An Extension of Beta Regression to Handle Scores at Boundaries

Jim Rogers, Ph.D. and Jonathan French, Sc.D.
Metrum Research Group

Objectives

To characterize pharmacokinetic / pharmacodynamic relationships in the population of interest, we want to fit a population PK/PD model that has a fairly conventional structure. However, the response variable is the PASI, which has the following features:

- By construction, the PASI can only take values in the range 0–72.
- Approx. 10% – 20% of data are exactly at the lower boundary (zero).
- The number of intermediate possible values is large (697 possible values, all of which are multiples of 0.1 between 0 and 72; some, e.g., 71.9 are not possible).

These features imply that a (conventional) Normal residual likelihood is unrealistic, such that inferences based on a Normal likelihood will be incorrect, and simulations from a model with Normal likelihood will go outside of the allowable range.

Outline

1 Motivation

2 Beta Regression Methodology
- Mathematics
- NONMEM Implementation

3 Augmented Beta Regression
- Methodology Mathematics
- NONMEM Implementation

Approaches to modeling constrained responses with boundary observations

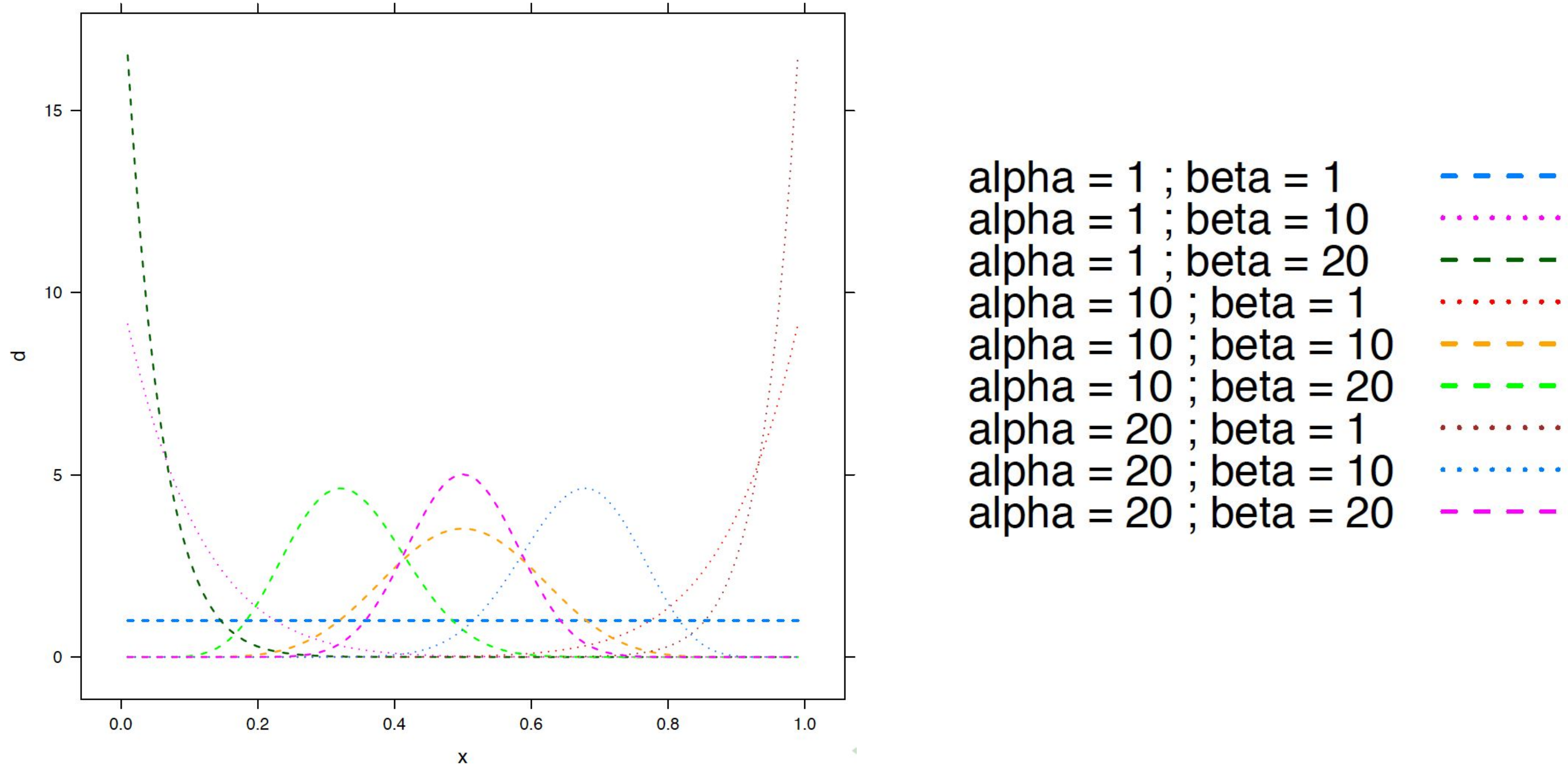
- Approaches assuming Normal likelihood for a *transformed dependent variable* and treating boundary values as censored data (Hutmacher et al, *Stat Med*, 2010).
- Approaches assuming Multinomial likelihood for the untransformed dependent variable, with probabilities structured by a link function to achieve parsimony when there are many possible outcomes (Hu et al, *JPKPD*, 2011).
- Approaches assuming Beta likelihood for the untransformed dependent variable (Beta regression, e.g., Samtani et al, *JPKPD*, 2013), with conditional means structured by a link function, and treating boundary values as censored data. This extension for handling boundary values has not been published, and is elaborated here.

