STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 353
(SP 88)

On mixtures of Distributions

by

W. Molenaar
W.R. van Zwet

(Annals of mathematical Statistics, 37(1966),
p 281-283)

1966

Reprinted from THE ANNALS OF MATHEMATICAL STATISTICS
Vol. 37, No. 1, February, 1966
Printed in U.S.A.

ON MIXTURES OF DISTRIBUTIONS[1]

BY W. MOLENAAR AND W. R. VAN ZWET[2]

Mathematisch Centrum, Amsterdam, and University of Oregon

1. Introduction. Let F_θ, $\theta \varepsilon T \subseteq R^1$, be a family of distribution functions on R^1, where F_θ is measurable in θ. For an arbitrary non-degenerate distribution function H that assigns probability 1 to T we consider the H-mixture of the family F_θ, i.e. the distribution function

$$F_H(x) = \int_T F_\theta(x)\, dH(\theta).$$

In 1943 W. Feller [1] proved that if F_θ is Poisson, then the variance of F_H is always larger than the variance of the Poisson distribution with the same expectation. This note is an attempt to generalize this and related results.

2. Convexity arguments. If g_1 and g_2 are integrable with respect to F_θ for all $\theta \varepsilon T$ and with respect to F_H we define for $i = 1, 2$

$$\chi_i(\theta) = \int g_i(x)\, dF_\theta(x)$$
$$\chi_i(H) = \int g_i(x)\, dF_H(x) = \int_T \chi_i(\theta)\, dH(\theta).$$

We note that the same symbols χ_i are used to denote both the functions $\chi_i(\theta)$ and the functionals $\chi_i(H)$.

We shall say that the function χ_2 is convex, concave, or linear with respect to χ_1 on T if there exists a convex, concave, or linear function φ on $\chi_1(T)$ such that

$$\chi_2(\theta) = \varphi(\chi_1(\theta)) \qquad \text{for all} \quad \theta \varepsilon T.$$

We note that this definition differs from the one usually given in that there is no monotonicity requirement for χ_1 involved.

The following version of Jensen's inequality will be needed in the sequel (cf. [2], p. 75).

LEMMA 1. *Let $\mathcal{H}$ denote the class of distribution functions H on T having*

$$\chi_1(H) = \chi_1(\theta_H) \qquad \textit{for some} \quad \theta_H \varepsilon T.$$

Then a necessary and sufficient condition for $\chi_2(H) \geqq \chi_2(\theta_H)$ to hold for all $H \varepsilon \mathcal{H}$ is that χ_2 is convex with respect to χ_1 on T.

The inequality is strict for all $H \varepsilon \mathcal{H}$ for which χ_1 is not constant a.e. *$[H]$ on T, if and only if the convexity is everywhere strict; there is equality for all $H \varepsilon \mathcal{H}$ if and only if χ_2 is linear with respect to χ_1 on T. If in the above "convex" is replaced by "concave" the inequality is reversed.*

Received 25 May 1965.

[1] Report SP 88, Statistische Afdeling, Mathematisch Centrum, Amsterdam.

[2] Now at the University of Leiden and Mathematisch Centrum, Amsterdam.

PROOF. If $\chi_2 = \varphi\chi_1$ and φ is convex we have $\chi_2(H) = \int_T \chi_2(\theta)\, dH(\theta) = \int_T \varphi(\chi_1(\theta))\, dH(\theta) \geqq \varphi(\int_T \chi_1(\theta)\, dH(\theta)) = \varphi(\chi_1(H)) = \varphi(\chi_1(\theta_H)) = \chi_2(\theta_H)$ by Jensen's inequality, which also yields the sufficiency of the conditions for strictness and equality. The converse follows by considering distributions H that concentrate at two points.

To apply Lemma 1 to a comparison of the variances of the distributions F_H and F_{θ_H} where F_H and F_{θ_H} have the same expectation, we set $g_1(x) = x, g_2(x) = x^2$ and define

$$\mu_1(\theta) = \int x\, dF_\theta(x),$$
$$\mu_1(H) = \int x\, dF_H(x),$$
$$\mu_2(\theta) = \int x^2\, dF_\theta(x),$$
$$\mu_2(H) = \int x^2\, dF_H(x),$$
$$\sigma^2(\theta) = \mu_2(\theta) - \mu_1^2(\theta),$$

and

$$\sigma^2(H) = \mu_2(H) - \mu_1^2(H).$$

COROLLARY 1. *A necessary and sufficient condition for* $\sigma^2(H) \geqq \sigma^2(\theta_H)$ *to hold for all* H *having* $\mu_1(H) = \mu_1(\theta_H)$, $\theta_H \varepsilon T$, *is that* μ_2 *is convex with respect to* μ_1 *on* T.

