35th SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE

# SOME REMARKS ON MIXTURES OF DISTRIBUTIONS

*by*

W. MOLENAAR

*Stichting Mathematisch Centrum, Amsterdam*

Beograd, 1965

## 1. Summary


The connection between one-parameter mixtures and generalizations (i.e. $\underline{y}$-fold [1] convolutions, with a random $\underline{y}$) given by GURLAND (1957), is used to derive some simple properties of both, partly found with different proofs in TEICHER (1960). The possible types of a mixture of continuous singular components are discussed. The general one-parameter mixture is compared with a representative component, and the generalization with a representative fixed convolution. Examples of mixtures and generalizations are listed in Appendix 1. Appendix 2 is a specification of the distributions used in the paper.


## 2. Definitions

A random variable has a compound Poisson distribution if it has a Poisson distribution with a parameter that is not a positive constant but a random variable assuming positive real values. A generalized Poisson distribution is the $\underline{n}$-fold convolution of an arbitrary distribution, where $\underline{n}$ has a Poisson distribution. These definitions introduced by FELLER (1943) have been extended by GURLAND (1957) and TEICHER (1960) to the Definitions 1 and 2 given below. As some authors, e.g. FELLER (1957), use "compound" for what is here called "generalized", we shall henceforth replace "compound" by the less ambiguous "mixed".

Definition 1. If $F_\theta(\circ)$ is a distribution function for each parameter value $\theta \in T \subset R^1$, such that $F_\theta(x)$ is Borel measurable on $T \times R^1$, and H is a distribution function which assigns probability 1 to T, then the H-mixture of $F_\theta$ is the distribution function $F_\theta \underset{\theta}{\bigwedge} H$ given by

$$(1) \qquad (F_\theta \underset{\theta}{\bigwedge} H)(x) = \int F_\theta(x)\, dH(\theta).$$

We shall denote by $\underline{x}_\theta$ the random variable with distribution function $F_\theta$

---

1) Random variables are underlined.

and by $\underline{x}_\theta$ the random variable with distribution function $F_\theta \underset{\theta}{\wedge} H$ [2].

This is a special case of TEICHER's definition of m-parameter mixtures. In this report the distribution function H is always one-dimensional, though $F_\theta$ may have more than one parameter. The extra $\theta$ under the sign "$\wedge$" is convenient in this case. Sometimes we shall write $F_{c\theta} \underset{\theta}{\wedge} H$, though we could have included the constant c in the distribution function H. The symbol $\underset{\theta}{\wedge}$ is also used between names of distributions. Several well-known mixtures are listed in Appendix 1. Two examples of mixtures are

$$\text{(2)} \qquad \begin{aligned} \text{Binomial } (n,p) \underset{n}{\wedge} \text{Poisson } (\mu) &= \text{Poisson } (\mu p); \\ \text{Poisson } (k\mu) \underset{k}{\wedge} \text{Poisson } (\lambda) &= \text{Neyman } A(\lambda,\mu). \end{aligned}$$

We shall call a mixture <u>non-trivial</u> if neither H nor all $F_\theta$ are degenerate distribution functions. For degenerate H the mixture is just one $F_\theta$; any distribution H could be written as a mixture by taking $F_\theta(x) = \varepsilon(x - \theta)$ [3].

It is obvious that for example the characteristic function and the moments about zero (if existing) are <u>mix-linear</u> with respect to (w.r.t.) $F_\theta \underset{\theta}{\wedge} H$, i.e. if $\phi_\theta$ is the characteristic function of $F_\theta$ and $\phi$ of $F_\theta \underset{\theta}{\wedge} H$, then

$$\text{(3)} \qquad \begin{aligned} \phi(t) &= \int \phi_\theta(t)\, dH(\theta); \\ \mathcal{E}\underline{x}_\theta^k &= \int \mathcal{E}\underline{x}_\theta^k\, dH(\theta). \end{aligned}$$

<u>Definition 2</u>. If the random variable $\underline{x}$ has distribution function F and the non-negative integer-valued random variable $\underline{y}$ has distribution function G, the <u>G-generalized F-distribution</u>, with distribution function denoted by $F^{G*}$, is the distribution of

---

2) GURLAND uses $F \wedge H$ and $\underline{x} \wedge \theta$. We have preferred to use another notation, more suggestive of the underlying idea and avoiding the possibly confusing suggestion of symmetry.

3) $\varepsilon$ denotes the unit stepfunction.

(4) $$\underline{x}^{\underline{y}*} \stackrel{\text{def}}{=} \underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_{\underline{y}},$$

where the $\underline{x}_i$ are independent and have common distribution function F. If the characteristic function $\phi$ of F is such that $\{\phi(t)\}^y$ is a uniquely defined [4] characteristic function for all values y in the carrier [5] of G, we extend the definition to arbitrary distribution functions G with G(0-) = 0 and define $\underline{x}^{\underline{y}*}$ and $F^{G*}$ by their characteristic function

(5) $$g(\phi(t)) \stackrel{\text{def}}{=} \int \{\phi(t)\}^y \, dG(y).$$

This clearly includes the definition by (4). GURLAND (1957) uses the notation $\underline{y} \vee \underline{x}$, and $G \vee F$, and defines it, for non-negative $\underline{x}$ and $\underline{y}$, by stating that its generating function is $g(f(z))$, where $f(z) = \mathcal{E} z^{\underline{x}}$ and $g(z) = \mathcal{E} z^{\underline{y}}$. In GURLAND's examples, where F is infinitely divisible, $g(f(z))$ is always a generating function, but this is not correct in all cases with F(0-) = 0 and G(0-) = 0. A counterexample is $f(z) = pz + q$ and $g(z) = z^{\frac{1}{2}}$. Compared with GURLAND's definition, our definition admits negative values for $\underline{x}$ but excludes the cases where $\{\phi(t)\}^y$ is not a characteristic function for some y in the carrier of G. In the list of examples in Appendix 1 we have always a non-negative integer-valued $\underline{y}$ or an infinitely divisible $\underline{x}$ with a non-negative $\underline{y}$; in both cases $\underline{x}^{\underline{y}*}$ is defined.

