Modelling reporting systems

Gerhard Neubauer and Gordana Đuraš

Institute of Applied Statistics, JOANNEUM RESEARCH, Graz, Austria

In registers for infectious diseases like HIV or chronic diseases like diabetes recording failures may occur as result of diagnostic errors or patients avoiding diagnosis. Hence we are confronted with underreporting and the estimation of the total number of cases is of particular interest. For situations when capture-recapture approaches are not applicable Neubauer & Friedl (2006) and Neubauer & Đuraš (2008) developed a binomial, a beta-binomial and a negative binomial model for the estimation of the total number of cases.

Consider y_t, $t = 1, \ldots, T$, a sample of counts, then under a binomial assumption, we have $Y_t \sim Binomial(\lambda_t, \pi)$, with both λ_t, the total number of cases, and π, the reporting probability, to be estimated. The model is identified if we take the regression approach $\lambda_t = \exp(x_t'\beta)$. For situations where the sample variance is larger than implied by $v(\mu) = \mu - \mu^2/\lambda$ random effects models are developed. For randomness in the reporting probability we take $P \sim Beta(\gamma, \delta)$ and obtain a marginal beta-binomial model, with $Var(Y_t) = (\mu_t - \mu_t^2/\lambda_t)\phi$ and $\phi > 1$. Applications have shown that this model works well for the situations when $Var(Y_t) \approx E(Y_t)$, while for $Var(Y_t) > E(Y_t)$ results are not satisfying. Another approach is to allow for randomness in λ_t: $Y_t|L_t \sim Binomial(L_t, \pi)$ together with $L_t \sim Poisson(\lambda_t)$ gives $Y_t \sim Poisson(\lambda_t\pi)$, where the parameters are not identified. Making the additional assumption $L_t|K_t \sim Poisson(k_t\lambda_t)$ and $K_t \sim Gamma(\omega_t, \omega_t)$ results in $Y_t \sim Negativ\text{-}Binomial(\omega_t, 1-\pi)$.

The approach has the implicit assumption that the reporting system must be deficient. Obviously for all models we have $Y \to \lambda$ as $\pi \to 1$, and consequently all randomness has to vanish from the observations. Taking an alternative approach we arrive at conditional Poisson distributions that allow for perfect reporting systems, i.e. $\pi \to 1$. For $\pi = 1$ the conditional and the marginal distribution of Y coincide. So Y is still a random variable even though the reporting system is perfect. To illustrate the model we consider three mixing assumption for a random Poisson parameter L: a binomial, a Poisson and negative binomial.

In all cases we have $Y|L \sim Poisson(l)$:

1. $L \sim Binomial(\lambda, \pi)$: $Y \sim Poisson\text{-}Binomial(\lambda, \pi)$
2. $L \sim Poisson(\lambda\pi)$: $Y \sim Neyman\ Type\ A(1, \lambda\pi)$
3. $L \sim Negative\ Binomial(\omega, 1-\pi)$: $Y \sim Poisson\text{-}Negative\ Binomial(\lambda, \pi)$

with $E(Y) = \mu = \lambda\pi$ in all cases and $Var(Y) = \mu\phi_j$ where $\phi_1 = 2 - \pi$, $(1 < \phi_1 < 2)$, $\phi_2 = 2$ and $\phi_3 = (2-\pi)/(1-\pi)$, $(2 < \phi_3 < \infty)$.

For the Poisson mixture we have the alternative $Y|L \sim Poisson(l\pi)$, $L \sim Poisson(\lambda)$ leading to $Y \sim Neyman\ Type\ A(\pi, \lambda)$, with $E(Y) = \lambda\pi$, as before, but $\phi = (1+\pi)$. An intermediate model is obtained from the beta-Poisson distribution which results as marginal for Y_t if $Y_t|P \sim Poisson(\lambda p)$ and $P \sim Beta(\gamma, \delta)$.

The performance of the approach is illustrated by simulations and an application to real data.

References

Neubauer, G. & Đuraš, G. (2008). On the use of the generalized Poisson distribution for estimating unreported cases. Unpublished technical report.

Neubauer, G. & Friedl, H. (2006). Modelling Sample Sizes of Frequencies. In Proceedings of the 21st International Workshop on Statistical Modelling, 3-7 July 2006, Galway, Ireland.