Statistics and Probability: 
Essays in Honor of C. R. Rao

Edited by

G. KALLIANPUR
P. R. KRISHNAIAH
J. K. GHOSH

1982

NORTH-HOLLAND PUBLISHING COMPANY—AMSTERDAM·NEW YORK·OXFORD
ON WEIGHTED DISTRIBUTIONS

M. MAHFOUD* and G. P. PATIL

Department of Statistics, The Pennsylvania State University, University Park, PA 16802, U.S.A.

1. Introduction and summary

The concept of weighted distributions can be traced to the study of "the effect of methods of ascertainment upon estimation of frequencies" by Fisher (1934). It was Rao (1965), however, who saw the need for a unifying concept and formulated it in his remarkable paper presented at the First International Symposium on Classical and Contagious Discrete Distributions held at McGill University in 1963. He identified various situations that can be modeled by what he called weighted distributions. These situations refer to instances where the recorded observations cannot be considered as a random sample from the original distribution. This may occur because of non-observability of some events or a damage caused to original observation resulting in a reduced value, or adoption of a sampling procedure which gives unequal chances to the units in the original. Rao's paper has stimulated considerable research on damage models, characterization of discrete distributions, and sampling mechanisms generating a wide variety of weighted distributions.

In all these cases, weighted distributions are justified because of the built-in probability sampling at some stage of data collection. But they arise from distinctly different recording mechanisms also; see for example, Patil and Rao (1976, 1978). Furthermore, the concept of weighted distributions can be extended to accommodate moment distributions known in economics (Hart, 1975) and mass-size distributions used in small particle physics and sedimentology (Herdan, 1960). For a comprehensive survey of examples of weighted distributions and how they arise in a wide range of scientific areas, see Patil and Rao (1976). Situations leading to discrete weighted distributions include the analysis of family data, the aerial survey involving visibility bias in wildlife ecology, and line transect sampling. Examples of continuous weighted distributions refer to cell kinetics, early disease detection, heart transplant statistics, etc.

In this paper, the properties of weighted distributions are studied in comparison with those of the original distributions. We examine how some parameters of the weighted distribution relate to those of the original distributions. As most often the interest is to make statements about the parameters or the form of the original distribution, several questions can be raised as to how the weight function affects the original distribution. Some relationships between the parameters of the two distribu-

*Present address: Institut National de la Statistique et de l'Economie Appliquee, Rabat, Morocco.
tions prove to be characteristics of specific pdf's such as log-normal, gamma and Poisson. These characterizations are examined and also the effect of size-biased sampling on the mixtures of specific distributions is investigated.

Further, we study bivariate weighted distributions with different w's. The w \( w(x, y) = x^y \) is considered. The effects of \( w(x, y) = x \) on the marginal and conditional variances of \( X \) are examined. The trinomial, the negative trinomial and the double Poisson distributions are characterized as the only bivariate SSPS D's for which \( \eta^2(X|Y) = \eta^2(X^*|Y^*) \) where \( w(x, y) = x \). Bivariate weighted distributions with \( w(x, y) = \max(x, y) \) are also studied.

2. Notation and terminology

In this section, we introduce the notation and terminology to be used in the paper. The statistical concepts are recorded as definitions. Table 1 provides a list of needed univariate distributions and related notation.

Consider a natural mechanism generating a random variable (rv) \( X \) with probability (density) function (pdf/pcf) \( f(x) \). Let \( w(x) \) be a nonnegative weight function, and assume that \( X \) is such that \( E[w(X)] \) exists. Denote a new pdf by:

\[
\tilde{f}(x) = \frac{w(x) f(x)}{E[w(X)]},
\]

and denote by \( X^* \) the rv whose pdf is \( f^*(x) \). \( X \) and \( X^* \) are referred to as original and weighted rv's and their respective distributions are called original and weighted distributions.

When sampling a mixture where the rv \( X \sim f(x) \), the recorded observation \( X^* \) turns out to be a weighted version of \( X \) with \( w(x) = x^\alpha \), we say that \( X^* \) is size-biased of order \( \alpha \) (Sb(\alpha)). Such a selection procedure is called size-biased sampling of order \( \alpha \) (Sb(\alpha)).

