Uncertainties in the Time Swapping Background Estimate: Roman's β Term

James T. Linnemann
Michigan State University and
Los Alamos National Laboratory
April 23, 2003

I study Roman Fleysher's correction in the uncertainty of the background estimate (Nb = $\hat{\alpha}$ Noff) for β, the number of times he re-uses an event in time swapping to estimate α. Roman's calculation is right. I perform an intuitive approximate calculation, and re-derive Roman's result two different ways, using some handy facts about combining distributions. The variance of Nb could perhaps be reduced very slightly by choosing the number of trials more deterministically; I show how, in case it might be more useful in another problem.

1. Introduction and Definitions

1.1 Motivation

Roman's implementation of the time-swapping method for an extended source requires an estimate of α = Ton/Toff. He estimates α by Monte Carlo, roughly by time-swapping each of his off-source events β times. This introduces a new source of uncertainty to the background estimate, so that the variance of the background estimate must be larger than the "standard" answer, α Nb = α^2 Noff. I decided to investigate further because:

- the derivation in Roman's thesis was hard to follow
- some were surprised the new term only had one factor of α, unlike the other term--it seemed to be too large
- the two terms might be comparable in size, so one ought to get it right
- more understanding might even suggest better methods (one can always hope)

The calculations involve distributions whose parameters are themselves random variables. It's a lot easier to calculate second moments than to work directly with full distributions, even with *Mathematica*'s help. I did bring back as a souvenir the interesting fact that a certain combination of Poisson and Binomial distributions leads to a Poisson distribution.

1.1.1 Var[Nb] with No Uncertainty in α

The standard answer for the uncertainty in the background estimate, ignoring any uncertainty in α, is

$$\text{Var[b]} = \text{Var[Nb]} = \alpha^2 n = \alpha\, b = \alpha^2\, \text{Noff} = \alpha\, \text{Nb} \qquad \text{(no uncertainty in } \alpha\text{)}$$

We should certainly never expect to find an estimate for Var[b] less than the above value.

■ 1.1.2 Roman's Answer

In Roman's scheme for background estimation, he firmly excludes the source region from the background estimate. A consistent application of the null hypothesis to the signal estimate, Ns = Non - Nb, assumes that Non and Nb have the same mean, and are uncorrelated, but different of variance:

$$\text{Var[Ns]} = \text{Var}[\mu\text{on}] + \text{Var[Nb]} \rightarrow \text{Nb} + \text{Var[Nb]}, \quad \text{or using Roman's estimated variance for the background,}$$
$$\text{Var[Ns]} = \text{Nb} + \alpha\,\text{Nb} + \text{Nb}/\beta = \text{Nb}(1 + \alpha + 1/\beta).$$

As expected, the uncertainties of the background are of second order: background is still well-estimated, and signal fluctuations dominate the significance calculation, though the $1/\beta$ term reduces the significance by of order 5%

This particular form, even before the inclusion of the β term, is neither Li and Ma's Eq 5 nor Eq 9--in fact Non appears nowhere. However, since the excess is regrettably small, using Li and Ma Eq 5 as the starting point would cause little problem, because Non ≈ Nb.

■ 1.2 Estimating Nb (A Bit More Detail)

Roman starts with a set of off-source events, Noff. He scales up (by re-using) the number of events in the off-source region by a factor ξ, to match the number of events he would have had if he could have measured purely background events over the entire period over which he does an analysis. He does a MC simulation based on replacing the time in these events (with spatial coordinates retained) with another time from the appropriate distribution; this is called time-swapping. In order to get better MC statistics, he does this β times; that is, he time-swaps each event $\gamma=\beta\xi$ times. Then he measures x, how many of these γ Noff time-swapped events fall into the source region. Finally, the estimated background in the source region is b =x/β, which we usually call Nb. Thus, to calculate Var[Nb] requires knowledge of the distribution of x.

■ 1.3 Definitions

For exposition purposes, I will ignore declination, and describe regions as times or angles as seems convenient. I also here elide the dependence of α etc. on local coordinates. I apologize for using a somewhat simplified notation, but it becomes wearing to type Noff in many long formulas; I restore traditional names like Noff and Nb at the end of the derivations.