---

4) Determine $\{\phi(t)\}^y = \exp\{y \log \phi(t)\}$ by selecting a branch of the logarithm. In every zero of $\phi$ one may switch from one branch to another. If one such selection gives a characteristic function, this selection is unique when $\phi$ is infinitely divisible and also when $\phi$ is regular on an open interval containing 0 (use $\{\phi(0)\}^y = 1$ and the extension theorem, LOÈVE (1963) p. 212). Whether the selection is unique under other circumstances seems to be unknown.

5) The carrier of a distribution function G is the set of all values y for which $\epsilon > 0$ implies $G(y + \epsilon) - G(y - \epsilon) > 0$.

<u>Definition 3</u>. The family $\{F_\theta \mid \theta \in T\}$ of distribution functions is <u>additively closed</u> (w.r.t. $\theta$) if we have, for all $\theta$, $\eta \in T$:

$$\theta + \eta \in T \text{ and } F_\theta(x) * F_\eta(x) \equiv F_{\theta+\eta}(x). \tag{6}$$

It is <u>strongly additively closed</u> if there exists a distribution function $F_1$ with characteristic function $\phi_1$ independent of $\theta$ such that for each $\theta \in T$ the characteristic function $\phi_\theta$ of $F_\theta$ is $\phi_\theta(t) = \{\phi_1(t)\}^\theta$.

If T consists of the positive integers or rationals the two notions coincide; if $T = (0,\infty)$ an additively closed family is strongly additively closed if $\phi_\theta(t)$ is a continuous function of $\theta$ or if $\phi_\theta(t)$ is real-valued for real t (see TEICHER (1954), where also additively closed families in more than one parameter are investigated). PYKE (1960) has shown for additively closed families with $T = [0,\infty)$ that $\phi_\theta(t) = \{\phi_1(t)\}^\theta \exp\{itc(\theta)\}$, where the real-valued function $c(\theta)$ on $[0,\infty)$ has $c(\theta) = 0$ for all rational $\theta$ and $c(\theta) + c(\eta) = c(\theta + \eta)$ for all $\theta \geq 0$ and $\eta \geq 0$. For a strongly additively closed family with for T the positive reals or rationals, $\phi_1$ is of course infinitely divisible.

If $\phi_1$ is not degenerate, the parameter of a strongly additively closed family can assume only non-negative values, otherwise $|\phi_\theta(t)| = |\phi_1(t)|^\theta$ would assume values larger than 1.

A few examples of strongly additively closed families are [6]:
Normal $(0,\sigma^2)$ w.r.t. $\sigma^2$ $(\sigma > 0)$;
Normal $(\mu\theta,\sigma^2\theta)$ w.r.t. $\theta$ $(\theta > 0)$;
Poisson $(\theta)$ w.r.t. $\theta$ $(\theta > 0)$;
Binomial $(n,p)$ w.r.t. n (integer $n > 0$);
Pascal $(\gamma,p)$ w.r.t. $\gamma$ $(\gamma > 0$, or integer $\gamma > 0)$;
in all cases the parameter value 0 corresponds to the degenerate distribution in 0 and may be included.

---

[6] A specification of the distributions is given in Appendix 2.

## 3. Simple properties

Lemma 1. Generalizing and mixing are associative operations, i.e.

$$(7) \qquad (H^{G**})^{F**} = H^{(G^{F**})**} \text{ and } (F_\theta \bigwedge_\theta G_\eta) \bigwedge_\eta H = F_\theta \bigwedge_\theta (G_\eta \bigwedge_\eta H)$$

in all cases where the distributions are defined.

Proof. If both sides of the first formula are meaningful, then $F(0-) = 0$ and $G(0-) = 0$ and $f(g(z))$ must be a generating function (corresponding to $G^{F**}$). Now if $\phi(t)$ is the characteristic function of H, then by (5) both sides have characteristic function $f(g(\phi(t)))$. For the second formula, one finds from TULCEA's theorem (LOÈVE (1963), p. 137)

$$(8) \qquad \iint F_\theta(x)\, dG_\eta(\theta)\, dH(\eta) = \int F_\theta(x)\, d_\theta \int G_\eta(\theta)\, dH(\eta).$$

Remark. It is trivial that for a two-parameter family $\{F_{\theta,\eta}\}$ we have

$$(9) \qquad (F_{\theta,\eta} \bigwedge_\theta G) \bigwedge_\eta H = (F_{\theta,\eta} \bigwedge_\eta H) \bigwedge_\theta G.$$

This might be extended by splitting any two-dimensional distribution of $\theta$ and $\eta$ in the two possible ways in conditional and marginal distributions.