<table>
<thead>
<tr>
<th>( X )</th>
<th>pdf/pdf</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Poisson ((\theta))</strong></td>
<td>( e^{-\theta} \cdot \theta^x \cdot x! ), ( x = 1, 2, \ldots, \theta &gt; 0 )</td>
</tr>
<tr>
<td><strong>Binomial ((n, \theta))</strong></td>
<td>( \binom{n}{x} \cdot \theta^x \cdot (1-\theta)^{n-x} ), ( x = 0, 1, \ldots, n ), ( 0 &lt; \theta &lt; 1 )</td>
</tr>
<tr>
<td><strong>Negative-binomial ((k, \theta))</strong></td>
<td>( \frac{-\theta}{1-\theta} \cdot (1-\theta)^{x-1} \cdot \theta^k ), ( x = 0, 1, \ldots ), ( 0 &lt; \theta &lt; 1 )</td>
</tr>
<tr>
<td><strong>Geometric ((\theta))</strong></td>
<td>( (1-\theta)^{\theta} \cdot x ), ( x = 0, 1, \ldots ), ( 0 &lt; \theta &lt; 1 )</td>
</tr>
<tr>
<td><strong>Log-series ((\theta))</strong></td>
<td>( \theta^{-1} \cdot \log(1-\theta) \cdot x ), ( x = 1, 2, \ldots ), ( 0 &lt; \theta &lt; 1 )</td>
</tr>
<tr>
<td><strong>Uniform ((0, \theta))</strong></td>
<td>( \theta^{-1} \cdot \theta, 0 &lt; x &lt; \theta )</td>
</tr>
<tr>
<td><strong>Beta ((a, b))</strong></td>
<td>( \Gamma(a+b)/\Gamma(a)\Gamma(b) \cdot x^a \cdot (1-x)^{b-1} ), ( 0 &lt; x &lt; 1 ), ( a &gt; 0, b &gt; 0 )</td>
</tr>
<tr>
<td><strong>Gamma ((k, \theta))</strong></td>
<td>( \Gamma(k) \cdot \theta^{-k} \cdot x^{k-1} \cdot \theta^{-x} \cdot x ), ( x &gt; 0 ), ( \theta &gt; 0 ), ( k &gt; 0 )</td>
</tr>
<tr>
<td><strong>Pareto ((\theta_0, \lambda))</strong></td>
<td>( \lambda / x \cdot (\theta_0/x)^{\lambda} \cdot x \cdot \theta_0, \theta_0 &gt; 0, \lambda &gt; 0 )</td>
</tr>
<tr>
<td><strong>Log-normal ((\mu, \sigma))</strong></td>
<td>( \left( \frac{1}{\sqrt{2\pi}} \cdot \sigma \right) \cdot \exp\left(-1/2((\ln x - \mu)/\sigma)^2\right) ), ( x &gt; 0 ), ( \sigma &gt; 0, -\infty &lt; \mu &lt; \infty )</td>
</tr>
</tbody>
</table>
Because of the wide use of the wf \( w(x) = x \), its corresponding weighted rv \( X^* \) and weighted distribution \( f^*(x) \) are denoted by \( X^* \) and \( f^*(x) \). Hence we have
\[
f^*(x) = \frac{xf(x)}{E(X)}, \quad x \geq 0.
\]

Throughout this study we will make use of the following notation:
(a) \( x^{(r)} = x(x-1) \cdots (x-r+1) \).
(b) \( \mu_1 = E(X^*), \mu_2 = E(X_1^2) \), \( \mu = E(X), \mu^* = E(X^*) \).
(c) \( \mu_1 = E(X^* - \mu)^2, \mu_2 = V(X) \).
(d) The harmonic mean of \( X \): \( H(X) = \frac{1}{E(X^{-1})} \).
(e) Coefficient of skewness: \( \gamma_1 = \mu_3 / \mu_2^{3/2} \). If \( \gamma_1 \) is negative, we say that the distribution of \( X \) is negatively skewed.
(f) Coefficient of kurtosis: \( \gamma_2 = \mu_4 / \mu_2^2 \).
(g) The cumulative distribution function (cdf) of \( X \) is \( F(x) = \int_{-\infty}^{x} f(t) \, dt \). \( \bar{F}(x) = 1 - F(x) \).

**Definition 1.1.** Consider the rv's \( X \sim f_X(x) \), \( Y \sim f_Y(y) \), \( Z \sim g_Z(z) \) and a real number, \( p \), such that \( 0 < p < 1 \), then \( g_Z \) is a mixture of \( f_X \) and \( f_Y \) if
\[ g_Z(x) = pf_X(x) + (1-p)f_Y(x). \]

**Definition 1.2.** Consider the rv's \( X \sim f_X(x) \), \( Y \sim f_Y(y) \), \( Z \sim g_Z(z) \) and a real number \( b > 1 \), then if
\[ g_Z(x) = bf_X(x) + (b-1)f_Y(x), \]
we say that \( g_Z \) is a negative mixture of \( f_X \) with \( f_Y \).

Observe that \( g_Z \) is a negative mixture of \( f_X \) with \( f_Y \) whenever \( f_X \) is a mixture of \( f_Y \) and \( g_Z \).

3. Univariate weighted distributions

3.1. Properties of the weighted distributions

Some properties of weighted distributions, which will be used later, are studied.

Since these properties are straightforward to establish, their proofs are omitted. Let a non-negative rv \( X \sim f(x) \), for which the first three moments exist and subject to sbs. So the rv \( X^* \sim f^*(x) \) and some of its properties are given below.

