Giri S. Lingappaiah
Generalized length biased distributions

Aplikace matematiky, Vol. 33 (1988), No. 5, 354--361

Persistent URL: http://dml.cz/dmlcz/104316

33 (1988) APLIKACE MATEMATIKY No. 5, 354—361

GENERALIZED LENGTH BIASED DISTRIBUTIONS

G. S. Lingappaiah

(Received September 10, 1986)

Summary. Generalized length biased distribution is defined as $h(x) = \phi_r(x) f(x)$, $x > 0$, where $f(x)$ is a probability density function, $\phi_r(x)$ is a polynomial of degree r, that is, $\phi_r(x) = a_1(x/\mu'_1) + \ldots + a_r(x^r/\mu'_r)$, with $a_i > 0$, $i = 1, \ldots, r$, $a_1 + \ldots + a_r = 1$, $\mu'_i = \mathsf{E}(x^i)$ for $f(x)$, $i = 1, 2, \ldots, r$. If $r = 1$, we have the simple length biased distribution of Gupta and Keating [1]. First, characterizations of exponential, uniform and beta distributions are given in terms of simple length biased distributions. Next, for the case of generalized distribution, the distribution of the sum of n independent variables is put in the closed form when $f(x)$ is exponential. Finally, Bayesian estimates of $a_1, \ldots, a_r$ are obtained for the generalized distribution for general $f(x)$, $x > 0$.

Key words: length biased distributions; exponential, beta and uniform distributions; Bayesian estimates

AMS Classification: 62E15.

1. CHARACTERIZATIONS

Consider the simple length biased distribution (SLBD) represented by

$$h(x) = (x/\mu) f(x), \quad x > 0 \tag{1}$$

with $\mu = \mathsf{E}(x)$ for a probability density function $f(x)$. Putting $f_1(x) = f(x)$, $f_2(x) = h(x)$, (1) can be rewritten as

$$f_2(x) = [x/\mu'_{(1)1}] f_1(x) \tag{2}$$

where $\mu'_{(1)1} = \mu = \mathsf{E}(x)$ for $f(x)$. Let the n-th order SLBD be

$$f_n(x) = [x/\mu'_{(n-1)1}] f_{n-1}(x) \tag{2a}$$

where $\mu'_{(i)1} = \mathsf{E}(x)$ for $f_i(x)$, $i = 1, 2, \ldots, n$.

(a) **Exponential:** $f(x)$ is $\mathsf{E}(0, \theta)$

Now

$$f_1(x) = \theta\, e^{-\theta x}, \quad x > 0 \tag{3}$$

and $\mu'_{(1)1} = 1/\theta$; from (2),

(3a) $$f_2(x) = (\theta x)\,\theta\, e^{-\theta x}$$

and $\mu'_{(2)1} = 2/\theta^2$. Hence

(4) $$f_3(x) = [x/\mu'_{(2)1}]\, f_2(x)$$

(5) $$= e^{-\theta x}(\theta x)^2\, \theta/2$$

Continuing in this way, one gets

(6) $$f_n(x) = e^{-\theta x}(\theta x)^{n-1}\, \theta/\Gamma(n)$$

Now, the characterization is as follows: $f(x)$ in (1) is $\mathsf{E}(0, \theta)$ if and only if $f_n(x)$ is of the same form as $f^{(n)}(x)$ where $f^{(n)}(x)$ is the n-fold convolution, $f(x) * f(x) * \ldots \ldots * f(x)$. "If" part is proved above. To prove "iff" part, let (6) be true. The from (2a), we have

(7) $$e^{-\theta x}(\theta x)^{n-1}\, \theta/\Gamma(n) = (x/\mu'_{(n-1)1})\, f_{n-1}(x)$$

From (7), it follows

(8) $$f_{n-1}(x) = \mu'_{(n-1)1}\, e^{-\theta x}(\theta x)^{n-2}\, \theta^2/\Gamma(n)$$

and since $f_{n-1}(x)$ is a pdf, we get from (8)

(9) $$(\mu'_{(n-1)1})\left(\frac{\theta}{n-1}\right) = 1$$

and hence $\mu'_{(n-1)1} = (n-1)/\theta$.