The following basic lemma is a slightly modified form of a theorem by GURLAND (1957):

Lemma 2. If $\{F_\theta \mid \theta \in T\}$ is strongly additively closed and H assigns probability 1 to T, then

$$(10) \qquad F_\theta \bigwedge_\theta H = F_1^{H**}.$$

Proof. On both sides the characteristic function is $\int \{\phi_1(t)\}^\theta\, dH(\theta)$. It is not necessary to assume $1 \in T$, as $\phi_1$ and $F_1$ are defined by Definition 3.

Several relations of the type (10) are found in Appendix 1.

As an example we mention

$$\text{Pascal}\ (\gamma, p) = \text{Poisson}\ (\lambda) \bigwedge_{\lambda} \text{Gamma}\ (pq^{-1}, \gamma) = \{\text{Poisson}\ (1)\}^{\text{Gamma}\ (pq^{-1}, \gamma)^{*}}$$

<u>Lemma 3</u>. If $\{F_\theta\}$ is strongly additively closed, the operations $F_\theta \bigwedge_\theta .$ and $.^{G*}$ may be interchanged.

<u>Proof</u>. By lemmas 1 and 2 we have

$$(11) \qquad F_\theta \bigwedge_\theta (H^{G*}) = F_1{}^{(H^{G*})*} = (F_1{}^{H*})^{G*} = (F_\theta \bigwedge_\theta H)^{G*}.$$

By lemma 2, each mixed Poisson distribution Poisson $(\theta) \bigwedge_\theta H$ is also $\{\text{Poisson}\ (1)\}^{H*}$. Some generalized Poisson distributions $G^{\text{Poisson}*}$ are at the same time mixed Poisson (e.g. Neyman A and Pascal). MACEDA (1948) proves that all distributions of the form $(\text{Poisson} \bigwedge H)^{\text{Poisson}*}$ are both generalized Poisson and mixed Poisson. This is now a consequence of lemma 3, as we have

$$(12) \qquad (\text{Poisson} \bigwedge H)^{\text{Poisson}*} = \text{Poisson} \bigwedge (H^{\text{Poisson}*}).$$

<u>Lemma 4</u>. One-sided distributivity of generalizing and mixing with respect to convolution holds in the following sense: whenever both sides are defined we have

$$(13) \qquad H^{(F*G)*} = (H^{F*}) * (H^{G*}),$$

and

$$(14) \qquad F_\theta \bigwedge_\theta (H * G) = (F_\theta \bigwedge_\theta H) * (F_\theta \bigwedge_\theta G)$$

if $\{F_\theta\}$ is strongly additively closed.

<u>Proof</u>. If H has characteristic function $\phi$, both sides of (13) have characteristic function $f(\phi(t))\ g(\phi(t))$. From (13) and lemma 2 follows (14).

(14) was already proved by TEICHER (1960), by writing out the integrals. For $F_\theta$ = Poisson $(\theta)$ it is mentioned by FELLER (1943) and MACEDA (1948). The last author proves also

$$(15) \qquad \{H_1^{\text{Poisson }(\lambda)*}\} * \{H_2^{\text{Poisson }(\mu)*}\} = \{\frac{\lambda H_1+\mu H_2}{\lambda+\mu}\}^{\text{ Poisson }(\lambda+\mu)*}.$$

Only the special form of the Poisson characteristic function makes this extension of (13) to different $H_i$ possible.

One can easily see that the other two distributive laws

$$(H^{F*}) * (G^{F*}) = (H * G)^{F*} \text{ and } (F_\theta * F_\eta) \bigwedge H = (F_\theta \bigwedge H) * (F_\eta \bigwedge H)$$

hold only in the trivial cases where at least one of the distributions is degenerate.

Lemma 5. If G is infinitely divisible, then so is $H^{G*}$ for each H, and also $F_\theta \underset{\theta}{\bigwedge} G$ for each strongly additively closed family $\{F_\theta\}$.

Proof. For each positive integer n there is a distribution function $G_n$ such that $G_n^{n*} = G$. Thus by lemmas 4 and 2

$$(16) \qquad H^{G*} = H^{(G_n^{n*})*} = (H^{G_n*})^{n*}$$

and

$$(17) \qquad F_\theta \underset{\theta}{\bigwedge} G = F_1^{G*} = (F_1^{G_n*})^{n*} = (F_\theta \underset{\theta}{\bigwedge} G_n)^{n*}.$$

The second half of lemma 5 is stated by TEICHER (1960), with a different proof.

## 4. Type of a mixture

As stated by TEICHER (1960), any mixture of absolutely continuous distributions is absolutely continuous, while a mixture of discrete distributions can have any type (take $F_\theta(x) = \varepsilon(x - \theta)$, where $\varepsilon$ is the unit stepfunction). From TEICHER's remark that $F_\theta \underset{\theta}{\bigwedge} H$ is discontinuous in $x_0$ if and only if the $\theta$-set for which $F_\theta$ is discontinuous in $x_0$ has positive H-measure, we see immediately that a mixture of continuous distributions has no discrete part. The problem for a mixture of continuous singular distributions, suggested by Dr. Maitra, is settled in the following lemma.

Lemma 6. (i) A mixture of continuous singular distributions is again continuous singular if the mixture is countable (i.e. if the mixing distribution H is discrete). (ii) It can be continuous singular or absolutely continuous or a mixture of both if the mixture is non-countable.

