 51 ~ 65.
- 4 孟宪宇. 使用Webull分布对人工油松林直径分布的研究. 北京林学院学报, 1985, 7(1): 30 ~ 39.
- 5 陈希孺. 柴根象. 非参数统计教程. 上海: 华东师范大学出版社, 1993.
- 6 易丹辉. 非参数统计——方法与应用. 北京: 中国统计出版社, 1996.
- 7 崔恒建, 王雪峰. 核密度估计及在直径分布研究中的应用. 北京林业大学学报, 1996, 18(2): 67 ~ 72.

## Applying Nonparametric Kernel Method to Probe the Diameter Structure of Natural Forest

*Wang Xuefeng*<sup>1)</sup> *Guan Qingjun*<sup>2)</sup> *Wu Lijun*<sup>2)</sup> *Dai Haiping*<sup>2)</sup>

( 1)The Research Institute of Forest Resource Information Techniques, CAF, 100091, Beijing, China;

2)Deerbuer Forestry Bureau, Inner Mongolia, 022357, Genhe, Inner Mongolia, China)

**Abstract** The paper applies Nonparametric Kernel Density Assessment to simulate yield diametric distribution, the source comes from 12 blocks of yield randoms in Wangching Jilin Province. The results show: (1) If randoms can be described by Weibull and other distributive functions, Kernel Assessment Method can also describe it, moreover, superior to other distributive functions; (2) Kernel Assessment Method can describe stand when distribution functions can not describe it, at the same time, the result is obvious; (3) Both yield and plantation can be described by Kernel Assessment Method to analysis stand diameter structure. Nonparametric method perhaps be useful method.

**Key words** diameter distribution; diameter structure of natural forest; kernel density assessment ; nonparametric statistic method; stand table