From (8) and (9), we get

(10) $$f_{n-1}(x) = e^{-\theta x}(\theta x)^{n-2}\, \theta/\Gamma(n-1)$$

Continuing in this way, one gets $f_1(x) = \theta\, e^{-\theta x}$, $x > 0$.

(b) $f(x)$ **is uniform** $U(0, \theta)$

Now let

(11) $$f(x) = 1/\theta\,, \quad 0 < x < \theta$$

The characterization is as follows: $f(x)$ in (1) is $U(0, \theta)$ if and only if $f_n(x)$ is of the same form as $f(x_{(n)})$ where $x_{(n)} = \max(x_1, \ldots, x_n)$

Proof. From (2), we get for this case that

(12) $$f_2(x) = [x/\mu'_{(1)1}]\,[1/\theta]$$

and from (11), $\mu'_{(1)1} = \theta/2$. Then

(13) $$f_2(x) = \frac{2x}{\theta^2}\,, \quad 0 < x < \theta\,,$$

and from (13), we get $\mu'_{(2)1} = \frac{2}{3}\theta$. Further, from (4)

(14) $$f_3(x) = 3x^2/\theta^3, \quad 0 < x < \theta$$

Continuing one gets

(15) $$f_n(x) = nx^{n-1}/\theta^n$$

and (15) is of the same form as $f(x_{(n)})$. To prove "iff," let (15) be true. Then

(16) $$f_n(x) = nx^{n-1}/\theta^n = [x/\mu'_{(n-1)1}]\, f_{n-1}(x)$$

hence

(16a) $$f_{n-1}(x) = (\mu'_{(n-1)1})\,(nx^{n-2}/\theta^n)$$

Since $f_{n-1}(x)$ is a pdf, (16a) gives

$$[\mu'_{(n-1)1}]\,[n/(n-1)\,\theta] = 1$$

Hence

(16b) $$\mu'_{(n-1)1} = (n-1)\,\theta/n$$

(16a) and (16b) give

(17) $$f_{n-1}(x) = (n-1)\,x^{n-2}/\theta^{n-1}$$

Continuing, one gets

(17a) $$f_1(x) = \frac{1}{\theta}, \quad 0 < x < \theta$$

(c) $f(x)$ **is beta distribution** $B_x(a, b)$

Let $f(x)$ in (1) be

(18) $$f(x) = [x^{a-1}(1-x)^{b-1}/B(a, b)], \quad 0 < x < 1,$$

where $B(a, b) = \Gamma(a)\,\Gamma(b)/\Gamma(a+b)$, $a, b > 0$. Then $\mu'_{(1)1} = B(a+1, b)/B(a, b)$ and (2) gives

(19) $$f_2(x) = x^a(1-x)^{b-1}/B(a+1, b)$$

(19) gives $\mu'_{(2)1} = B(a+2, b)/B(a+1, b)$ and now from (4), one gets

(20) $$f_3(x) = x^{a+1}(1-x)^{b-1}/B(a+2, b)$$

Continuing, one gets,

(21) $$f_n(x) = x^{(a+n-1)-1}(1-x)^{b-1}/B(a+n-1, b)$$

Hence the characterization is as follows: $f(x)$ in (1) is $B_x(a, b)$ if, and only if, $f_n(x)$ is $B_x(a+n-1, b)$. "If" part is shown above. "Iff" can be proved similarly to cases (a) and (b).

