Making sense of . . .

Mixture distributions

Excerpt #1 from Jim Daniel's LTAM seminars

James W. Daniel
Jim Daniel's Actuarial Seminars
www.actuarialseminars.com

February 21, 2020

Foreword

This document briefly describes mixture distributions and how to compute with them. Essentially it's part of one of the lessons I gave in my face-to-face LTAM exam-prep seminars. While mixture distributions appear far less often on Exam LTAM than on Exam STAM, the Mixing Method makes it simple to solve the problems that do appear (often in models similar to those in the Examples presented here).

NOTE: The numbered items—Examples, Equations, Definitions, Sections, *et cetera*—can be reached using a modern pdf reader or browser by clicking on the number in the reference. For example, if some text mentions Equation 1.4, if you click on the 1.4 you should be taken directly to that equation; some readers and browsers will return you to where you were if you click on the Back arrow.

Chapter 1

Mixture distributions and calculations

1.1 Mixtures and the Mixing Method

I'll start with a simple example to illustrate the ideas.

Example 1.1 Imagine a box full of fair dice; 20% of the dice are four-sided with the faces numbered one through four, and 80% are six-sided with the faces numbered one through six. If you repeatedly roll a four-sided die, you'll be seeing values N from a Uniform random variable on the integers 1, 2, 3, 4, all equally likely. And if you repeatedly roll a six-sided die, you'll be seeing values from a Uniform random variable on the integers 1, 2, 3, 4, 5, 6, all equally likely.

But suppose that you reach into the box, grab a die at random (so a 20% chance of grabbing a four-sided die), roll it once, record the number, and return the die to the box. And then you repeat that process over and over, grabbing and rolling a random die each time. The results N you will see are not from a Uniform random variable on 1, 2, 3, 4, nor from a Uniform random variable on 1, 2, 3, 4, 5, 6. Rather, they are from a 20/80 ***mixture*** of those two distributions—on average, 20% of the time N is Uniform on 1, 2, 3, 4 and 80% of the time it's Uniform on 1, 2, 3, 4, 5, 6.

A more formal way of viewing the above situation is to say that there is a random variable S, the number of sides on the die, with $\Pr[S=4]=0.2$ and $\Pr[S=6]=0.8$. The *conditional* random variable $N\big|S$ describing the result of a roll of a die, given the value of S, is a (discrete-type) Uniform random variable on the values $1,2,\ldots,S$. We're interested in the *unconditional* distribution of N—that is, in the behavior of the mixture distribution N that is the result of a random roll of a random die.

Suppose that you want to compute the probability of rolling a 3 with a randomly chosen die—that is, $\Pr[N=3]$ where N follows the mixture distribution. The rigorous mathematical computation of this proceeds as follows:

$$\begin{aligned}
\Pr[N=3] &= \Pr[N=3 \text{ and } S=4]+\Pr[N=3 \text{ and } S=6] \\
&= \Pr[S=4]\Pr[N=3 \,|\, S=4]+\Pr[S=6]\Pr[N=3 \,|\, S=6] \\
&= (0.2)\left(\frac{1}{4}\right)+(0.8)\left(\frac{1}{6}\right)=\frac{11}{60}=0.18333.
\end{aligned}$$

The more intuitive Mixing Method reasons as follows: 20% of the time the answer is $1/4$ and 80% of the time it's $1/6$, so on average the answer is $(0.2)(1/4)+(0.8)(1/6)=11/60$. That is, pretend you know which case you're in and compute the answers [1/4 and 1/6 in this case], and then compute the expected value of those answers as the cases vary randomly [so $(0.2)(1/4)+(0.8)(1/6)=11/60$

as above]. Let's use this approach to compute the expected value of N. If you knew you had a four-sided die, the expected value would be 2.5, while it would be 3.5 for a six-sided die—on average 20% of the time the mean is 2.5 and 80% of the time it's 3.5. So what's the expected value for a random die? Just the expected value of those two answers as the die-type varies randomly: $(0.2)(2.5) + (0.8)(3.5) = 3.3$.
End of Example 1.1

For the Mixing Method used in Example 1.1 to be valid, the quantity you are computing for the mixture distribution must involve only a ***linear*** computation with the probability function (or density function for a continuous-type random variable) of the mixture—that is, no squaring of probabilities, no dividing by probabilities, et cetera. This is certainly true for an unconditional probability for the mixture, as just illustrated with $\Pr[N = 3]$ above. It's also valid for the expected value, which is just a sum of probabilities times values of the variable. Likewise for, say, the second moment, since that is just a sum of probabilities times squared values of the variable; in Example 1.1, for example, $\mathrm{Mom}_2[N] = (0.2)(30/4) + (0.8)(91/6) = 409/30 = 13.6333$ since $30/4$ is the second moment for the four-sided die and $91/6$ is the second moment for the six-sided die. *But not so for the variance,* since the variance involves the square of the mean and thus the squaring of probabilities. In Example 1.1, for instance, $\mathrm{Var}[N]$ would be correctly computed as the mixture second moment $409/30$ minus the square of the mixed mean 3.3: $\mathrm{Var}[N] = 409/30 - 3.3^2 = 2.7433$. Mixing the two variances $15/12$ and $35/12$ in the 20/80 proportion gives the incorrect value 2.5833.

