

Making sense of . . .

Mixture distributions
in Exams MLC/3L and C/4

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

February 1, 2012

©Copyright 2012 by James W. Daniel; reproduction in whole or in part without the express permission of the author is forbidden.

Foreword

This document briefly describes the ideas behind the use of mixture distributions in some of the material for Exams MLC and C of the Society of Actuaries and Exams 3L and 4 of the Casualty Actuarial Society. Not a traditional exam-prep study manual, it concentrates on explaining key ideas so that you can then understand the details presented in the textbooks or study manuals. In order to conserve space, rather than containing problems it instead lists problems for practice that can be downloaded from the SoA website starting (as of this date) at <http://www.soa.org/education/resources/edu-multiple-choice-essay-examinations.aspx> and from the CAS website starting at <http://www.casact.org/admissions/studytools/>.

One piece of notation I should mention. This note is full of Examples, and each example ends with the symbol ¶ at the left in a line by itself.

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 $N|S$ 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, ..., 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 are 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 [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 is the expected value for a random die? Just the expected value of those two answers as the die-type varies: $(0.2)(2.5) + (0.8)(3.5) = 3.3$.

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, $\text{Mom}_2[N] = (0.2)(30/4) + (0.8)(91/6) = 409/30 = 13.6333$. *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, $\text{Var}[N]$ is correctly computed as the mixture second moment $409/30$ minus the square of the mixed mean 3.3: $\text{Var}[N] = 409/30 - 3.3^2 = 2.7433$. Mixing the two variances $15/12$ and $35/12$ in the $20/80$ proportion gives the incorrect value 2.5833.

KEY \Rightarrow **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.*

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 $E[X]$ 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— $E[X] = 0.2(2) + 0.8(3) = 2.8$. [That is, $E\left[\frac{\omega}{2}\right] = 0.2\left(\frac{4}{2}\right) + 0.8\left(\frac{6}{2}\right) = 2.8$.]

Suppose next that you want to compute $\Pr[X \leq 3]$. 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 $\Pr[X \leq 3] = E\left[\frac{3}{\omega}\right] = 0.2\left(\frac{3}{4}\right) + 0.8\left(\frac{3}{6}\right) = 0.55$.

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

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 $E[X]$ 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— $E\left[\frac{\omega}{2}\right] = \frac{E[\omega]}{2} = 6$.

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

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

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. In Chapter 3 you'll see a variety of *implicit* mixtures in which the mixture nature is less obvious, but here's a simple example now.

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

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

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} —MLC/3 candidates should recognize those as densities for future-lifetime random variables with constant forces of mortality of 4 and 8, while Exam C/4 candidates should recognize them as densities for Exponential random variables with means $\frac{1}{4}$ and $\frac{1}{8}$. In fact, f_X can be written as 0.25 times one of those densities plus 0.75 times the other:

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

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 $\int_0^\infty f_X(z) z dz$.

Problems 1.1

[See my Foreword on page 2 for the web links.]

From the SoA Exam MLC/M/3 archives: Fall 2006 #39; Fall 2005 #16; Spring 2000 #17.

From the CAS Exam 3 archives: Fall 2005 #32.

Chapter 2

Explicit mixtures

2.1 Exam MLC/3L explicit mixtures

I'll illustrate some types of explicit mixture distributions encountered on Exam MLC/3 with examples involving survival probabilities and expectations, single benefit premiums (also called actuarial present values, net single premiums, and expected present values), annual benefit premiums, and Poisson processes.

Example 2.1 (survival-model probabilities and expectations) Since probabilities in survival-model calculations are given by integrals of the density function for the future-lifetime random variable, just as for expectations they are linear in the density function and so the Mixing Method can be used.

For example, suppose that 20% of 40-year-olds are smokers with constant force of mortality $\mu^S = 0.05$ while 80% are non-smokers with constant force of mortality $\mu^N = 0.02$. Compute both ${}_{10}p_{40}$ and \dot{e}_{40} for a random 40-year-old.