■ 1.3.1 α and α'

Ton	Source search region: depending on analysis in question, 4 (± 2) to 10 degrees	
Twindow	Region over which signal and background are studied = 8 hours = 120 degrees	
Tx	Source "contaminated" region, eXcluded from background measurement; notice Tx > Ton	= 14 degrees (± 7)
Toff = Twindow - Tx	Off-source region, = 8h - Tx	
α = Ton/Toff	(as usual)	Typically ~ 1/7 in this analysis (you'd guess 10/106 before declination). Actual values of α are higher than nominal: incomplete runs have Twindow < 8h
α' = Tx/Toff		Note $\alpha' > \alpha$; $\alpha' \sim .2$ or so

■ 1.3.2 Variables Related to the Background Estimate

p = Ton/Twindow	Fraction of background events across Twindow which fall into the on region
$p = \alpha/(1+\alpha') = \alpha/\xi$	
$\xi = 1+\alpha' =$ Twindow/Toff	Factor by which background needs to be scaled up
n = Noff	Events from off-source region available for time swapping to estimate α.
	I use n in mid-calculation, and Noff in final results
$\mu = \mu$off	True mean of events in the off-source region.
$\beta = 10$	Chosen (integer) factor of MC oversampling in α estimate.
k	Number of time-swapped events
$\langle k\rangle = \beta\, \xi\, \mu = \gamma\, \mu$	we must understand how k is distributed
$\gamma = \beta\, \xi$	The overall number of times the events are re-used.
x	Number of time-swapped events which land in the on region.
$\langle x\rangle = p\langle k\rangle = \alpha\, \beta\, \mu$	That is, a fraction p of the k generated events land in the on region.

■ 1.3.3 Performing the Background Estimate

b = Nb	the background estimate itself, in my, and standard, notation
	I use b in mid-calculation for ease, and Nb in final results
$b = x/\beta$	how it's related to x, the Monte Carlo output
$b = \alpha n$	another way of expressing the background estimate (needs an α estimate)
$\alpha \to b/n = x/(n\,\beta)$	how alpha was estimated.
$\alpha\, \mu \to b = x/\beta$	from the above definitions; handy to use in final substitutions
$p\, \gamma \to \alpha\, \beta$	another handy substitution

I'll write final answers in terms of n=Noff or b=Nb.

■ 2. Approximate Calculation of Var[Nb]

■ 2.1 x is not a Poisson Variable

Treating x as a simple Poisson variable gives a wrong result. Since $\langle x\rangle = \alpha\, \beta\, \mu$, we'd expect $Var[b] = \langle x\rangle/\beta^2 = \alpha\, \mu\, /\beta \to b/\beta$ after inserting estimates for the parameters. This can't be the right answer: the uncertainty on the background cannot fall to 0 as the number of MC repetitions β becomes large, because n, the number of available background-estimating events, is finite. So we have to work harder, because the distribution of x is *not* a simple Poisson.

■ 2.2 An approximate calculation

Having failed to understand Roman's calculation as described in his thesis, I did an approximate calculation. I start with the basic equation for the background estimate b:

b = α n

In the time swapping method, α must be estimated by a MC integration. Let a be the estimate.

a = x / (n β), which corresponds to the ultimate estimate for the background, b = x / β.

Since we saw x events in the on region, of the k events generated, we can think of the distributions as:

x ~ Binomial(p,k) (that is, x is distributed as a Binomial)

n ~ Poisson(μ)

Now I will do an *approximate* (first order error propagation) variance calculation for b, *ignoring* correlations between a and n. Since a = x/n, that's a big assumption, but separating the two sources of fluctuation is a useful approximation which captures an important part of what's going on. We consider the uncertainty in n separately from the uncertainty of a. We estimate the uncertainty of a by replacing the actual, fluctuating Binomial parameter k, the number of trials, by its expected value, which we hold fixed for this purpose--since k peaks around its expected value.

Var(b) = Var(a n) → Var(a) n^2 + Var(n) a^2 *assuming a and n independent.*

Var(b) = (Var(x) / (nβ)^2) n^2 + Var(n) a^2.

Next we insert the variances of the random variables x and n in terms of the distribution parameters (imagining k fixed):

Var(b) = k p (1-p) / β^2 + μ α^2 → α β μ (1-p) / β^2 + μ α^2 = μ (α^2 + α (1-p)/β) (substituting ⟨k⟩ for k)

Using the estimate μ → n leads to the main result.