Proof. (i) If $\lambda$ denotes Lebesque measure, and $\mu_k$ the measure corresponding to a continuous singular distribution $F_k$, then there exists a set $A_k$ with $\lambda(A_k) = 0$ and $\mu_k(A_k) = 1$, while $\mu_k(\{x\}) = 0$ for all x. Now the countable mixture $\Sigma\rho_k F_k$ is continuous singular, since $\lambda(\cup A_k) = 0$, $\Sigma\rho_k\mu_k(\cup A_k) = 1$ and $\Sigma\rho_k\mu_k(\{x\}) = 0$ for all x.

(ii) Let $F_0$ denote any continuous singular distribution with carrier contained in $[0,1]$, and let $F_\theta$ denote the same distribution shifted over a positive distance $\theta$ to the right but modulo 1, i.e. $F_\theta(x) = F_0(1 + x - \theta) - F_0(1 - \theta)$ if $0 < x < \theta$ and $F_\theta(x) = F_0(x - \theta) + 1 - F_0(1 - \theta)$ if $\theta \leq x \leq 1$. If for fixed $\rho$ $(0 \leq \rho < 1)$ we choose $H(\theta) = \rho\ \varepsilon(\theta) + (1 - \rho)\ U(\theta)$, where U denotes the uniform distribution on $[0,1]$, then

$$\int F_\theta(x)\ dH(\theta) = \rho F_0(x) + (1 - \rho)\ U(x). \tag{18}$$

If $F_\theta$ does not modify the location of $F_0$, but distributes the mass 1 in a way depending on $\theta$ over the same set A with $F_0$-measure 1 and $\lambda(A) = 0$, one obtains a non-countable mixture that is again continuous singular.

## 5. Moments of mixtures

For the investigation of the moments of the random variable $\underline{x}_\theta$ (with distribution function $F_\theta \underset{\theta}{\wedge} H$) we introduce

$$\begin{aligned} m_\theta &\overset{\text{def}}{=} \mathcal{E}\underline{x}_\theta = \int x\ dF_\theta(x); \\ s_\theta^2 &\overset{\text{def}}{=} \sigma^2(\underline{x}_\theta) = \int (x - m_\theta)^2\ dF_\theta(x); \end{aligned} \tag{19}$$

we shall assume that all moments mentioned exist. By well-known formulae we have

$$
(20)\qquad
\begin{aligned}
&\mathcal{E}\,\underline{x}_\theta = \mathcal{E}\{\mathcal{E}(\underline{x}_\theta|\underline{\theta})\} = \mathcal{E}\,m_{\underline{\theta}};\\
&\sigma^2(\underline{x}_\theta) = \sigma^2\,\{\mathcal{E}(\underline{x}_\theta|\underline{\theta})\} + \mathcal{E}\{\sigma^2(\underline{x}_\theta|\underline{\theta})\} = \sigma^2(m_{\underline{\theta}}) + \mathcal{E}\,s^2_{\underline{\theta}}.
\end{aligned}
$$

Let us assume that $F_\theta$ has expectation $m_\theta = \theta$, then $\mathcal{E}\,\underline{x}_\theta = \mathcal{E}\,\underline{\theta}$ and $\sigma^2(\underline{x}_\theta) = \sigma^2(\underline{\theta}) + \mathcal{E}\,s^2_{\underline{\theta}}$. If $F_\theta$ is non-degenerate for some set of $\theta$-values assumed with positive probability, then the variance of $\underline{x}_\theta$ is larger than the variance of $\underline{\theta}$. In the special case $F_\theta$ = Poisson $(\theta)$ we have also $s^2_\theta = \theta$ and, as proved by FELLER (1943), $\sigma^2(\underline{x}_\theta) = \sigma^2(\underline{\theta}) + \mathcal{E}\,\underline{x}_\theta$: the variance of a non-trivial mixture of Poisson distributions is always larger than its expectation.

Another special case is $F_\theta \bigwedge_\theta$ Poisson $(\lambda)$. Here it is not very realistic to assume $m_\theta = \theta$, as in most mixtures the expectation of $F_\theta$ will not be integer-valued for all $\theta$. If $m_\theta = k\theta$ for some constant $k$ and integer $\theta$, then

$$
(21)\qquad \sigma^2(\underline{x}_\theta) = k^2\lambda + \mathcal{E}\,s^2_{\underline{\theta}} > k\lambda = \mathcal{E}\,\underline{x}_\theta,
$$

provided we have $k > 1$. This proviso is not necessary for a Poisson-mixture of distributions $F_\theta$ with expectation $k\theta$ to have larger variance than expectation. The Poisson Pascal distribution Pascal $(\gamma\theta, p) \bigwedge_\theta$ Poisson $(\lambda)$ has expectation $\gamma\lambda p q^{-1}$ and variance $\gamma\lambda p q^{-2}(\gamma p + 1)$; here the variance always exceeds the expectation even if the constant $k = \gamma p q^{-1}$ is $\leq 1$.