**Property 1.** \( \mu^* - \mu = \mu_2 / \mu \).

**Property 2.** \( \mu_1^* - \mu_2 = (\mu_2 / \mu)(\gamma_1 / c - 1) \).
Remark. While the mean of $X^*$ is always greater than the mean of $X$, the variance of $X^*$ is greater than the variance of $X$ if $\gamma_1$ exceeds $c$. Also for all negatively skewed original distributions, $\mu_{X}^*$ is less than $\mu_{X^*}$.

Property 3. $(\mu_{X^*}^-)^{-1} = \mu$.

The harmonic mean of $X^*$ is equal to the mean of $X$.

Property 4. $\mu_{X} = (\mu_{X^*}^-)^{-1}[\mu_{X^*} - (\mu_{X^*}^-)^{-1}].$

From Properties 1 and 4, the variance to mean ratio of $X$ is given by the difference between the arithmetic mean and the harmonic mean of $X^*$.

3.2. Effects of the weight function on the original distribution

The effects of the w.f on the original distribution are discussed for commonly used forms of w(-). We also study the effect of sbs on the hazard rate function.

Definition. Let the rv $X \sim f(x; \theta)$, then we say that $f(x; \theta)$ is form invariant under the w.f $w(x) = x^\alpha$, if $X^\alpha \sim f(x; \gamma)$, where $\gamma$ is the new parameter depending on $\theta$ and $\alpha$.

First, consider a rv $X \sim \Gamma(k, \theta)$. The size-biased rv $X^*$ is also distributed as $\Gamma(k + 1, \theta)$. However, if the rv $X \sim \text{LSD}(\theta)$, then $X^* \sim \text{Geometric}(\theta)$. In this case the rv $X^*$ is not distributed as another log-series distribution. In connection with this result, Rao (1965) noticed that the geometric distribution is sometimes found to provide a good fit to an observed distribution of family size, but this may be due to a sampling procedure giving large probability of selection to families with large sizes and, it may well be that the original distribution of the family size is logarithmic. In the light of these two examples, the question arises as to under which conditions the original distribution is form invariant under sbs($\alpha$). Under some regularity conditions, Patil and Ord (1975) prove the following result.

Theorem 3.1. Let the rv $X \sim f(x; \theta)$ such that $\mu^*_\alpha$ exists and the w.f be $w(x) = x^\alpha$. If $X$ satisfies the regularity conditions of continuity and expectation relative to $\alpha$, then $X$ subject to the w.f $w(x) = x^\alpha$ is form invariant if and only if its pdf is given by

$$f(x; \theta) = a(x)x^\theta/m(\theta) = \exp[\theta \ln x + A(x) - B(\theta)]$$

where $a(x) = \exp[A(x)]$ and $m(\theta) = \exp[B(\theta)]$.

Table 2 gives some well-known distributions which belong to the log-exponential family of distributions. It also defines the parameter $\theta$ and the range for $\alpha$.

Sampling components which are operating in a system at a fixed time and then observing the life lengths of these components leads to a size-biased sample (Blumenhal, 1967; Cox, 1969; Schaeffer, 1972). As in life length studies the concept
Table 2

<table>
<thead>
<tr>
<th>Distribution</th>
<th>pdf</th>
<th>$\theta$</th>
<th>Range for $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-normal</td>
<td>$(2\pi)^{-1/2}\exp[-(1/2\alpha^2)(\ln x-\mu)^2-\ln x]$</td>
<td>$\alpha&gt;0$</td>
<td>$\alpha&gt;0$</td>
</tr>
<tr>
<td>Pareto</td>
<td>$(\alpha/x^\alpha)\exp(-\alpha x)$, $x&gt;\alpha, \alpha&gt;0, \lambda&gt;0$</td>
<td>$-\lambda$</td>
<td>$\alpha&lt;\lambda$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$x^{k-1}e^{-x}/\Gamma(k)$, $x&gt;0, k&gt;0$</td>
<td>$k-1$</td>
<td>$\alpha&gt;k$</td>
</tr>
<tr>
<td>Beta int</td>
<td>$x^{\alpha-1}(1-x)^{\beta-1}/\beta(a,b)$, $0&lt;x&lt;1$</td>
<td>$\alpha-1$</td>
<td>$\alpha&gt;a$</td>
</tr>
<tr>
<td>kind</td>
<td>$x^{\alpha-1}(1-x)^{\beta-1}/\beta(a,b)$, $x&gt;0$</td>
<td>$\alpha-1$</td>
<td>$\alpha&lt;k-a$</td>
</tr>
<tr>
<td>Beta 2nd</td>
<td>$x^{\alpha-1}(1-x)^{\beta-1}/\beta(a,b)$, $x&gt;0$</td>
<td>$\alpha-1$</td>
<td>$\alpha&lt;k-a$</td>
</tr>
<tr>
<td>kind</td>
<td>$x^{\alpha-1}(1-x)^{\beta-1}/\beta(a,b)$, $x&gt;0$</td>
<td>$\alpha-1$</td>
<td>$\alpha&lt;k-a$</td>
</tr>
<tr>
<td>Pearson</td>
<td>$x^{-k-1}\exp(-x^{-1})/\Gamma(k)$</td>
<td>$k-1$</td>
<td>$\alpha&lt;k$</td>
</tr>
<tr>
<td>type V</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The hazard rate function, $h(x)$, is an alternative way to describe the life distribution of the components, we propose to examine the effect of sex on $h(x)$.

