COMMUN. STATIST.-THEOR. METH., 15(3), 907-918 (1986)

ON DISCRETE WEIGHTED DISTRIBUTIONS AND THEIR USE IN MODEL CHOICE FOR OBSERVED DATA

G. P. Patil
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

C. R. Rao
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

M. V. Ratnaparkhi
Department of Mathematics and Statistics
Wright State University
Dayton, Ohio 45435

Key Words and Phrases: Discrete weighted distributions, bivariate weighted distributions, weight functions, size biased distributions, model specification, form invariance, mixtures, stochastic ordering, Bayesian inference, posterior and weighted posterior distributions, bivariate Poisson-negative hypergeometric distribution.

ABSTRACT

This paper provides a brief structural perspective of discrete weighted distributions in theory and practice. It develops a unified view of previous work involving univariate and bivariate models with some new results pertaining to mixtures, form-invariance and Bayesian inference.

1. INTRODUCTION AND SUMMARY

The concept of weighted distributions can be traced to the study of the effects of methods of ascertainment upon estimation

0361-0926/86/1503-0907$3.50/0

of frequencies by Fisher in 1934 and it was formulated in general terms and developed by Rao in a paper presented at the First International Symposium on Classical and Contagious Discrete Distributions held in Montreal in 1963. Since then, a number of papers have appeared on the subject. The bibliography at the end of this paper provides a few references. A more comprehensive bibliography is under preparation in conjunction with an ongoing advanced research conference on weighted distributions and related weighted methods for statistical analysis and interpretation of encountered data, observational studies, representativeness issues, and resulting inferences. See, for example, Patil, Rao, and Zelen (1985).

R. A. Fisher posed the problem of statistical inference as that of specification, estimation, testing of hypothesis and interpretation of results for future course of action. He laid great stress on specification, which in modern terms may be described as the choice of a family of probability measures on the sample space. A wrong choice may lead to wrong inference, which is sometimes described as the third kind of error. The concept of weighted distributions was developed during the last 25 years as a useful tool in the selection of appropriate models for observed data, specially when samples are drawn without a proper frame.

2. UNIVARIATE DISCRETE WEIGHTED DISTRIBUTIONS

Consider a mechanism generating a non-negative random variable (rv) X with probability (density) function (pdf/pf) $f(x) = f(x;\theta)$. Let $w(x) = w(x,\alpha)$ be a non-negative weight function (wf) and assume that $E[w(X)]$ exists. Denote a new pdf by

$$f^w(x;\theta,\alpha) = f^w(x) = \frac{w(x)f(x)}{E[w(X)]},$$

and the corresponding rv by X^w. Then X and X^w are referred to as original and weighted rv's and their respective distributions are called original and weighted distributions.

When sampling a population with $X \sim f(x)$, the recorded observation X^w turns out to be a weighted version of X with wf $w(x) = x^\alpha$, $\alpha > 0$, X^w is said to be size-biased of order

α, sometimes written as $X^{*\alpha}$. See, for example, Rao (1965, 1985) where $\alpha = 1/2$ was found to provide a good fit to data on the number of albino children in a family ascertained through affected children. When $\alpha = 1$, X^w is written as $X^* = X^{*1}$ and is simply called size-biased, and has size-biased pdf

$$f^*(x) = \frac{xf(x)}{E[X]} .$$

The following tabulation gives certain discrete distributions and their size-biased forms.

Forms of weight functions useful in scientific and statistical literature include:

(1) $w(x) = x^\alpha$ for $x = 1,2,3,1/2,2/3,\ 0 < \alpha < 1.$

(2) $w(x) = \binom{x}{2}^\alpha$ for $\alpha = 1,1/2,\ 0 < \alpha < 1,$

(3) $w(x) = x^{(\alpha)} = x(x-1)\ldots(x-\alpha+1).$

(4) $w(x) = e^{\alpha x},\ \theta^x.$

Table 1: Certain Discrete Distributions and their Size-Biased Versions

rv*	pf	size-biased version
Binomial, $B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	$1+B(n-1,p)$
Poisson, $P(\lambda)$	$e^{-\lambda}\lambda^x/x!$	$1+P(\lambda)$
Negative Binomial, $NB(k,p)$	$\binom{k+x-1}{x}p^k(1-p)^x$	$1+NB(k+1,p)$
Logarithmic Series, $L(\theta)$	$\theta^x/x[-\log(1-\theta)]$	$1+NB(1,p)$
Hypergeometric, $H(n,M,N)$	$\binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$	$1+H(n-1,M-1,N-1)$
Binomial Beta, $BB(k,\delta,\gamma)$	$\binom{n}{x}B(\delta+z,\gamma+n-z)$	$1+BB(k-1,\delta,\gamma)$

(5) $w(x) = x+1, \alpha x+\beta$.