## 6. Comparison between a mixture and one of its components

If $\mathcal{E}\,\underline{\theta} = \int \theta\, dH(\theta)$ exists and is an admissible parameter value, it is interesting to compare the mixture $F_\theta \bigwedge_\theta H$ with the single component $F_{\mathcal{E}\underline{\theta}}$. When the expectation $m_\theta$ of $F_\theta$ is proportional to $\theta$, this amounts to a comparison of the mixture to the component with the same expectation. Starting-point is the result of FELLER (1943) that for each non-trivial Poisson $(\theta) \bigwedge_{\theta} H$ we have larger $\sigma^2(\underline{x})$, larger $P\{\underline{x} = 0\}$ and smaller $P\{\underline{x} = 1\}/P\{\underline{x} = 0\}$ than for the Poisson distribution with the same expectation. Two possible extensions are stated in the following lemmas.

Lemma 7. For a non-trivial mixture $F_\theta \bigwedge_{\theta} H$, where $\mathcal{E}\underline{\theta}$ is an admissible parameter value, each of the following conditions is sufficient for $\sigma^2(\underline{x}_{\underline{\theta}}) \geq \sigma^2(\underline{x}_{\mathcal{E}\underline{\theta}})$:

(a) $s_\theta^2 \overset{\text{def}}{=} \int (x - m_\theta)^2 \, dF_\theta(x)$ is a convex [7] function of $\theta$;

(b) $F_\theta$ is Binomial $(n,\theta)$ for fixed n.

We have $\sigma^2(\underline{x}_{\underline{\theta}}) > \sigma^2(\underline{x}_{\mathcal{E}\underline{\theta}})$ if (a) holds with strictly convex $s_\theta^2$, or if (b) holds with $n > 1$, or if we add to (a) that $m_{\underline{\theta}} \overset{\text{def}}{=} \int x \, dF_{\underline{\theta}}(x)$ has positive variance.

Proof. In case (a) we have

(22) $$\sigma^2(\underline{x}_{\underline{\theta}}) = \sigma^2(m_{\underline{\theta}}) + \mathcal{E}\, s^2_{\underline{\theta}} \geq s^2_{\mathcal{E}\underline{\theta}},$$

by Jensen's inequality and the fact that the first term is non-negative. In case (b) $m_\theta = n\theta$, $s_\theta^2 = n\theta(1 - \theta)$, and

(23) $$\sigma^2(\underline{x}_{\underline{\theta}}) = n^2\sigma^2(\underline{\theta}) + n\,\mathcal{E}\underline{\theta} - n\mathcal{E}\underline{\theta}^2 \geq n\,\mathcal{E}\underline{\theta} - n(\mathcal{E}\underline{\theta})^2 = s^2_{\mathcal{E}\underline{\theta}},$$

because $(n^2 - n)\sigma^2(\underline{\theta}) \geq 0$. The statements on strict inequality follow directly.

Remark. The lemma holds for Binomial $(n,\theta)$ though its variance is a strictly concave function of $\theta$. Cases (a) and (b) together cover all mixtures listed in Appendix 1.

---

[7] As usual "convex" includes "linear"; a function f is called "strictly convex" when $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$ for all x and y in the interval where f is defined and all $\lambda \in (0,1)$.

Definition 4. A functional v mapping a class of distribution functions into the real numbers is mix-concave w.r.t. $F_\theta \bigwedge_\theta H$, if it is defined at least for $F_\theta \bigwedge_\theta H$ and all $F_\theta$, and satisfies

(24) $$v(F_\theta \bigwedge_\theta H) \geq \int v(F_\theta)\, dH(\theta).$$

It is strictly mix-concave if we have strict inequality and mix-linear if we have equality in (24). The definition of mix-convex is analogous.

Examples. For all mixtures and all real a and b functionals like $v(F) = P\{\underline{x} = a \mid F\}$ and $v(F) = P\{a < \underline{x} \leq b \mid F\}$ are mix-linear. The expectation is mix-linear; as we have proved $\sigma^2(\underline{x}_\theta) = \sigma^2(m_\theta) + \mathcal{E}\, s_\theta^2$, the variance is always mix-concave, and for any non-trivial mixture it is strictly mix-concave unless $m_\theta$ has zero variance.

Lemma 8. Let $F_\theta \bigwedge_\theta H$ be a non-trivial mixture and $\mathcal{E}\,\underline{\theta}$ an admissible parameter value. If the functional v is mix-concave w.r.t. $F_\theta \bigwedge_\theta H$, and $v(F_\theta)$ is a convex function of $\theta$, then $v(F_\theta \bigwedge_\theta H) \geq v(F_{\mathcal{E}\underline{\theta}})$. The last inequality is strict as soon as v is strictly mix-concave or strictly convex in $\theta$.

Proof. $v(F_\theta \bigwedge_\theta H) \geq \int v(F_\theta)\, dH(\theta) \geq v(F_{\mathcal{E}\underline{\theta}})$.

Example 1. Pascal $(\gamma, p)$ = Poisson $(\theta) \bigwedge_\theta$ Gamma $(pq^{-1}, \gamma)$.
The variance $s_\theta^2 = \theta$ is linear in $\theta$, so lemma 7 is applicable. In fact the variance $\gamma p q^{-2}$ for Pascal $(\gamma, p)$ is larger than the variance of Poisson $(\mathcal{E}\underline{\theta})$, which is $\mathcal{E}\underline{\theta} = \gamma p q^{-1}$.
$v(F_\theta) = P\{\underline{x} = 0 \mid F_\theta\} = e^{-\theta}$ is strictly convex in $\theta$ and mix-linear, so lemma 8 is applicable. In fact $P\{\underline{x} = 0\} = q^\gamma$ for Pascal $(\gamma, p)$; this is larger than $\exp(-\gamma p q^{-1})$ for Poisson $(\mathcal{E}\underline{\theta})$ as we have $\gamma \log(1 - p) > - \gamma p (1 - p)^{-1}$.
$v(F) = P\{\underline{x} = 1 \mid F\} / P\{\underline{x} = 0 \mid F\}$ can be shown to be strictly mix-convex and $v(F_\theta) = \theta$ is linear, so by an obvious modification of lemma 8