**Theorem 3.2.** Consider a non-negative real-valued $X$ subject to six and such that $E(X)$ exists. If $h(x)$ is the hazard rate function of $X$, then

$$h^*(x) < h(x), \quad x > 0.$$ 

**Proof.** By definition we have

$$h(x) = \frac{f(x)}{\int_{x}^{\infty} f(t) \, dt},$$

$$h^*(x) = x f(x) / \int_{x}^{\infty} t f(t) \, dt.$$ 

Consider the ratio $h^*(x)/h(x)$ and note that

$$\left[ \int_{x}^{\infty} t f(t) \, dt / \int_{x}^{\infty} f(t) \, dt \right] > x.$$ 

**Remark.** On intuitive grounds, one can expect such a result since the sampling mechanism tends to select a unduly low proportion of components which become defective at early age.

3.3. **Characterizations based on weighted distributions**

Some relationships between parameters of the original distribution and those of the weighted distribution turn out to be characteristic properties of specific distributions. Thus we characterize log-normal, gamma and Poisson distributions.

The characterization of the log-normal distribution came about by connecting an empirical remark made by Krumbine and Pettijohn (1938) with the fact that this distribution provides a good fit to the observed particle sizes. They noticed from experimental data that when plotting, separately, the logarithm of the particle sizes, $X$, versus their frequencies, and versus their total weights, $X^w$, the variability in in $X$ and in $X^w$ appears to be the same from the two histograms.
Theorem 3.3. Consider a non-negative real-valued random variable \( X \sim f(x) \) and such that \( E(X^*) = \mu \) exists for \( \mu > 0 \). Let \( X^* \) be the weighted form of \( X \) with \( w(x) = x^\alpha \). If \( Y = \ln X \) and \( Z = \ln X^* \), then \( X \) is log-normally distributed if and only if \( V(Y) = V(Z) \) for all \( \alpha > 0 \).

Proof. If \( X \sim LN(\mu, \sigma^2) \), then \( X^* \sim LN(\mu + a\sigma^2, \sigma^2) \) and therefore \( V(Y) = V(Z) \).

Conversely, if \( X \sim f(x) \) and \( \mu \) exists for \( \mu > 0 \), then with \( Y = \ln X \) and \( Z = \ln X^* \) we have

\[
Y \sim g(y) = e^y f(e^y)
\]

and

\[
Z \sim g'(z) = e^z g(z) / M_f(a), \tag{3.1}
\]

where \( M_f(a) \) is the mgf of \( Y \). From (3.1) it can be shown (Patil and Shorrock, 1963) that

\[
V(Z) = \frac{d}{da} \left[ \frac{M_f'(a)}{M_f(a)} \right]
\]

where \( M_f'(a) = (d/da) M_f(a) \). Now let \( V(Y) = \sigma^2 \) and assume that \( V(Z) = \sigma^2 \), that is

\[
\frac{d}{da} \left[ \frac{M_f'(a)}{M_f(a)} \right] = \sigma^2. \tag{3.2}
\]

With \( M_f(0) = 1 \) and \( M_f'(0) = E(Y) = \mu \), the solution for (3.2) is given by

\[
M_f(a) = e^{a + a^2 / 2},
\]

which implies that \( X \sim LN(\mu, \sigma^2) \).

The next theorem gives a characterization of the Poisson distribution based on the equality of the variances of the rv's \( X \) and \( X^* \).

Theorem 3.4. Consider a non-negative real-valued random variable \( X \sim f(x; \theta) \) where

\[
f(x; \theta) = a(x) e^{x/\theta} / M(\theta), \tag{3.3}
\]

and subject to \( \theta > 0 \), then \( X \) has Poisson distribution if and only if \( V(X) = V(X^*) \).

Theorem 3.5. Consider a non-negative real-valued random variable \( X \) with pdf

\[
f(x; \theta) = \frac{a(x) e^{x/\theta}}{M(\theta)} \tag{3.4}
\]

then the rv \( X \) has gamma distribution if and only if the coefficient of variation of \( X \) does not depend on \( \theta \).