We note that if $\mu_1(\theta)$ is linear, the above condition reduces to convexity of $\mu_2(\theta)$ on T. The result is then a direct generalization of Feller's theorem.

3. Total positivity. It turns out that for most well-known families F_θ, $\mu_2(\theta)$ is convex in $\mu_1(\theta)$ and hence mixing increases the variance. As will be seen the reason for this is that many of these families possess totally positive densities, a concept that was extensively investigated by S. Karlin et al. (cf. e.g. [3], [4], [5]).

Suppose that the family F_θ possesses densities $p(x, \theta)$ with respect to a σ-finite measure ν with spectrum $X \subseteq R^1$, i.e.

$$F_\theta(x) = \int_{-\infty}^{x} p(u, \theta)\, d\nu(u).$$

The density $p(x, \theta)$ (or the family F_θ) is called totally positive of order $k(TP_k)$, if for all $x_1 < x_2 < \cdots < x_m$ in X, $\theta_1 < \theta_2 < \cdots < \theta_m$ in T and all $1 \leqq m \leqq k$, the determinant

$$\begin{vmatrix} p(x_1, \theta_1) & p(x_1, \theta_2) & \cdots & p(x_1, \theta_m) \\ p(x_2, \theta_1) & p(x_2, \theta_2) & \cdots & p(x_2, \theta_m) \\ \vdots & & & \vdots \\ p(x_m, \theta_1) & p(x_m, \theta_2) & \cdots & p(x_m, \theta_m) \end{vmatrix} \geqq 0.$$

Densities of this type possess the following variation diminishing property (cf. [5] for a statement of the general result as it is mentioned here; cf. [3] for a proof of a special case. The determinantal inequality on which this proof is based may also be used to provide a direct proof of the general result). Let $V(g)$ denote the number of changes of sign of g. If χ is given by the absolutely convergent integral

$$\chi(\theta) = \int g(x)p(x, \theta)\, d\nu(x) = \int g(x)\, dF_\theta(x),$$

here $p(x, \theta)$ is TP_k and $V(g) \leqq k - 1$ then $V(\chi) \leqq V(g)$. If $V(\chi) = V(g)$ on g and χ change sign in the same order.

The following lemma may be obtained by exploiting this variation diminishing property. It is a slightly different formulation of a result obtained by Karlin in [4], p. 343.

LEMMA 2. *If $p(x, \theta)$ is TP_3, g_1 is monotone on X, and g_2 is convex with respect to g_1 on X, then χ_2 is convex with respect to χ_1 on T.*

Karlin proved the lemma for the case where $g_1(x) = x$ and χ_1 is linear. He also showed (cf. [4], pp. 342–343) that if g_1 is non-decreasing (non-increasing) on X, then so is χ_1 on T. Since the property of total positivity is preserved under non-decreasing (non-increasing) transformations g_1 and χ_1 of the random variable and the parameter simultaneously, the lemma may be proved by reducing it to the special case considered by Karlin.

Combining Lemmata 1 and 2, we have

THEOREM 1. *If F_θ is TP_3, g_1 is monotone on X, g_2 is convex with respect to g_1 on X, and $\chi_1(H) = \chi_1(\theta_H)$ for some $\theta_H \varepsilon T$,* then $\chi_2(H) \geqq \chi_2(\theta_H)$.

Since $g_1(x) = x$ is monotone and $g_2(x) = x^2$ is convex, we have from Corollary 1

COROLLARY 2. *If F_θ is TP_3 and $\mu_1(H) = \mu_1(\theta_H)$, $\theta_H \varepsilon T$, then $\sigma^2(H) \geqq \sigma^2(\theta_H)$.*

Finally we remark that the conclusions of Lemma 2, Theorem 1 and Corollary 2 are obviously independent of a particular parametrization of the family F_θ. It is therefore sufficient to require that there exists a parametrization such that the family is TP_3.

REFERENCES

[1] FELLER, W. (1943). On a general class of contagious distributions. *Ann. Math. Statist.* **14** 389–400.

[2] HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1934). *Inequalities*. Cambridge Univ. Press.

[3] KARLIN, S. (1955). Decision theory for Pólya type distributions. Case of two actions, I. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 115–128, Univ. of California, Berkeley.

[4] KARLIN, S. (1963). Total positivity and convexity preserving transformations. *Proc. Symp. Pure Math.* **7** 329–347, Amer. Math. Soc., Providence.

[5] KARLIN, S., PROSCHAN, F., and BARLOW, R. E. (1961). Moment inequalities of Pólya frequency functions. *Pacific J. Math.* **11** 1023–1033.