Bayesian Modeling of Overdispersed Data with the Tilted Beta Binomial Distribution

Eugene D. Hahn

Department of Information and Decision Sciences
Salisbury University

Joint Statistical Meetings 2020. August 5, 2020

Modeling Success/Failure Data

- Binary data with x successes and $n - x$ failures in n trials is common and important category of data.
- Often modeled with the binomial distribution, $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$.
- Binomial may be inadequate for real-world data because of overdispersion.
- Beta binomial distribution is often used for such cases.
- Unconditional modeling of data with beta binomial distribution dates back to Skellam (1948).

Modeling Success/Failure Data

- Binary data with x successes and $n - x$ failures in n trials is common and important category of data.
- Often modeled with the binomial distribution, $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$.
- Binomial may be inadequate for real-world data because of overdispersion.
- Beta binomial distribution is often used for such cases.
- Unconditional modeling of data with beta binomial distribution dates back to Skellam (1948).

Modeling Success/Failure Data

- Binary data with x successes and $n - x$ failures in n trials is common and important category of data.
- Often modeled with the binomial distribution, $p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$.
- Binomial may be inadequate for real-world data because of overdispersion.
- Beta binomial distribution is often used for such cases.
- Unconditional modeling of data with beta binomial distribution dates back to Skellam (1948).

Modeling Success/Failure Data

- Binary data with x successes and $n - x$ failures in n trials is common and important category of data.
- Often modeled with the binomial distribution, $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$.
- Binomial may be inadequate for real-world data because of overdispersion.
- Beta binomial distribution is often used for such cases.
- Unconditional modeling of data with beta binomial distribution dates back to Skellam (1948).

Modeling Success/Failure Data

- Binary data with x successes and $n - x$ failures in n trials is common and important category of data.
- Often modeled with the binomial distribution, $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$.
- Binomial may be inadequate for real-world data because of overdispersion.
- Beta binomial distribution is often used for such cases.
- Unconditional modeling of data with beta binomial distribution dates back to Skellam (1948).

Background

- Beta rectangular distribution proposed by Hahn (2008) for project management.
- Overdispersed alternative to Beta.
- Tilted beta distribution proposed by Hahn & López Martín (2015).
- Permits excesses at the extremes.

$$p(p|\alpha,\beta,v,\theta) = (1-\theta)\big(2-2v+(4v-2)p\big)+\theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

- Tilted beta also has p parameter. Accordingly, tilted beta binomial proposed by Hahn (2012).

Background

- Beta rectangular distribution proposed by Hahn (2008) for project management.
- Overdispersed alternative to Beta.
- Tilted beta distribution proposed by Hahn & López Martín (2015).
- Permits excesses at the extremes.

$$p(p|\alpha,\beta,v,\theta) = (1-\theta)\big(2-2v+(4v-2)p\big)+\theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \quad (1)$$

- Tilted beta also has p parameter. Accordingly, tilted beta binomial proposed by Hahn (2012).

Background

- Beta rectangular distribution proposed by Hahn (2008) for project management.
- Overdispersed alternative to Beta.
- Tilted beta distribution proposed by Hahn & López Martín (2015).
- Permits excesses at the extremes.

$$p(p|\alpha,\beta,v,\theta) = (1-\theta)\big(2-2v+(4v-2)p\big)+\theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

- Tilted beta also has p parameter. Accordingly, tilted beta binomial proposed by Hahn (2012).

Background

- Beta rectangular distribution proposed by Hahn (2008) for project management.
- Overdispersed alternative to Beta.
- Tilted beta distribution proposed by Hahn & López Martín (2015).
- Permits excesses at the extremes.

$$p(p|\alpha,\beta,v,\theta) = (1-\theta)\big(2-2v+(4v-2)p\big)+\theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \quad (1)$$

- Tilted beta also has p parameter. Accordingly, tilted beta binomial proposed by Hahn (2012).

Background

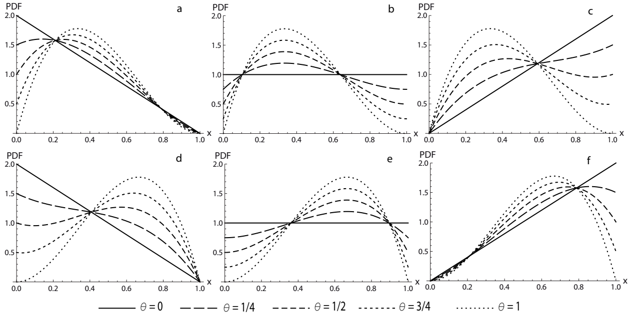
- Beta rectangular distribution proposed by Hahn (2008) for project management.
- Overdispersed alternative to Beta.
- Tilted beta distribution proposed by Hahn & López Martín (2015).
- Permits excesses at the extremes.

$$p(p|\alpha,\beta,v,\theta) = (1-\theta)\big(2-2v+(4v-2)p\big)+\theta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

- Tilted beta also has p parameter. Accordingly, tilted beta binomial proposed by Hahn (2012).

