Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry 2015; **88**, 143–151, doi:10.1093/forestry/cpu038
Advance Access publication 28 October 2014

Size-biased distributions in the generalized beta distribution family, with applications to forestry

Mark J. Ducey[1]* and Jeffrey H. Gove[2]

[1]*Department of Natural Resources and the Environment, University of New Hampshire, 114 James Hall, Durham, NH 03824, USA*
[2]*USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, USA*

*Corresponding author. E-mail: mark.ducey@unh.edu

Received 27 July 2014

Size-biased distributions arise in many forestry applications, as well as other environmental, econometric, and biomedical sampling problems. We examine the size-biased versions of the generalized beta of the first kind, generalized beta of the second kind and generalized gamma distributions. These distributions include, as special cases, the Dagum (Burr Type III), Singh-Maddala (Burr Type XII) and Fisk distributions as well as better-studied distributions such as the Weibull, lognormal, beta (of the first and second kind), gamma and exponential. Our results indicate that specification and estimation of the size-biased forms of these distributions can be viewed within a unified framework. This should facilitate broader application of size-biased distributions in forestry sampling, modeling and analysis.

Introduction

Size-biased distributions arise naturally in a range of sampling and modeling problems in forestry (Gove, 2003a). They also occur in applications spanning domains including environmental sciences, econometrics, human demography and biomedical sciences (Scheaffer, 1972; Patil and Rao, 1978). Suppose that X is a variable of interest, such as tree diameter, and we wish to represent the distribution of X using a parametric model $X \sim f(x;\theta)$. In the ordinary case, this is usually straightforward and moment- or maximum likelihood estimators, as well as random variable generators, are available for a wide range of distributions. However, if sample units are weighted or selected with probability proportional to X^α, then the density is the so-called size-biased density of order α (Patil and Ord, 1976; Patil, 1981):

$$f^*_\alpha(x;\theta) = \frac{x^\alpha f(x;\theta)}{\int x^\alpha f(x;\theta)\,\mathrm{d}x} \tag{1}$$

For example, if $f(x;\theta)$ is the distribution of diameter at breast height (DBH) in a forest stand, then the distribution of basal area by diameter will be $f^*_2(x;\theta)$, because basal area of a tree is proportional to DBH squared.

Gove (2003a) provides a review of applications of size-biased distributions in forestry, in which geometric approaches to sampling often induce size-biased distributions of order 1 (length-biased sampling) or 2 (area-biased sampling). Perhaps the first mention of size-bias sampling applications in forestry and wood products was the example given by Warren (1975) related to sampling wood cell cross-sectional area using a microscope. A more recent application to wood products studies is found in Svensson *et al.* (2007), where a sample of wood fiber length samples from increment cores were recognized as arising from length-biased selection. In field applications, size-biased distributions can arise either because individuals are sampled with unequal probability by design or because of unequal detection probability. For example, both line intersect sampling and horizontal line sampling often induce size-biased distributions of order 1, while horizontal point sampling of standing trees (also known as prism sampling and Bitterlich sampling) induces size-biased distributions of order 2 (Van Deusen, 1986; Gove, 2003a). Ducey and Valentine (2008) and Ducey (2009) suggest other sampling approaches for standing trees that would generate size-biased distributions of non-integer order with $\alpha \approx 1.6$ and $2 \le \alpha \le 3$, respectively. Size-biased distributions arise in a number of other social, environmental and biological science applications. For example, many inferences from on-site survey data involve size-biased distributions of order 1 (Moeltner and Shonkwiler, 2005). Length-biased sampling, which induces size-biased distributions of order 1, also arises in linkage mapping of genome scans (Terwilliger *et al.*, 1997). Size-biased distributions come into play when organisms occur in groups, and group size influences the probability of detection (e.g. Drummer and McDonald, 1987). Other applications, including environmental applications, are suggested by Patil and Ord (1976) and Patil and Rao (1978).

© Institute of Chartered Foresters, 2014. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

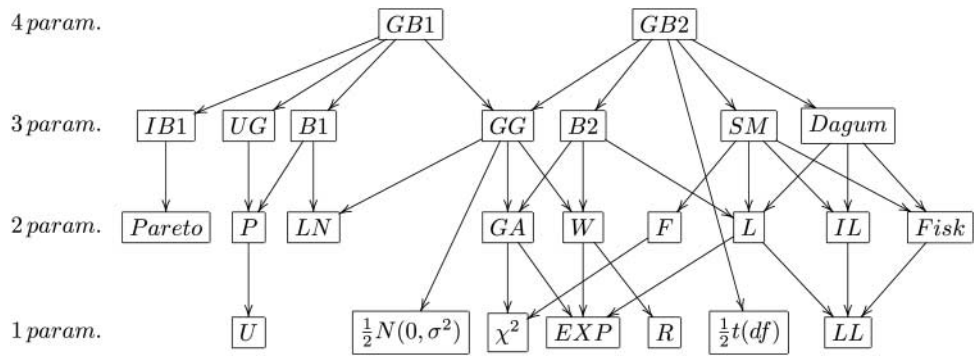

Figure 1 Relationships between members of the generalized beta family, following McDonald (1984) and McDonald and Xu (1995). For symbols, see Table 1.

A great deal of research on size-biased distributions has focused on the question of form-invariance. A distribution is said to be form-invariant under size-biasing if the size-biased distribution has the same mathematical form as the original distribution, only with altered parameter values. If a distribution is form-invariant that simplifies both estimation and simulation because the same software that was used for the original distribution can also be used for the size-biased distribution (albeit with some attention to possible constraints on parameter values). In a key paper, Patil and Ord (1976) demonstrated form-invariance for size-biased distributions within the log-exponential family, and Alavi and Chinipardaz (2009) provide generalization of those results to additional weight functions. Other studies have tended to focus on elucidating the relationships between single distributions and their size-biased counterparts, often focused on one (or perhaps two) specific orders motivated by specific applications. For example, previous work on size-biased distributions in forestry has focused on the Weibull distribution, because of its long history in modeling tree diameter distributions (Bailey and Dell, 1973). Gove (2000, 2003b) presented results and useful formulae for the size-biased Weibull under length- and area-biased sampling. However, size-biased versions of some distributions that have seen widespread practical use outside forestry, such as those of Dagum (1977) and Singh and Maddala (1976), have seen limited study (Ye *et al.*, 2012). Because size-biased distributions arise so commonly in forest sampling and modeling, understanding size-biasing in these other distributions could make them more useful in the forestry context.

