Handout on Mixtures of Densities and Distributions

The purpose of this handout is to clarify and summarize the definitions and motivations for the topic of mixture densities, probability mass functions, and distributions. The topic arose in connection with the discussion of **mixed-type distributions** in Casella and Berger's *Example 1.5.6, p. 33,* and later in the discussion of **Hierarchical Models and Mixture Distributions** in Casella and Berger's *Section 4.5*.

One can talk about mixtures of probability densities, mass functions, or distributions on any space of data values, e.g., any vector space $\mathbf{R}^k$, but for concreteness in this handout we will talk only about densities, mass functions, or distribution functions on the line, i.e., for scalar random variables.

A. DISCRETE MIXTURES

Mixture mass functions, densities or distribution functions with a finite or countable number of components are simply **convex combinations** of probability mass functions, density functions or distribution functions. Let L be the number of mixture components, either a finite integer or a countable ∞, and let $\{w_j : j = 1, \ldots, L\}$ be a system of positive weights summing to 1. Next let $\{p_j(\cdot) : j = 1, \ldots, L\}$ be an arbitrary set of (distinct) probability mass functions, and $\{f_j(\cdot) : j = 1, \ldots, L\}$ an arbitrary set of (distinct) probability density functions on the real line. The mixture with weights w_j of the probability mass functions p_j or of the probability density functions f_j are respectively defined as the probability mass function $p(x)$ or density function $f(x)$

$$p(x) = \sum_{j=1}^{L} w_j \, p_j(x) \quad , \qquad f(x) = \sum_{j=1}^{L} w_j \, f_j(x) \tag{1}$$

In either of these two cases, with the component distribution functions $F_j(x)$ defined respectively by

$$F_j(x) = \sum_{t:\, p_j(t) \geq 0,\, t \leq x} p_j(t) \qquad \text{or} \qquad F_j(x) = \int_{-\infty}^{x} f_j(t)\, dt \tag{2}$$

we can say that the mixture distribution function F associated with the

mass function $p(\cdot)$ or density function $f(\cdot)$ in equation (1) has the form

$$F(x) \;=\; \sum_{j=1}^{L} w_j \, F_j(x) \tag{3}$$

B. Mixtures as Two-stage Hierarchically Specified Distributions

The statistical motivation for mixture densities or distributions is that a population used to sample individuals from which to take measurements may naturally fall into *subpopulations* with different characteristics. So the population Ω from which we sample may naturally be partitioned into disjoint *subpopulations* Ω_j, with $\Omega = \cup_{j=1}^{L} \Omega_j$. Suppose that when an individual is sampled randomly from the full population Ω, the probability with which that individual falls into Ω_j is w_j, but that it is not necessarily known to the experimenter which subpopulation Ω_j contains that individual. (The probabilities w_j are assumed to be positive and sum to 1, so each individual in Ω belongs to one and only one of the subpopulations Ω_j.) Suppose we can say that the random-variable measurement X that the experimenter takes on that individual will have the distribution function F_j if that individual belongs to population j, i.e, that $P(X \leq x \,|\, \Omega_j) \;=\; F_j(x)$. Then our experiment has two stages: first, the individual is sampled, and has the unobservable discrete random-variable label J which takes the value $J = j$ whenever the sampled individual belongs to subpopulation Ω_j; then the second stage is to take the measurement X which has the conditional probability distribution $P(X \leq x \,|\, J = j) \;=\; F_j(x)$. In this case, the law of total probability along with countable additivity say that

$$F(x) \;=\; P(X \leq x) \;=\; P(X \leq x,\, J \in \{1,\dots,L\}) \;=\; \sum_{j=1}^{L} P(X \leq x,\, J = j)$$

$$= \; \sum_{j=1}^{L} P(J = j) \cdot P(X \leq x \,|\, J = j) \;=\; \sum_{j=1}^{L} w_j \, F_j(x)$$

The description in the previous paragraph might be summarized as in the definition of *hierarchical models* in Casella and Berger's Section 4.5 by saying that in the two-stage sampling and measurement experiment, there are underlying variables J and $\mathbf{X} = (X_j \; : \; j = 1,\dots,L)$, with J independent of the vector $\mathbf{X}$, such that

$$p_J(j) \;=\; w_j \,, \quad 1 \leq j \leq L \qquad \text{and} \qquad X_j \sim F_j$$

and that the ultimate measurement can be viewed as

$$X \;=\; X_J \;=\; \sum_{j=1}^{L} X_j \, I_{[J=j]}$$

<h2 style="text-align:center">C. Distribution Functions of Mixed Type</h2>

However, the mixture of distribution functions in (3) can be more general than the distribution function associated with $p(\cdot)$ or $f(\cdot)$ in (1), because the distribution functions F_j need not all be pure-jump distribution functions — the type associated with discrete random variables — or all continuously differentiable distribution functions – the type associated with continuous random variables possessing continuous densities with respect to dx or Lebesgue measure. In this more general setting, the so-called **mixed-type** distribution functions (3) have at least one pure-jump component distribution function $F_j(x) \;=\; \sum_{t:t\le x} I_{[p_j(t)>0]} \, p_j(t)$ and at least one differentiable component distribution $F_k(x) \;=\; \int_{\infty}^{x} f_k(x)\,dx$, and such distribution functions have points t to which they assign positive probability $F(t) - F(t-) \;=\; F(t) - \lim_{h\to 0+} F(t-h)$, but also have intervals (a,b) on which they are strictly increasing.

It is clear that a mixture distribution function (3) has *mixed type* — some jump points and some intervals of everywhere strict increase — whenever at least one of the F_j corresponds to a discrete random variable and at least one to a random variable with continuous density f_j. Conversely, any distribution function $G(x)$ with the property that for every point x

either G is continuously differentiable at x, or $G(x) - G(x-) > 0$

can be represented (for $K \le \infty$) in the form

$$G(x) \;=\; \sum_{k=1}^{K} q_k \, I_{[x_k \le x]} \;+\; (1-p) \int_{\infty}^{x} g_0(s)\,ds$$

where $\sum_{k=1}^{L} q_k = p \in (0,1)$, and where g_0 is a density function which is piecewise continuous between the consecutive jump-points x_k if either K is finite or the jump-points x_k are ordered monotonically increasing on the line. In this case, it is easy to see that $G = p\,F_1 + (1-p)\,F_2$ is a two-component mixture of the distribution function $F_1(x) \;=\; \sum_{k=1}^{K} (q_j/p)\, I_{[x_k \le x]}$ and the differentiable distribution function $F_2(x) = \int_{-\infty}^{x} g_0(t)\,dt$ with density g_0.