(6) $w(x) = 1-(1-\beta)^x$ for $0 < \beta < 1$.

(7) $w(x) = (\alpha x+\beta)/(\delta x+\gamma)$.

(8) $w(x) = G(x) = \text{Prob}(Y \leq x)$ for some rv Y.

(9) $w(x) = \bar{G}(x) = \text{Prob}(Y > x)$ for some rv Y.

(10) $w(x) = r(x)$, where $r(x)$ is the probability of "survival" of observation x.

<u>Remark</u>: If $r(x) = 1$, $x \in T$ and $r(x) = 0$ otherwise, we generate truncated distributions with a given support T. Thus the study of truncated distributions can be brought under the framework of weighted distributions.

Note that these weight functions are monotone increasing or decreasing functions of x. The following Theorem allows us a useful comparison of X^w with X. The Theorem can be proved using a monotone likelihood ratio type argument between the pdf's of X and X^w.

<u>Theorem 1</u>: The weighted version X^w is stochastically greater or smaller than the original rv X according as the weight function $w(x)$ is monotone increasing or decreasing in x.

As a consequence, we have the next result which compares the expected value of X^w with that of X.

<u>Theorem 2</u>: The expected value of the weighted version X^w is greater or smaller than the expected value of the original rv X according as the weight function $w(x)$ is monotone increasing or decreasing in x.

The following results will be of general interest.

<u>Theorem 3</u>: (Patil and Rao, 1978): Let a non-negative rv X have pdf $f(x)$. Let the weight functions $w_i(x) > 0$ have $E[w_i(X)] < \infty$ for $i = 1,2$, defining the corresponding w_i-weighted rv's of X denoted by X^{w_i}. Then the expected value of the weighted version X^{w_2} is greater or smaller than the expected value of

the weighted version X^{w_1} according as the ratio of the weight functions defined by $r(x) = w_2(x)/w_1(x)$ is monotone increasing or decreasing in x.

Theorem 4: (Patil and Rao, 1978): The expected value of the size-biased version X^* exceeds the expected value of the original rv X by the variance to mean ratio of X. Further, the harmonic mean of X^* is equal to the mean of X.

Theorem 5: (Mahfoud and Patil, 1982): The variance to mean ratio of the original rv X is given by the difference between the arithmetic mean and the harmonic mean of the size-biased version X^*.

Theorem 6: (Mahfoud and Patil, 1982): The hazard rate function of the size-biased version X^* is less than or equal to the hazard rate function of the original rv X.

Theorem 7: (Patil and Ord, 1975): Define original rv X with pdf $f(x;\theta)$ to be form invariant under size-bias of order α if the pdf of the size-biased version of order α given by $f^w(x;\theta,\alpha) \equiv f(x;\eta)$ for some $\eta = \eta(\theta,\alpha)$. Then, under certain mild regularity conditions, X is form invariant under size-bias of order α if and only if X is a member of the log-exponential family defined by a pdf of the form $f(x;\theta) = x^\theta a(x)/m(\theta)$.

3. INVARIANCE AND FORM INVARIANCE UNDER SIZE-BIASED SAMPLING AND MIXTURES

We consider first individual examples, and then formulate the theorem. It is straightforward to establish the individual results.

Result 1: The size-biased version of the mixture of the common-index-parameter binomial distributions, when their binomial parameter has beta type I distribution, arises also as the mixture of the size-biased versions of the binomial distributions, when their binomial parameter has the size-biased version of the original beta type I distribution. In symbols, where $X|p \sim$ Binomial(n,p), $P \sim$ Beta I, $X \wedge P \sim$ Mixture of X on its parameter P, and $\underset{=}{d}$ means equal in distribution.

Result 2: The size-biased version of the mixture of the Poisson distributions, when their parameter has gamma distribution, arises also as the mixture of the size-biased version of the original gamma distribution. In symbols, $(X \wedge \lambda)^* \underset{=}{d} (X^* \wedge \lambda^*)$, where $X|\lambda \sim$ Poisson(λ), and $\lambda \sim$ Gamma.