Here, we take up the generalized beta distributions of the first and second kind and the generalized gamma, as developed by McDonald (1984). These families are extremely flexible and include a great many familiar (and not-so-familiar) distributions as special cases. We show that these distributions are form-invariant under size-biased sampling, with a specific and simple mapping of parameters from their original forms to their size-biased forms. Because these two distributions contain a large number of other useful distributions as special cases, our results lead immediately to simple mappings between all these special cases and their size-biased versions (even when those special cases are not form-invariant under size-biased sampling themselves). Thus, the entire family of distributions can be placed within a unified framework for estimation, inference and simulation.

The generalized beta and gamma distributions

McDonald (1984) introduced two generalized beta distributions, the generalized beta of the first kind (GB1) and of the second kind (GB2), which contain numerous other distributions as special cases (Figure 1, Table 1). The GB1 has probability density function (pdf)

$$\mathrm{GB1}(y; a, b, p, q) = \frac{|a| y^{ap-1}[1-(y/b)^a]^{q-1}}{b^{ap}\mathrm{B}(p,q)} \tag{2}$$

for $0 < y^a < b^a$, and $b > 0, p > 0, q > 0$. Here, $\mathrm{B}(x, y)$ denotes the beta function, i.e. $\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$. The raw moments of order h are given by

$$\mu'_h = E_{\mathrm{GB1}}[y^h] = \frac{b^h \mathrm{B}(p + h/a, q)}{\mathrm{B}(p, q)} \tag{3}$$

and are defined for $p + h/a > 0$ (McDonald, 1984). As McDonald (1984) and McDonald and Xu (1995) show, the GB1 includes several widely used distributions as special or limiting cases, including the beta of the first kind, inverse beta of the first kind, Pareto and power, generalized gamma, uniform and unit gamma distributions.

The GB2 has pdf

$$\mathrm{GB2}(y; a, b, p, q) = \frac{|a| y^{ap-1}}{b^{ap}\mathrm{B}(p,q)[1+(y/b)^a]^{p+q}} \tag{4}$$

for $0 < y < \infty$, with the same restrictions on parameters. The raw moments of order h of the GB2 are given by

$$\mu'_h = E_{\mathrm{GB2}}[y^h] = \frac{b^h \mathrm{B}(p + h/a, q - h/a)}{\mathrm{B}(p, q)} \tag{5}$$

and are defined for $-p < h/a < q$. The GB2 includes many more distributions as special or limiting cases such as the

Table 1 Special cases of the generalized beta family

Symbol	Distribution name	Relationship
IB1	Inverse beta of the first kind	$\mathrm{GB1}(-1, b, p, q)$
$\mathrm{UG}(b, d, q)$	Unit gamma	$\lim_{a\to 0} \mathrm{GB1}(a, b, p = d/a, q)$
$B1$	Beta of the first kind	$\mathrm{GB1}(1, b, p, q)$
$B2$	Beta of the second kind	$\mathrm{GB2}(1, b, p, q)$
SM	Singh-Maddala (Burr Type XII)	$\mathrm{GB2}(a, b, 1, q)$
Dagum	Dagum (Burr Type III)	$\mathrm{GB2}(a, b, p, 1)$
Pareto	Pareto	$\mathrm{IB1}(q = 1) = \mathrm{GB1}(-1, b, p, 1)$
P	Power	$\mathrm{UG}(q = 1) = B1(q = 1) = \mathrm{GB1}(1, b, p, 1)$
$\mathrm{LN}(\mu, \sigma^2)$	Log-normal	$\lim_{a\to 0} \mathrm{GG}(a, (\sigma^2 a^2)^{1/a}, (a\mu + 1)/(\sigma^2 a^2))$
GA	Gamma	$\mathrm{GG}(1, \beta, p) = \lim_{q\to\infty} B1 = \lim_{q\to\infty} B2$
W	Weibull	$\mathrm{GG}(a, \beta, 1) = \lim_{q\to\infty} \mathrm{SM}$
$F(u, v)$	F	$B2(v/u, u/2, v/2) = \mathrm{GB2}(1, v/u, u/2, v/2)$
L	Lomax	$B2(p = 1) = \mathrm{GB2}(1, b, 1, q)$
IL	Inverse Lomax	$\mathrm{SM}(a = -1) = \mathrm{Dagum}(a = 1) = \mathrm{GB2}(1, b, p, 1)$
Fisk	Fisk	$\mathrm{SM}(q = 1) = \mathrm{Dagum}(p = 1) = \mathrm{GB2}(a, b, 1, 1)$
U	Uniform	$P(p = 1) = \mathrm{GB1}(1, b, 1, 1)$
$\frac{1}{2}N(0, \sigma^2)$	Half-normal	$\mathrm{GG}(2, \beta, 1/2)$
χ^2	χ^2	$\mathrm{GA}(\beta = 2) = \lim_{q\to\infty} F = \mathrm{GG}(1, 2, p)$
EXP	Exponential	$\mathrm{GA}(p = 1) = W(a = 1) == \lim_{q\to\infty} L = \mathrm{GG}(1, \beta, 1)$
R	Rayleigh	$W(a = 2) = \mathrm{GG}(2, \beta, 1)$
$\frac{1}{2}t(\mathrm{df})$	Half-t	$\mathrm{GB2}(2, \sqrt{\mathrm{df}}, 1/2, \mathrm{df}/2)$
LL	Log-logistic	$L(q = 1) = \mathrm{IL}(p = 1) = \mathrm{Fisk}(a = 1) = \mathrm{GB2}(1, b, 1, 1)$

Symbols refer to Figure 1.

generalized gamma, Dagum (also known as the Burr Type III and as the Kappa), Singh-Maddala (or Burr Type XII), lognormal, Weibull, gamma, Lomax, F, Rayleigh (or Fisk), chi-square, half-normal, half-t, exponential and log-logistic (McDonald, 1984; McDonald and Xu, 1995).

