Weighted distributions

G.P. Patil

Volume 4, pp 2369–2377

in

Encyclopedia of Environmetrics
(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

Weighted distributions

Traditional environmetric theory and practice have been occupied largely with **randomization** and replication. But in environmental and ecological work, observations also fall in the nonexperimental, nonreplicated, and nonrandom categories. The problems of model specification and data interpretation then acquire special importance and great concern. The theory of weighted distributions provides a unifying approach for these problems. Weighted distributions take into account the method of ascertainment, by adjusting the probabilities of actual occurrence of events to arrive at a specification of the probabilities of those events as observed and recorded. Failure to make such adjustments can lead to incorrect conclusions.

The concept of weighted distributions [26] can be traced to the study of the effect of methods of ascertainment upon estimation of frequencies by Fisher [9]. In extending the basic ideas of Fisher, Rao [33, 34] saw the need for a unifying concept and identified various sampling situations that can be modeled by what he called weighted distributions. Within the context of cell kinetics and the early detection of disease, Zelen [41] introduced weighted distributions to represent what he broadly perceived as length-biased sampling (introduced earlier in [4]). In a series of papers with his co-workers, Patil has pursued weighted distributions for purposes of **encountered data** analysis, equilibrium population analysis subject to harvesting and predation, **meta-analysis** incorporating publication bias and heterogeneity, modeling **clustering** and extraneous variation, etc. (see, for example, [5], [12], [17]–[19], [21], [22], [24], [27] and [37]). For more references, see [20].

To introduce the concept of a weighted distribution, suppose X is a non-negative random variable (rv) with its natural probability density function (pdf) $f(x;\theta)$, where the natural parameter is $\theta \in \Omega$ (Ω is the parameter space). Suppose a realization x of X under $f(x;\theta)$ enters the investigator's record with probability proportional to $w(x,\beta)$, so that

$$\frac{\Pr(\text{recording}|X=y)}{\Pr(\text{recording}|X=x)} = \frac{w(y,\beta)}{w(x,\beta)} \qquad (1)$$

Here the recording (weight) function $w(x,\beta)$ is a non-negative function with the parameter β representing the recording (sighting) mechanism. Clearly, the recorded x is not an observation on X, but on the rv X^w, say, having a pdf

$$f^w(x;\theta,\beta) = \frac{w(x,\beta)f(x;\theta)}{\omega} \qquad (2)$$

where ω is the normalizing factor obtained to make the total probability equal to unity by choosing $\omega = \mathrm{E}[w(X,\beta)]$. The rv X^w is called the weighted version of X, and its distribution in relation to that of X is called the weighted distribution with weight function w. Note that the weight function $w(x,\beta)$ need not lie between zero and one, and actually may exceed unity, as, for example, when $w(x,\beta) = x$, in which case $X^* = X^w$ is called the *size-biased version* of X. The distribution of X^* is called the *size-biased distribution* with pdf

$$f^*(x;\theta) = \frac{xf(x;\theta)}{\mu} \qquad (3)$$

where $\mu = \mathrm{E}[X]$. The pdf f^* is called the length-biased or size-biased version of f, and the corresponding observational mechanism is called length-or **size-biased sampling**. Weighted distributions have seen much use as a tool in the selection of appropriate models for observed data drawn without a proper frame. In many situations the model given above is appropriate, and the statistical problems that arise are the determination of a suitable weight function, $w(x,\beta)$, and drawing inferences on θ. Appropriate statistical modeling helps accomplish unbiased inference in spite of the biased data and, at times, even provides a more informative and economic setup; see [24].

The following examples may help illustrate a few situations generating weighted distributions and their applications.

Example 1 *Analysis of family data*, $w(x,\beta) = w(x) = x$. Various demographic studies involve family size and sex ratio as important factors that can have some bearing on the main study. This example shows how a weighted distribution arises as a result of the size-biased sampling.