Proof. If \( X \sim Gamma(k, \theta) \) with \( k \) known, then the coefficient of variation of \( X \) is \( c = k^{-1/2} \), which does not depend on \( \theta \). Conversely, suppose that the pdf of \( X \) is
given by (3.4) and that
\[ \frac{\mu_2}{\mu^2} = a, \quad a > 0 \] (3.5)
and does not depend on \( \theta \). (3.5) can be rewritten as
\[ \frac{M''(\theta)}{M'(\theta)} = (1+a) \frac{M'(\theta)}{M(\theta)} \]
implying that
\[ M'(\theta) = b [M(\theta)]^{(r+1)}, \quad b > 0. \] (3.6)
The solution of (3.6) is given by
\[ \frac{1}{-a[M(\theta)]^a} = b\theta + d. \]
With \( M(0) = 1 \) we have \( d = -a^{-1} \) and therefore \( M(\theta) \) can be written as
\[ M(\theta) = (1 - ab\theta)^{-1/a} \]
which implies that \( X \) has a gamma distribution.

Warren (1974) observed that, under shs, the relationship between the coefficient of skewness, \( \gamma_1 \), and the coefficient of variation, \( c \), is the same for the rv's \( X \) and \( X^* \), when the rv \( X \) has a gamma distribution. Also if the original distribution is log-normal, \( \gamma_1, c \) and their relationship are invariant under shs. In the following theorems we generalize the above results to the case of shs(\( \alpha \)) and including its implications on the coefficients of kurtosis, \( \gamma_2 \).

**Theorem 3.6.** Consider a rv \( X \) subject to shs(\( \alpha \)). If \( X \) has a gamma distribution, then the following relationships
\[ \gamma_1 = 2c, \]
\[ \gamma_2 = 3(1+2c^2) \] (3.7) (3.8)
hold for \( X \) and their form is invariant for \( X^* \).

**Proof.** If \( X \sim \text{Gamma}(k, \theta) \), then \( X^* \sim \text{Gamma}(k+\alpha, \theta) \). Without loss of generality, the parameter \( \theta \) can be taken to be one. Using the fact that \( E(X^*) = \Gamma(k+\alpha)/\Gamma(k) \), the relationships (3.7) and (3.8) hold for \( X \). Also for \( X^* \) we have
\[ \gamma_1^* = 2c^* \]
and
\[ \gamma_2^* = 3\left[1 + 2(c^*)^2\right]. \]

**Theorem 3.7.** Consider a rv \( X \) subject to shs(\( \alpha \)). If \( X \) has a log-normal distribution, then \( c^* = c \), \( \gamma_1^* = \gamma_1 \) and \( \gamma_2^* = \gamma_2 \).
Proof. By observing that if \( X \sim \text{LN}(\mu, \alpha^2) \), \( X^* \sim \text{LN}(\mu + \alpha \sigma^2, \sigma^2) \) and that \( E(X^*) = \exp(\mu + \frac{1}{2} \alpha^2 \sigma^2) \), the theorem is readily proved.

As a consequence of the above result, for the log-normal distribution the relationship \( y = c(e^x + 3) \) hold for the rv \( X \) as well as the rv \( X^* \) with \( w(x) = x^a \).

4. Multivariate weighted distributions

4.1. Introduction

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

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

The pair of rv's whose joint pdf is \( f^*(x, y) \) is denoted by \((X, Y)^*\).

The extension to s-variate weighted distributions is straightforward.

4.2. Bivariate weighted distributions with \( w(x, y) = x^a \)

Let \((X, Y)\) be a pair of non-negative rv's with joint pdf \( f(x, y) \) such that \( E(X^*) \) exists. With \( w(x, y) = x^a \), the joint pdf of \((X, Y)^*\) is

\[
 f_{*}^*(x, y) = \frac{x^a f(x, y)}{E(X^*)}.
\]  \( \quad (4.1) \)

From (4.1) we obtain the following marginal and conditional distributions

\[
 f_{*}^*(x) = \frac{x^a f_x(x)}{E(X^*)} \quad \text{(4.2)}
\]

\[
 f_{*}^*(y) = \frac{E(X^* | y)f_y(y)}{E(E(X^* | y))} \quad \text{(4.3)}
\]

\[
 f_{*}^*(x | y) = \frac{x^a f_x(x | y)}{E(X^* | y)} \quad \text{(4.4)}
\]

\[
 f_{*}^*(y | x) = f_y(y | x) \quad \text{(4.5)}
\]

When the \( w(x, y) = x^a \), the marginal distributions of \( X^* \) and \( Y^* \) are given by (4.2) and (4.3) with \( \alpha = 1 \). So the marginal distribution of \( Y^* \) is the weighted marginal distribution of \( Y \) with \( w \) being the regression of \( X \) on \( Y \), i.e. \( w(y) = E(Y | y) \). Furthermore, if the regression of \( X \) on \( Y \) is linear, i.e \( E(X | y) = ay + b \), then the
marginal distribution of \( Y^* \) can be written as

\[
f^{*}(y) = a \frac{E(Y)}{E(X)} f^{*}(y) + \left[ 1 - a \frac{E(Y)}{E(X)} \right] f_r(y).
\]

We distinguish three cases: (i) \( aE(Y)/E(X)<0, f^{*} \) is a negative mixture of \( f_r \) with \( f^{*} \) (Patil et al. 1975); (ii) \( 0<aE(Y)/E(X)<1, f^{*} \) is a mixture of \( f_r \) and \( f^{*} \), and (iii) \( aE(Y)/E(X)>1, f^{*} \) is a negative mixture of \( f_r \) with \( f^{*} \).