Of particular interest because of its centrality in Figure 1 is the generalized gamma distribution (GG), originally introduced by Stacy (1962) with a slightly different parameterization. The GG has pdf

$$\mathrm{GG}(y; a, \beta, p) = \frac{a y^{ap-1}\, \mathrm{e}^{-(y/\beta)^a}}{\beta^{ap}\Gamma(p)} \tag{6}$$

and can be obtained from the GB1 in the limit as $q \to \infty$ with $b = (p+q)^{1/a}\beta$, or from a GB2 in the limit as $q \to \infty$ with $b = q^{1/a}\beta$ (McDonald, 1984). The raw moments of order h of the GG are given by

$$\mu'_h = E_{\mathrm{GG}}[y^h] = \frac{\beta^h \Gamma(p + h/a)}{\Gamma(p)} \tag{7}$$

Size-biased generalized beta and gamma distributions

Here, we derive expressions for the size-biased generalized beta distributions of the first and second kind. We do so also for the generalized gamma for the sake of completeness. Recall that an equivalent expression of the definition in equation (1) is

$$f^*_\alpha(x; \theta) = \frac{x^\alpha f(x; \theta)}{\mu'_\alpha}. \tag{8}$$

Substitution of the relevant pdfs and moments into equation (8), followed by algebraic simplification, leads directly to the following results.

Size-biased generalized beta of the first kind

Substituting equation (2) for the pdf, and equation (3) for the αth raw moment, into equation (8), we obtain the pdf of the size-biased GB1 (SBGB1):

$$\begin{aligned}
&\mathrm{SBGB1}(y; a, b, p, q, \alpha) \\
&\quad = \left[\frac{\mathrm{B}(p, \alpha/a)}{b^\alpha \mathrm{B}(p+q, \alpha/a)\mathrm{B}(p, q)}\right]\left[\frac{a y^{ap+\alpha-1}[1-(y/b)^a]^{q-1}}{b^{ap}}\right] \\
&\quad = \left[\frac{\mathrm{B}(p, \alpha/a)}{\mathrm{B}(p+q, \alpha/a)\mathrm{B}(p, q)}\right]\left[\frac{a y^{ap+\alpha-1}[1-(y/b)^a]^{q-1}}{b^{ap+\alpha}}\right].
\end{aligned}$$

Recalling that $\mathrm{B}(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$, substituting and simplifying we obtain

$$\begin{aligned}
&\mathrm{SBGB1}(y; a, b, p, q, \alpha) \\
&\quad = \left[\frac{\Gamma(p)\Gamma(\alpha/a)/\Gamma(p+\alpha/a)}{[\Gamma(p+q)\Gamma(\alpha/a)/\Gamma(p+q+\alpha/a)][\Gamma(p)\Gamma(q)/\Gamma(p+q)]}\right]
\end{aligned}$$

$$\times \left[\frac{ay^{ap+\alpha-1}[1-(y/b)^a]^{q-1}}{b^{ap+\alpha}}\right]$$
$$= \left[\frac{\Gamma(p+q+\alpha/a)}{\Gamma(p+\alpha/a)\Gamma(q)}\right]\left[\frac{ay^{ap+\alpha-1}[1-(y/b)^a]^{q-1}}{b^{ap+\alpha}}\right]$$
$$= \left[\frac{1}{\mathrm{B}(p+\alpha/a)}\right]\left[\frac{ay^{ap+\alpha-1}[1-(y/b)^a]^{q-1}}{b^{ap+\alpha}}\right]$$
$$= \mathrm{GB1}(a, b, p+\alpha/a, q) \tag{9}$$

and we see that the GB1 is form-invariant under size-biasing, for all orders $\alpha > -ap$, i.e. for all orders for which the corresponding αth raw moment of the original distribution exists.

Size-biased generalized beta of the second kind

Substituting equation (4) for the pdf, and equation (5) for the αth raw moment, into equation (8), we obtain the pdf of the size-biased GB2 (SBGB2):

$$\mathrm{SBGB2}(y; a, b, p, q, \alpha)$$
$$= \left[\frac{\mathrm{B}(p,q)}{b^{\alpha}\mathrm{B}(p+\alpha/a, q-\alpha/a)}\right]\left[\frac{ay^{ap+\alpha-1}}{b^{ap}\mathrm{B}(p,q)[1+(y/b)^a]^{p+q}}\right]$$
$$= \frac{ay^{ap+\alpha-1}}{b^{ap+\alpha}\mathrm{B}(p+\alpha/a, q-\alpha/a)[1+(y/b)^a]^{p+q}}.$$

Letting $p' = p+\alpha/a$ and $q' = q-\alpha/a$, and noting that $p'+q' = p+q$ we have

$$\mathrm{SBGB2}(y; a, b, p, q, \alpha) = \frac{ay^{ap'-1}}{b^{ap'}\mathrm{B}(p', q')[1+(y/b)^a]^{p'+q'}}$$
$$= \mathrm{GB2}(y; a, b, p+\alpha/a, q-\alpha/a) \tag{10}$$

and we see that the GB2 is form-invariant under size-biasing, for all orders such that $-ap < \alpha < aq$, i.e. for all orders for which the corresponding αth raw moment of the original distribution exists.

Size-biased generalized gamma

Substituting equation (6) for the pdf, and equation (7) for the αth raw moment, into equation (8), we obtain the pdf of the size-biased GG (SBGG):

$$\mathrm{SBGG}(y; a, \beta, p, \alpha) = \left[\frac{\Gamma(p)}{\beta^{\alpha}\Gamma(p+\alpha/a)}\right]\left[\frac{ay^{ap+\alpha-1}\,\mathrm{e}^{-(y/b)^a}}{\beta^{ap}\Gamma(p)}\right]$$
$$= \frac{ay^{ap+\alpha-1}\,\mathrm{e}^{-(y/b)^a}}{\beta^{ap+\alpha}\Gamma(p+\alpha/a)}$$
$$= \mathrm{GG}(y; a, \beta, p+\alpha/a) \tag{11}$$

and we see that the GG is form-invariant under size-biasing, for all orders $\alpha > -ap$, i.e. for all orders for which the corresponding αth raw moment of the original distribution exists.