Pretending that you know the value of the constant force μ , you compute the survival probability as $e^{-10\mu}$ and the complete expectation of life as $\frac{1}{\mu}$. The Mixing Method computes the values for random 40-year-olds as the expected values of these answers. That is,

$${}_{10}p_{40} = E[e^{-10\mu}] = 0.2e^{-10 \times 0.05} + 0.8e^{-10 \times 0.02} = 0.77629,$$

and

$$\dot{e}_{40} = E\left[\frac{1}{\mu}\right] = 0.2\left(\frac{1}{0.05}\right) + 0.8\left(\frac{1}{0.02}\right) = 44.$$

Example 2.2 (single benefit premiums) Since single benefit premiums can be computed as the integral of the density function for future lifetime times the formula for the present-value-of-benefit random variable, they are linear in the density function and so the Mixing Method can be used.

For example, suppose that 20% of 40-year-olds are smokers with constant force of mortality $\mu^S = 0.05$ while 80% are non-smokers with constant force of mortality $\mu^N = 0.02$. Compute both \bar{A}_{40} and \bar{a}_{40} for a random 40-year-old, using $\delta = 0.1$.

Pretending that you know the value of the constant force μ , you compute the single benefit premium for the insurance as $\frac{\mu}{\mu + \delta}$ and for the annuity as $\frac{1}{\mu + \delta}$. The Mixing Method computes the values for random 40-year-olds as the expected values of these answers. That is,

$$\bar{A}_{40} = E\left[\frac{\mu}{\mu + \delta}\right] = 0.2\frac{0.05}{0.05 + 0.1} + 0.8\frac{0.02}{0.02 + 0.1} = 0.2,$$

and

$$\bar{a}_{40} = E\left[\frac{1}{\mu + \delta}\right] = 0.2\frac{1}{0.05 + 0.1} + 0.8\frac{1}{0.02 + 0.1} = 8.$$

Alternatively, after finding $\bar{A}_{40} = 0.2$ you could simply compute \bar{a}_{40} as $\bar{a}_{40} = \frac{1-\bar{A}_{40}}{\delta} = \frac{1-0.2}{0.1} = 8$.

Example 2.3 (annual benefit premiums) Annual benefit premiums are computed as the quotient of the single benefit premium for the benefit and the single benefit premium for the premium paying pattern; for example, $\bar{P}(\bar{A}_x) = \frac{\bar{A}_x}{\bar{a}_x}$. You saw in Example 2.2 that both the numerator single benefit premium and the denominator single benefit premium are linear in the density function, which means that the quotient is **not** linear in the density function and so the Mixing Method **cannot** be used directly on the premium rates.

For example, suppose that 20% of 40-year-olds are smokers with constant force of mortality $\mu^S = 0.05$ while 80% are non-smokers with constant force of mortality $\mu^N = 0.02$, and that $\delta = 0.1$. In Example 2.2 you saw that $\bar{A}_{40} = 0.2$ and $\bar{a}_{40} = 8$ for a random 40-year-old. Therefore, for a random 40-year-old the annual benefit premium is $\bar{P}(\bar{A}_x) = \frac{\bar{A}_x}{\bar{a}_x} = \frac{0.2}{8} = 0.025$.

Note that if you computed the annual rate for a smoker, you'd get 0.05, while for a non-smoker you'd get 0.02; if—**incorrectly**—you applied the Mixing Method to these two values, you'd produce the incorrect value $(0.2)(0.05) + (0.8)(0.02) = 0.026$ for the annual benefit premium for a random 40-year-old.

Example 2.4 (Poisson processes) [NOTE: This material is for CAS Exam 3L; it is no longer on the MLC syllabus.] I'm going to cop out a little here and refer you to more detail on mixtures and Poisson processes in my free study note "Poisson processes (and mixture distributions)" that you can download for free from the Jim Daniel's Study Note section of my website <http://www.actuarialseminars.com>, especially numbered items 1.19 through 1.24. But here's part of one example from that note.

Suppose that, given the value of λ , N is a homogeneous Poisson process with rate λ , but that λ itself is random, namely a Uniform random variable on $[1, 1.5]$. Let's use the Mixing Method to find $E[N(2)]$ for the mixed Poisson process—that is, for random values of $N(2)$ from random λ 's.

If we pretend that we know λ , then N is a Poisson process with rate λ and so $N(2)$ —which equals the increment $N(2) - N(0)$ —is a Poisson random variable M with mean $\Lambda = \int_0^2 \lambda dz = 2\lambda$. Since λ 's values are uniformly distributed over $[1, 1.5]$, it's clear that Λ 's values, which are double λ 's, are uniformly distributed over $[2, 3]$. That is, $M | \Lambda$ is a Poisson random variable with mean Λ while Λ is a Uniform random variable on $[2, 3]$.