Result 3: The size-biased version of the mixture of the hypergeometric distributions H(n,M,N), when their parameter M has binomial distribution B(N,p), arises also as the mixture of the size-biased versions of the hypergeometric distributions, when their parameter M has the size-biased version of the original binomial distribution. In symbols, $(X \wedge M)^* \underset{=}{d} (X^* \wedge M^*)$, where $X \sim$ Hypergeometric(n,M,N) and $M \sim$ Binomial(N,p).

Result 4: The size-biased version of the mixture of the common-index-parameter negative binomial distributions, when their odds ratio parameter has beta type II distribution, arises also as the mixture of the size-biased versions of the negative binomial distributions, when their odds ratio parameter has the size-biased version of the original beta type II distribution. In symbols, $(X \wedge P)^* \underset{=}{d} (X^* \wedge P^*)$, where $X \sim$ Negative Binomial(k,p), where $p = 1/(1+P)$, and $P \sim$ Beta II.

We now formulate the following Theorem. The proof is straightforward.

Theorem 8: Let $X|\theta$ be form invariant under size-bias of order $\alpha > 0$. Let $E[X|\theta] = \lambda\theta^\beta$. Let $\theta|\gamma$ be form invariant under size-bias of order $\alpha > 0$. Then $X \wedge \theta$ is form invariant under size-bias of order $\alpha > 0$, and furthermore $(X \wedge \theta)^{*\alpha} \underset{=}{d} (X^{*\alpha} \wedge \theta^{*\beta})$, where $Y^{*\alpha}$ is size-biased version of Y of order α.

4. BIVARIATE WEIGHTED DISTRIBUTIONS

Let (X,Y) be a pair of non-negative rv's with a joint pdf $f(x,y)$ and let $w(x,y)$ be a non-negative wf. The joint pdf of (X,Y) is such that $E[w(X,Y)]$ exists. The weighted form of $f(x,y)$ is

$$f^w(x,y) = \frac{w(x,y)f(x,y)}{E[w(X,Y)]}.$$

The corresponding weighted version of (X,Y) is denoted by $(X,Y)^w$. The following forms of the marginal and conditional distributions of $(X,Y)^w$ are obvious.

$$f^w(x) = \frac{E[w(x,Y)|x]f(x)}{E[w(X,Y)]}$$

$$f^w(y|x) = \frac{w(x,y)f(y|x)}{E[w(x,Y)|x]}.$$

Clearly, both are weighted versions of the corresponding marginal and conditional distributions of (X,Y).

Special cases of weight functions of practical interest are

(1) $w(x,y) = x^\alpha$,

(2) $w(x,y) = x$,

(3) $w(x,y) = \max(x,y)$,

(4) $w(x,y) = x + y$,

(5) $w(x,y) = xy$,

(6) $w(x,y) = x^\alpha y^\beta$

in conjunction with various bivariate discrete and continuous distributions. We record the following results, and then discuss the effects of the weight functions $w(x,y) = w(y)$ and $w(x,y) = w(x)$ as they arise in sample survey situation and in Bayesian inference respectively.

Theorem 9: (Mahfoud and Patil, 1982): Let (X,Y) be a pair of non-negative rv's with joint pdf $f(x,y)$ such that $E[X]$ and $E[X|y]$ exist for all $y > 0$. Let $w(x,y) = x$. Then

$$E[X|Y=y] = H(X^w|Y^w=y), \text{ and}$$

$$\eta^2(X|Y) = \frac{E[H(X^w|Y^w)]-H(X^w)}{E[X^w]-H(X^w)},$$

where H stands for the harmonic mean and η^2 is the correlation ratio of X on Y.

If X and Y are independent r.v.'s, a weight function $w(x,y)$ in general makes X^w and Y^w dependent, which provides a general method of generating bivariate distributions. It would be of interest to examine the nature of dependence introduced by different weight functions. For instance, it is easy to establish the following theorem.

Theorem 10: If X and Y are non-negative independent random variables, then $w(x,y) = x+y$ induces a negative correlation between X^w and Y^w.

