DOES SPLITTING BY SEX HURT YOUR CV?

MVB, AUG 22

In CKMR, sometimes you might be tempted to just fix the sex ratio at 0.5 and use a single-sex model (remembering to multiply by 2 in the right places). Of course, it's a gamble if the true sex ratio is *not* 0.5, as shown by one of the examples. But it saves you a parameter (no need to estiamte the sex ratio) so, *if* your gamble is lucky, do you gain much in precision for the *combined* abundance estimate ?

Here's some hyper-simple code. (I first got this result algebraically, via the Delta-method approximation, then thought I'd better check...)

Suppose you collect Male adults and Female adults (in equal numbers), and the sex ratio in the population really is 0.5, and you estimate Male & Female abundance separately from the number of POPs $Y1$ and $Y2$ respectively (1 & 2 are labels for Male & Female here). Each estimate is reciprocal to the number of POPs found; obviously there's a "constant" but let's just call it 1.

Then the estimated total for separate-sex models is $1/Y_1 + 1/Y_2$. For an amalgamated single-sex model, it would be $4/(Y_1 + Y_2)$— remember what I said about multiplying by 2, and it's "left as an exercise" to work out why the 4 is needed!

Y_1 & Y_2 are IID Poisson, so we can compare the variances of these two expressions empirically with a tiny simulation, as follows. Here I generate 1000 cases where $\mathbb{E}[Y] = 100$, and then repeat it to check:

```
test2> set.seed( 1)
test2> Y1 <- rpois( 1000, lambda=100)
test2> Y2 <- rpois( 1000, lambda=100)
test2> var( 1/Y1 + 1/Y2)
[1] 0.000002227
test2> var( 4/(Y1+Y2))
[1] 0.000002142
test2> Y1 <- rpois( 1000, lambda=100)
test2> Y2 <- rpois( 1000, lambda=100)
test2> var( 1/Y1 + 1/Y2)
[1] 0.000002155
test2> var( 4/(Y1+Y2))
[1] 0.00000206
```

The variance of the amalgamated one is only *slightly* smaller. Basically there's no diff. This is with a total of around 200 POPs, so it's not crazy.

If we drop to only 20 POPs total, things do change:

```
test2> Y1 <- rpois( 1000, lambda=10)
test2> Y2 <- rpois( 1000, lambda=10)
test2> var( 1/Y1 + 1/Y2)
[1] 0.004225
test2> var( 4/(Y1+Y2))
[1] 0.002605
```

The amalgamated version does have lower variance this time, but the dataset will be pretty bad with just 20 POPs; CV would be at least $1/\sqrt{20} = 22\%$ in this case (actually that sounds OK, but in practice other parameters and messy data usually make it worse).

With a reasonable total of 60 POPs, what do we get?

```
test2> Y1 <- rpois( 1000, lambda=30)
test2> Y2 <- rpois( 1000, lambda=30)
test2> var( 1/Y1 + 1/Y2)
[1] 0.00009499
test2> var( 4/(Y1+Y2))
[1] 0.00008474
```

Very little difference. (Bear in mind also that the "meaningful" comparison should be on the square-root-variance scale, not of variance itself.)

I won't include the algebraic version 'cos it's getting late. Anyway, it looks like this is a "sub-asymptotic effect"; you get a tiny gain from fitting amalgamated model *if* you happen to be exactly right, *but* only in situations where your dataset is verging on useless anyway...

Moral of the story: always put both sexes into your CKMR model unless you've got an outstanding excuse!