■ 2.3 The Approximate Answer

Var(b) = α (1-p) n/ β + n α^2

It is easy to see that with the further approximation Binomial → Poisson, the result would have been simpler still, with p(1-p) → p, raising the variance somewhat. Writing Var(b) in terms of the familiar Noff and Nb, and understanding we would use the estimates for α, I find:

Var(Nb) = α^2 Noff + α(1-α/(1+α')) Noff / β

= α Nb + Nb (1-α/(1+α')) / β (approximate, x ~Binomial uncorrelated with n~Poisson)

→ α^2 Noff + α Noff / β = α Nb + Nb/β (more approximate: two independent Poissons)

Compared to the result α Nb for no uncertainty in α, there is a new term inversely proportional to β, the factor by which events are re-used in the MC estimate of α. The second approximate answer above is identical to the answer Roman arrived at by a more complex calculation. It is gratifying that the approximate calculation seems to have captured the essence of the problem. Further, notice that **Roman's calculation gives the same answer as if *x* and *n* were two independent Poisson variables.**

However, I am not sure whether any of the calculations I present in this note represent the full picture: I am still uneasy with a calculation that says all uncertainty in α disappears with enough MC (big enough β), because it somehow forgets that we have only a finite sample of input events with which to perform time swapping (a finite sample of local coordinates), though this is really the same in both time swapping and direct integration. On the other hand, direct integration makes that same assumption.

■ 3. Exact Calculation of Var[Nb]

■ 3.1 Some Useful Facts About Complex Distributions

3.1.1. A sum of a fixed number of Poisson variables has a Poisson distribution (Stuart & Ord, Kendall's Advanced Theory of Statistics, Vol 1, §11.8, 1994))

$x = \sum_{i=0}^{n} m_i$ $\quad m_i \sim$ Poisson(a) $\quad \Longrightarrow \quad$ x ~ Poisson(a · n)

and more generally,

$x = \sum_{i=0}^{n} m_i$ $\quad m_i \sim$ Poisson(μ_i) $\quad \Longrightarrow \quad$ x ~ Poisson(c), $c = \sum_{i=0}^{n} \mu_i$

3.1.2. Definition of Compound Poisson Distribution (Eadie et al, Statistical Methods in Experimental Physics, North Holland (1971), §4.1.4)

$$x \sim CP(a,b) = \sum_{j=0}^{\infty} \frac{(j\,a)^x\, e^{-a\,j}}{x!} \frac{b^j\, e^{-b}}{j!}$$

⟨x⟩= a b

Var[x] = a b (1 + a)

A useful derivation of the Compound Poisson is given in 3.1.3 below:

3.1.3. A sum of Poisson variables, with the number of terms Poisson-distributed, is distributed as Compound Poisson.

$x = \sum_{i=1}^{j} m_i$, $m_i \sim$ Poisson(a), j ~ Poisson(b), $\Longrightarrow$ x ~ CPoisson(a, b).

This can be seen (Eadie 4.1.4) by using 3.1.1, giving x ~ Poisson(a j) for j fixed. Then substituting into

$P(x) = \sum_{j=0}^{\infty} P(x \mid j, a)\, P(j \mid b)$

yields the form given in 3.1.2 for the Compound Poisson.

3.1.4 A Binomial for a number of successes with a Poisson-distributed number of trials is the same as a Poisson-distributed number of successes. That is: [Rose and Smith, Mathematical Statistics with *Mathematica*, Springer (2002), §3.4B]

x ~ Binomial(p,k) ⊗ k ~ Poisson(a) $\quad \Longrightarrow \quad$ x ~ Poisson(a p)

3.1.5. Two linked Poisson processes are described by the Compound Poisson distribution. (shown in 3.1.3 above)

x ~ Poisson(a n) ⊗ n ~ Poisson(μ) $\quad \Longrightarrow \quad$ x ~ CPoisson(a, μ)

These relations can also be demonstrated by use of generating functions, which are described in all the references above. I've checked them with *Mathematica* and mathstatica (the software which comes with Rose and Smith).