2. MOMENTS

(i): From (2), we have

(22) $$\mu'_{(2)1} = \mu'_{(1)2}/\mu'_{(1)1}$$

where $\mu'_{(1)i} = \mathsf{E}(x^i)$ for $f_1(x)$ and (22) gives

(22a) $$\mu'_{(1)2} = \mu'_{(1)1}\mu'_{(2)1} \, .$$

Continuing, we get

(23) $$\mu'_{(1)n} = \mu'_{(1)1}\mu'_{(2)1} \dots \mu'_{(n)1}$$

In (23), the right-hand side is a product of $\mathsf{E}(x)$ for $f_1(x), \dots, f_n(x)$ respectively while left-hand side is $\mathsf{E}(x^n)$ for $f_1(x)$. If $f_1(x)$ is $\mathsf{E}(0, \theta)$ then

(23a) $$\mu'_{(1)n} = \mathsf{E}(x^n) \quad \text{for} \quad f_1(x) = n!/\theta^n$$

and

(24) $$\mu'_{(i)1} = (i/\theta)$$

From (23a) and (24), (23) follows. Similarly, one can check (23) easily for $U(0, \theta)$ and $B_x(a, b)$ cases.

(ii): Now, consider the generalized length biased distribution (GLBD) whose pdf is

(25) $$h(x) = [(x/\mu'_1)\, a_1 + \dots + (x^r/\mu'_r)\, a_r]\, f(x)$$

with $x > 0, a_i > 0, i = 1, 2, \dots, r$, $\mathsf{E}(x^i) = \mu'_i$ for some pdf $f(x)$, and $a_1 + \dots + a_r = 1$. Let $\mu'_{0(i)}$ be $\mathsf{E}(x^i)$ for $h(x)$. Then (25) gives

(26) $$\underset{\sim}{\boldsymbol{U}}_0 = \boldsymbol{A}'\underset{\sim}{\boldsymbol{U}}$$

where $\boldsymbol{A}' = (a_1, \dots, a_r)$ and $\underset{\sim}{\boldsymbol{U}}'_0 = (\mu'_{(0)1}, \dots, \mu'_{(0)s})$, and the (i, j)-th element of $\underset{\sim}{\boldsymbol{U}}$ is U_{ij} where

(26a) $$U_{ij} = (\mu'_{i+j}/\mu'_j)\, , \quad i = 1, 2, \dots, s\, ; \quad j = 1, 2, \dots, r$$

For $U(0, \theta)$, $\mathsf{E}(0, \theta)$ cases, U_{ij}'s can be evaluated very easily.

3. DISTRIBUTION OF THE SUM

(i): Consider (25). Suppose $f(x)$ is $\mathsf{E}(0, \theta)$. Since $\mu'_i = i!/\theta^i$, (25) can be written as

(27) $$h(x) = \sum_{i=1}^{r} a_i\, G(i + 1, \theta, x)$$

where $G(i, \theta)$ is the gamma pdf

(27a) $$\mathrm{e}^{-\theta x}(\theta x)^{i-1}\, \theta/\Gamma(i)$$

So, in this case of $\mathsf{E}(0, \theta)$, $h(x)$ is a mixture of gamma pdf's. If $\phi(t)$ is the characteristic function of $h(x)$, then

(28) $$\phi(t) = \sum_{j=1}^{r}\left[a_j \Big/ \left(1 - \frac{it}{\theta}\right)^{j+1}\right]$$

and hence

(28a) $$\phi^n(t) \propto \sum a_1^{j_1} \dots a_r^{j_r} \Big/ \left(1 - \frac{it}{\theta}\right)^{2j_1 + 3j_2 + \dots + (r+1)j_r}$$

where $j_1 + \dots + j_r = n$ and the sum $\sum$ is over all permutations of $j_1, \dots, j_r$. On inverting, (28a), one gets the distribution of $y = x_1 + \dots + x_n$, where x_i are i.i.d. each having $h(x)$ as pdf. Hence pdf of y is

(29) $$h(y) = a_0 \sum a_1^{j_1} a_2^{j_2} \dots a_r^{j_r}\, G(2j_1 + \dots + (r+1)\, j_r, \theta; y)$$

with $j_1 + \dots + j_r = n$ and $a_0 = n!/j_1! \dots j_r!$.