An example of a bivariate distribution using a weight function is the following. Let $X \sim P(\lambda p)$, $Y \sim P(\lambda(1-p))$, where $0 < p < 1$, and

$$w(x,y) = B(x+m,y+n)/B(m,n)p^x(1-p)^y.$$

The resulting bivariate distribution of X^w and Y^w is

$$e^{-\lambda} \cdot \frac{\lambda^{x+y}}{(x+y)!} \frac{\binom{-m}{x}\binom{-n}{y}}{\binom{-m-n}{x+y}}$$

which may be called bivariate Poisson-negative hypergeometric distribution (BPNHD). It is interesting to note that this distribution can also be obtained as a mixture of the joint distribution of $X \sim P(\lambda p)$ and $Y \sim P(\lambda(1-p))$ with p having a beta distribution. It is easy to check that the correlation between X^W and Y^W is negative.

Theorem 11: (Mahfoud and Patil, 1982): Let (X,Y) be a pair of non-negative independent rv's with joint pdf $f(x,y) = f_X(x)f_Y(y)$ and let $w(x,y) = \max(x,y)$. Then the rv's $(X,Y)^W$ are dependent. Further, the regression of Y^W on X^W given by $E[Y^W|X^W=x]$ is a decreasing function of x.

Theorem 12: (Patil and Rao, 1978): Let (X,Y) be a pair of non-negative rv's with pdf $f(x,y)$. Let wf $w(x,y) = w(y)$ as is the case in a sampling proportional size type of sample survey. Then the random variables X and X^W are related by

$$f^W(x) = \frac{E[w(Y)|x]f(x)}{E[w(Y)]}.$$

Note that X^W is a weighted version of X in which case the regression function of $w(Y)$ on x serves as the weight function.

Consider the usual Bayesian inference in conjunction with (X,θ) having joint pdf $f(x,\theta) = f(x|\theta)f(\theta) = f(\theta|x)f(x)$. The posterior pdf $f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{\ell(\theta|x)f(\theta)}{E[\ell(\theta|X)]}$ can be interpreted as a weighted version of the prior pdf $f(\theta)$. The weight function is the likelihood function of θ for the observed x.

The following table provides the well known conjugate pairs involving Poisson, binomial, and negative binomial distributions. It should be interesting to revisit Theorem 7 of form invariance of original and weighted distributions in the context of the form invariance of the prior and the posterior distributions and related results in Bayesian literature.

Theorem 13: Consider the usual Bayesian inference in conjunction with (X,θ) with pdf $f(x,\theta) = f(x|\theta)f(\theta) = f(\theta|x)f(x)$. Let $w(x,\theta) = w(x)$ be the weight function for the distribution of $X|\theta$, so that the pdf of $X^W|\theta$ is $w(x)f(x|\theta)/\omega(\theta)$, where $\omega(\theta) = E[w(X)|\theta]$. Then the original and the weighted posteriors are related by

Table 2: Distributions* of $X|\Theta = \theta$, θ, and $\theta|X = x$

$X\|\Theta=\theta$	θ	$\theta\|X=x$
Poisson(θ)	Gamma(k,λ)	Gamma($k+x,\frac{\lambda}{\lambda+1}$)
Binomial(n,θ)	Beta(a,b)	Beta($x+a,n-x+b$)
Negative Binomial(k,θ)	Beta(a,b)	Beta($k+a,x+b$)

*For more examples and for relevant notation and terminology, see Patil et al (1984).

$$f^w(\theta|x) = f(\theta|X^w=x) = \frac{1}{\omega(\theta)}f(\theta|x)/E[\frac{1}{\omega(\theta)}|x]$$

and

$$f(\theta|x) = \frac{\omega(\theta)f^w(\theta|x)}{E[w(\theta)|X^w=x]}.$$

Further the weighted posterior rv $\theta^w|X^w = x$ is stochastically greater or smaller than the original posterior rv $\theta|X = x$ according as $\omega(\theta)$ is monotonically decreasing or increasing function of θ.

Proof: Straightforward, using Theorem 1.

REFERENCES

Chakraborty, R., and Hanis, C. L., (1984). Discussion: Implications of nonrandom sampling in genetic epidemiology. *In Genetic Epidemiology of Coronary Heart Disease: Past, Present, and Future*. New York: Alan R. Liss, Inc., 213-215.

Ewens, W. J., and Asaba, B., (1984). Estimating parameters of the family-size distribution in ascertainment sampling schemes: Numerical results. *Biometrics*, 40, 367-374.

Fisher, R. A., (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics*, 6, 13-25.

Gupta, R. C., (1975). Some characterizations of discrete distributions by properties of their moment distributions. *Comm. Stat.*, 4(8), 761-765.