■ 3.4 Exact Calculation for Roman's Model

Roman's model of his background calculation is:
x ~ Binomial(p,k)
k ~ Poisson(γn) Roman re-uses each event k_i ~ Poisson(γ) times, k = $\sum_{i=1}^{n} k_i$. But by 3.1.1, k ~ Poisson(γ n).
n ~ Poisson(μ)
By 3.1.4, the first two can be combined as x ~ Poisson(pγ n); by 3.1.5, this combined with the n ~ Poisson(μ) gives :
x ~ CPoisson(pγ, μ)

Then $\langle x \rangle$ = p$\gamma\mu$ = $\alpha\,\beta\,\mu$ and Var[x] = p$\gamma\mu$ (1+ pγ) = $\alpha\beta\mu$ (1+$\alpha\beta$). Finally,
Var[b] = Var[x]/β^2 = $\alpha\,\mu(\alpha + 1/\beta)$ = b(α + 1/β) = Nb(α + 1/β) as advertised.

While this derivation is delightfully brief, it depends on some non-so-obvious facts. It is also instructive to look at Roman's derivation in something close to its original form. I'll do that in the next section.

■ 3.5 Roman's Derivation

For those of you who read (or wrote!) his thesis, I apologize for not having followed the notation in Roman's: he uses m, M, Nout→ where I use x, k, n; and uses D(x) where I use Var(x). He also uses E[x] for an expectation operator, so Var(x) = E[x^2] - $E[x]^2$. If you read this, I hope you will understand his derivation, and learn something about calculating variances of complex distributions where the parameters of distributions are themselves random variables. His model is :
x ~ Binomial(p,k) k ~ Poisson(γn) n ~ Poisson(μ)

■ 3.5.1 Multi-stage Distributions Can Be Thought of as Expectations

Roman kindly supplied me with a derivation more detailed than in his thesis. The starting point is to write the full distribution for x. It can be built up in stages from the basic rule:

P(x) = $\Sigma_{j=0}^{\infty}$ P(x | j) P(j) = E[P(x|j), j ~ P(j)] (where I write the distribution explicitly)

Roman wrote out those sums in the expectations explicitly to give the distribution of x:

x ~ $\sum_{k=0}^{\infty} \sum_{n=0}^{\infty}$ Binomial($x \mid p, k$) Poisson($k \mid \gamma$n) Poisson($n \mid \mu$)

Actually, his k sum started from x. Since this is not a derivation, but an explanation, suffice it to say you can get by with ignoring this in the calculations below, particularly if you are careful enough in your definition of the functions.

I now outline his proof in terms of expectations, which is equivalent to his manipulation of sums, but opening the door to a more powerful view of his methods. In terms of expectations, the probability can be built up in stages:

x ~ Binomial(p,k) ~ E[Binomial(p,k), k~Poisson(γn)] ~ E[E[Binomial(p,k), k~Poisson(γn)], n~Poisson(μ)]

In words: The Binomial's parameter k is actually a random variable. To get the distribution of x, you need to average over (get the expectation of) the Binomial for all values of its parameter, weighted by their probability. And since the Poisson for k also has a random parameter n, you must repeat.

■ 3.5.2 The Substitution Method for Calculating Variances

Roman's proof elegantly concentrates on the variance calculation, for which one needs only the first two moments of all the other distributions involved. This is general, and I will use it later to look at another complex distribution.

Now I follow the essence of his derivation (he wrote out all the sums carefully).
For compactness, rewrite the distribution for x using B for Binomial, P for Poisson.
x ~ E[E[Bpk, k~Pγn], n~Pμ]
Calculate variance by taking expectations of x and x^2. Expectations are linear operators, so we can slide around the expectations, corresponding to flipping orders of summation.
First, the mean:
E[x] = E[x~E[E[[Bpk,k~Pγn],n~Pμ]]= E[E[E[x~Bpk],k~Pγn],n~Pμ]]=E[E[pk,k~Pγn],n~Pμ]=E[pγn,n~Pμ]=p$\gamma\mu$,
more or less what you expected, though you have to squint a bit.

The next step is to calculate E[x^2] and subtract the square of the mean. Having understood how to manipulate the expectation operators, it is even clearer to view this as a problem in successive algebraic substitution (which my friend *Mathematica* is pretty good at). I can drop the E[ ] and just write the substitutions.

With this substitution scheme, you can't just calculate the variances directly. For example, the Binomial variance cancels out terms quadratic in k. Because k^2 is not a constant, but a random variable, you should take its expectation with respect to the distribution of k. Thus the way to proceed is to calculate separately the first two moments, and subtract to get the variance.