If $r = n = 2$, then, we get (29) as

(29a) $$h(y) = a_1^2\, G(4, \theta; y) + 2a_1 a_2\, G(5, \theta; y) + a_2^2\, G(6, \theta; y)$$

and for $n = 3$, $r = 2$, one gets

(29b) $$h(y) = a_1^3\, G(6, \theta; y) + 3a_1^2 a_2\, G(7, \theta, y) + 3a_1 a_2^2\, G(8, \theta; y) + a_2^3\, G(9, \theta; y)$$

(ii): Now suppose x_i's are independent but x_s has the parameter θ_s, $s = 1, \dots, n$. Then

(30) $$\phi_y(t) = \prod_{s=1}^{n} \phi_s(t) = \prod_{s=1}^{n} \left[\frac{a_1}{\left(1 - \frac{it}{\theta_s}\right)^2} + \dots + \frac{a_r}{\left(1 - \frac{it}{\theta_s}\right)^{r+1}} \right]$$

where $\phi_s(t)$ is the characteristic function of x_s, that is

(31) $$\phi_s(t) = \sum_{j=1}^{r} \left[a_j \Big/ \left(1 - \frac{it}{\theta_s}\right)^{j+1} \right]$$

Hence (30) is

(32) $$\phi_y(t) = \prod_{s=1}^{n} \left[a_1 b_{s1} + \dots + a_r b_{sr} \right]$$

with

$$b_{sk} = 1 \Big/ \left(1 - \frac{it}{\theta_s}\right)^{k+1}, \quad s = 1, 2, \dots, n\,; \quad k = 1, 2, \dots, r\,.$$

We see that

(33) $$\phi_y(t) = \sum \prod_{k=1}^{r} a_k^{j_k} \left[\prod_{m=1}^{j_k} b_{(k_m)k} \right]$$

with $k_m \neq k_l$ for $m \neq l$; $k_m, k_l = 1, 2, \dots, n$. That is, $k_m \neq k_l$ for *same* k or *different* k. For example, if $n = 4$, $r = 2$, for $a_1^2 a_2^2$, we have $\prod_{m=1}^{2} b_{(1_m)1} b_{(2_m)2}$. All (1_m)'s and (2_m)'s are different *within* as well as *between* the products. Suppose $r = 2$, $n = 3$, then (33) is,

(34) $$a_1^3 b_{11} b_{21} b_{31} + a_2^3 b_{12} b_{22} b_{32} + a_1 a_2^2 [b_{11} b_{22} b_{32} + b_{21} b_{32} b_{12} + b_{31} b_{12} b_{22}] + \\ + a_1^2 a_2 [b_{12} b_{31} b_{21} + b_{22} b_{11} b_{31} + b_{32} b_{11} b_{21}]\,.$$

That is, in (33), the outside subscript k in $(k_m)\,k$ repeats j_k times when the exponent of a_k is j_k.

Now, we have from (33),

$$\phi_y(t) = \sum \prod_{k=1}^{r} a_k^{j_k} \left[\prod_{m=1}^{j_k} \left(1 \Big/ \left(1 - \frac{it}{\theta_{(k_m)}} \right)^{k+1} \right) \right] \tag{35}$$

For example, if $r = 2$, $n = 2$, (35) is

$$\begin{aligned} \phi_y(t) = & \left[a_1^2 \Big/ \left(1 - \frac{it}{\theta_1}\right)^2 \left(1 - \frac{it}{\theta_2}\right)^2 \right] + \left[a_2^2 \Big/ \left(1 - \frac{it}{\theta_1}\right)^3 \left(1 - \frac{it}{\theta_2}\right)^3 \right] + \\ & + \left[a_1 a_2 \Big/ \left(1 - \frac{it}{\theta_1}\right)^3 \left(1 - \frac{it}{\theta_2}\right)^2 \right] + \left[a_1 a_2 \Big/ \left(1 - \frac{it}{\theta_1}\right)^2 \left(1 - \frac{it}{\theta_2}\right)^3 \right] \end{aligned} \tag{36}$$

On inverting (35), one gets the pdf of y.