Consider the data in Table 1 relating to brothers and sisters in families of 104 boys admitted to a postgraduate course. Assume that in families of given size n, the probability of a family with x boys coming into the record is proportional to x. Also, suppose that the number of boys follows a binomial distribution with probability parameter

Table 1 Family data from Example 1

Family size:	1	2	3	4	5	6	7	8	9	10	11	12	13	15	Total
No. of families:	1	6	6	13	12	7	14	11	12	8	6	5	2	1	104
Brothers:	1	8	12	34	34	29	59	50	54	46	32	31	16	8	414
Sisters:	0	4	6	18	26	13	39	38	54	34	34	29	10	7	312

π. Then $f(x;\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$, $w(x) = x$, $\omega = n\pi$, $f^w(x;\pi) = \binom{n-1}{x-1}\pi^{x-1}(1-\pi)^{n-x}$, $\mathrm{E}\left[X^W/n\right] = \pi + (1-\pi)/n > \pi$, and $\mathrm{E}\left[(X^W - 1)/(n-1)\right] = \pi$.

If k boys representing families of size $n_1, n_2, \ldots, n_k$ report $x_1, x_2, \ldots, x_k$ boys, an unbiased estimate of π is

$$\tilde{\pi} = \frac{\sum x_i - k}{\sum n_i - k} = \frac{414 - 104}{726 - 104} = \frac{1}{2} \qquad (4)$$

Example 2 *Analysis of intervention data*, $w(x, \beta) = w(x) = x$. The expected value of the duration to completion of a random event sampled randomly at the end of its duration turns out to be approximately equal to the expected duration of its random intervention. This can be explained using the concept of size-biased/length-biased sampling with weight function $w(x, \beta) = w(x) = x$, where x represents the duration of the random event in the life cycle assessment (LCA). Applications in medicine and environmental health include: (a) cell cycle analysis and pulse labeling [41]; (b) efficacy of early screening for disease and scheduling of examinations [41]; (c) cardiac transplantation [39]; (d) estimation of antigen frequencies [35]; and (e) ascertainment studies in genetics [36].

Example 3 *Modeling clustered sampling, heterogeneity, and extraneous variation*, $w(x, \beta) = w(x, \beta, \theta)$. During their examination of the problem of toxoplasmosis, Diaconis and Efron [6, 8] studied the data and found that there was more dispersion in the data than existing models could accommodate. They therefore introduced a model called the double-exponential family (DEF). This family allows the data analyst to model overdispersion while carrying out usual regression analyses for the mean as a function of the predictors (*see* **Dispersion parameter**). The overdispersion may be due to one or more possible causes, such as clustered sampling, heterogeneity, selection bias, etc.

Interestingly, the DEF can be seen as a weighted distribution; see [18] and [24]. The weight function has the form

$$w(x, \beta) = w(x, \beta, \theta) = \exp(1 - \beta)I(x, \mu(\theta)) \qquad (5)$$

where $I(x, \mu)$ is the *Kullback–Leibler distance function* between x and $\mu(\theta) = \mu = \mathrm{E}[X]$ of the usual exponential family density function f with parameter θ. The Kullback–Leibler distance increases with the distance from the mean μ, thus allowing a more distant observation, larger weight, and accommodating extra dispersion in the dataset when $1 - \beta > 0$.

Example 4 ***Meta-analysis*** *incorporating heterogeneity and publication bias*. Meta-analysis consists of quantitative methods for combining evidence from different studies on a particular issue. Its objective is to summarize quantitatively a research literature with respect to a particular question and to examine systematically the manner in which a collection of studies contributes to knowledge about that question.

The weight function enters the analysis in order to represent the publication/selection bias and the heterogeneity among different studies. It also helps model the overdispersion/underdispersion in the data caused by publication bias and any inherent heterogeneity.

The weight functions examined include:

1. critical value model: $w(x) = (x/x_{\mathrm{crit}})^{\beta}$ if $|x| < x_{\mathrm{crit}}$, and $= 1$ otherwise;
2. half-normal model: $w(x) = \exp[-\beta p(x)^2]$;
3. negative exponential model: $w(x) = \exp[-\beta p(x)]$.

Here $p(x)$ is the ***P* value** when the test statistic takes value x, and x_{crit} stands for the critical value under the test statistic; see [10], [12] and [24].