Special cases

Form-invariance of the GB1, GB2 and GG, and the simple mappings of parameters in equations (9)–(11), lead directly to expresssions for the size-biased distributions of the special cases listed in Table 1, some of which are form-invariant while others map to other distributions within the generalized beta family. These results are given in Table 2.

As Patil and Ord (1976) proved the lognormal, Pareto, gamma and betas of the first and second kind are form-invariant under size-biased sampling. To these we may add the inverse beta of the first kind, unit gamma and χ^2 distributions, which had not previously been identified as form-invariant. Form-invariance is useful in that it simplifies estimation and simulation of size-biased distributions.

A small number of other special cases have previously been identified in the literature. For example, Gove and Patil (1998) had shown that the size-biased exponential is a gamma deviate. Gove and Patil (1998) had also shown that the size-biased Weibull of order 2, following a change of variables, was related to the standard gamma deviate. From the results presented here, we can clearly see that the size-biased Weibull (of any order for which the size-biased distribution exists) is in fact a GG. Ye *et al.* (2012) identify special cases of the size-biased GB2 and provide equations for their moments but fail to identify the size-biased GB2 as a GB2. To our knowledge, no previous study has been made of many of the remaining special cases. As Table 2 shows that most of these distributions map to either a GB1 or a GB2 under size-biasing. When this occurs, one or more constraints is imposed on the parameters such that the number of free parameters is equal to the number of free parameters in the original distribution.

A result given on McDonald and Xu (1995, p. 140) implies that the CDF of either a GB1 or a GB2 can be obtained by a change of variables using the incomplete beta function. Specifically, if $Y \sim GB1(y; a, b, p, q)$, then $X = (Y/b)^a \sim B1(x; b' = 1, p, q)$. Likewise, if $Y \sim GB2(y; a, b, p, q)$, then $X = (Y/b)^a/[1+(Y/b)^a] \sim B1(x; b' = 1, p, q)$. Ready availability of the CDF greatly simplifies working with grouped, censored or truncated versions of any of the size-biased distributions discussed here. Moreover, if x is the output of a competent B1 random deviate generator, then $y = bx^{1/a}$ will be a GB1 random deviate, and $y = b(x/(1-x))^{1/a}$ will be a GB2 random deviate. Thus, random deviates from all the size-biased distributions whose original forms are GB1 or GB2 can be generated easily, either in support of modeling or of Monte Carlo approaches to inference.

Examples

Diameter distributions from horizontal point sampling

As noted above, horizontal point sampling of standing trees (also known as prism sampling and Bitterlich sampling) induces size-biased distributions of order 2 (Van Deusen, 1986; Gove, 2003a). The exponential distribution is often used to model uneven-aged stands (Meyer, 1952; Leak, 1964), but perhaps the most common model for tree diameter distributions is the Weibull (Bailey and Dell, 1973). However, the Weibull can only fit a restricted range in the skewness-kurtosis plane

Table 2 Size-biased distributions for special cases of the generalized beta family

Symbol	Size-biased pdf of order α	Existence condition
IB1	$\mathrm{GB1}(-1, b, p-\alpha, q) = \mathrm{IB1}$	$\alpha < p$
$\mathrm{UG}(b, d, q)$	$\lim_{a\to 0} \mathrm{GB1}(a, b, (d+\alpha)/a, q) = \mathrm{UG}(b, d+\alpha, q)$	$\alpha > -d$
$B1$	$\mathrm{GB1}(1, b, p+\alpha, q) = B1$	$\alpha > -p$
$B2$	$\mathrm{GB2}(1, b, p+\alpha, q-\alpha) = B2$	$-p < \alpha < q$
SM	$\mathrm{GB2}(a, b, 1+\alpha/a, q-\alpha/a)$	$-a < \alpha < aq$
Dagum	$\mathrm{GB2}(a, b, p+\alpha/a, 1-\alpha/a)$	$-ap < \alpha < a$
Pareto	$\mathrm{GB1}(-1, b, p-\alpha, 1) = \mathit{Pareto}$	$\alpha < p$
P	$\mathrm{GB1}(1, b, p+\alpha, 1) = P$	$\alpha > -p$
$\mathrm{LN}(\mu, \sigma^2)$	$\lim_{a\to 0} \mathrm{GG}(a, (\sigma^2 a^2)^{1/2}, [a(\mu+\alpha\sigma^2)+1]/(\sigma^2 a^2))$	$\forall\alpha$
GA	$\mathrm{GG}(1, \beta, p+\alpha) = \mathrm{GA}$	$\alpha > -p$
W	$\mathrm{GG}(a, \beta, 1+\alpha/a)$	$\alpha > -a$
$F(u, v)$	$\mathrm{GB2}(1, v/u, u/2+\alpha, v/2-\alpha)$	$-u/2 < \alpha < v/2$
L	$\mathrm{GB2}(1, b, 1+\alpha, q-\alpha) = B2$	$-1 < \alpha < q$
IL	$\mathrm{GB2}(1, b, p+\alpha, 1-\alpha) = B2$	$-p < \alpha < 1$
Fisk	$\mathrm{GB2}(a, b, 1+\alpha/a, 1-\alpha/a)$	$-a < \alpha < a$
U	$\mathrm{GB1}(1, b, 1+\alpha, 1) = P$	$\alpha > -1$
$\frac{1}{2}N(0, \sigma^2)$	$\mathrm{GG}(2, \beta, \alpha+1/2) = \mathrm{GA}$	$\alpha > -1/2$
χ^2	$\mathrm{GG}(1, 2, p+\alpha) = \chi^2$	$\alpha > -p$
EXP	$\mathrm{GG}(1, \beta, 1+\alpha) = \mathrm{GA}$	$\alpha > -1$
R	$\mathrm{GG}(2, \beta, 1+\alpha)$	$\alpha > -1$
$\frac{1}{2}t(\mathrm{df})$	$\mathrm{GB2}(2, \sqrt{\mathrm{df}}, (1+\alpha)/2, (\mathrm{df}-\alpha)/2)$	$-1 < \alpha < \mathrm{df}$
LL	$\mathrm{GB2}(1, b, 1+\alpha, 1-\alpha)$	$-1 < \alpha < 1$

Symbols refer to Figure 1 and Table 1.