If we pretend we know the value of Λ , then $E[M | \Lambda] = \Lambda$. We get the mean for the mixture distribution by the Mixing Method—taking the expectation of this value as Λ varies:

$$E[M] = E[\Lambda] = \frac{2+3}{2} = 2.5.$$

So $E[N(2)] = 2.5$.

Problems 2.1

[See my Foreword on page 2 for the web links.]

From the SoA Exam MLC/M/3 archives: Fall 2006 #5, 17, 39; Fall 2005 #16, 20, 32; Fall 2004 #2, 4; Fall 2003 #15, 18; Fall 2001 #17, 27; Spring 2001 #28; Spring 2000 #8, 17, 35.

2.2 Exam C/4 explicit mixtures

Note that the broad Exam C/4 syllabus areas of Bayesian models and of credibility all involve explicit mixtures—random quantities of interest depend on risk parameters that are themselves random. I'll illustrate some types of explicit mixture distributions encountered on Exam C/4 with examples I describe as Poisson-Gamma, Exponential-Inverse Gamma, and Normal-Normal.

Example 2.5 (Poisson-Gamma) A key model in the Exam C/4 material assumes that $N | \Lambda$ is a Poisson random variable with mean Λ and that Λ is a Gamma random variable with parameters α and θ , all in the notation of the exam tables and the *Loss Models* textbook. A key fact you need for Exam C/4 is that this makes the unconditional distribution of N —that is the mixture distribution—a Negative Binomial random variable with parameters r and β with $r = \alpha$ and $\beta = \theta$. In particular, this makes $E[N] = r\beta = \alpha\theta$.

You could also compute $E[N]$ by the Mixing Method. Pretending Λ was known rather than random, you'd compute the mean as Λ . The Mixing Method would then find $E[N]$ as the expected value of that result, namely $E[\Lambda] = \alpha\theta$.

Example 2.6 (Exponential-Inverse Gamma) A less common model in the Exam C/4 material assumes that $X | \Theta$ is an Exponential random variable with mean Θ and that Θ is an Inverse Gamma random variable with parameters α and θ , all in the notation of the exam tables and the *Loss Models* textbook. A fact from the Exam C/4 syllabus is that this makes the unconditional distribution of X —that is the mixture distribution—a two-parameter Pareto random variable with parameters $\tilde{\alpha}$ and $\tilde{\theta}$ with $\tilde{\alpha} = \alpha$ and $\tilde{\theta} = \theta$. In particular, this makes $E[X] = \frac{\tilde{\theta}}{\tilde{\alpha}-1} = \frac{\theta}{\alpha-1}$.

You could also compute $E[X]$ by the Mixing Method. Pretending Θ was known rather than random, you'd compute the mean as Θ . The Mixing Method would then find $E[X]$ as the expected value of that result, namely $E[\Theta] = \frac{\theta}{\alpha-1}$.

Example 2.7 (Normal-Normal) An even less common model in the Exam C/4 material assumes that $X | \Theta$ is a Normal random variable $\mathcal{N}(\Theta, v)$ with mean Θ and known variance v and that Θ is a Normal random variable $\mathcal{N}(\mu, a)$ with known mean μ and known variance a . A fact from the Exam C/4 syllabus is that this makes the unconditional distribution of X —that is the mixture distribution—a Normal random variable $\mathcal{N}(\mu, v + a)$ with mean μ and variance $v + a$.

You could also compute $E[X]$ by the Mixing Method. Pretending Θ was known rather than random, you'd compute the mean as Θ . The Mixing Method would then find $E[X]$ as the expected value of that result, namely $E[\Theta] = \mu$.

Problems 2.2

[See my Foreword on page 2 for the web links.]

From the SoA Exam MLC/M/3 archives: Fall 2005 #17; Spring 2005 #10, 34; Fall 2004 #32; Fall 2002 #5; Spring 2001 #3; Fall 2000 #13; Spring 2000 #4.

From the SoA Exam C/4 archives: Spring 2007 #3; Fall 2006 #9; Fall 2004 #13.

From the CAS Exam 3 archives: Fall 2006 #19, 20; Spring 2005 #10; Fall 2004 #29.