Hanis, C. L., and Chakraborty, R., (1984). Nonrandom sampling in human genetics: Familial correlations. *J. Math. Appl. Med. & Biol.*, 1, 193-213.

Janardan, K. G., and Rao, B. R., (1984). Certain Inter-relationships between Two classes of probability models for ascertainment of families in segregation analysis. *J. Ind. Statist. Assoc.*, Vol. 22.

Mahfoud, M., and Patil, G. P., (1981). Size biased sampling, weighted distributions, and Bayesian estimation. In *Statistics in Theory and Practice: Essays in Honor of Bertil Matern.* Bo Ranneby, (ed.). Umea, Sweden: Swedish Univ. Agri. Sci., 173-187.

Mahfoud, M., and Patil, G. P., (1982). On weighted distributions. *In Statistics and Probability: Essays in Honor of C. R. Rao*, G. Kallianpur et al (eds.), Amsterdam: North Holland, 479-492.

Patil, G. P., (1981). Studies in statistical ecology involving weighted distributions. In *Statistics Applications and New Directions: Proceedings of ISI Golden Jubilee International Conference*, J. K. Ghosh and J. Roy, (ed.s). Calcutta, India, *Statistical Publ. Soc.*, 478-503.

Patil, G. P., and Ord, J. K., (1975). On size-biased sampling and related form-invariant weighted distributions. *Sankhya*, 37, 48-61.

Patil, G. P., and Rao, C. R., (1977). Weighted distributions: a survey of their applications, In *Applications of Statistics*, P. R. Krishnaiah, (ed.). Amsterdam: North-Holland Publishing Company, 383-405.

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.

Patil, G. P., Boswell, M. T., Joshi, S. W., and Ratnaparkhi, M. V., (1984). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work Vol. 1: Discrete Models.* Burtonsville, Maryland: International Co-operative Publishing House.

Patil, G. P., Boswell, M. T., and Ratnaparkhi, M. V., (1984). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work Vol. 2: Continuous Univariate Models.* Burtonsville, Maryland: International Co-operative Publishing House.

Patil, G. P., Boswell, M. T., Ratnaparkhi, M. V., and Roux, J. J. J., (1984). *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work Vol. 3: Multivariate Models.* Burtonsville, Maryland: International Co-operative Publishing House.

Patil, G. P., Rao, C. R., and Zelen, M., (1985). A computerized bibliography of weighted distributions and related weighted methods for statistial analysis and interpretation of encountered data, observational studies, representativness issues, and resulting inferences. (under preparation). Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University.

Rao, C. R., (1965). On discrete distributions arising out of methods of ascertainment. In *Classical and Contagious Discrete Distributions*, G. P. Patil, (ed.). Calcutta: Pergamon Press and Statistical Publishing Society, 320-332.

Rao, C. R., (1985). Weighted distributions arising out of methods of ascertainment: What population does a sample represent? Chapter 24 in *Celebration of Statistics, The 151 Centenary Volume*, A. C. Atkinson and S. E. Fienberg (eds.), 543-569.

Rosenzweig, M., Zelen, M., VonHoff, D. D., et al. (1978). Waiting for a bus: Does it explain age-dependent differences in response to chemotherapy of early breast cancer. *New England J. Med.*, 299, 1363-1364.

Schotz, W. E., and Zelen, M., (1971). Effect of length sampling bias on labeled mitotic index waves. *J. Theo. Biol.*, 32, 383-404.

Simon, R., (1980). Length biased sampling in etiologic studies. *Amer. J. Epidemiol.*, 111, 444-452.

Stene, J., (1981). Probability Distributions Arising from the Ascertainment and the Analysis of Data on Human Families and Other Groups, *Statistical Distributions in Scientific Work Volume 6-Applications in Physical, Social, and Life Sciences*, C. Taillie, G. P. Patil, and B. A. Baldessari (eds.). Dordrecht: D. Reidel Publishing Company, 233-264.

Vardi, Y., Shepp, L. A., and Logan, B. F., (1981). Distribution functions invariant under residual-lifetime and length biased sampling. Z. Wahrscheinlichtkeit theorie verw, *Geberte*, 56, 427-441.

Zelen, M., (1969). On the theory of screening for chronic disease. *Biometrika*, 56, 601-614.

Zelen, M., (1974). Problems in cell kinetics and the early detection of disease. In *Reliability and Biometry*, F. Proschan and R. J. Serfling (eds.). Philadelphia: SIAM, 701-706.

Revised December, 1985.

Refereed Anonymously.