(Hafley and Schreuder, 1977). Other distributions in the generalized beta family that have occasionally been used to fit diameter distributions include the gamma (Nelson, 1964), lognormal (Bliss and Reinker, 1964), beta (Clutter and Bennett, 1965; Lenhart and Clutter, 1971; Gorgosa-Varela *et al.*, 2008) and Dagum or Burr III (Lindsay *et al.*, 1996; Gove *et al.*, 2008). Li *et al.* (2002) use a bivariate generalized beta distribution to model the joint height-diameter distribution of trees, but the univariate version of their distribution is not a member of the generalized beta class explored here. Of these distributions, only the exponential and Weibull have seen detailed attention to the size-biased form (Gove and Patil, 1998; Gove, 2003b).

To illustrate fitting members of the generalized beta family to horizontal point sampling data, we used data from the Mendum's Pond parcel owned by the University of New Hampshire. The parcel includes ~76 ha of forested land. In July, 2011, the parcel was inventoried using a systematic grid of 70 horizontal point samples, with a basal area factor of 4.59 m^2/ha and a minimum measurable DBH of 5.08 cm. DBH was recorded to the nearest 0.25 cm. Overall, 531 live trees of 13 species were included in the inventory, with a maximum DBH of 89.7 cm. Basal area of all species combined was 34.8 m^2/ha, led by eastern hemlock (*Tsuga canadensis*), eastern white pine (*Pinus strobus*), northern red oak (*Quercus rubra*) and red maple (*Acer rubrum*).

We fit all the members of the GB2 portion of the generalized beta family, except for the half-normal, χ^2, Rayleigh and half-*t*. Although in principle one could attempt a model fit recognizing the hierarchical nature of the data, we followed conventional practice for illustrative purposes and treated each measured diameter as an independent observation. All distributions were fit using their size-biased versions of order 2 by maximum likelihood, with the distribution left-truncated at 5.08 cm. Asymptotic standard errors for the model parameters were calculated as the square root of the diagonals of the inverse of the Hessian matrix of the likelihood. Model fits were assessed using the Akaike Information Criterion (AIC), with ΔAIC computed as the difference in AIC between any given model and the best-fitting model (the model with minimum AIC).

Results of the model fits are shown in Table 3. Three of the distributions, the B2, Lomax and LL, did not converge despite attempts to fit the model using multiple initial estimates of the parameters, so no results are shown. Although several of the models had competitive or at least plausible fits ($\Delta\mathrm{AIC} < 10$), the best performer overall was the two-parameter Gamma. The fit of the size-biased Gamma (i.e. the Gamma with parameters β, $p+2$) compared with a histogram of the raw tree tallies, and of the ordinary Gamma in comparison to a histogram illustrating the estimated stand table from the tree tallies, is shown in Figure 2. The size-biased Gamma fits the raw tally quite well. Some discrepancy can be seen between the estimated stand table and the original version of the Gamma in the smallest size classes. However, it is notable that while the estimates of trees per hectare in the smallest classes are large, they are based on very few trees and have a high sampling error. For example, the total number of trees tallied in the smallest size class is only 13, on 70 plots, and the estimate of density in that class has a standard error of ~35 percent. Forcing the distribution to fit the smallest size classes more closely would necessarily compromise the fit in larger size classes. Those are associated with fewer trees per hectare, but more basal area per hectare and tallied trees, and

Table 3 Members of the GB2 and GG sub-families fit to a horizontal point sample tally of 531 trees from a forested parcel in New Hampshire

Distribution	Parameters	Log-likelihood	AIC	Δ AIC
GB2	$a = 2.1472 \pm 0.0178$ $b = 60.4036 \pm 1.2185$ $p = 0.7079 \pm 0.0516$ $q = 7.1576 \pm 0.0413$	−2112.607	4233.214	3.058
GG	$a = 1.1643 \pm 0.1253$ $\beta = 9.9859 \pm 2.5578$ $p = 2.2184 \pm 0.2555$	−2112.901	4231.802	1.646
SM	$a = 1.8162 \pm 0.0878$ $b = 77.0589 \pm 2.7351$ $q = 11.2604 \pm 0.7225$	−2112.645	4231.289	1.133
Dagum	$a = 6.6093 \pm 1.2822$ $b = 36.0052 \pm 1.3386$ $p = 0.0726 \pm 0.0234$	−2116.340	4238.680	8.523
Gamma	$\beta = 6.0790 \pm 0.3071$ $p = 3.2524 \pm 0.1635$	−2113.078	4230.157	0.000
Weibull	$a = 1.5339 \pm 0.1562$ $\beta = 19.3953 \pm 0.2837$	−2113.774	4231.548	1.392
Fisk	$a = 4.4867 \pm 0.3928$ $b = 20.8290 \pm 0.6335$	−2161.312	4326.624	96.468
Exp	$\beta = 10.5001 \pm 0.5991$	−2144.502	4291.004	60.848

the estimates in those classes are therefore better constrained. Indeed, the inefficiency of fitting assumed diameter distributions to the stand table is a key reason why fitting the size-biased form of the distribution to the empirical distribution of the tallied trees is preferred (Van Deusen, 1986; Gove, 2000).

Piece length distributions from line intersect sampling

In comparison with diameter distributions of live trees, size distributions of dead downed wood have not been as widely studied. However, as Gove (2003a) notes, many probability proportional to size methods exist for downed wood sampling, and these induce size-biased distributions as well. For example, line intersect sampling (LIS) samples downed wood with probability proportional to piece length, so the distribution of sample lengths is the size-biased distribution of order 1 for the original distribution in the population. LIS is one of the most widely used techniques for downed wood sampling, having been introduced over 50 years ago (Warren and Olsen, 1964). For the theory of LIS, see de Vries (1979); Kaiser (1983).