4. BAYESIAN ESTIMATES

Let the priors for $a_1, \ldots, a_r$ be the Dirichlet distribution given in Lingappaiah [2], [3]. That is, $f(a) = D(d_1, \ldots, d_r)$ where

$$f(\boldsymbol{a}) = a_1^{d_1 - 1} \ldots a_{r-1}^{d_{r-1} - 1} (1 - a)^{d_r - 1} / B(d_1, \ldots, d_r) \tag{37}$$

where $d_i > 0$, $i = 1, \ldots, r$ and $B(d_1, \ldots, d_r) = \Gamma(d_1) \ldots \Gamma(d_r)/\Gamma(d)$ with $d = d_1 + \ldots \ldots + d_r$; $\boldsymbol{a} = (a_1, \ldots, a_r)$, $a = a_1 + \ldots + a_{r-1}$.

Now from (25), the likelihood is

$$L(\boldsymbol{x} \mid \boldsymbol{a}) = \prod_{s=1}^{n} h(x_s) = \prod_{s=1}^{n} [a_1(x_s/\mu_1') + \ldots + a_r(x_s^r/\mu_r')] f(x_s) \tag{38}$$

$$= \prod_{s=1}^{n} [a_1 C_{s1} + \ldots + a_r C_{sr}] f(x_s) \tag{39}$$

Similarly to (32), formula (39) can be written as

$$= \sum \prod_{k=1}^{r} a_k^{j_k} [\prod_{m=1}^{j_k} C_{(k_m)k}] [\prod_{s=1}^{n} f(x_s)] \tag{40}$$

where $C_{(k_m)k} = x_{(k_m)}^k / \mu_k'$, $k_m = 1, 2, \ldots, n$.

Then from (37) and (40), we get,

$$\begin{aligned} L(\boldsymbol{x}) &= \int L(\boldsymbol{x} \mid \boldsymbol{a}) f(\boldsymbol{a}) \, \mathrm{d}\boldsymbol{a} = \\ &= A \sum \prod_{k=1}^{r} [\prod_{m=1}^{j_k} C_{(k_m)k}] B(j_1 + d_1, \ldots, d_r + j_r) \end{aligned} \tag{41}$$

with

$$A = [\prod_{s=1}^{n} f(x_s)] / B(d_1, \ldots, d_r) . \tag{41a}$$

From (41), we get the Bayesian estimate $\hat{a}_t$ as

(42) $$\hat{a}_t = \int a_t\, L(\mathbf{x} \mid \mathbf{a}) f(\mathbf{a})\, d\mathbf{a} / \int L(\mathbf{x} \mid \mathbf{a}) f(\mathbf{a})\, d\mathbf{a} =$$

(43) $$= \frac{A \,.\, \sum_{j_1 + \ldots + j_r = n} \left[\prod_{k=1}^{r} \prod_{m=1}^{j_k} C_{(k_m)k}\right] B(j_1 + d_1, \ldots, j_t + d_t + 1; \ldots j_r + d_r)}{L(\mathbf{x})}$$

For example if $n = r = 2$, then (40) may be written as

(43a) $$a_1^2 C_{11} + a_1 a_2 (C_{12} + C_{21}) + a_2^2 C_{22}$$

where

$$C_{11} = x_1 x_2 / (\mu_1')^2\,, \quad C_{22} = x_1^2 x_2^2 / (\mu_2')^2$$

$$C_{12} = x_1 x_2^2 / (\mu_1' \mu_2')\,, \quad C_{21} = x_1^2 x_2 / (\mu_1' \mu_2')$$

and for this case, the Bayesian estimate of a_1 is

(44) $$\hat{a}_1 =$$

$$= \frac{[B(d_1 + 3, d_2)]\, C_{11} + [B(d_1 + 2, d_2 + 1]\, (C_{12} + C_{21}) + [B(d_1 + 1, d_2 + 2)]\, C_{22}}{[B(d_1 + 2, d_2)]\, C_{11} + [B(d_1 + 1, d_2 + 1]\, (C_{12} + C_{21}) + [B(d_1, d_2 + 2)]\, C_{22}}$$

For example, for the $U(0, \theta)$ case, with $x_1 = \cdot 25$, $x_2 = \cdot 5$, $\theta = 1$, $d_1 = d_2 = 1$, we get (with $\mu_1' = 1/2$, $\mu_2' = 1/3$), (44) as