KEY ⇒ **Fact 1.2 (Mixing Method)** *Suppose you want to compute a quantity for a mixture distribution, and that quantity involves only linear computations with the probability function or density function of the mixture distribution. Then that quantity can be computed using the Mixing Method:*

1. *Pretend that you know which case of the mixture holds, and compute the quantity; and*
2. *compute the expected value of the quantity as the case varies randomly to get the final result.*

You already saw this method applied in Example 1.1 and its following paragraph. Here are some more examples.

Example 1.3 Suppose that, given the value of ω, a certain random variable X is a Uniform random variable on $(0, \omega)$ but that ω itself is random, with $\Pr[\omega = 4] = 0.2$ and $\Pr[\omega = 6] = 0.8$. That is, on average 20% of the time X is Uniform on $(0, 4)$ and 80% of the time it's Uniform on $(0, 6)$. You want to understand the behavior of the mixture distribution X defined by random values of the variable coming from random values of ω.

More formally, you know that the conditional random variable $X|\omega$ is Uniform on $(0, \omega)$ and that ω is the random variable described above, and you want to understand the behavior of the unconditional random variable X—that is, the behavior of the mixture distribution. The Mixing Method in Key Fact 1.2 lets you do so.

Suppose, for example, that you want to compute the mean $\mathrm{E}[X]$ of X. Use the Mixing Method. Since the mean of a Uniform random variable on an interval is the midpoint of the interval, if you pretend you know $\omega = 4$ you get a mean of $\frac{0+4}{2} = 2$, while you get $\frac{0+6}{2} = 3$ if you pretend you know $\omega = 6$; so on average 20% of the time the answer is 2 and 80% of the time it's 3. [That is, pretending you know ω gives a mean of $\frac{0+\omega}{2} = \frac{\omega}{2}$.] Then you compute the expected value as your two answers vary—$\mathrm{E}[X] = 0.2(2) + 0.8(3) = 2.8$. [That is, $\mathrm{E}\left[\frac{\omega}{2}\right] = 0.2(\frac{4}{2}) + 0.8(\frac{6}{2}) = 2.8$.]

Suppose next that you want to compute $\Pr[X \le 3]$. You can use the Mixing Method since this probability is linear in the density function for X—it's just the integral of that density from 0 to 3. Pretending you know ω gives a probability of $\frac{3-0}{\omega-0} = \frac{3}{\omega}$. The desired probability is the the expected value of these answers, so $\Pr[X \le 3] = \mathrm{E}\left[\frac{3}{\omega}\right] = 0.2(\frac{3}{4}) + 0.8(\frac{3}{6}) = 0.55$.

But you need to be careful—as you saw previously—in computing the variance $\mathrm{Var}[X]$ of X. You use the Mixing Method to get the second moment $\mathrm{E}[X^2]$ and subtract the square of the mean 2.8. Do this; you should get 2.8267.
End of Example 1.3

Example 1.4 Suppose that, given the value of ω, a certain random variable is a Uniform random variable on $(0, \omega)$ but that ω itself is a Uniform random variable on $(6, 18)$. That is, you know that the conditional random variable $X|\omega$ is Uniform on $(0, \omega)$ and that ω is Uniform on $(6, 18)$. You want to understand the behavior of the mixture distribution X defined by random values of the variable coming from random values of ω.

Suppose, for example, that you want to compute the mean $\mathrm{E}[X]$ of X. Use the Mixing Method. Pretending you know ω gives a mean of $\frac{\omega}{2}$. Then you compute the expected value as your answers vary—$\mathrm{E}\left[\frac{\omega}{2}\right] = \frac{\mathrm{E}[\omega]}{2} = 6$.

Suppose next that you want to compute $\Pr[X \leq 3]$. Pretending you know ω gives a probability of $\frac{3}{\omega}$. The desired probability is the the expected value of these answers, so

$$\Pr[X \leq 3] = \mathrm{E}\left[\frac{3}{\omega}\right] = \int_6^{18} f_\omega(z)\,\frac{3}{z}\,dz = \int_6^{18} \frac{1}{18-6}\,\frac{3}{z}\,dz = 0.27465.$$

End of Example 1.4

I think of each of the preceding three examples as *explicit* mixtures: you were explicitly told that a certain random variable depended on a parameter that was itself random. Sometimes you'll encounter *implicit* mixtures in which the mixture nature is less obvious; here's a simple example now.

