Re: Is it true that a zero-inflated Gamma model is not possible

Dale McLerran <stringplaye...@yahoo.com>
--- "Swank, Paul R" <Paul.R.Sw...@UTH.TMC.EDU> wrote:

> I thought you couldn't have zeroes in a gamma distribution?

> Paul R. Swank, Ph.D.
> Professor and Director of Research
> Children's Learning Institute
> University of Texas Health Science Center - Houston

That is correct. Thus, a zero-inflated gamma model is nothing more than modeling the zero values as a binary random r.v. and modeling the positive values employing the gamma distribution. Mathematically, a zero-inflated gamma could be described as

\[ g(y) \sim B(1, f(x)) \]
\[ y \sim \text{gamma}(\theta/h(x), \theta), \quad y>0 \]

The first line indicates that we should compute and use as a response variable the function \( g(y) \) which takes on the value 1 if \( y=0 \) and is 0 if \( y>0 \). (Note that any values of \( y<0 \) including values of \( y \) which are missing would become missing for \( g(y) \).) This response is then modeled as a binary outcome. The function \( f(x) \) is inverse logit function which constructs \( p \) from \( x \). Employing a logit link, we would have

\[ f(x) = \frac{\exp(b_0_f + b_1_f x_1 + b_2_f x_2 + ...)}{1 + \exp(b_0_f + b_1_f x_1 + b_2_f x_2 + ...)} \]

The second line indicates that all of the positive values of \( y \) can be modeled as a gamma response. This is just what Amy has been doing up to this point, so there is nothing new here. The function \( h(x) \) yields the expectation. Employing a log link, we would have

\[ h(x) = \exp(b_0_h + b_1_h x_1 + b_2_h x_2 + ...) \]

The following NLMIXED code would allow one to fit a zero-inflated gamma distribution:

```sas
proc nlmixed data=mydata;
p parms b0_f=0 b1_f=0 b2_f=0 ...
    b0_h=0 b1_h=0 b2_h=0 ...
    log_theta=0;
eta_f = b0_f + b1_f*x1 + b2_f*x2 + ...;
p_yEQ0 = 1 / (1 + exp(-eta_f));
eta_h = b0_h + b1_h*x1 + b2_h*x2 + ...;
mu = exp(eta_h);
theta = exp(log_theta);
r = mu/theta;
```
if y=0 then
   ll = log(p_yEQ0);
else
   ll = log(1 - p_yEQ0)
       - lgamma(theta) + (theta-1)*log(y) - theta*log(r) - y/r;

model y ~ general(ll);
predict (1 - p_yEQ0)*mu out=expect_zig;
predict r out=shape;
estimate "scale" theta;
run;

Given the single class predictor GROUP with three values Control, Intervention1, and Intervention2, we would write:

```
proc nlmixed data=mydata;
parms b0_f=0 b1_f=0 b2_f=0
      b0_h=0 b1_h=0 b2_h=0
      log_theta=0;
x1 = 0;  if group='Intervention1' then x1=1;
x2 = 0;  if group='Intervention2' then x2=1;
eta_f = b0_f + b1_f*x1 + b2_f*x2;
p_yEQ0 = 1 / (1 + exp(-eta_f));
eta_h = b0_h + b1_h*x1 + b2_h*x2;
mu = exp(eta_h);
theta = exp(log_theta);
r = mu/theta;
if y=0 then
   ll = log(p_yEQ0);
else
   ll = log(1 - p_yEQ0)
       - lgamma(theta) + (theta-1)*log(y) - theta*log(r) - y/r;
model y ~ general(ll);
predict (1 - p_yEQ0)*mu out=expect_zig;
predict r out=shape;
contrast "Diff across trts in p(y=0)" b1_f, b2_f;
contrast "Diff across trts in E(Y|y>0)" b1_h, b2_h;
contrast "Diff across trts in E(Y)"
   exp(b0_h + b1_h)*(1 - 1/(1 + exp(b0_f + b1_f)))
   - exp(b0_h)*(1 - 1/(1 + exp(b0_f)));
   exp(b0_h + b2_h)*(1 - 1/(1 + exp(b0_f + b2_f)))
   - exp(b0_h)*exp(b0_f));
estimate "scale" theta;
run;
```

Note that I added contrast statements in order to test the effect of treatment condition on each of the two components of this model: 1) the probability of a zero value, 2) the expectation of the nonzero values given the gamma distribution, and 3) the overall difference in expectations across the three treatment groups. I don't know the properties of the latter test. It may or may not be a good test for its purpose.

It should be pointed out that the only reason for estimating the zero-inflated gamma is in order to compute an expected response value (as output to the data set expect_zig) along with standard
errors of the expected value and/or to construct a test of whether
there are differences across groups in the expected value of the
mean response (not just the nonzero response values). If you do
not want the expected value with standard errors or the particular
test statistic which is obtained by the third contrast statement
above, then it is easier to fit separate models estimating the
probability of a zero value and then estimating the parameters of
the gamma distribution.

A couple of final notes about a zero-inflated gamma model. The zero
values are indicators of observations which are not generated under
the gamma distribution. In the cost model, I would assume that these
zero values indicate individuals who were service nonusers. Note,
too, that the parameters of the gamma portion of the zero-inflated
gamma are just the parameters of the gamma distribution which you
can verify by fitting the NLMIXED code given above and then fitting
a gamma response model using GENMOD with log link.

Dale

---------------------------------------
Dale McLerran
Fred Hutchinson Cancer Research Center
mailto: dmclerra@NO_SPAMfhcrc.org
Ph: (206) 667-2926
Fax: (206) 667-5977
---------------------------------------

Be a better friend, newshound, and
know-it-all with Yahoo! Mobile. Try it now.  http://mobile.yahoo.com/
_ylt=Ahu06i62sR8HDTDypao8Wcj9tAcJ