Chapter 3

Implicit mixtures

3.1 Exam 3L implicit mixtures

Other than implicit mixtures as in Example 1.5, about the only implicit mixtures you encounter in Exam 3L are compound Poisson processes. That is, you are interested in

$$S(t) = \sum_{j=1}^{N(t)} X_j,$$

where N is a Poisson process, all the X_j have the same distribution as a given random variable X , and all the X_j and $N(t)$ form an independent set. Since, for each t , $N(t)$ is a Poisson *random variable*, it follows that $S(t)$ for each t is a compound Poisson *random variable*. Since I treat compound random variables as mixture distributions in Example 3.1 in the following section, I suggest you take a look at that example.

3.2 Exam C/4 implicit mixtures

I'll illustrate some types of implicit mixture distributions encountered on Exam C/4 with examples involving compound random variables, individual risk models, and kernel smoothing.

Example 3.1 (compound random variables) A key Exam C/4 model is the compound random variable S , where

$$S = \sum_{j=1}^N X_j,$$

with all the random variables X_j having the same distribution as a single random variable X , and with the random variable N and all the X_j 's forming an independent set. [This also arises in Exam MLC/3, with the random variable N replaced by the random variable $N(t)$ for a Poisson process N .] I'll use the Mixing Method to derive the crucial formulas for $E[S]$ and $\text{Var}[S]$ —formulas that arise so often you should memorize them. And I'll also use it to derive the formula (rarely needed on exams) for the density function f_S of S [or the probability function f_S if X is a discrete type random variable].

S can be viewed as a mixture since for each random value of N , say $N = n$, S becomes just

$$S_n = \sum_{j=1}^n X_j,$$

the sum of a fixed non-random number n of X_j 's. Thus S is a mixture of the variables S_0, S_1, S_2, \dots .

Now for the Mixing Method. Pretending that I know the value of N —call it n , say—I can easily compute the expectation $E[S_n] = n E[X]$ and variance $\text{Var}[S_n] = n \text{Var}[X]$; and I can less easily compute the density (or probability) function f_{S_n} , which the *Loss Models* textbook denotes by f_X^{*n} , by recursion (see the textbook) from $f_X^{*(n-1)}$.

Since I now know that $E[S | N] = N E[X]$ and the Mixing Method can be applied to obtain means, I get $E[S]$ as the mean of these answers: $E[S] = E[N E[X]] = E[N] E[X]$, the key formula for the mean of a compound random variable.

The Mixing Method of course cannot be applied directly to obtain the variance of the mixture S ; instead I apply the Mixing Method to obtain the second moment $E[S^2]$ and obtain the variance as $E[S^2] - E[S]^2$. To find $E[S^2]$ by the Mixing Method, I first find it pretending N is fixed and then take the expected value of those answers as N varies. But

$$E[S^2 | N] = \text{Var}[S | N] + E[S | N]^2 = N \text{Var}[X] + \{N E[X]\}^2.$$

Applying the Mixing Method by taking the mean of this as N varies gives

$$E[S^2] = E[N] \text{Var}[X] + E[N^2] E[X]^2.$$

Subtracting from this the formula for $E[S]^2$ —that is, the square of the key formula for $E[S]$ —and recalling that $E[N^2] - E[N]^2$ is just $\text{Var}[N]$ gives

$$\text{Var}[S] = E[N] \text{Var}[X] + \text{Var}[N] E[X]^2,$$

the key formula for the variance of a compound random variable.

Finally, consider the density function (or probability function) f_S . Since the density function certainly is linear in the density function, the Mixing Method can be applied. Pretending we know N , we found the function to be f_X^{*N} . All we need do next to take the expected value of this as N varies. If we let p_k denote the probabilities for N , $p_k = \Pr[N = k]$, then this expected value f_S is simply $f_S = p_0 f_X^{*0} + p_1 f_X^{*1} + p_2 f_X^{*2} + \dots$.