Example 1.5 Suppose that the random variable X has density function f_X given by

$$f_X(z) = e^{-4z} + 6e^{-8z}, \qquad z > 0$$

and you need to compute expectations or probabilities involving X. You might wish that the density had been proportional to only e^{-4z} or to only e^{-8z}—you should recognize them as densities for Exponential random variables with means $\frac{1}{4}$ and $\frac{1}{8}$ [or, equivalently, as densities for future-lifetime random variables with forever constant forces of mortality 4 and 8]. In fact, f_X can be written as 0.25 times one of those densities plus 0.75 times the other:

$$f_X(z) = 0.25\left(4e^{-4z}\right) + 0.75\left(8e^{-8z}\right).$$

So what? Suppose you had been told explicitly that some random variable X is a 25/75 mixture of two distributions—on average 25% of the time the density is $4e^{-4z}$ and 75% of the time it's $8e^{-8z}$. Since the density function certainly depends linearly on the density function, you can use the Mixing Method to get the density for the mixture distribution—and you'll get precisely the density in this Example! So you could also use the Mixing Method to compute, say, $E[X]$ as $E[X] = 0.25\left(\frac{1}{4}\right) + 0.75\left(\frac{1}{8}\right) = 0.15625$. Noticing that X is a mixture distribution and using the Mixing Method for $E[X]$ is certainly simpler than performing the two integrations by parts necessary to evaluate the mean as $\int_0^\infty f_X(z)\,z\,dz$.
End of Example 1.5

1.2 Ogive distributions as mixtures

You probably are familiar with the use of an *ogive* to estimate the cumulative distribution function of a random variable based on grouped data; if not, you might take a look at my LTAM Seminar Excerpt *Empirical estimation with complete grouped data.* I'll just briefly sketch the setup here.

In general, you're told that for a random variable X there were n_1 observations of X between c_0 and c_1, n_2 between c_1 and c_2, and so on up through n_k between c_{k-1} and c_k, where $c_0 < c_1 < c_2 <$

$\cdots < c_k$ and the total number of observation is n. The empirical cumulative distribution function F_n at a group endpoint c_j is just the fraction of observations up through c_j, so

$$F_n(c_j) = \frac{n_1 + n_2 + \cdots + n_j}{n}.$$

Between consecutive group endpoints, F_n, the ogive, is assumed to be a linear function interpolating the values at the group endpoints, and the empirical random variable $\widehat{X}$ is that random variable whose cumulative distribution function equals F_n. Since the cumulative distribution function F_n is linear between group endpoints, the density function must be constant between group endpoints, and a constant density function is associated with a Uniform random variable. That is, when $\widehat{X}$ falls within a particular group, its values are uniformly distributed there; since there are k different groups, there are k different Uniform random variables—some of the time you have one particular Uniform random variable, and some of the time you have another, which sure sounds like a mixture of Uniforms!

Example 1.6 Suppose that you observe 10 values of X: none were below 0, 2 were between 0 and 10, 3 were between 10 and 50, 5 were between 50 and 100, and none were beyond 100. Then the empirical random variable $\widehat{X}$ defined by the ogive is just a mixture:

1. 20% of the time [since 2 of the 10 values fell between 0 and 10] $\widehat{X}$ is a Uniform random variable on $[0, 10]$,

2. 30% of the time [since 3 of the 10 values fell between 10 and 50] $\widehat{X}$ is a Uniform random variable on $[10, 50]$, and

3. 50% of the time [since 5 of the 10 values fell between 50 and 100] $\widehat{X}$ is a Uniform random variable on $[50, 100]$.

So you could use the Mixing Method to compute some quantity for $\widehat{X}$ as an estimate of the corresponding quantity for X. For example, you could estimate $\mathrm{E}[X]$ by $\mathrm{E}[\widehat{X}]$, where $\mathrm{E}[\widehat{X}]$ is computed by the Mixing Method as follows. If you pretend you know that $\widehat{X}$ is a Uniform random variable on $[0, 10]$, then its expected value would be 5; if you pretend you know that $\widehat{X}$ is a Uniform random variable on $[10, 50]$, then its expected value would be 30; and if you pretend you know that $\widehat{X}$ is a Uniform random variable on $[50, 100]$, then its expected value would be 75; then, the true expected value of $\widehat{X}$ is the expected value of those values as the case varies randomly, namely: $\mathrm{E}[\widehat{X}] = (0.2)(5) + (0.3)(30) + (0.5)(75) = 47.5$.
End of Example 1.6