■ 3.5.3 Roman's Derivation As Successive Substitution

Mean calculation:

epxression	substitution	justification
x	x → pk	Binomial mean
pk	k → γ n	Poisson
pγn	n → μ	Poisson
p γ μ		

Second moment $\langle x^2 \rangle$ calculation:

x^2	$x^2 \to p\,(1-p)\,k + p^2\,k^2$	Binomial $\langle x^2 \rangle$
$k\,(1-p)\,p + k^2\,p^2$	$k^2 \to (\gamma\,n)^2 + \gamma\,n,\ k \to \gamma\,n$	Poisson
$n\,(1-p)\,p\,\gamma + p^2\,(n\,\gamma + n^2\,\gamma^2)$	$n^2 \to \mu^2 + \mu,\ n \to \mu$	Poisson
$(1-p)\,p\,\gamma\,\mu + p^2\,(\gamma\,\mu + \gamma^2\,(\mu + \mu^2))$		
p γ μ (1+p γ (1+μ))	Simplified	

Now Var[x] = E[x^2] - $E[x]^2$ yields Var[x] = p γ (1+p γ) μ . We can also recognize p γ = $\alpha\,\beta$, so:
Var[b] = Var[x]/β^2 =$\alpha\,\mu$ (α + 1/β). Inserting the estimate $\alpha\,\mu$→ b, we find
Var[b] = b(α + 1/β) = Var[Nb] = Nb (α + 1/β) as advertised, where α → x/β when we need the numerical estimate.

■ 4. Can Roman's Nb Calculation be Improved?

Can Var[Nb] be reduced while still performing a MC estimate of α? Answer: maybe, but not much. I'll show a minimum variance method to generate integers with a given floating point mean. This is handy if you aren't forced to use Poisson.

■ 4.1 Deriving k From n With Less Variance than a Poisson Process

One concern in Roman's calculation is that the choice of the events going into the MC evaluation of x is random rather than deterministic. This can be thought of as either inducing more variance in k, the number of trials, than need be; or more fundamentally, not using all the n events available equally in trying to estimate the distribution of spatial coordinates before time-swapping. I assume the time coordinates for all events are available for swapping: no hidden parameter such as a smaller number of time coordinates remembered at any point in passing through the data. In this section I consider two alternative ways to get γ n events from n events with less variance than the Poisson process.

■ 4.2 Binomial - Factor - Poisson Model

The best case is if k could be deterministically derived from n. Imagine γ is an integer and one can re-use each event exactly γ times. Then k = γn exactly. So the model is

x ~ Binomial(p,γ n) n ~ Poisson(μ) .

We will attack this by the same substitution method discussed for Roman's derivation; the details are in an Appendix.
The mean for x is again p γ μ.
E[x^2] = p γ μ (1+p (-1 + γ + γ μ)) (from $n^2 \to \mu^2 + \mu$ in the expectation), so
Var[x] = p (1+p (-1+γ)) γ μ (subtracting the square of the mean)
Var[b] = α b + b (1-p) / β γ Factor model (divided by β^2)
which we recognize as identical to the uncorrelated Binomial - Poisson model in the approximate calculation. This model has a slightly lower Var[b] than Roman's model (by a factor of 1-p = 1 - α/(1+α')) in the 1/β term. This is the best that can be done while still using a MC estimate of α, but it's a modest improvement over k ~ Poisson(γn). For typical parameters, the 1/β variance term is lowered by perhaps 10%, the overvall variance of Nb by 5%, and the significance hardly at all. So I'm afraid the improvement I'm about to describe doesn't relly help much, but the technique might be useful in another context.

■ 4.3 How to (Nearly) Achieve the Reduced Variance

Since γ is not an integer, how could we come close to the Factor model above (if it was important)? Define ν to be the chosen number of reuses per event. Consider the variance in ν, assuming for the moment n to be fixed:
1) k ~ k=γn $\Rightarrow$ $\nu = \gamma$ $\Rightarrow$ Var[ν] = 0 in the Factor model.
2) k ~ P(γn) and Var[k] = γ^2n for fixed n, in Roman's Binomial Poisson Poisson model. So for this model,
Var[ν] = Var[k]/n = ⟨k⟩/n = γ.
I propose here a minimum-variance method of achieving the correct mean re-use of each event, $\gamma = \beta(1+\alpha')$ times, even when γ doesn't happen to be an integer. (For reference, γ ~ 12 for Roman's analysis). From the properties of the Q function defined below, it turns out that this can be done with Var[ν] < 1/4 << γ (the Poisson value). Thus, one would expect that using k ~ $\sum_{i=1}^{n}$ Q[ν_i| γ] would be rather close to the k=γn variance limit, as I'll show below.