To illustrate fitting members of the generalized beta family to LIS data, we used data from the U.S. Forest Service, Forest Inventory and Analysis (FIA) program. FIA measures downed dead wood on a cluster of LIS lines at a subset of its permanent plot locations. For details of the FIA field protocol and associated estimators, see Woodall and Williams (2005). In addition to variables such as diameter and decay class, the length of each tallied piece is recorded to the nearest foot (0.3048 m), with a minimum measured length of 3 ft (0.9144 m). Note that the smallest length class contains only pieces from 3 to 3.5 ft in length; all other classes are obtained by simple rounding (e.g. the nominal 4 foot class contains pieces from 3.5 to 4.5 ft in length). We obtained the LIS data for the states of Maine, New Hampshire

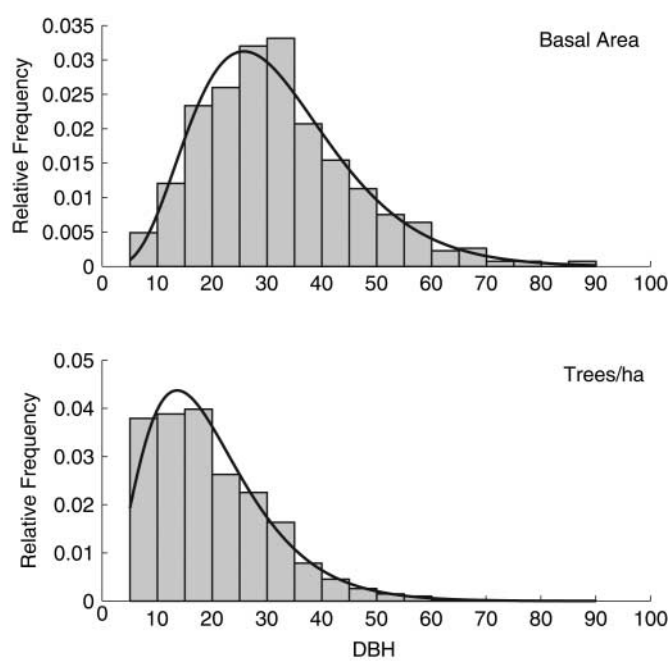

Figure 2 Size-biased gamma distribution fit to a raw horizontal point sample tally of 531 trees from a forested parcel in New Hampshire (top panel), and the corresponding stand table (bottom panel).

Table 4 Members of the GB1 sub-family fit to line intersect sampling data on downed wood piece length from the United States Forest Inventory and Analysis, for Vermont, New Hampshire and Maine

Distribution	Parameters	Log-likelihood	AIC	Δ AIC
GB1	$a = 0.0487 \pm 0.0009$ $p = 7.2814 \pm 0.0590$ $q = 3.4762 \pm 0.0198$	−9514.354	19 034.708	2.004
IB1	$p = 1.6826 \pm 0.0117$ $q = 1.7322 \pm 0.0124$	−9711.505	19 427.010	394.306
UG	$d = 0.4130 \pm 0.0105$ $q = 3.4752 \pm 0.0296$	−9514.352	19 032.704	0.000
Pareto	$p = 1.2805 \pm 0.0047$	−9796.922	19 595.844	563.140

and Vermont from the FIA database, and combined these for inventory years 2006–2010, resulting in data for 2657 pieces from 275 non-empty plots. Recorded piece length ranged from the minimum (0.9144 m) to 24.4 m.

We fit selected members of the GB1 portion of the generalized beta family, including the GB1, IB1, UG, B1 and Pareto. All distributions were fit as size-biased distributions of order 1, by maximum likelihood. Because of the unusual pattern of binning at the small end of the length class range, we fit the discretized version of each distribution, treating the probability associated with each length class as the difference between the CDF at the upper boundary of the class and the CDF at the lower boundary. All distributions were left-truncated at 0.9144 m, and the scale parameter b was set so that the maximum length equaled the upper boundary of the largest length class. Asymptotic standard errors for the model parameters were calculated as the square root of the diagonals of the inverse of the Hessian matrix of the likelihood, and model fits were assessed using the AIC.

Results of the model fits are shown in Table 4. The B1 did not converge for this particular dataset. Strictly speaking, the GB1 attained numerical convergence, but at very small value of the parameter a, and at a log-likelihood that was slightly worse than that of the UG, which is the limiting case of the GB1 as $a \to 0$. The UG, which has rarely if ever been used in forestry, was the best-fitting distribution and (other than the GB1, which converges on it) the only plausible distribution among those evaluated (ΔAIC < 10). The fit of the size-biased UG compared with a histogram of the raw piece tallies, and of the ordinary UG in comparison to a histogram illustrating the estimated distribution of piece lengths in the population, is shown in Figure 3.

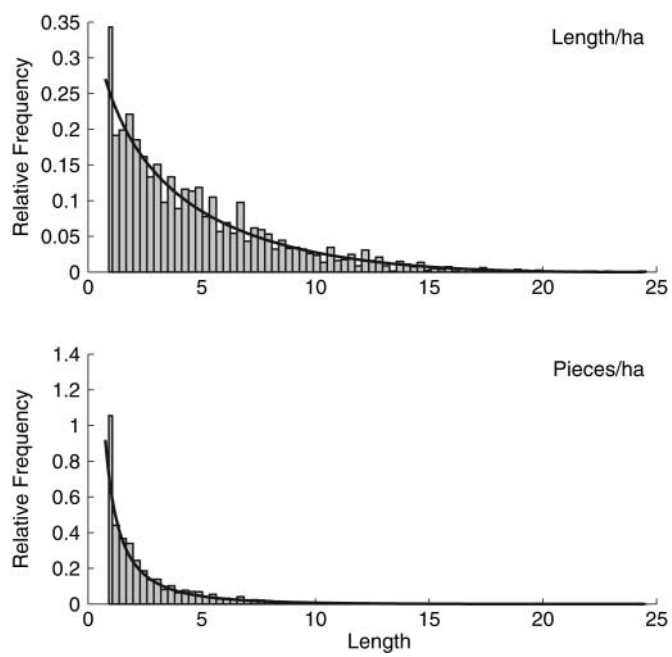

Figure 3 Size-biased unit gamma distribution fit to a raw tally of downed wood lengths from line-intercept sampling in Vermont, New Hampshire and Maine (top panel), and the corresponding estimates of relative frequency in the original population (bottom panel). Note that abundance in the smallest length class is depicted at double value in both graphs because that class is only one-half its nominal width.