Example 3.2 (individual risk models) The compound random variable of the preceding example can be viewed as representing total insurance claims as the sum, over the random number of people having claims, of the random amount of each claim. An alternative approach is the individual risk model that views total insurance claims as the sum, over the known non-random number of people insured, of the random amount (including zero) of each claim. That is, if n is the number of people insured, we consider $S_n = \sum_{j=0}^n X_j$. In the *Loss Models* textbook, the X_j 's are assumed to form an independent set but not necessarily to all have the same distribution. Instead, each X_j is assumed to be of the form $X_j = I_j B_j$, where each I_j can be only 0 or 1 (a Bernoulli random variable, that is, a Binomial random variable with one trial) and each severity B_j has some known distribution. Since $E[S_n] = \sum_{j=1}^n E[X_j]$ and $\text{Var}[S_n] = \sum_{j=1}^n \text{Var}[X_j]$, to compute the mean and variance of S_n requires only a method to find the mean and variance of each X_j . I'll use the Mixing Method to find the mean and variance of each such variable of the form $X = IB$, where I 's probabilities are given by $p = \Pr[I = 1]$ and $1 - p = \Pr[I = 0]$.

Why can X be viewed as a mixture? Since I is either 0 or 1, X is either 0 or B , depending on the value of I . If we pretend we know that $I = 0$, then the mean and second moment of $X | I = 0$, which is just 0, are both 0. On the other hand, if we pretend we know that $I = 1$, then the mean and second moment of $X | I = 1$, which is just B , are just the mean $E[B]$ and second moment $E[B^2]$ of B . The Mixing Method then gives the mean and second moment of the mixture X as the expected values as I varies of the means and of the second moments. That is, $E[X] = (1-p)0 + p E[B] = p E[B]$ and $E[X^2] = (1-p)0 + p E[B^2] = p E[B^2]$. The formula for $\text{Var}[X]$ then becomes $\text{Var}[X] = E[X^2] - E[X]^2 = p E[B^2] - \{p E[B]\}^2$. If you prefer formulas involving variances, you should be able to see that this is the same as $\text{Var}[X] = p \text{Var}[B] + p(1-p) E[B]^2$.

KEY \Rightarrow **Example 3.3 (kernel smoothing)** The method of kernel smoothing (or kernel density) is a statistical estimation procedure for estimating various quantities for a continuous-type random variable X based on n observations of individual values x_1, x_2, \dots, x_n of X . [For example, the observations might be 3, 5, 5, 8, 9.] The method starts with the empirical random variable \widehat{X} approximating X from those observations and “smooths” it to obtain another approximation to X I’ll denote by \widetilde{X} . To estimate anything to do with X , you calculate it exactly for \widetilde{X} .

Recall the definition of the random variable \widehat{X} . Roughly speaking, it is the discrete-type random variable whose only possible values are the observed x_j , with each x_j having equal probability $\frac{1}{n}$; to deal with repeated observations, the precise definition is as follows. Let $y_1 < y_2 < \dots < y_k$ be the *distinct* observations of X , and let s_j equal the number of observations of X that equal y_j . [If the observations are 3, 5, 5, 8, 9 as above, then the y_j ’s are 3, 5, 8, 9 and the s_j ’s are 1, 2, 1, 1, respectively.] Then \widehat{X} is precisely defined as the discrete-type random variable whose only possible values are the k y_j ’s, with $\Pr[\widehat{X} = y_j] = \frac{s_j}{n}$; I’ll denote that probability by p_j : $p_j = \frac{s_j}{n}$. [Continuing my numerical example, I see that \widehat{X} can only equal 3, 5, 8, or 9, with p_j probabilities 0.2, 0.4, 0.2, and 0.2, respectively.]

I’m now going to think of \widehat{X} as a mixture distribution; this may seem an overly complicated way to think of a simple discrete-type random variable, but bear with me for a bit. [In my numerical example, on average 20% of the time \widehat{X} is 3, 40% of the time it’s 5, 20% of the time it’s 8, and 20% of the time it’s 9.] Let Y_j denote the not-very-random random variable whose only possible value is y_j —that is, $\Pr[Y_j = y_j] = 1$. Then \widehat{X} is nothing other than a mixture of all the Y_j ’s, using the probabilities p_j , respectively. [In my numerical example, on average 20% of the time \widehat{X} is Y_1 , 40% of the time it’s Y_2 , 20% of the time it’s Y_3 , and 20% of the time it’s Y_4 .] More formally, think of J as a random variable with $\Pr[J = j] = p_j$; then \widehat{X} is just the unconditional distribution of Y_J —the random value of Y_J for random values of J .