Example 5 *Statistical analysis incorporating overdispersion and heterogeneity in teratological binary data*. The problem of overdispersion and heterogeneity in **binary data** arises quite naturally in **developmental toxicity studies**. Since a pregnant female is exposed to the chemical dose, litter becomes the primary unit. The random effect of the litter, i.e. biological response of the mother to the chemical dose, affects the toxic responses of the fetuses. This potential random **litter effect** causes heterogeneity, excess variation, and also intra-litter correlation between the responses of fetuses within the litter.

In order to incorporate this random litter effect in the analysis of such binary data, research workers have introduced the **beta-binomial distribution**. The beta-binomial model, regarded as a binomial mixture model, incorporates overdispersion, heterogeneity, and also the clustering of observations within the litter. This clustering occurs because of the tendency of the fetuses within the litter to 'behave' alike. For this reason, the use of a double-binomial family model (a form of DEF distribution) is a useful alternative model in the analysis of developmental toxicity data.

The problem of analysis of overdispersed binary data stemming from different litters at a given dose or from different cities with given rainfall can also be viewed as the problem of encounter data where one is trying to combine the observational data coming from different sources. Therefore one can view both the beta-binomial and also the double binomial as weighted binomial distributions where each model has its own separate weight function.

The two distributions are quite comparable in their capability to describe overdispersed binary data. It appears that the choice between them may be made on such grounds as parameter interpretation and inferential convenience; see [27] and [38].

Example 6 *Extra-distance sources of encounter bias in transect sampling*. The method of **line-transect sampling** has been used to estimate the **abundance** of plants or animals of a particular species in a given region. The line-transect method consists of drawing a baseline across the region to be surveyed and then drawing a line transect through a randomly selected point on the baseline. The surveyor views the surroundings while walking along the line transect and includes the sighted objects of interest in the sample.

It is obvious that the nearer the object or the larger its size, the higher is the probability of sighting the object. Similarly, when the individuals cluster in groups, such as schools or herds, it is appropriate to regard clusters as the basic sampling units; their encounter probabilities are affected by cluster size. Estimates of cluster abundance can be adjusted to individual abundance using the recorded cluster sizes.

Encounter probabilities in transect sampling can be influenced by numerous other factors, such as varying terrain and vegetation cover, weather conditions, time of day, systematic responsive movement toward or away from the transect, etc. Some of these factors are characteristics of the objects themselves and will vary from object to object. 'Size' is an example of such a factor. Other factors are generally survey characteristics but can vary from segment to segment in a multisegmented survey. We refer to members of these two classes as *object factors* and *survey factors*, respectively.

It is clearly desirable to account for as many of these factors as possible, and much of the recent transect sampling literature has been concerned with this issue.

Deep-sea Red Crab

Patil et al. [28, 29] describe a photographic survey in which features of optical geometry partially masked the sighting-distance bias by inflating the recorded counts at larger distances from the transect. The survey's purpose was to determine the abundance of the deep-sea red crab. The data were analyzed using a composite weight function of form

$$w(x) = (a + bx) \cdot v(x; \theta) \tag{6}$$

where the sighting function $v(x; \theta)$ represented the pure sighting-distance bias and would typically involve the unknown parameters, θ. For the red crab study, the exponential-power sighting function gave a reasonable fit to the data and yielded abundance estimates that were consistent with the results of other survey methods [40].

Ramsey [32] has suggested a general technique for constructing parametric visibility functions in which the effective half width ω appears explicitly as a scale parameter. Starting with a *kernel* $h(t; \theta)$, $0 < t < \infty$, which is monotone decreasing in t and satisfies $h(0; \theta) = 0$ and $\int_0^\infty h(t; \theta)\,\mathrm{d}t = 1$, Ramsey's family of sighting functions is defined by

$$w(x; \omega, \theta) = h\left(\frac{x}{\omega; \theta}\right) \tag{7}$$

Here, θ is a vector of nuisance parameters that regulates the shape of the sighting function. A common choice of kernel is the exponential-power form

$$h(t; \gamma) = \exp[-\Gamma(1 + \gamma^{-1})t^{\gamma}] \tag{8}$$

which includes the negative exponential and the half-normal as special cases.