Tilted Beta Distribution

Figure: Examples of Tilted Beta Distributions for: $\alpha = 2, \beta = 3, v = 0$ (**A**); $\alpha = 2, \beta = 3, v = 0.5$ (**B**), $\alpha = 2, \beta = 3, v = 1$ (**C**), $\alpha = 3, \beta = 2, v = 0$ (**D**); $\alpha = 3, \beta = 2, v = 0.5$ (**E**), $\alpha = 3, \beta = 2, v = 1$ (**F**)



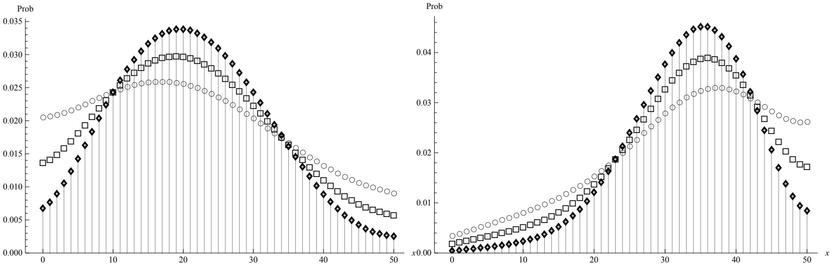
Tilted Beta Binomial Distribution

- Compounding the tilted beta with the binomial produces the distribution.

$$p(y|\alpha,\beta,\theta,v) = (1-\theta)\frac{2(y(2v-1)+n(1-v)+1)}{(n+1)(n+2)} + \theta\binom{n}{y}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}. \tag{2}$$

Tilted Beta Binomial Distribution

Figure: Examples of Tilted Beta Binomial Distributions



$v = 0.3, \alpha = 3, \beta = 4$ $v = 0.9, \alpha = 3, \beta = 4$

Both plots: $\circ, \theta = 0.25$; $\square, \theta = 0.5$; $\diamond, \theta = 0.75$.

CDF for Tilting Binomial

- The tilting mixture component, preceded by $1-\theta$ in (2), can be called tilting binomial.
- Unlike beta binomial, this distribution has a relatively simple form for CDF.
-

$$F(y|v,n) = \begin{cases} 1 & \text{if } y = n, \\ \frac{\left(y(2v-1)+2n(1-v)+2\right)(y+1)}{(n+2)(n+1)} & \text{if } 0 \le y \le n, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

CDF for Tilting Binomial

- The tilting mixture component, preceded by $1 - \theta$ in (2), can be called tilting binomial.
- Unlike beta binomial, this distribution has a relatively simple form for CDF.
- $$F(y|v,n) = \begin{cases} 1 & \text{if } y = n, \\ \frac{\left(y(2v-1)+2n(1-v)+2\right)(y+1)}{(n+2)(n+1)} & \text{if } 0 \le y \le n, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

CDF for Tilting Binomial

- The tilting mixture component, preceded by $1-\theta$ in (2), can be called tilting binomial.
- Unlike beta binomial, this distribution has a relatively simple form for CDF.
- $$F(y|v,n) = \begin{cases} 1 & \text{if } y = n, \\ \frac{\left(y(2v-1)+2n(1-v)+2\right)(y+1)}{(n+2)(n+1)} & \text{if } 0 \leq y \leq n, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Tilted Beta Binomial Distribution

- Several ways to estimate a compounded mixture using Bayesian methods.
- Approach A: Use (2) as *marginalized* likelihood, place priors on α, β, θ, v.
- Approach B: Use binomial distribution as *conditional* likelihood, give p a tilted beta 'quasi-prior', place 'real' priors on α, β, θ, v.
- Approach C: Use the approach familiar from Dempster, Laird & Rubin (1977). For a two component mixture with densities $f_1(\cdot)$ and $f_2(\cdot)$, form the likelihood as $f_1(\cdot)^{z_i} f_2(\cdot)^{1-z_i}$ where z_i is a latent binary variable with Bernoulli parameter θ.
- Approach D: Combination of Approach B and C.