(25) $$v(F_{\mathcal{E}\underline{\theta}}) = \int v(F_\theta)\, dH(\theta) > v(F_\theta \bigwedge_\theta H),$$

namely

(26) $$\gamma p q^{-1} = \mathcal{E}\underline{\theta} > \gamma p.$$

Example 2. Polya $(n,r,s)$ = Binomial $(n,p) \bigwedge_p$ Beta $(r,s)$.
By lemma 7, the variance of the Polya distribution must exceed for $n > 1$ that of Binomial $(n, r/(r+s))$. In fact we have

$$(27) \qquad \frac{nrs(n+r+s)}{(r+s)^2(r+s+1)} > \frac{nrs}{(r+s)^2}.$$

$v(F_p) = P\{\underline{x} = 0 \mid F_p\} = (1-p)^n$ is convex in $p$ (strictly so for $n > 1$), and mix-linear. In fact $P\{\underline{x} = 0\}$ equals

$$(28) \qquad \frac{(s+n-1)(s+n-2)\ \ldots\ s}{(r+s+n-1)(r+s+n-2)\ \ldots\ (r+s)} \quad \text{and} \quad \left(\frac{s}{r+s}\right)^n$$

for Polya $(n,r,s)$ and Binomial $(n,r/(r+s))$ respectively. They are clearly equal for $n = 1$ and for $n > 1$ the Polya distribution has larger $P[\underline{x} = 0]$.

## 7. Analogous results for generalizations

We shall compare the sum $\underline{x}^{\underline{y}*}$ of a random number of independent $\underline{x}_i$ with the sum $\underline{x}^{(\mathcal{E}\underline{y})*}$ of a fixed (not necessarily integer) number of $\underline{x}_i$; we assume that all moments mentioned in the following exist. Let $\underline{x}$ have characteristic function $\phi$, and suppose that, for some real $a \geq 0$, $\phi^a$ is a characteristic function. From the expansion of $\log \phi^a(t)$ in powers of $t$ it is clear that, also for non-integer $a$,

$$(29) \qquad \mathcal{E}\underline{x}^{a*} = a.\mathcal{E}\underline{x} \quad \text{and} \quad \sigma^2(\underline{x}^{a*}) = a.\sigma^2(\underline{x}).$$

In analogy to (20) we have

$$(30) \qquad \begin{aligned} \mathcal{E}\underline{x}^{\underline{y}*} &= \mathcal{E}\underline{y}.\,\mathcal{E}(\underline{x}^{\underline{y}*} \mid \underline{y}) = \mathcal{E}\underline{y}\,.\,\mathcal{E}\underline{x}; \\ \sigma^2(\underline{x}^{\underline{y}*}) &= \mathcal{E}\sigma^2(\underline{x}^{\underline{y}*} \mid \underline{y}) + \sigma^2(\mathcal{E}(\underline{x}^{\underline{y}*} \mid \underline{y})) = \\ &= \mathcal{E}(\underline{y}\sigma^2(\underline{x})) + \sigma^2(\underline{y}\mathcal{E}\underline{x}) = \mathcal{E}\underline{y}.\sigma^2(\underline{x}) + \sigma^2(\underline{y}).\mathcal{E}\underline{x}. \end{aligned}$$

Lemma 9. If both generalizations are defined, $\underline{x}^{\underline{y}*}$ has the same expectation and no smaller variance than $\underline{x}^{(\mathcal{E}\underline{y})*}$; the variance is larger unless $\sigma^2(\underline{y}) = 0$ or $\mathcal{E}\underline{x} = 0$.

Lemma 10. If $P\{\underline{x} \geq 0\} = 1$ and $0 < p_0 \overset{\text{def}}{=} P\{\underline{x} = 0\} < 1$, and the distribution of $\underline{y}$ is non-degenerate, then if both generalizations are defined we have

$$(31) \qquad P\{\underline{x}^{\underline{y}*} = 0\} > P\{\underline{x}^{(\mathcal{E}\underline{y})*} = 0\}.$$

Proofs. Lemma 9 follows from (29) and (30). For lemma 10 note that $p_0^{\,y}$ is a strictly convex function of y and apply Jensen's inequality to $P\{\underline{x}^{(\mathcal{E}\underline{y})*} = 0\} = p_0^{\mathcal{E}\underline{y}}$ and $P\{\underline{x}^{\underline{y}*} = 0\} = \int P\{\underline{x}^{y*} = 0\}\, dG(y) = \mathcal{E}p_0^{\underline{y}}$.

Remark 1. We have equality in (31) if $p_0 = 0$, $p_0 = 1$ or $\underline{y}$ is degenerate.