For widely used bivariate distributions, the regression of \( X \) on \( Y \) is linear. Some of these bivariate distributions are multinomial, negative multinomial, hypergeometric, logarithmic, Poisson, Dirichlet, gamma, Pareto and F-distribution (Murdia, 1970).

The effect of the wf \( w(x, y)=x \) on the regression of \( X \) on \( Y \) is given by

\[
E(X|Y^*=y) - E(X|Y=y) = V(X|Y=y)/E(X|Y=y).
\]

In the following theorem we show how the correlation ratio and the regression of \( X \) on \( Y \) can be obtained from the bivariate weighted distribution.

**Theorem 4.1.** Let \( (X, Y) \) be a pair of non-negative r.v's with joint pdf \( f(x, y) \) such that \( E(X) \) and \( E(X|Y) \) exist for all \( y>0 \). Consider the wf \( w(x, y)=x \), then we have

\[
E(X|Y=y) = H(X^*|Y^*=y) \tag{4.6}
\]

and

\[
\eta^2(X|Y) = \frac{E[H(X^*|Y^*)] - H(X^*)}{E(X^*) - H(X^*)}. \tag{4.7}
\]

**Proof.** (4.6) is trivial. To establish (4.7) we have

\[
V(X) = H(X^*)[E(X^*) - H(X^*)],
\]

and

\[
H(X^*|Y^*=y) = E(X|Y=y),
\]

therefore,

\[
V(E(X|Y)) = H(E(X|Y))\{E[E(X|Y)] - H[E(X|Y)]\}
= H[H(X^*|Y^*)\{E[H(X^*|Y^*)] - H[H(X^*|Y^*)]\}
= H(X^*)\{E[H(X^*|Y^*)] - H(X^*)\}.
\]

Hence

\[
\eta^2(X|Y) = \frac{V(E(X|Y))}{V(X)} = \frac{E[H(X^*|Y^*)] - H(X^*)}{E(X^*) - H(X^*)}.
\]

Table 3 gives the effect of the wf \( w(x, y)=x \) on the correlation ratio of \( X \) on \( Y \). The bivariate distributions considered are multinomial, negative multinomial, logarithmic and Dirichlet.
Table 3
Comparisons between $\pi^*(X|Y)$ and $\pi^*(X^*|Y^*)$ when $w(x, y) = x$.

| Distribution | $\pi^*(X|Y)$ | $\pi^*(X^*|Y^*)$ | $\pi^*(X|Y)/\pi^*(X^*|Y^*)$ |
|--------------|-------------|-----------------|-------------------|
| $BM(\theta_1, \theta_2)$ | $\theta_1^2/(1-\theta_1)(1-\theta_2)$ | $\theta_2^2/(1-\theta_1)(1-\theta_2)$ | 1 |
| $BM(k_1, \theta_2)$ | $\theta_2^2/(1-\theta_1)(1-\theta_2)$ | $\theta_2^2/(1-\theta_1)(1-\theta_2)$ | 1 |
| $BLSD(\gamma_1, \theta_2)$ | $\gamma_1^2/(1-\theta_1)(1-\theta_2)$ | $\gamma_1^2/(1-\theta_1)(1-\theta_2)$ | $\pi^*(X|Y)/\pi^*(X^*|Y^*)$ |
| Dirichlet | $bn/(m+n)(l+n)$ | $(l+1)(l+1+n)(m+n)$ | $(l+1)(l+n)$ |

4.3. Bivariate weighted SSPSD's with $w(x, y) = x$

Patil (1968) has shown that some probability models arising from sampling with replacement from population with multiple characteristics can be expressed in the form of what he has called "Sum-Symmetric Power Series Distributions" (SSPSD's).

**Definition.** A pair of non-negative integer valued rv's $(X, Y)$ with joint pf:

$$p(x, y) = a(x, y) x^y \theta_1^x \theta_2^y / f(\theta_1, \theta_2),$$

where $\theta_1, \theta_2 > 0$, and

$$f(\theta_1, \theta_2) = \sum_{x, y} a(x, y) x^y \theta_1^x \theta_2^y,$$

is said to have a bivariate SSPSD($\theta_1, \theta_2, f(\theta_1, \theta_2)$).

$$\pi^*(X) = (\theta_1 \gamma_{12} - \theta_1 + \theta_2 \gamma_{12}) \theta_1 / (\theta_1 \gamma_{12})^2,$$

$$\pi^*[E(X|Y)] = \frac{\theta_1}{(1-\theta_1) \gamma_{12}} \frac{\theta_1}{\theta_2 \gamma_{12}} \frac{\theta_1}{\gamma_{12}} + \frac{\theta_2}{(1-\theta_1) \gamma_{12}} \frac{\theta_2}{\gamma_{12}} + \frac{\theta_1 \gamma_{12}}{(1-\theta_1) \gamma_{12}} \frac{\theta_2}{\gamma_{12}}$$

where

$$\theta_i = 1-\theta_1-\theta_2, \quad \gamma_i = -\ln(1-\theta_i), \quad i = 1, 2,$$

$$\gamma_{12} = -\ln(1-\theta_1-\theta_2).$$

The following theorem characterizes double Poisson, bivariate multinomial, and bivariate negative binomial.