Drummer and McDonald [7] have considered the problem of estimating minke whale abundance where the group size, an object-specific factor, becomes important. They take the group as the basic sighting object and include the group size y as a scaling factor in the exponential-power sighting function

$$\begin{aligned} w(x, y) &= w(x, y; \lambda, \alpha, \gamma) \\ &= \exp\left[\Gamma(1+\gamma^{-1})\left(\frac{x}{\lambda y^{\alpha}}\right)^{\gamma}\right] \end{aligned} \tag{9}$$

Example 7 *Biased sampling, model weight function, and the amount of information.* While it might be supposed that 'biased' sampling is disadvantageous for purposes of inference on the population parameters, this is not always the case. The sampling bias, when properly accounted for, can lead to stronger inferences than would be possible with an equal number of observations from the original distribution. Patil and Taillie [23] consider observations from a one-parameter *exponential family*

$$f(x) = \frac{c(x)\mathrm{e}^{-\theta A(x)}}{M(\theta)} \tag{10}$$

and an equal number of observations from the $w(x)$-weighted version of this family

$$f_w(x) = \frac{w(x)c(x)\mathrm{e}^{-\theta A(x)}}{M_w(\theta)} \tag{11}$$

The intent is to compare the relative information content of the two sets of observations for inferences concerning θ. In particular, which of the two families is more informative for θ? Also, for given $w(x)$, when is it the case that f and f_w are equally informative for θ (a situation that may be described as *Fisher neutral*)? The results are:

1. The weighted version of f is uniformly more informative for θ if and only if $M_w(\theta)/M(\theta)$ is log convex.
2. The weighted version of f is uniformly less informative for θ if and only if $M_w(\theta)/M(\theta)$ is log concave.
3. f and f_w are uniformly equally informative (Fisher neutral) if and only if $M_w(\theta)/M(\theta)$ is log-linear. For given w, this characterizes Fisher neutrality by a functional equation involving M.

Presumably there would be examples in which f is more informative for some θ and less informative for other θ.

For the most part, we consider the size-biased weight function $w(x) = x$. Other weight functions do not introduce new concepts though they may be of practical interest. For example, the weight function $w(0) = 0$, $w(x) = 1$, $x > 0$, allows one to compare the information content of zero-truncated and untruncated observations from a discrete distribution. Bayarri and DeGroot [1] give an extensive treatment of single-parameter truncation models from the perspective of Blackwell sufficiency as well as Fisher information.

Log Exponential Family

The log exponential family has a pdf given by

$$f(x) = \frac{c(x)\mathrm{e}^{\theta(\log x)}}{M(\theta)} = \frac{x^{\theta}\cdot c(x)}{M(\theta)} \tag{12}$$

where $M(\theta)$ is the Mellin transform of $c(x)$. For this family one can show that for arbitrary weight function w, the difference in Fisher information is

$$J_w(\theta) - J(\theta) = \mathrm{var}(\log X_w) - \mathrm{var}(\log X) \tag{13}$$

In particular, the variable with larger logarithmic variance is more informative for θ.

For the remainder of this section we consider only the size-biased weight function, $w(x) = x$, for which $M_w(\theta) = M(\theta + 1)$.

Lognormal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right] \tag{14}$$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\log x)^2\right] \\ &\quad\times \frac{\exp[(\mu/\sigma^2)\log x]}{\exp(\mu^2/2\sigma^2)} \end{aligned} \tag{15}$$

When σ^2 is known, (15) is a log exponential family with $\theta = \mu/\sigma^2$ and

$$M(\theta) = \exp\left[\frac{\mu^2}{2\sigma^2}\right] = \exp\left[\left(\frac{\sigma^2}{2}\right)\theta^2\right] \tag{16}$$

Thus

$$M_w(\theta) = M(\theta+1) = \exp\left[\left(\frac{\sigma^2}{2}\right)(\theta+1)^2\right] \quad (17)$$

and

$$\frac{M_w(\theta)}{m(\theta)} = \exp\left[\left(\frac{\sigma^2}{2}\right)(2\theta+1)\right] \quad (18)$$

Since this is log-linear in θ, it follows that the lognormal is Fisher neutral for μ/σ^2. This has a simple interpretation since the effect of the weight function is to translate the corresponding normal curve to the right by an amount equal to σ^2.