The beta distribution

The probability density function for a Beta(α,β) distribution is:

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where Γ is the gamma function, defined as: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t.$



Parameterization of beta distribution for regression

- Beta distribution typically expressed in terms of parameters α and β, as shown on previous slide.
- For purposes of regression it is useful to re-parameterize in terms of
  $\mu = \alpha/(\alpha+\beta)$ and $\tau = \alpha + \beta$
- Under this parameterization, the mean and variance of the distribution are expressed as:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{Var}[X] &= \frac{\mu(1-\mu)}{\tau+1} \end{aligned}$$

- To show the correspondence to the usual parameterization, we refer to this as a $\mathrm{Beta}(\mu\tau, (1-\mu)\tau)$ distribution.

Beta regression approach differs only in residuals

1. As in the usual (Normal residual) case, we express our model as:

$$Y_{ij} = f(t, \theta, \eta_i, x_i(t)) + \epsilon_{ij}$$

2. We do NOT apply any transformations to Y_{ij} (the `DV`)

3. We DO apply transformations (or use alternative tricks, e.g. bounded parameter spaces) to keep $f(t, \theta, \eta_i, x_i(t))$ between zero and `SMAX`, just as we might do with a Normal residual model (this bit is not unique to Beta regression).

4. The defining difference is in how we model the residuals. We set $\mu = f(t, \theta, \eta_i, x_i(t))$ and then specify

$$\epsilon_{ij} \sim \mathrm{Beta}(\mu\tau, (1-\mu)\tau)$$

τ is a free parameter (just as σ is in the usual Normal residual implementation).
This residual distribution needs to be modified to handle 0s and 1s. We get to that later.

METRUM RESEARCH GROUP

NONMEM implementation, step 1

Find a strategy to keep $f(t, \theta, \eta_i, x_i(t))$ in the desired range.

For example, we might have something like:

1 Use the logit transform to get the baseline in range:

```
$PK
COBS = THETA(1) + THETA(8)*CGR1 ;; start adding covariate effects
LBAS = COBS + ETA(1)
SMAX = 10                          ;; maximum score on constrained scale
A_0(3) = SMAX*EXP(LBAS)/(1+EXP(LBAS)) ;; effect compartment constrained to [0,SMAX]
```

2 Define the differentials to keep post-baseline values in range (need to define rate constants accordingly; details not shown)

```
$DES
DADT(3)= KINN*INHD*(1-BEFF) - KOWT*A(3)
```