(45) $$\hat{a}_1 = 31/48$$

Comments: 1. The origin and the usefulness of length biased distributions are well explained in Gupta and Keating [1]. 2. By using (40), one can get the Bayesian estimate of θ (with a proper prior), which is included in $f(x)$ and in μ_i''s, $i = 1, \ldots, r$. 3. Though for large r the summation $\sum$ over $j_1 + j_2 + \ldots + j_r = n$, is somewhat cumbersome, it can be easily performed on computers. 4. Only in the case of exponential and a few other cases, $h(x)$ may turn out to be a mixture, while in general $h(x) = \phi_r(x) f(x)$ where $\phi_r(x)$ is a polynomial of degree r. 5. For arbitrary $f(x)$ in $h(x)$, it is quite difficult to find the distribution of $y = x_1 + \ldots + x_n$.

References

[1] *Ramesh Gupta, Jerome P. Keating:* Relations for reliability measures under length biased sampling. Scand. J. Stat. 13 (1986), 49–56.

[2] *G. S. Lingappaiah:* On the Dirichlet Variables. J. Stat. Research, 11 (1977), 47–52.

[3] *G. S. Lingappaiah:* On the generalized inverted Dirichlet distribution. Demonstratio Math. 9 (1976), 423–433.

Souhrn

ZOBECNĚNÁ DÉLKOVĚ ZKRESLENÁ ROZLOŽENÍ

G. S. LINGAPPAIAH

Zobecněné délkově zkreslené rozložení je definováno jako $h(x) = \phi_r(x) f(x)$, $x > 0$, kde $f(x)$ je hustota pravděpodobnosti, $\phi_r(x)$ je polynom stupně r, to jest $\phi_r(x) = a_1(x/\mu_1') + \ldots$

$\ldots + a_r(x^r/\mu'_r)$, kde $a_i > 0$, $i = 1, \ldots, r$, $a_1 + \ldots + a_r^r = 1$, $\mu'_i = \mathsf{E}(x^i)$ pro $f(x)$, $i = 1, 2, \ldots, r$. Pro $r = 1$ máme jednoduché délkově zkreslené rozložení Gupty a Keatinga [1]. V článku se nejprve charakterizují exponenciální, rovnoměrné a beta rozložení pomocí jednoduchých délkově zkreslených rozložení. Dále pro zobecněná rozložení pro případ exponenciálního $f(x)$ je odvozeno rozložení součtu n nezávislých veličin. Konečně jsou uvedeny bayesovské odhady $a_1, \ldots, a_r$ pro zobecněná rozložení.

ОБОБЩЕННЫЕ ДИСТАНЦИОННО СМЕЩЕННЫЕ РАСПРЕДЕЛЕНИЯ

G. S. Lingappaiah

Обобщенное дистанционно смещенное распределение определятся формулой $h(x) = \phi_r(x) \cdot f(x)$, $x > 0$, где $f(x)$ — плотность вероятности и $\phi_r(x)$ — многочлен степени r вида $\phi_r(x) = a_1(x/\mu'_1) + \ldots + a_r(x^r/\mu'_r)$, где $a_i > 0$ для $i = 1, \ldots, r$, $a_1 + \ldots + a_r = 1$ и $\mu'_i = \mathsf{E}(x^i)$ для $f(x)$ и $i = 1, 2, \ldots, r$. В случае $r = 1$ получается простое дистанционно смещенное распределение Гупта и Китинга. В статье прежде всего с помощью простых дистанционно смещенных распределений характеризуются экспоненциальное, равномерное и бета распределения. Потом для обобщенных распределений и экспоненциального $f(x)$ найдено разложение суммы n независимых величин. И наконец проведены оценки Байеса констант $a_1, \ldots, a_r$ в случае обобщенных распределений.

Author'a address: Prof. G. S. Lingappaiah, Department of Mathematics, Sir George Williams Campus, Concordia University, 1455 De Maisonneuve Blvd. West Montreal, Quebec Canada. H3G 1M8.