Tilted Beta Binomial Distribution

- Several ways to estimate a compounded mixture using Bayesian methods.
- Approach A: Use (2) as *marginalized* likelihood, place priors on α, β, θ, v.
- Approach B: Use binomial distribution as *conditional* likelihood, give p a tilted beta 'quasi-prior', place 'real' priors on α, β, θ, v.
- Approach C: Use the approach familiar from Dempster, Laird & Rubin (1977). For a two component mixture with densities $f_1(\cdot)$ and $f_2(\cdot)$, form the likelihood as $f_1(\cdot)^{z_i} f_2(\cdot)^{1-z_i}$ where z_i is a latent binary variable with Bernoulli parameter θ.
- Approach D: Combination of Approach B and C.

Tilted Beta Binomial Distribution

- Several ways to estimate a compounded mixture using Bayesian methods.
- Approach A: Use (2) as *marginalized* likelihood, place priors on α, β, θ, v.
- Approach B: Use binomial distribution as *conditional* likelihood, give p a tilted beta 'quasi-prior', place 'real' priors on α, β, θ, v.
- Approach C: Use the approach familiar from Dempster, Laird & Rubin (1977). For a two component mixture with densities $f_1(\cdot)$ and $f_2(\cdot)$, form the likelihood as $f_1(\cdot)^{z_i} f_2(\cdot)^{1-z_i}$ where z_i is a latent binary variable with Bernoulli parameter θ.
- Approach D: Combination of Approach B and C.

Tilted Beta Binomial Distribution

- Several ways to estimate a compounded mixture using Bayesian methods.
- Approach A: Use (2) as *marginalized* likelihood, place priors on α, β, θ, v.
- Approach B: Use binomial distribution as *conditional* likelihood, give p a tilted beta 'quasi-prior', place 'real' priors on α, β, θ, v.
- Approach C: Use the approach familiar from Dempster, Laird & Rubin (1977). For a two component mixture with densities $f_1(\cdot)$ and $f_2(\cdot)$, form the likelihood as $f_1(\cdot)^{z_i} f_2(\cdot)^{1-z_i}$ where z_i is a latent binary variable with Bernoulli parameter θ.
- Approach D: Combination of Approach B and C.

Tilted Beta Binomial Distribution

- Several ways to estimate a compounded mixture using Bayesian methods.
- Approach A: Use (2) as *marginalized* likelihood, place priors on α, β, θ, v.
- Approach B: Use binomial distribution as *conditional* likelihood, give p a tilted beta 'quasi-prior', place 'real' priors on α, β, θ, v.
- Approach C: Use the approach familiar from Dempster, Laird & Rubin (1977). For a two component mixture with densities $f_1(\cdot)$ and $f_2(\cdot)$, form the likelihood as $f_1(\cdot)^{z_i} f_2(\cdot)^{1-z_i}$ where z_i is a latent binary variable with Bernoulli parameter θ.
- Approach D: Combination of Approach B and C.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. x is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1-\mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and v. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. x is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1 - \mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and v. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. x is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1-\mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and v. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. x is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1 - \mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and v. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. x is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1 - \mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and v. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Modeling

- Data for state of Maryland from the 2010 U.S. Census. There are 518 Census places in Maryland. *x* is number of people indicating African-American and of one race.
- Bayesian MCMC.
- 5000 MCMC burn-in iterations and 200,000 estimation. WinBUGS.
- Use transformations $\alpha = \mu\phi$ and $\beta = \phi(1 - \mu)$ for beta component parameters.
- Uniform (0,1) priors on μ, θ, and *v*. Uniform (0,100) prior on ϕ.
- We examine Monte Carlo (MC) error and estimation time.

Tilted Beta Binomial Results

Table: Parameter Estimates: Posterior Means

	Approach A	Approach B	Approach C	Approach D
α	0.5437	0.5439	0.5439	0.5437
β	5.911	5.925	5.925	5.906
μ	0.08674	0.08668	0.08668	0.08684
ϕ	6.454	6.469	6.469	6.45
v	0.6478	0.647	0.647	0.6484
θ	0.7461	0.7458	0.7458	0.7464

Approaches agree with regard to posterior means.

Tilted Beta Binomial Model Estimation Time

- As mentioned, 200,000 iterations.
- Approach B (Binomial) was the fastest, 40% less time than Approach A.
- Approach C ('EM') took approximately 40% more time than Approach A.
- Approach D ('EM') took approximately 17% more time than Approach B.
- Conclusion: Binomial likelihood approach is faster for WinBUGS.

Tilted Beta Binomial Model Estimation Time

- As mentioned, 200,000 iterations.
- Approach B (Binomial) was the fastest, 40% less time than Approach A.
- Approach C ('EM') took approximately 40% more time than Approach A.
- Approach D ('EM') took approximately 17% more time than Approach B.
- Conclusion: Binomial likelihood approach is faster for WinBUGS.