Remark 2. With lemma 2 we could translate lemmas 9 and 10 into results for strongly additively closed $\{F_\theta\}$, requiring $P\{\underline{x}_\theta \geq 0\} = 1$ in the case of lemma 10. But then $\sigma^2(\underline{x}_\theta) = \theta\sigma^2(\underline{x}_1)$ is linear in $\theta$ and the variance is mix-concave, while $P\{\underline{x}_\theta = 0\} = (P\{\underline{x}_1 = 0\})^\theta$ is convex and $P\{\underline{x} = 0\}$ is mix-linear. If $\sigma^2(\underline{\theta}) > 0$ and $\mathcal{E}\underline{x}_1 \neq 0$ the variance is strictly mix-concave; if $0 < P\{\underline{x}_1 = 0\} < 1$ and $\sigma^2(\underline{\theta}) > 0$ we have strictly convex $P\{\underline{x}_\theta = 0\}$.
Thus the results follow already from the (stronger) lemma 8; lemmas 9 and 10 could also have been derived in this way.

Acknowledgment

I want to thank Professor Hemelrijk and Dr. Van Zwet for some helpful suggestions.

## Appendix 1

### List of mixtures and generalizations

For the definitions of the distributions see Appendix 2.
This list gives some well-known examples; it is far from being complete.

As usual, q denotes 1 - p.

Neyman A$(\lambda,\mu)$ = Poisson $(k\mu) \bigwedge_k$ Poisson $(\lambda)$ = $[\text{Poisson } (\mu)]^{\text{Poisson } (\lambda)*}$ =
= $[\text{Binomial } (n,p) \bigwedge_n \text{Poisson } (\mu p^{-1})]^{\text{Poisson } (\lambda)*}$

Pascal $(\gamma,p)$ = Poisson $(\lambda) \bigwedge_\lambda$ Gamma $(pq^{-1},\gamma)$ = $[\text{Poisson } (1)]^{\text{Gamma } (pq^{-1},\gamma)*}$ =
= $[\text{Poisson } (pq^{-1})]^{\text{Gamma } (1,\gamma)*}$ = $[\text{Log } (p)]^{\text{Poisson } (-\gamma \log q)*}$

Poisson Pascal $(\lambda,\gamma,p)$ = Pascal $(k\gamma,p) \bigwedge_k$ Poisson $(\lambda)$ =
= $[\text{Pascal } (\gamma,p)]^{\text{Poisson } (\lambda)*}$

Poisson Binomial $(\lambda,n,p)$ = Binomial $(kn,p) \bigwedge_k$ Poisson $(\lambda)$ =
= $[\text{Binomial } (n,p)]^{\text{Poisson } (\lambda)*}$

Pascal $(c\gamma,p) \bigwedge_\gamma$ Gamma $(\beta,\gamma)$ = $[\text{Log } (p)]^{\text{Pascal } (\gamma,\frac{-c\beta \log q}{1-c\beta\log q})*}$ =
= $[\text{Pascal } (c,p)]^{\text{Gamma } (\beta,\gamma)*}$

Poisson $(\mu p)$ = Binomial $(n,p) \bigwedge_n$ Poisson $(\mu)$ = $[\text{Binomial } (1,p)]^{\text{Poisson } (\mu)*}$

Polya $(n,r,s)$ = Binomial $(n,p) \bigwedge_p$ Beta $(r,s)$

Gamma $(1,\gamma)$ = Gamma $(a^{-1},\gamma+j) \bigwedge_j$ Pascal $(\gamma,1-a^{-1})$

Gamma $(\lambda,1)$ = $[\text{Gamma } (\lambda p^{-1},1)]^{[1 + \text{Pascal } (1,q)]*}$

Gurland $(\alpha,\beta,\mu)$ = Poisson $(\mu p) \bigwedge_p$ Beta $(\alpha,\beta)$ = Polya $(n,\alpha,\beta) \bigwedge_n$ Poisson $(\mu)$ =
= (Binomial $(n,p) \bigwedge_p$ Beta $(\alpha,\beta)) \bigwedge_n$ Poisson $(\mu)$

Laplace (1) = Normal $(0,\sigma^2) \bigwedge_{\frac{1}{2}\sigma^2}$ Gamma $(1,1)$

## Appendix 2

### List of distributions

It will be obvious what is meant by Binomial $(n,p)$, Degenerate in a, Normal $(\mu,\sigma^2)$ or Poisson $(\lambda)$. The other distributions used are listed here.