We conclude this section with a curious example in which size biasing sometimes increases information and sometimes decreases information depending upon the value of a second parameter. Consider again the **lognormal distribution** (15) and let $\theta = \mu/\sigma^2$ and $\phi = 1/\sigma^2$. Regard θ as known and ϕ as unknown. The pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{x}\frac{\exp\left[-\phi\dfrac{(\log x)^2}{2} + \theta\log x\right]}{(1/\sqrt{\phi})\exp(\theta^2/2\phi)} \quad (19)$$

so

$$\begin{aligned} M(\phi) &= \frac{\exp(\theta^2/2\phi)}{\sqrt{\phi}} \\ M_w(\phi) &= \frac{\exp[(\theta+1)^2/2\phi]}{\sqrt{\phi}} \quad (20) \\ \frac{M_w(\phi)}{m(\phi)} &= \exp\left(\frac{2\theta+1}{2\phi}\right) \end{aligned}$$

and

$$\frac{d^2}{d\phi^2}\log\frac{M_w(\phi)}{m(\phi)} = \frac{d^2}{d\phi^2}\left(\frac{2\theta+1}{2\phi}\right) = \frac{2\theta+1}{\phi^3} \quad (21)$$

Positivity is equivalent to $2\theta+1>0$ which is equivalent to $(\mu/\sigma^2) > -\frac{1}{2}$. Note that Fisher neutrality occurs when $(\mu/\sigma^2) = -\frac{1}{2}$ which happens when the normal curves corresponding to X and X_w are symmetrically located with respect to the origin.

Example 8 *Informative weighted multiparameter distributions.* Bayarri and DeGroot [1] posed the question of whether 'weighted' observations could be more informative than the corresponding 'natural' observations especially from the point of view of Blackwell's [2, 3] comparison of experiments. Intrigued by the idea that encountered, and biased, data might sometimes be more scientifically revealing than experimental data, Patil and Taillie [25] investigated the issue further.

Let $f(x;\theta)$ be the natural distribution and $f_w(x;\theta) \propto w(x)\cdot f(x;\theta)$ be the corresponding weighted distribution, where θ is a vector of parameters. Also let $J(\theta)$ and $J_w(\theta)$ be the respective Fisher information matrices for θ (*see* **Information matrix**). Intrinsic comparison between the weighted and natural observations is possible when the difference, $J_w(\theta) - J(\theta)$, is either positive definite or negative definite. For example, positive definiteness would mean that *every* scalar-valued function of θ could be estimated with smaller asymptotic **standard error** under the weighted observation. The natural observations are similarly favored when the difference is negative definite. However, definiteness of $J_w(\theta) - J(\theta)$ appears to be a rare occurrence (unless θ is a scalar), and one usually has to conclude that relative informativeness depends upon the question being asked.

When $J_w(\theta) - J(\theta)$ is indefinite, comparison of weighted and natural observations can be made in at least two ways:

1. In terms of a suitable scalar-valued measure of joint information for θ. Here the reciprocal of the generalized variance, det $[J(\theta)]$, naturally suggests itself.
2. In terms of the standard error of estimation for the scalar-valued function of θ that is considered most relevant to the scientific problem at hand.

The first possibility has been studied by Patil and Taillie [24] for the two-parameter gamma, **negative binomial**, and lognormal distributions and with known weight function $w(x) = x^{\alpha}$. In all these cases, $J_w(\theta) - J(\theta)$ is indefinite. For the gamma and negative binomial distributions, the weighted observations are jointly less informative for θ when generalized variance is the criterion. However, the relative efficiency is quite close to unity unless the shape parameter is small (less than 0.5, say). The lognormal distribution is an interesting case because det $[J_w(\theta)]$ = det $[J(\theta)]$ for all θ so that weighted and unweighted observations are equally informative in terms of the generalized variance.

The following species abundance example applies the second possibility to the truncated lognormal distribution with parameters μ and σ^2. The example is motivated by Preston's [30, 31] lognormal distribution of species abundance. Scientific interest resides primarily in the parameter σ^2 which is related to the evenness component of species diversity [24].