Discussion and conclusions

Several studies have examined size-biased distributions, but these studies have most often been restricted to one or a few specific distributions (e.g. Scheaffer, 1972; Gove and Patil, 1998), have examined one or two integer orders of one or two distributions (e.g Gove, 2000, 2003b; Correa and Wolfson, 2007), or have been concerned primarily with identifying form-invariance (e.g. Patil and Ord, 1976; Alavi and Chinipardaz, 2009). Conversely, more recent studies, including those of Priyadarshani (2011) and Ye *et al.* (2012) examine the weighted GG and GB2, respectively, with the size-biased form as a special case, but fail to identify either of those distributions as form-invariant.

Here, we show that size-biased distributions arising from the family of continuous distributions encompassed by the GB1, GB2 and GG as described by McDonald (1984) can be treated within a unified framework. This unified framework should simplify estimation, interpretation and simulation of size-biased distributions within this family. As Patil and Ord (1976) showed, given certain

regularity conditions, membership in the log-exponential family of distributions is a necessary and sufficient condition for a distribution to be form-invariant under size biasing. Therefore, the GB1, GB2 and GG are all members of this family. However, neither the generalized beta of McDonald and Xu (1995), which includes the GB1 and GB2 as special cases nor their closely related exponential generalized beta appear to belong to this family.

Although size-biased versions of individual distributions within this family have previously been studied (Patil and Ord, 1976; Gove, 2000, 2003b), size biased versions of some distributions that are widely used in econometrics and other disciplines have seen little study. Ye *et al.* (2012) enumerate the special cases of the weighted GB2 under a flexible weight function, including the Singh–Maddala (Singh and Maddala, 1976), Dagum (Dagum, 1977) and Fisk (Fisk, 1961), and give moments for these when the weight takes the form x^{α}, but do not identify either the size-biased GB2 or its special cases as being GB2 themselves. The Singh–Maddala distribution, also known as the Burr XII (Burr, 1942), has been widely used for modeling income distributions. The Dagum distribution, also known as the Burr III (Burr, 1942) and the kappa (Mielke, 1973), is less well known than the Singh–Maddala but has also seen wide use for modeling income distributions. Kleiber (2008) provides a recent review of properties, estimation, and application for the Dagum with an emphasis on econometrics; Lindsay *et al.* (1996) and Gove *et al.* (2008) employ this distribution for modeling tree diameter distributions. As Table 2 shows that both the size-biased Singh–Maddala and the size-biased Dagum are GB2 with restricted parameters. The Fisk distribution was also developed for modeling wealth and income distributions. A special case of both the Singh–Maddala and Dagum, it also called the log-logistic in survival analysis and environmental applications (e.g. Bennett, 1983; Ahmad *et al.*, 1988). Properties of the length-biased Fisk were explored by Correa and Wolfson (2007). In general, the size-biased Fisk is also GB2, with parameter restrictions combining those of the Singh–Maddala and Dagum.

None of the size-biased distributions in Table 2 includes a shift (or location) parameter, but for some applications the introduction of a shift parameter is important. Let $Y \sim g(y; [\theta\psi])$ be a shifted version of a distribution $X \sim f(x; \theta)$, such that $Y \sim f(y - \psi; \theta)$. Then the size-biased distribution of order α is

$$g^*_{\alpha}(y; [\theta\psi]) = \frac{y^{\alpha} f(y - \psi; \theta)}{\int y^{\alpha} f(y - \psi; \theta)\, \mathrm{d}y} \tag{12}$$

We can see that in general, $g^*_{\alpha}(y; [\theta\psi]) \neq f^*_{\alpha}(y - \psi, \theta)$ because the weight $w = y^{\alpha} \neq (y - \psi)^{\alpha}$. Gove (2003b) demonstrates how formulae for the size-biased Weibull of integer order can be obtained by expansion using the binomial theorem. We speculate that approach will work equally well for the GB1 or GB2 and their special cases.

Size-biased distributions have a wide range of potential applications (Patil and Rao, 1978), and the family of continuous distributions represented by the GB1 and GB2 includes many special cases with great flexibility (McDonald, 1984). We hope that a unified framework for the use of size-biased distributions within this family will facilitate new work in this area.

Funding

Partial funding was provided by the New Hampshire Agricultural Experiment Station. This is Scientific Contribution Number 2576. This work was supported by the USDA National Institute of Food and Agriculture McIntire Project NH-00067-M, #226391.

Conflict of interest statement

None declared.

References

Ahmad, M.I., Sinclair, C.D. and Werritty, A. 1988 Log-logistic flood frequency analysis. *J. Hydrol.* **98**, 205–224.

Alavi, S.M.R. and Chinipardaz, R. 2009 Form-invariance under weighted sampling. *Statistics* **43**, 81–90.

Bailey, R.L. and Dell, T.R. 1973 Quantifying diameter distributions with the Weibull function. *For. Sci.* **19**, 97–104.

Bennett, S. 1983 Log-logistic regression models for survival data. *Appl. Stat.* **32**, 165–171.

Bliss, C.I. and Reinker, K.A. 1964 A log-normal approach to diameter distributions in even-aged stands. *For. Sci.* **10**, 350–360.

Burr, I.W. 1942 Cumulative frequency functions. *Ann. Math. Stat.* **13**, 215–232.

Clutter, J.L. and Bennett, F.A. 1965 *Diameter distributions in old-field slash pine plantations*. Ga. For. Res. Counc. Rep. 13, 9 pp.

Correa, J.A. and Wolfson, D.B. 2007 Length-bias: some characterizations and applications. *J. Stat. Comp. Sim.* **64**, 209–219.

Dagum, C. 1977 A new model of personal income distribution: specifications and estimation. *Econ. Appliqueé* **30**, 413–436.

Dennis, B. and Patil, G.P. 1984 The gamma distribution and weighted multimodal gamma distributions as models of animal abundance. *Math. Biosci.* **68**, 187–212.

de Vries, P.G. 1979 Line intersect sampling – statistical theory, applications, and suggestions for extended use in ecological inventory. In *Sampling Biological Populations*. R.M. Cormack, G.P. Patil and D.S. Robson (eds). International Cooperative Publishing House.