■ 4.3.1 The Q Function

The idea is to use only one of two integer values surrounding the correct number of re-uses, in a proportion that gives the right mean. I here define a probability function for integers $\nu \sim Q[\nu \mid \gamma]$ that does this.

Q[$\nu \mid \gamma$], $\gamma \geq 0$ is defined by

$\gamma = j + q$, where we have defined subsidiary parameters

$j = \lfloor \gamma \rfloor$ = Least integer[γ] $\quad j \geq 0$

$q = \gamma - j \quad 0 \leq q < 1$ and $q \leq \gamma$.

$Q(\nu \mid \gamma)$ has two possible values:

$\nu = j$, with probability 1-q

$\nu = j+1$, with probability q

The properties of Q are

$\langle \nu \rangle = \gamma$ $\quad E[\nu,Q] = (1 - q)\, j + q\,(j + 1) = j + q = \gamma$ as required

$E[\nu^2] = (1 - q)\, j^2 + q\,(j + 1)^2 = j^2 + 2\, q\, j + q$

$Var[\nu] = q(1-q) = \varepsilon \leqslant 1/4$ so $Var[\nu]$ is independent of j, bounded by [0, 1/4], and < $Var[\nu,P(\gamma n)]$ for all γ.

■ 4.3.2 Calculation of Variance In the Poisson - Q Poisson Model

To calculate the achieved variance for x, we need

$k = \sum \nu_i$ and its moments.

$\langle k \rangle = \sum \langle \nu_i \rangle = \gamma\, n$

$Var[k] = E[k^2] - E[k]^2 = \sum Var[\nu] = n\, Var[\nu]$ (because the ν's are independent; Eadie §2.4.3; and assume γ_i all equal)

$E[k^2] = n\, Var[\nu] + (\gamma\, n)^2 = (\gamma\, n)^2 + n\, \varepsilon$

Then we can proceed with the substitution calculation:

$\langle x \rangle = p\, \gamma\, \mu$

$E[x^2] = k\,(1 - p)\, p + k^2\, p^2$ $\quad k^2 \to n^2\, \gamma^2 + n\, \epsilon,\ k \to \gamma\, n$

$E[x^2] = n\,(1 - p)\, p\, \gamma + p^2\,(n^2\, \gamma^2 + n\, \epsilon)$ $\quad n^2 \to \mu\,(\mu + 1),\ n \to \mu$

$E[x^2] = (1 - p)\, p\, \gamma\, \mu + p^2\,(\epsilon\, \mu + \gamma^2\, \mu\,(1 + \mu))$ subtract square of mean

$Var[x] = ((1 - p)\, p\, \gamma + p^2\, \gamma^2 + p^2\, \epsilon)\, \mu$ divide by β^2, substitute for p, $\alpha\mu$

$$Var[b] = \alpha\, b + \frac{b}{\beta}\left[\, 1 - p\left(1 - \frac{\varepsilon}{\beta\, \xi}\right)\right]$$

Using k ~ Q instead k = γn gives the same result as the Factor model, plus a small term; since $\varepsilon \leqslant 1/4$ and $1/\beta$ term increases by less than 1% for typical values in this analysis. So the Q function is almost as good as the Factor Model.

■ Appendix 1: History

This note developed by a rather convoluted path. First, after failing to understand Roman's derivation of Var[b] in his thesis, I performed an approximate calculation of the variance; this was very useful. Next, I tried to reproduce Roman's results with the assistance of *Mathematica* and matstatistica. This was fairly hard work, because the natural inclination of these tools was to find representations of the entire probability distribution, when the distribution is built up in stages of related random variables. Even when my goal was to use these distributions to calculate the variance, it was not so easy to persuade *Mathematica* to put answers in nice closed forms. However, I learned useful things about combinations of distributions. None of this work survived in the final version of this note.