METRUM RESEARCH GROUP

NONMEM implementation, step 2

Define scale parameter for residual distribution. Where you would normally do this:

```
$SIGMA
1.1
```

Do this instead:

```
$PK
[...]
TAU = EXP(THETA(47))
```

NONMEM implementation, step 3

Specify residual distribution. Where you would normally do this:

```
$ERROR
IPRED = A(3)
Y = IPRED + ERR(1)
```

Do this instead (this uses approximation noted in Samtani et al):

```
$ERROR
MU = A(3) / SMAX ; assumes modeling done on [0,SMAX] scale

;Approximation of the log(gamma) function
ALPHA=MU*TAU
BETA=(1-MU)*TAU
X1=ALPHA+BETA
X2=ALPHA
X3=BETA
LG1=0.5*(LOG(2*3.1415)-LOG(X1)) + X1 * (LOG(X1)-1) + (5/4)* X1 * ( LOG ( 1 +  (1/(15*X1
     **2))));
LG2=0.5*(LOG(2*3.1415)-LOG(X2)) + X2 * (LOG(X2)-1) + (5/4)* X2 * ( LOG ( 1 +  (1/(15*X2
     **2))));
LG3=0.5*(LOG(2*3.1415)-LOG(X3)) + X3 * (LOG(X3)-1) + (5/4)* X3 * ( LOG ( 1 +  (1/(15*X3
     **2))));

;Log Likelihood of the beta distribution
LOGL = LG1 - LG2 - LG3 +  (ALPHA-1)*LOG(DV/SMAX) + (BETA-1)*LOG(1-DV/SMAX)
Y = -2 * LOGL
```

NONMEM implementation, step 4

Where you would normally do this:

```
$EST MAXEVAL=99999  NOABORT METHOD=1 INTER NOABORT
```

Do this instead:

```
$EST MAXEVAL=99999  NOABORT METHOD=1 -2LOGLIK NUMERICAL LAPLACIAN
```

NONMEM / R implementation, step 5

Simulation. Not easy to generate Beta random variates in NONMEM, so we do part of it in R.

Where you would normally do this:

```
$TABLE NOHEADER NOPRINT NOAPPEND FILE=./1050.tab
 TRL STUD ID TIME IPRED DV
```

Do this instead (note that ALPHA and BETA vary over time within individuals):

```
$TABLE NOHEADER NOPRINT NOAPPEND FILE=./1050.tab
 TRL STUD ID TIME  ALPHA BETA
```

And then in R:

```
simres$IPRED <- SMAX * simres$ALPHA / (simres$ALPHA + simres$BETA)
simres$DV <- SMAX * rbeta(nrow(simres), shape1 = simres$ALPHA, shape2 = simres$BETA)
```

A class of augmented beta distributions

In general, one may define a 0-1-augmented $\mathrm{Beta}(\alpha, \beta, p_0, p_1)$ distribution as one with density[1]:

$$p(x) = \begin{cases} p_0 & : \quad x = 0 \\ p_1 & : \quad x = 1 \\ (1 - p_0 - p_1)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & : \quad 0 < x < 1 \end{cases}$$

In our application, we assume that the same conditions that make *low*-valued responses (or high-valued responses, respectively) likely also make *zero*-valued (or one-valued, respectively) responses likely. It therefore makes sense for p0 and p1 to be a function of $\mu = \alpha/(\alpha+\beta)$. One approach is to let:

$$p_0 = f_0(\mu) = \mathrm{logit}^{-1}\left(-\gamma_0 - \gamma_1 \cdot \mathrm{logit}(\mu)\right)$$

$$p_1 = f_1(\mu) = \mathrm{logit}^{-1}\left(-\gamma_0 + \gamma_1 \cdot \mathrm{logit}(\mu)\right)$$

Other choices for f_0 and f_1 are possible, as long as they force the condition $p_0 + p_1 \leq 1$.

[1] technically, we should use Dirac δ functions to make *p* a true *density* that integrates to one, but our less accurate notation is probably easier to follow.