**Theorem 4.2.** Let $(X, Y) \sim SSPSD(\theta_1, \theta_2, f(\theta_1, \theta_2))$ and let the $w(x, y) = x$, then $\pi^*(X|S) = \pi^*(X^*|S^*)$ if and only if $(X, Y)$ is distributed as double Poisson, bivariate multinomial, or bivariate negative multinomial.
4.4. Bivariate weighted distribution with \( w(x, y) = \max(x, y) \)

A situation where the \( w(x, y) = \max(x, y) \) arises is when the recorded observation is obtained from a renewal process sampled at a given time \( t \). Let \( (X, Y) \sim f(x, y) \) be the lifetimes of two components in parallel and forming a kit. The lifetime of a kit is \( Z = \max(X, Y) \). Consider a renewal system of kits. If at time \( t \), one records the lifetime, \( Z^* \), of the working kit, the pdf of \( Z^* \) is the weighted pdf of \( Z \) with \( w(z) = \max(x, y) \) (Cox, 1962). If the lifetimes of both components in the sampled kit are recorded, their joint pdf is the weighted form of \( f(x, y) \) with \( w(x, y) = \max(x, y) \).

In general, let \( (X, Y) \) be a pair of non-negative rv's with joint pdf \( f(x, y) \) such that \( E(\max(X, Y)) \) exists, and let \( (X, Y)^* \) be distributed as \( f^*(x, y) \), where

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

(4.8)

The marginal pdf of \( Y^* \) is proportional to

\[
 \int_0^\infty \max(x, y) f(x, y) dx = \int_0^\infty y f^*(x, y) dx + \int_y^\infty x f(x, y) dx
\]

\[
= F_{Y^*}(y) \cdot y f_Y(y) + \bar{F}_{X^*}(y) \cdot \bar{F}_{Y^*}(y) \cdot f_Y(y)
\]

where \( F_{Y^*}(y) = \int_0^y f(x, y) dx \), \( F_{X^*}(y) = 1 - F_{X^*}(y) \) and \( \bar{F}_{X^*}(y) = \int_0^y f(x, y) dx \). So the marginal pdf of \( Y^* \) is

\[
f_{Y^*}(y) = \frac{yF_{X^*}(y) + \bar{F}_{X^*}(y) F_{Y^*}(y)}{E[YF_{X^*}(Y) + \bar{F}_{X^*}(Y) F_{Y^*}(Y)]} f_Y(y).
\]  

(4.9)

Assuming that \( X \) and \( Y \) are independent, the following theorems examine the effects of the \( w(x, y) = \max(x, y) \) on the independence of \( X^* \) and \( Y^* \) and the regression of \( Y^* \) on \( X^* \).

**Theorem 4.3.** Let \( (X, Y) \) be a pair of non-negative independent rv's with joint pdf \( f(x, y) = f_X(x) f_Y(y) \); with \( w(x, y) = \max(x, y) \), the rv's \( (X, Y)^* \) with joint pdf given by (4.8) are dependent.

**Proof.** Using the independence of \( X \) and \( Y \), the conditional pdf of \( Y^* \mid X^* = x \) can be written as

\[
f_{Y^* \mid X^*}(y \mid x) = \begin{cases} u(x, y) y f_Y^*(y), & x \geq y, \\ u(x, y) y f_Y^*(y), & x < y \end{cases}
\]

(4.10)

where

\[
u(x, y) = E[\max(X, Y)] \left[ y f_X(y) + \bar{F}_X(y) F_Y(y) \right] \left[ x F_Y(x) + \bar{F}_Y(x) F_Y(x) \right],
\]

hence \( X^* \) and \( Y^* \) are dependent.
Lemma. Let \((X, Y)\) be a pair of non-negative independent rv's with joint pdf \(f(x, y) = f_x(x)f_y(y)\). With the wff \(w(x, y) = \max(x, y)\), the regression of \(Y^w\) on \(X^w\) is given by

\[
E(Y^w | X^w = x) = \frac{xf_x(x)F_y(x) + yf_y(y)\bar{F}_y(x)}{xF_x(x) + F_y(x)\bar{F}_y(x)}
\]

(4.11)

where \(\mu_y = \int_0^y yf_y(y)\,dy/F_y(x)\) and \(\bar{\mu}_y = \int_y^\infty yf_y(y)\,dy/F_y(x)\).