| NAME<br>RESTRICTIONS | DENSITY OR PROBABILITIES [8] | EXPECTATION; VARIANCE;<br>CHARACTERISTIC FUNCTION |
|---|---|---|
| Beta $(r,s)$<br>$r > 0,\ s > 0$ | $\frac{x^{r-1}(1-x)^{s-1}}{B(r,s)}\ (0 \leq x \leq 1)$ | $\frac{r}{r+s}$; $\frac{rs}{(r+s+1)(r+s)^2}$;<br>$\frac{\Gamma(p+q)}{\Gamma(p)} \sum_{j=0}^{\infty} \frac{\Gamma(p+j)(it)^j}{\Gamma(p+q+j)\Gamma(j+1)}$ |
| Gamma $(\beta,\gamma)$<br>$\beta > 0,\ \gamma > 0$ | $\frac{e^{-x/\beta}x^{\gamma-1}}{\beta^{\gamma}\Gamma(\gamma)}\ (x \geq 0)$ | $\beta\gamma$; $\beta^2\gamma$;<br>$(1-\beta it)^{-\gamma}$ |
| Gurland $(\alpha,\beta,\mu)$<br>$\alpha > 0,\ \beta > 0,\ \mu > 0$ | [9] $(x = 0, 1, \ldots)$ | $\frac{\alpha\mu}{\alpha+\beta}$; $\frac{\alpha\mu}{\alpha+\beta} + \frac{\alpha\beta\mu^2}{(\alpha+\beta+1)(\alpha+\beta)^2}$<br>${}_1F_1(\alpha,\alpha+\beta,\mu(e^{it}-1))$ |
| Laplace $(\beta)$<br>$\beta > 0$ | $\frac{1}{2\beta}\exp\{-\frac{\lvert x\rvert}{\beta}\}$ | $0$; $2\beta^2$;<br>$(1+\beta^2t^2)^{-1}$ |
| Log $(p)$<br>$0 < p < 1$ | $\frac{p^x}{x\lvert\log q\rvert}\ (x = 1, 2, \ldots)$ | $\frac{p}{q\lvert\log q\rvert}$; $\frac{p(p+\log q)}{q^2\lvert\log q\rvert^2}$;<br>$\log(1-pe^{it})/\log q$ |
| Neyman $A(\lambda,\mu)$<br>$\lambda > 0,\ \mu > 0$ | $\frac{e^{-\lambda}\mu^x}{x!}\sum_{k=0}^{\infty}\frac{k^x\lambda^k e^{-\mu k}}{k!}$<br>$(x = 0, 1, \ldots)$ | $\lambda\mu$; $\lambda\mu(\mu+1)$;<br>$\exp(\lambda\{e^{\mu(e^{it}-1)}-1\})$ |
| Pascal $(\gamma,p)$ [10]<br>$\gamma > 0,\ 0 < p < 1$ | $\frac{\Gamma(\gamma+x)}{\Gamma(\gamma)x!}p^x q^{\gamma}\ (x = 0, 1, \ldots)$ | $\gamma pq^{-1}$; $\gamma pq^{-2}$;<br>$q^{\gamma}(1-pe^{it})^{-\gamma}$ |

8) For arguments not mentioned the value is zero.

9) Recurrence relations for $P[\underline{x} = x]$ are given by GURLAND (1958).

10) This is the negative binomial, but we prefer the shorter name. For integer $\gamma$ it is the distribution of the number of successes preceding the $\gamma^{th}$ failure.

| NAME RESTRICTIONS | DENSITY OR PROBABILITIES [8] | EXPECTATION; VARIANCE; CHARACTERISTIC FUNCTION |
|---|---|---|
| Poisson Binomial $(\lambda, n, p)$ $\lambda > 0$, integer $n > 0$, $0 < p < 1$ | $e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \binom{nk}{x} p^x q^{nk-x}$ $(x = 0, 1, \ldots)$ | $\lambda np$; $\lambda n^2 p^2 + \lambda npq$; $\exp(\lambda\{(pe^{it}+q)^n - 1\})$ |
| Poisson Pascal $(\lambda, \gamma, p)$ $\lambda > 0$, $\gamma > 0$, $0 < p < 1$ | [11] $(x = 0, 1, \ldots)$ | $\gamma\lambda pq^{-1}$; $\gamma\lambda pq^{-2}(\gamma p+1)$; $\exp(\lambda\{q^{\gamma}(1-pe^{it})^{-\gamma} - 1\})$ |
| Polya $(n, r, s)$ | $\binom{n}{j} \frac{B(r+x, s+n+x)}{B(r,s)}$ | $\frac{nr}{r+s}$; $\frac{nrs(n+r+s)}{(r+s)^2(r+s+1)}$ |

## References


W. FELLER (1943), On a general class of "contagious" distributions, Ann. Math. stat. 14, 389-400.

W. FELLER (1957), An introduction to probability theory and its applications, Vol. 1, Second edition, Wiley, New York.

J. GURLAND (1957), Some interrelations among compound and generalized distributions, Biometrika 44, 265-268.

J. GURLAND (1958), A generalized class of contagious distributions, Biometrics 14, 229-249.

S.K. KATTI & J. GURLAND (1961), The Poisson Pascal distribution, Biometrics 17, 527-538.

M. LOÈVE (1963), Probability theory, Third edition, Van Nostrand, New York.

E. CANSADO MACEDA (1948), On the compound and generalized Poisson distributions, Ann. Math. Stat. 19, 414-416.


[11] Recurrence relations for $P[\underline{x} = x]$ are given by KATTI & GURLAND (1961).

R. PYKE (1960), On centering infinitely divisible processes, Ann. Math. Stat. 31, 797-800.

H. TEICHER (1954), On the convolution of distributions, Ann. Math. Stat. 25, 775-778.

H. TEICHER (1960), On the mixture of distributions, Ann. Math. Stat. 31, 55-73.

## RESUME

Comme l'a démontré GURLAND (1957), il existe une relation entre le mélange de fonctions de répartition à un paramètre, et la généralisation d'une distribution (c'est à dire la convolution de $y$ facteurs identiques, où $y$ est une variable aléatoire). Avec cette relation, quelques propriétés simples du mélange et de la généralisation sont dérivées. Une partie de ces résultats a été obtenue déjà par TEICHER (1960), avec une démonstration différente. Quelques observations sont faites sur le type d'un mélange dont toutes les composantes sont singulières. Le mélange général à un paramètre est comparé avec une composante représentative, et la généralisation est comparée avec une convolution d'un nombre fixe de facteurs. Dans deux appendices on trouve des exemples et une liste des distributions mentionnées.