Species Abundance

Preston [30, 31] argues that in large ecological communities, the various species have their abundance X' distributed according to a lognormal curve with parameters μ' and σ^2. If an investigator samples a fraction t of the individuals in the community, then a species with natural abundance X' has recorded abundance $X = t \cdot X'$. (As Preston points out, the analysis is only approximate since it ignores sampling fluctuation. With Poisson sampling, for example, the Poisson–lognormal would be a more appropriate sampling distribution.) Now $X = tX'$ follows a lognormal distribution with parameters μ and σ^2, where $\mu = \mu' + \ln t$. The parameter of primary interest is $\varphi = 1/\sigma^2$ since it is a measure of species evenness.

Preston also notes that species with $X < 1$ will go unrecorded so that the observed data actually follow a truncated lognormal with truncation point located at $X = 1$ (or at $\ln X = 0$). Preston refers to the truncation point as the *veil line*: when the sampling intensity increases, the frequency curve moves to the right relative to the veil line and additional species become 'unveiled'.

Example 9 *Modeling bivariate overdispersion.* In developmental toxicity studies, one sometimes has more than two possible responses. For example, the responses might be *death, malformation*, and *normal* (*see* **Multiple outcomes**). Here, the simplest model would be the trinomial, but this would probably have to be generalized to account for overdispersion. Two such generalizations suggest themselves immediately: (a) the Dirichlet-trinomial (*see* **Dirichlet-multinomial distribution**) and (b) a multivariate DEF with $f_\mu(x)$ as the trinomial distribution and μ and x multivariate. For the double bivariate–normal distribution, in the univariate case, the double normal is again normal with an altered variance, making it rather uninteresting. Similarly, the double bivariate–normal distribution remains bivariate normal but with the interesting feature that the correlation is affected by the overdispersion. In particular, the recorded observations can exhibit correlation even when the natural observations are independent. In a certain overall sense, observations from the double bivariate–normal are more informative in spite of overdispersion.

Example 10 *Propagation functions for monitoring distributional changes.* Status and trend are the two important components in the assessment and monitoring of the environmentally important variables. The former is an indicator of the overall state of the variable of interest whereas trend describes the changes occurring in it over time (*see* **Trend, detecting**). These two indicators are typically distinguished in the environmental data by their spatial or temporal components. The first is a single snapshot which may arise for studying the spatial patterns, e.g. to assess the damage caused by the abrupt rupture in a hazardous waste storage facility; here the emphasis is on the rapid assessment of the damage spatially and the temporal sparseness is not important. The other is a multiple snapshot type required for **long-term environmental monitoring**. After a rapid assessment of damage, impact of rupture is monitored over a period of time in the vicinity of the site of incident. Monitoring is done by collecting the data at suitably selected, fixed monitoring stations. Typically these networks are spatially sparse, but generate multiple snapshots over a specified time period. A primary objective of the investigator in such monitoring studies is to describe changes in the frequencies over time in the values of the variable of interest.

Consider the situation where a set of monitoring stations is chosen by a random process from a pool of stations, and then an environmentally important 'indicator' X is measured at baseline time t_1 and at later times $t_2, t_3, \ldots$. The main interest is in the change(s) in the distribution of X through time for the population stations as reflected by the sampled stations. McDonald et al. [15] proposed the concept of a *selection function* to describe changes in the population. This idea has been used primarily in the context of biological populations and has applications in the study of natural selection [13], and **resource selection** [14]. In natural selection, the variable under study does not change. The changes, if any, are due to removal of some individuals from the population due to some natural reasons, e.g. predation. Formally, the selection function is defined as a function that describes how the individuals in the population must be selected in order to produce individuals in the second population.

In environmental monitoring programs, the units in the population remain constant, but the values of the variables measured on those units change. Hence, the term selection function may be misleading. Kaur et al. [11] use the term *propagation function* (PF) to emphasize that it is the **distribution function** of the variable X for the population of interest that is changing over time. Let X and Y be the values of an environmental variable on a population unit at two consecutive times. Let the marginal densities of X and Y be denoted by $f_X(\cdot)$ and $g_Y(\cdot)$. The *selection function* or PF is defined as

$$w(x) = \frac{g_Y(x)}{f_X(x)} \tag{22}$$

Often, the log of the PF

$$u(x) = \log(w(x)) \tag{23}$$

is more convenient than the PF itself. Formally, $g_Y(x) = w(x) f_X(x)$ is a weighted form of $f_X(x)$ with $w(x)$ as the weight function. Also, from the definition