Drummer, T.D. and McDonald, L.L. 1987 Size bias in line transect sampling. *Biometrics* **43**, 13–21.

Ducey, M.J. 2009 Sampling trees with probability nearly proportional to biomass. *For. Ecol. Manage.* **258**, 2110–2116.

Ducey, M.J. and Valentine, H.T. 2009 Direct sampling for stand density index. *West. J. Appl. For.* **23**, 78–82.

Fisk, P.R. 1961 The graduation of income distributions. *Econometrica* **29**, 171–185.

Gorgosa-Varela, J.J., Rojo-Alboreca, A., Afif-Khouri, E. and Barrio-Anta, M. 2008 Modelling diameter distributions of birch (*Betula alba* L.) and pedunculate oak (*Quercus robur* L.) stands in northwest Spain with the beta distribution. *Inv. Agr: Sist. Recurs. For.* **17**, 271–281.

Gove, J.H. 2000 Some observations on fitting assumed diameter distributions to horizontal point sampling data. *Can. J. For. Res.* **30**, 521–533.

Gove, J.H. 2003a Estimation and applications of size-biased distributions in forestry. In *Modeling Forest Systems*. A. Amaro, D. Reed and P. Soares (eds). CABI Publishing, pp. 201–212.

Gove, J.H. 2003b Moment and maximum likelihood estimators for Weibull distributions under length- and area-biased sampling. *Environ. Ecol. Stat.* **10**, 455–467.

Gove, J.H. and Patil, G.P. 1998 Modeling the basal area-size distribution of forest stands: a compatible approach. *For. Sci.* **44**, 285–297.

Gove, J.H., Ducey, M.J., Leak, W.B. and Zhang, L. 2008 Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr Type III distribution. *Forestry* **81**, 161–176.

Hafley, W.L. and Schreuder, H.T. 1977 Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* **7**, 481–487.

Kaiser, L. 1983 Unbiased estimation in line-intercept sampling. *Biometrics* **39**, 965–976.

Kleiber, C. 2008 A guide to the Dagum distributions. In *Modeling Income Distributions and Lorenz Curves*. D. Chotikapanich (ed.). Springer, pp. 97–117.

Leak, W.B. 1964 The J-shaped probability distribution. *For. Sci.* **11**, 405–409.

Lenhart, J.D. and Clutter, J.L. 1971 *Cubic-foot yield tables for old field loblolly pine plantations in the Georgia piedmont*. Ga. For. Res. Counc. Rep., 22, series 3, 13 pp.

Li, F., Zhang, L. and Davis, C.J. 2002 Modeling the joint distribution of tree diameters and heights by bivariate generalized Beta distribution. *For. Sci.* **48**, 47–58.

Lindsay, S.R., Good, G.R. and Woolons, R.C. 1996 Modelling the diameter distribution of forest stands using the Burr distribution. *J. Appl. Stat.* **23**, 609–619.

McDonald, J.B. 1984 Some generalized functions for the size distribution of income. *Econometrica* **52**, 647–664.

McDonald, J.B. and Xu, Y.J. 1995 A generalization of the beta distribution with applications. *J. Econometrics* **66**, 133–152.

Meyer, H.A. 1952 Structure, growth, and drain in balanced uneven-aged forests. *J. For.* **50**, 85–92.

Mielke, P.W. 1973 Another family of distributions for describing and analyzing precipitation data. *J. Appl. Meteorol.* **12**, 275–280.

Moeltner, K. and Shonkwiler, J.S. 2005 Correcting for on-site sampling in random utility models. *Am. J. Agric. Econ.* **87**, 327–339.

Nelson, T.C. 1964 Diameter distribution and growth of loblolly pine. *For. Sci.* **10**, 105–115.

Patil, G.P. 1981 Studies in statistical ecology involving weighted distributions. In *Applications and New Directions*. J.K. Ghosh and J. Roy (eds). Proc. Indian Stat. Inst. Golden Jubilee, Statistical Publishing Society, pp. 478–503.

Patil, G.P. and Ord, J.K. 1976 On size-biased sampling and related form-invariant weighted distributions. *Sankhya Ser. B* **38**, 48–61.

Patil, G.P. and Rao, C.R. 1978 Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179–189.

Priyadarshani, H.A. 2011 Statistical properties of weighted generalized gamma distribution. MS Thesis, Georgia State University, Statesboro, Georgia.

Scheaffer, R.L. 1972 Size-biased sampling. *Technometrics* **14**, 635–644.

Singh, S.K. and Maddala, G.S. 1976 A function for the size distribution of incomes. *Econometrica* **44**, 963–970.

Stacy, E.W. 1962 A generalization of the gamma distribution. *Ann. Math. Stat.* **33**, 1187–1192.

Svensson, I., Sjostedt-de Luna, S., Morling, T., Fries, A. and Ericsson, T. 2007 Adjusting for fibre length-biased sampling probability using increment cores from standing trees. *Holzforschung* **61**, 101–103.

Terwilliger, J.D., Shannon, W.D., Lathrop, G.M., Nolan, J.P., Goldin, L.R., Chase, G.A. *et al.* 1997 True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Am. J. Hum. Genet.* **61**, 430–438.

Van Deusen, P.C. 1986 Fitting assumed distributions to horizontal point sample diameters. *For. Sci.* **32**, 146–148.

Warren, W.G. 1975 Statistical distributions in forestry and forest products research. In *Statistical Distributions in Scientific Work*, vol. 2. G.P. Patil, S. Kotz and J.K. Ord (eds). D. Reidel, pp. 369–384.

Warren, W.G. and Olsen, P.F. 1964 A line intersect technique for assuming logging waste. *For. Sci.* **10**, 267–276.

Woodall, C.W. and Williams, M.S. 2005 *Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program*. USDA For. Serv., Gen. Tech. Rep. NC-256. 48 p.

Ye, Y., Oluyede, B.O. and Pararai, M. 2012 Weighted generalized beta distribution of the second kind and related distributions. *J. Stat. Econometric Methods* **1**, 13–31.