Proof. Using (4.10), we have

\[
E[Y^w | X^w = x] = \int_0^x u(x, y)xf_x(x)f_y(y)\,dy + \int_x^\infty u(x, y)yf_y(y)\,dy
\]

\[
= \int_0^x \frac{x\,yf_x(x)\,dy}{xF_x(x) + \bar{\mu}_y F(x)} + \int_x^\infty \frac{yf_y(y)\,dy}{xF_x(x) + \bar{\mu}_y F(x)}
\]

\[
= \frac{x\mu_y(x) - F_y(x) + \bar{\mu}_y \bar{F}_y(x)}{xF_x(x) + \bar{\mu}_y F(x)}
\]

Four properties of the regression of \(Y^w\) on \(X^w\) are given as Theorem 4.4.

Theorem 4.4. 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 such that \(E[\max(X, Y)]\) exists, then, with the wff \(w(x, y) = \max(x, y)\), the regression of \(Y^w\) on \(X^w\) is such that

(i) \(E(Y^w | X^w = x)\) is a decreasing function of \(x\),
(ii) \(E(Y^w | X^w = 0) = E(Y) + \frac{\mu_y}{E(X)}\),
(iii) \(E(Y^w | X^w = x) > E(Y)\),
(iv) \(\lim E(Y^w | X^w = x) = E(Y)\) as \(x \to \infty\).

Proof. Part (i) follows by differentiating (4.11) with respect to \(x\). We then, obtain the derivative as

\[
\int_0^x yf_y(y)\,dy \int_x^\infty yf_y(y)\,dy - \int_x^\infty y^2f_y(y)\,dy
\]

\[
\left[xf_x(x) + \bar{F}_y(x)\bar{\mu}_y\right]^2
\]

which is negative since

\[
\int_0^x yf_y(y)\,dy/F_y(x) < \int_0^\infty yf_y(y)\,dy/F_y(x).
\]

Parts (ii), (iii) and (iv) are straightforward.

As an example, let

\( (X, Y) \sim f(x, y) = \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y}, \quad x, y > 0, \)
then, with $w(x,y) = \max(x,y)$, we have

$$f_{X,Y}(x,y) = \begin{cases} 
\lambda_1 e^{-\lambda_1 y} \lambda_2 e^{-\lambda_2 x} / \max(X,Y), & x \geq y, \\
\lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y} / \max(X,Y), & x < y,
\end{cases} \quad (4.12)$$

where $E[\max(X,Y)] = \lambda_1^{-1} + \lambda_2^{-1} + \lambda^{-1}$, $\lambda = \lambda_1 + \lambda_2$. The marginal pdf of $X^w$ is

$$f^w_x(x) = \frac{\lambda_1^{-1}}{\lambda_1 (\lambda_2 \lambda_1)^{-1} + \lambda_1^{-1}} \lambda_1^2 e^{-\lambda_1 x} + \frac{\lambda_1 (\lambda_1 \lambda_2)^{-1}}{\lambda_2 (\lambda_2 \lambda_1)^{-1} + \lambda_1^{-1}} \lambda_2 e^{-\lambda_2 x}$$

which is a mixture of $\text{Gamma}(2, \lambda_1^{-1})$ and $\text{Gamma}(1, \lambda^{-1})$. Using part (ii) of Theorem 4.4, we have

$$E(Y^w | X^w = 0) - E(Y) = E(Y).$$

The moment generating function of $(X, Y)^w$ is

$$M_{X,Y}(s,t) = \frac{1}{\lambda_1^2 + \lambda_2^2} \left( \frac{\lambda_1^2}{(1-\lambda_1 s)^2 [1-\lambda_1^{-1}(x+t)]} + \frac{\lambda_1^2}{(1-\lambda_2 s)[1-\lambda_2^{-1}(y+t)]} \right)$$

which can be written as

$$M_{X,Y}(s,t) = \frac{1}{\lambda_1^2 + \lambda_2^2} \left[ \frac{\lambda_1^2}{1+2 \frac{s}{\lambda_1} + \left( \frac{s}{\lambda_1} \right)^2 + \cdots} \right] \frac{\lambda_1^2}{1+2 \frac{r}{\lambda_1} + \left( \frac{r}{\lambda_1} \right)^2 + \cdots}$$

$$+ \lambda_2^2 \left( \frac{1}{1+2 \frac{t}{\lambda_2} + \left( \frac{t}{\lambda_2} \right)^2 + \cdots} \right) \frac{1+2 \frac{t}{\lambda_2} + \left( \frac{t}{\lambda_2} \right)^2 + \cdots}{1+2 \frac{t}{\lambda_2} + \left( \frac{t}{\lambda_2} \right)^2 + \cdots}.$$

Hence,

$$\text{Cov}(X^w, Y^w) = -\lambda_1 \lambda_2 \left( \lambda_1^2 + \lambda_2^2 \right) / \lambda_1 \lambda_2 \left( \lambda_1^2 + \lambda_2^2 \right),$$

which implies that $X^w$ and $Y^w$ are negatively correlated.

References