$$\mathrm{E}_f[w] \equiv \mathrm{E}[w(X)] = 1 \tag{24}$$

McDonald et al. [15, 16] give the basic properties of the PF, including invariance to transformations of the environmental variable. Kaur et al. focus on the possible shapes of the PF as related to the marginal distributions of X and Y. Of particular interest are monotonicity and modality of the PF. When the PF is monotone increasing, for example, X is stochastically smaller than Y. A second focus is the parametric structure of the PF as related to the parametric structure of the joint distribution of (X, Y). Although the PF depends only on the marginal distributions, these marginal distributions may have some parameters in common. In this case, the PF is more efficiently estimated from the joint dataset than from the two marginal datasets. Accordingly, Kaur et al. examine the form of the PF for three families of bivariate distributions: bivariate normal, bivariate gamma, and bivariate logistic.

References

[1] Bayarri, M.J. & DeGroot, M.H. (1986). Information in selection models. Technical Report Number 365, Department of Statistics, Carnegie-Mellon University.

[2] Blackwell, D. (1951). Comparison of experiments, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 93–102.

[3] Blackwell, D. (1953). Equivalent comparisons of experiments, *Annals of Mathematical Statistics* **24**, 265–272.

[4] Cox, D.R. (1962). *Renewal Theory*, Barnes & Noble, New York.

[5] Dennis, B. & Patil, G.P. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abundance, *Mathematical Biosciences* **68**, 187–212.

[6] Diaconis, P. & Efron, B. (1985). Testing the independence of a two-way table: new interpretations of the chi-square statistic (with discussion and rejoinder), *The Annals of Statistics* **13**, 845–913.

[7] Drummer, T.D. & McDonald, L.L. (1987). Size-bias in line transect sampling, *Biometrics* **13**, 13–21.

[8] Efron, B. (1986). Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association* **81**, 709–721.

[9] Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *The Annals of Eugenics* **6**, 13–25.

[10] Iyengar, S. & Greenhouse, J.B. (1988). Selection models and the file drawer problem, *Statistical Science* **1**, 109–135.

[11] Kaur, A., di Consiglio, L., Patil, G.P. & Taillie, C. (1995). Propagation functions for monitoring distributional changes, *Environmental and Ecological Statistics* **2**, 239–269.

[12] Laird, N., Patil, G.P. & Taillie, C. (1988). Comment on S. Iyengar & J.B. Greenhouse, selection models and the file drawer problem, *Statistical Science* **3**, 126–128.

[13] Manly, B.F.J. (1985). *The Statistics of Natural Selection by Animal Populations*, Chapman & Hall, London.

[14] Manly, B.F.J., McDonald, L. & Thomas, D. (1993). *Resource Selection by Animals*, Chapman & Hall, London.

[15] McDonald, L., Gonzalez, L. & Manly, B.F.J. (1992). Statistical inference procedures for probability selection functions in long term monitoring programs. EMAP Final Report, USEPA, Corvallis, Oregon.

[16] McDonald, L., Gonzalez, L. & Manly, B.F.J. (1995). Using selection functions to describe changes in environmental variables, *Environmental and Ecological Statistics* **2**, 225–237.

[17] 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 & J. Roy, eds, Statistical Publishing Society, Calcutta, pp. 478–503.

[18] Patil, G.P. (1991). Encountered data, statistical ecology, environmental statistics, and weighted distribution methods, *Environmetrics* **2**, 377–423.

[19] Patil, G.P. (1996). Statistical ecology, environmental statistics, and risk assessment, in *Advances in Biometry*,

P. Armitage & H.A. David, eds, Wiley, New York, pp. 213–240.

[20] Patil, G.P. (1997). Weighted distributions, in *Encyclopedia of Biostatistics*, Vol. 6, P. Armitage & T. Colton, eds, Wiley, Chichester, pp. 4735–4738.

[21] Patil, G.P. & Ord, J.K. (1976). On size-biased sampling and related form-invariant weighted distributions, *Sankhyā* **38**, 48–61.

[22] Patil, G.P. & Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families, *Biometrics* **34**, 179–184.