METRUM
RESEARCH GROUP

A class of augmented beta distributions (continued)

Our particular choice of f_0 and f_1 has the following reasonable properties:

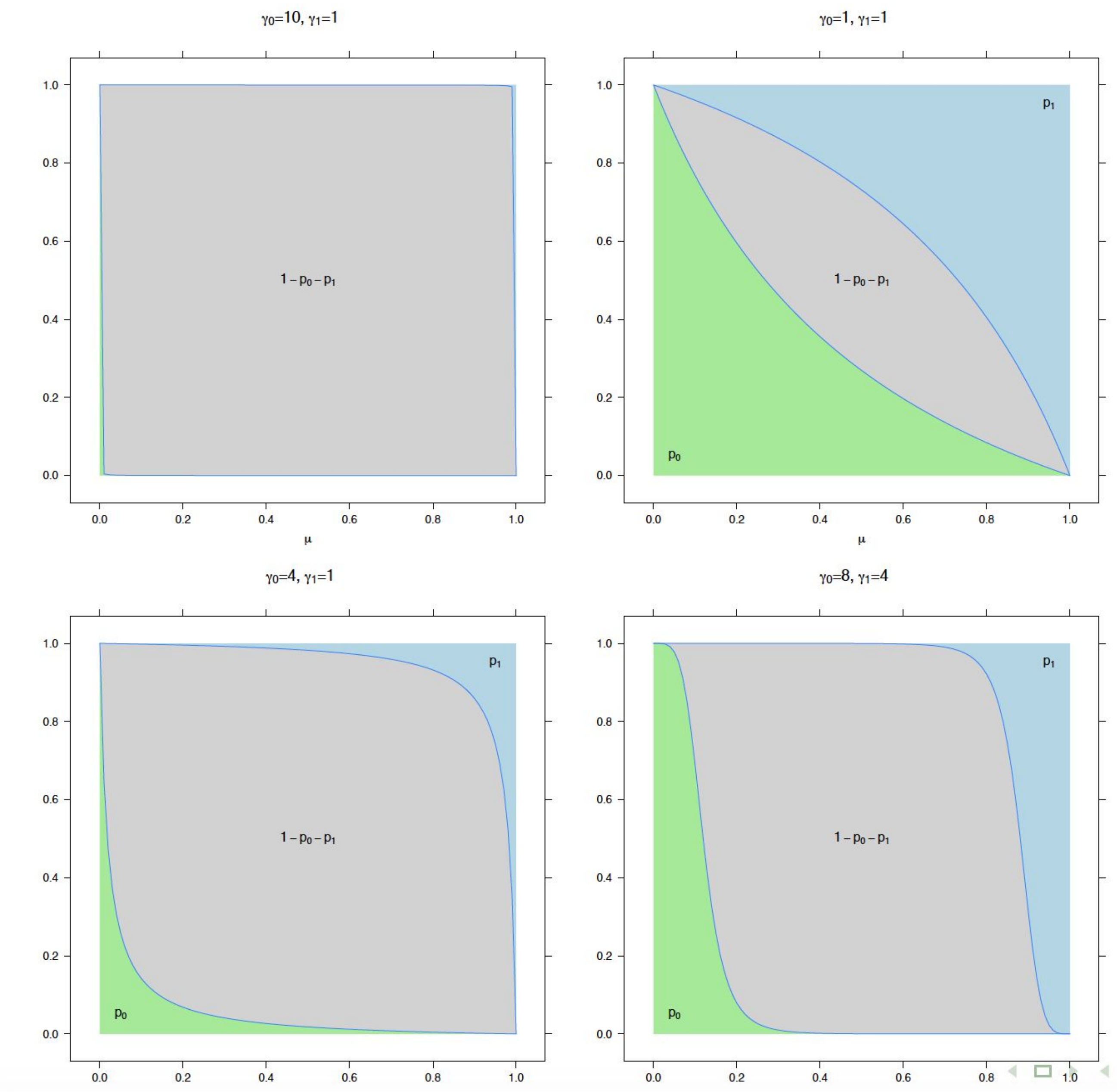
$$f_0(\mu) = f_1(1-\mu)$$

$$f_0(\mu) \overset{\mu\to 0}{\longrightarrow} 1 \quad ; \quad f_1(\mu) \overset{\mu\to 0}{\longrightarrow} 0$$