Thus empirical estimation uses a discrete-type random variable $\widehat{X} = Y_J$ to approximate a continuous-type random variable X . It seems more reasonable to approximate by a continuous-type random variable. To do so, just replace each not-very-random discrete-type random variable Y_j that equals y_j with probability 1 by a continuous-type random variable \widetilde{Y}_j that is on average equal to y_j , that is, with $E[\widetilde{Y}_j] = y_j$. For example, the “Uniform kernel with bandwidth b ” makes \widetilde{Y}_j a Uniform random variable on the interval $[y_j - b, y_j + b]$. The kernel-smoothed (or kernel density) approximation \widetilde{X} to X is just the mixture of the \widetilde{Y}_j with probabilities p_j ; that is, \widetilde{X} is just the (unconditional) mixture distribution \widetilde{Y}_J —the random values of \widetilde{Y}_J for random values of J .

This may sound complicated, but the Mixing Method makes computations easy. For example, to estimate a probability for X by that probability for \widetilde{X} , just pretend that you know the value of J , compute that probability for each \widetilde{Y}_j , and then take the expected value of those probabilities as J varies.

For example, suppose the observations are 3, 5, 5, 8, 9 as above and that the Uniform kernel with bandwidth $b = 2$ is used. How do I estimate something for X by kernel smoothing? By computing it exactly for the mixture distribution \widetilde{X} that on average is \widetilde{Y}_1 20% of the time and \widetilde{Y}_2 40% of the time and so on—using the p_j probabilities 0.2, 0.4, 0.2, 0.2 I found above. I know that \widetilde{Y}_1 is Uniform on $[y_1 - b, y_1 + b] = [1, 5]$, that \widetilde{Y}_2 is Uniform on $[3, 7]$, and so on. So on average \widetilde{X} is Uniform on $[1, 5]$ 20% of the time and Uniform on $[3, 7]$ 40% of the time and so on.

Specifically, how do I estimate $\Pr[X \leq 4.6]$, say, by kernel smoothing? I estimate it by computing $\Pr[\widetilde{X} \leq 4.6]$ exactly, and I compute that by using the Mixing Method—taking the expected value as j varies of $\Pr[\widetilde{Y}_j \leq 4.6]$. I know that \widetilde{Y}_1 is Uniform on $[1, 5]$, so $\Pr[\widetilde{Y}_1 \leq 4.6] = \frac{3.6}{4} = 0.9$. Similarly, $\Pr[\widetilde{Y}_2 \leq 4.6] = 0.4$ since \widetilde{Y}_2 is Uniform on $[3, 7]$, while $\Pr[\widetilde{Y}_3 \leq 4.6] = 0$ and $\Pr[\widetilde{Y}_4 \leq 4.6] = 0$ since those variables are Uniform on $[6, 10]$ and $[7, 11]$. Again, on average 20% of the time my value is 0.9 and 40% of the time it’s 0.4 and the other 20% + 20% of the time it’s 0. Then $\Pr[\widetilde{X} \leq 4.6]$ is just the expected value of these values as j varies. Thus the kernel-smoothed estimate of $\Pr[X \leq 4.6]$ is

$\Pr[\tilde{X} \leq 4.6] = (0.2)(0.9) + (0.4)(0.4) + (0.2)(0) + (0.2)(0) = 0.34$. The Mixing Method can be used similarly to estimate other quantities.

The *Loss Models* textbook describes two other choices of the distributions \tilde{Y}_j to mix. The “Gamma kernel” makes each \tilde{Y}_j a Gamma random variable with fixed parameter α for all j and parameter $\theta_j = \frac{y_j}{\alpha}$ so that its mean is y_j . The “triangular kernel with bandwidth b ” makes the density function for \tilde{Y}_j equal 0 to the left of $y_j - b$ and to the right of $y_j + b$, equal $\frac{1}{b}$ at y_j , and be linear in $[y_j - b, y_j]$ and linear in $[y_j, y_j + b]$ —which makes the graph of that density function form a triangle. Probabilities involving a \tilde{Y}_j are integrals of the density function and so are given by areas of parts of these triangles.

Problems 3.2

[See my Foreword on page 2 for the web links.]

From the SoA Exam MLC/M/3 archives: Fall 2004 #17.

From the SoA Exam C/4 archives: Spring 2007 #16; Fall 2006 #24; Fall 2004 #20;

Fall 2003 #4.

From the CAS Exam 3 archives: Spring 2006 #37.