[23] Patil, G.P. & Taillie, C. (1988). Weighted distributions and the effects of weight functions on Fisher information. Unpublished manuscript, Pennsylvania State University, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, University Park.

[24] Patil, G.P. & Taillie, C. (1989). Probing encountered data, meta analysis and weighted distribution methods, in *Statistical Data Analysis and Inference*, Y. Dodge, ed., Elsevier, Amsterdam, pp. 317–345.

[25] Patil, G.P. & Taillie, C. (1993). Environmental sampling, observational economy, and statistical inference with emphasis on ranked set sampling, encounter sampling, and composite sampling, in *Bulletin of the ISI, Proceedings of Forty-Ninth Session*, Florence, pp. 295–312.

[26] Patil, G.P., Rao, C.R. & Zelen, M. (1988). Weighted distributions, in *Encyclopedia of Statistical Sciences*, Vol. 9, S. Kotz & N.L. Johnson, eds, Wiley, New York, pp. 565–571.

[27] Patil, G.P., Taillie, C. & Talwalker, S. (1993). Encounter sampling and modelling in ecological and environmental studies using weighted distribution methods, in *Statistics for the Environment*, V. Barnett & K.F. Turkman, eds, Wiley, Chichester, pp. 45–69.

[28] Patil, G.P., Taillie, C. & Wigley, R.L. (1979). Transect sampling methods and their application to the deep-sea red crab, in *Environmental Biomonitoring, Assessment, Prediction, and Management – Certain Case Studies and Related Quantitative Issues*, J. Cairns, G.P. Patil & W.E. Waters, eds, International Cooperative Publishing House, Fairland, pp. 51–75.

[29] Patil, G.P., Taillie, C. & Wigley, R.L. (1980). Transect sampling methods and their applications to the deep-sea red crab, in *Advanced Concepts in Ocean Measurements for Marine Biology*, F.P. Diemer, F.J. Vernberg & D.Z. Mirkes, eds, University of South Carolina Press, Columbia, pp. 493–505.

[30] Preston, F.W. (1948). The commonness, and rarity, of species, *Ecology* **29**, 254–283.

[31] Preston, F.W. (1962). The canonical distribution of commonness and rarity, *Ecology* **43**, 186–215, 410–432.

[32] Ramsey, F.L. (1979). Parametric models for line transect surveys, *Biometrika* **66**, 505–512.

[33] Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment, in *Classical and Contagious Discrete Distributions*, G.P. Patil, ed., Pergamon Press and Statistical Publishing Society, Calcutta, pp. 320–332.

[34] Rao, C.R. (1985). Weighted distributions arising out of methods of ascertainment, in *A Celebration of Statistics*, A.C. Atkinson & S.E. Fienberg, eds, Springer-Verlag, New York, Chapter 24, pp. 543–569.

[35] Simon, R. (1980). Length biased sampling in etiologic studies, *American Journal of Epidemiology* **111**, 444–452.

[36] Stene, J. (1981). Probability distributions arising from the ascertainment and the analysis of data on human families and other groups, in *Statistical Distributions in Scientific Work, Applications in Physical, Social, and Life Sciences*, Vol. 6, C. Taillie, G.P. Patil & B. Baldessari, eds, Reidel, Boston, pp. 233–264.

[37] Taillie, C., Patil, G.P. & Hennemuth, R.C. (1995). Modelling and analysis or recruitment distributions, *Environmental and Ecological Statistics* **2**, 315–330.

[38] Talwalker, S., Patil, G.P. & Taillie, C. (1995). Qualitative and quantitative assessment of the risk from the exposure to fetotoxic chemical compounds, *Environmental and Ecological Statistics* **2**, 71–79.

[39] Temkin, N. (1976). Interactive information and distributional length biased survival models. Unpublished Ph.D. dissertation, University of New York at Buffalo.

[40] Wigley, R.L., Theroux, R.B. & Murray, H.E. (1975). Deep-sea red crab, *Marine Fisheries Review* **37**, 1–21.

[41] Zelen, M. (1974). Problems in cell kinetics and the early detection of disease, in *Reliability and Biometry*, F. Proschan & R.J. Serfling, eds, SIAM, Philadelphia, pp. 701–706.

(*See also* **Bivariate distributions**; **Sampling, environmental**)

G.P. PATIL