Tilted Beta Binomial Model Estimation Time

- As mentioned, 200,000 iterations.
- Approach B (Binomial) was the fastest, 40% less time than Approach A.
- Approach C (‘EM’) took approximately 40% more time than Approach A.
- Approach D (‘EM’) took approximately 17% more time than Approach B.
- Conclusion: Binomial likelihood approach is faster for WinBUGS.

Tilted Beta Binomial Model Estimation Time

- As mentioned, 200,000 iterations.
- Approach B (Binomial) was the fastest, 40% less time than Approach A.
- Approach C ('EM') took approximately 40% more time than Approach A.
- Approach D ('EM') took approximately 17% more time than Approach B.
- Conclusion: Binomial likelihood approach is faster for WinBUGS.

Tilted Beta Binomial Model Estimation Time

- As mentioned, 200,000 iterations.
- Approach B (Binomial) was the fastest, 40% less time than Approach A.
- Approach C (‘EM’) took approximately 40% more time than Approach A.
- Approach D (‘EM’) took approximately 17% more time than Approach B.
- Conclusion: Binomial likelihood approach is faster for WinBUGS.

Tilted Beta Binomial Model MC Error

- ‘Digestible’ comparison of MC errors possibly difficult.
- However, note that μ, θ and v are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C (‘EM’) had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D (‘EM’) had about a 3% greater MC error than Approach C (‘EM’).

Tilted Beta Binomial Model MC Error

- ‘Digestible’ comparison of MC errors possibly difficult.
- However, note that μ, θ and v are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C (‘EM’) had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D (‘EM’) had about a 3% greater MC error than Approach C (‘EM’).

Tilted Beta Binomial Model MC Error

- 'Digestible' comparison of MC errors possibly difficult.
- However, note that μ, θ and v are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C ('EM') had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D ('EM') had about a 3% greater MC error than Approach C ('EM').

Tilted Beta Binomial Model MC Error

- 'Digestible' comparison of MC errors possibly difficult.
- However, note that μ, θ and v are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C ('EM') had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D ('EM') had about a 3% greater MC error than Approach C ('EM').

Tilted Beta Binomial Model MC Error

- ‘Digestible’ comparison of MC errors possibly difficult.
- However, note that μ, θ and *v* are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C (‘EM’) had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D (‘EM’) had about a 3% greater MC error than Approach C (‘EM’).

Tilted Beta Binomial Model MC Error

- ‘Digestible’ comparison of MC errors possibly difficult.
- However, note that μ, θ and v are on common $(0, 1)$ range.
- We average the ratios of MC errors for these parameters.
- Approach C (‘EM’) had about a 52% greater MC error than Approach A.
- Approach B had about a 9% greater MC error than Approach A.
- Approach D (‘EM’) had about a 3% greater MC error than Approach C (‘EM’).

Conclusions and Summary

Estimation Time Conclusions

- ‘EM’ approaches take more run time due to latent z_i, but less problematic with binomial conditional likelihood due to ease of computation.

MC Error Conclusions

- ‘EM’ approaches increase MC error. Marginalizing reduces MC error.
- However, using conditional approach only increases MC error modestly vs. marginalized.

Summary: multiple approaches to estimation of tilted beta binomial models exist and data analysis can proceed with different approaches based on goals.

Conclusions and Summary

Estimation Time Conclusions

- ‘EM’ approaches take more run time due to latent z_i, but less problematic with binomial conditional likelihood due to ease of computation.

MC Error Conclusions

- ‘EM’ approaches increase MC error. Marginalizing reduces MC error.
- However, using conditional approach only increases MC error modestly vs. marginalized.

Summary: multiple approaches to estimation of tilted beta binomial models exist and data analysis can proceed with different approaches based on goals.

Conclusions and Summary

Estimation Time Conclusions

- ‘EM’ approaches take more run time due to latent z_i, but less problematic with binomial conditional likelihood due to ease of computation.

MC Error Conclusions

- ‘EM’ approaches increase MC error. Marginalizing reduces MC error.
- However, using conditional approach only increases MC error modestly vs. marginalized.

Summary: multiple approaches to estimation of tilted beta binomial models exist and data analysis can proceed with different approaches based on goals.

Selected References

- Hahn, E.D. (2008). Mixture densities for project management activity times: A robust approach to PERT. European Journal of Operational Research, 188(2), 450-459.
- Hahn, E.D. (2012). A new flexible distribution for discrete outcomes. Presented at INFORMS Annual Meeting 2012, Phoenix, AZ.
- Hahn, E.D. and López Martín, M.d.M. (2015). Robust project management with the tilted beta distribution. SORT-Statistics and Operations Research Transactions. 39(2), 253-272.
- Skellam, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. Journal of the Royal Statistical Society. Series B (Methodological). 10(2), 257–261.