$$f_0(\mu) \overset{\mu\to 1}{\longrightarrow} 0 \quad ; \quad f_1(\mu) \overset{\mu\to 1}{\longrightarrow} 1$$

$$f_0(0.5) = f_1(0.5) \quad = \quad 1/(1+\exp(\gamma_0))$$

METRUM
RESEARCH GROUP



Our choice for f_0 and f_1 offers reasonable flexibility as a function of γ_0 and γ_1, ranging from almost a pure Beta distribution (top left) to almost a pure binomial distribution (top right), with a variety of shapes for intermediate possibilities (bottom left and bottom right).

METRUM RESEARCH GROUP

NONMEM Implementation, step 2

We follow the same recipe as for basic beta regression, with modifications to steps 2, 3, and 5.

In step 2, we just need to define the γ parameters. They need to be constrained to be positive.

```
$PK
[...]
TAU = EXP(THETA(47)) ; (same as before)
GAMMA0 = EXP(THETA(48))
GAMMA1 = EXP(THETA(49))
```

NONMEM Implementation, step 3

The additional components to the likelihood are easy to add:

```
$ERROR
MU = A(3) / SMAX ; assumes modeling done on [0,SMAX] scale

MULGT = LOG(MU / (1-MU))
P0LGT = - GAMMA0 - GAMMA1*MULGT
P1LGT = - GAMMA0 + GAMMA1*MULGT
P0 = EXP(P0LGT)/(1+EXP(P0LGT))
P1 = EXP(P1LGT)/(1+EXP(P1LGT))

;Approximation of the log(gamma) function
ALPHA=MU*TAU
BETA=(1-MU)*TAU
X1=ALPHA+BETA
X2=ALPHA
X3=BETA
LG1=0.5*(LOG(2*3.1415)-LOG(X1)) + X1 * (LOG(X1)-1) + (5/4)* X1 * ( LOG ( 1 +  (1/(15*X1
      **2))));
LG2=0.5*(LOG(2*3.1415)-LOG(X2)) + X2 * (LOG(X2)-1) + (5/4)* X2 * ( LOG ( 1 +  (1/(15*X2
      **2))));
LG3=0.5*(LOG(2*3.1415)-LOG(X3)) + X3 * (LOG(X3)-1) + (5/4)* X3 * ( LOG ( 1 +  (1/(15*X3
      **2))));

;Log Likelihood of the 0-1-augmented beta distribution
IF(DV.GT.0.AND.DV.LT.SMAX) LOGL = LOG(1-P0-P1) + LG1 - LG2 - LG3 +  (ALPHA-1)*LOG(DV/SMAX
      ) + (BETA-1)*LOG(1-DV/SMAX)
IF(DV.EQ.0) LOGL = LOG(P0)
IF(DV.EQ.SMAX) LOGL = LOG(P1)
Y = -2 * LOGL
```

METRUM
RESEARCH GROUP

NONMEM Implementation, step 5

Now we need to additionally output the values for `P0` *and* `P1`

```
$TABLE NOHEADER NOPRINT NOAPPEND FILE=./1050.tab
 TRL STUD ID TIME  ALPHA BETA P0 P1
```

And then (in R) :

```
  # IPRED is a weighted average of zero, mean of Beta dist., and SMAX
simres$IPRED <- with(simres, 0 * P0 + SMAX * ALPHA / (ALPHA + BETA) + SMAX * P1

# To get DV, first simulate class variable ...
sim50$DV <- apply(sim50[c('P0', 'P1')], 1,
                  function(p) sample(0:2, 1, prob = c(p[1], p[2], 1-sum(p)))
                  )
# ... and then replace ``2'' category with Beta random variates:
sim50 <- within(sim50, DV[DV==2] <- rbeta(nrow(sim50), shape1 = ALPHA, shape2 = BETA)[DV
     ==2])
# Finally, scale up by SMAX:
sim50$DV <- SMAX * sim50$DV
```

METRUM
RESEARCH GROUP