Then Roman helped me in two ways, by sending me a less-terse version of his derivation, and explaining his variables carefully over the phone. Thinking about these led me to a better understanding of how I could perform the variance calculations by successive substitution for these complex models, and even for alternative models. Here, *Mathematica* was more helpful, even if it is tricky to persuade to put answers in a form I found useful. Finally, by thinking about both the analytical results on multistage distributions, and the substitution methods, I realized I could go back and view Roman's results from a higher level. At the same time, I was led to consider whether the variance could be reduced by a variant of Roman's process.

■ Appendix 3.5.3 BPP Algebra with Mathematica

```
ex = x^2
```

x^2

```
exk = ex /. x^2 → p (1 - p) k + p^2 k^2
```

$k\,(1-p)\,p + k^2\,p^2$

```
exkn = exk /. {k^2 → (γ n)^2 + γ n, k → γ n}
```

$n\,(1-p)\,p\,\gamma + p^2\,(n\,\gamma + n^2\,\gamma^2)$

```
exsq = exkn /. {n^2 → μ^2 + μ, n → μ} // Simplify
```

$p\,\gamma\,\mu\,(1 + p\,\gamma\,(1+\mu))$

```
varb = exsq - (p γ μ)^2 // Simplify
```

$p\,\gamma\,(1 + p\,\gamma)\,\mu$

```
varb7 = (varb / β^2); varb7 /. {p → α β / γ}; % /. {α μ → b}; Collect[%, b]
```

$b\left(\alpha + \frac{1}{\beta}\right)$

■ Appendix 4.2 Algebra for B Factor P model with Mathematica

```
ex1 = x^2
```

x^2

```
ex2 = ex1
```

$k\,(1-p)\,p + k^2\,p^2$

```
ex3 = ex2 /. { k → γ n}
```

$$n\,(1-p)\,p\,\gamma + n^2\,p^2\,\gamma^2$$

```
ex4 = ex3 /. {n^2 → μ^2 + μ, n → μ}
```

$$(1-p)\,p\,\gamma\,\mu + p^2\,\gamma^2\,(\mu+\mu^2)$$

```
Simplify[ex4]
```

$$p\,\gamma\,\mu\,(1+p\,(-1+\gamma+\gamma\,\mu))$$

```
vx1 = ex4 - (p γ μ)^2 // Simplify
```

$$p\,(1+p\,(-1+\gamma))\,\gamma\,\mu$$

```
vxx2 = vx1 / β^2 /. { p γ → α β} /. {α μ → b} /. {γ → β ξ, p → α / ξ}
```

$$\frac{b\,\left(1+\frac{\alpha\,(-1+\beta\,\xi)}{\xi}\right)}{\beta}$$

```
vxx3 = Collect[vxx2, β] /. {α / ξ → p}
```

$$b\,\alpha + \frac{b\,(1-p)}{\beta}$$

■ Appendix 4.3.2 Algebra for Variance of Poisson - Q - Poisson Model

```
eq1 = x^2 /. x^2 → p (1 - p) k + p^2 k^2
```

$$k\,(1-p)\,p + k^2\,p^2$$

```
eq2 = eq1 /. { k → γ n, k^2 → n^2 γ^2 + n ϵ}
```

$$n\,(1-p)\,p\,\gamma + p^2\,(n^2\,\gamma^2 + n\,\epsilon)$$

```
eq3 = eq2 /. {n^2 → μ (μ + 1), n → μ}
```

$$(1-p)\,p\,\gamma\,\mu + p^2\,(\epsilon\,\mu + \gamma^2\,\mu\,(1+\mu))$$

```
vq1 = Collect[eq3 - (p γ μ)^2 , μ]
```

$$((1-p)\,p\,\gamma + p^2\,\gamma^2 + p^2\,\epsilon)\,\mu$$

```
vbq1 = Collect[vq1 / β^2 /. { p γ → α β} /. {μ → b / α, (p γ)^2 → (α β)^2}, β]
```

$$b\,\alpha + \frac{b\,(1-p)}{\beta} + \frac{b\,p^2\,\epsilon}{\alpha\,\beta^2}$$

```
vbq2 = Collect[ Collect[vbq1, b]  /. {p^2 → p α β / γ}, β]
```

$$b\,\alpha + \frac{b\,\left(1-p+\frac{p\,\epsilon}{\gamma}\right)}{\beta}$$