We have $i = 1, ..., n_j$ observations in $j = 1, ..., J$ groups. We are modeling each group with its own intercept.

$$\begin{aligned}
\mu_{ij} &= \exp(\beta_{0j} + \beta_1 x_{ij}) && (1)\\
y_{ij} &\sim \text{Poisson}(\lambda_{ij}) && (2)\\
\lambda_{ij} &\sim \text{lognormal}(\log(\mu_{ij}), \sigma^2_\lambda) && (3)\\
\beta_{0j} &\sim \text{normal}(\mu_{\beta_o}, \sigma^2_{\beta_0}) && (4)\\
&+ \quad \text{priors} && (5)
\end{aligned}$$

The question is how to properly make predictions of the mean μ_{all} across all groups for a predictor value x. It seems there are two possibilities. At each iteration (k) of the MCMC chain we could compute

$$\mu_{all} = \exp(\mu^{(k)}_{\beta_0} + \beta^{(k)}_1 x) \tag{6}$$

or we could compute

$$\mu_{all} = \frac{1}{J}\sum_{j=1}^{J} \exp(\beta^{(k)}_{0j} + \beta^{(k)}_1 x). \tag{7}$$

These, of course, do not give the same results. I believe that the second formulation is the correct one because it approximates the posterior predictive distribution. he problem outlined in the pdf I sent arises from Jensen's inequality. When the model is linear, the two approaches (equations 6 and 7) to the mean of a new prediction across groups are the same. But when the function is non-linear, they are *not* the same. We desire the expected value of the output of a nonlinear function, not the output of a non-linear function at the expected value of the parameters. Therefore, we use the second approach (equation 7) instead of the first approach (equation 6) to obtain a new prediction across groups.