In the realm of data science, information is transformed, interpreted, and applied to real-world problems by new methods of machine learning every day. One such method, known as Bayesian generalized linear modelling, has helped to create high-resolution maps of disaggregated development in Southeast Asia, predicted indicators of coronary incidents with an incredibly low margin of error, and inculcated preexisting data to profile and protect potential carriers of tuberculosis in South Africa. Originally developed as a solution to a problem of inverse probability in the 18th century, British statistician Thomas Bayes’ theorem of “Bayesian probability” has since evolved to meet the demands of geospatial plotting projects and survey-based studies on human development. Though it is regarded as “one of the most theoretically and computationally challenging problems encountered in practice” (Chen et. al), it is also recognized as an incredibly accurate and efficient form of predictive modelling. In the simplest of terms, the Bayesian modelling approach utilizes a combination of prior – recorded knowledge and data – and likelihood – inferring which value of an unknown parameter is most likely to generate the data on record – to calculate the posterior distribution of a productive input. This technique is considered an extension of linear mixed and generalized linear models, as it considers “dependent variables from [non-normal] distributions” as well as “both fixed and random” outcomes (Ojo et. al). The actual equation involved with this process relies on the connection of a random, or stochastic, component with a separate, systemic component through a link function of *θ*. This can all be represented as a function in the form *P(θ|D) = (P(D|θ)* X *P(θ))/P(D)*, wherein *P(θ)* stands for the prior, *(P(D|θ)* stands for the likelihood of a result given the distribution of *θ*, *P(D)* stands for the probability of a certain data point, and *P(θ|D)* stands for the posterior parameters determined by the equation’s result (“Bayesian Statistics…”).

Bayesian generalized linear modelling has recently found a new purpose in the realm of human development, with data scientists across the globe currently using the method as a basal form of quantifying social progress in locales around the world. For instance, in 2017, a report published by the peer-reviewed *Journal of the Royal Society Interface* detailed the extensive applications of Bayes’ method in the context of development in their study of “gender-disaggregated development indicators” in Kenya, Tanzania, Nigeria, and Bangladesh (Bosco, C. et. al 1). By combining prior geospatial covariates like distances between settlements, health facilities, and schools; satellite indices, topography, pregnancy and neonatal mortality rates, livestock densities, and gross cell products with an “integrated nested Laplace approximations (INLA) approach” (4) to a Bayesian-modelled simulation, the authors of the report were able to accurately map geographic hot spots of low literacy, child stunting, and varied rates of contraceptive usage in four unique locations. The resulting data was accompanied by low measures of uncertainty despite the prior’s original breadth, demonstrating the efficiency and precision of the Bayesian generalized linear model in even the most “computationally intensive” of datasets (4). In a separate study published in the Public Library of Science’s international journal, *PLOS One*, a group of data scientists used Bayesian generalized linear mixed modelling to determine predictive, lifestyle-based indicators of tuberculosis in South Africa. Bayesian-based research of this manner – specialized, intensive, geo-specific, and most importantly, accurate to a letter – could potentially help healthcare facilities and programs identify and care for at-risk individuals, provide evidence for better legislative protections for women in underdeveloped areas, and identify social or political bases for localized malnutrition and stunting in children. Without overstating its importance, Bayesian generalized linear modelling is definitely a data science tool that the world should watch out for.

Works Cited

“Bayesian Statistics explained to Beginners in Simple English.” *Analytics Vidhya*, 20 Jun 2016, <https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/>.

Bakker, Ryan. “Bayesian Methods: Review of Generalized Linear Models.” <https://spia.uga.edu/faculty_pages/rbakker/bayes/Day2.applied.bayes.pdf>.

Bosco, C., et. al. “Exploring the high-resolution mapping of gender-disaggregated development indicators.” *Journal of the Royal Society Interface*, 5 Apr 2017, <http://dx.doi.org/10.1098/rsif.2016.0825>.

Chen, Ming Hui, et. al. “Bayesian Variable Selection and Computation for Generalized Linear Models with Conjugate Priors.” *Bayesian Analysis*, vol. 3, no. 3, 2007, pp. 586-613. *Project Euclid*, <https://projecteuclid.org/euclid.ba/1340370439>.

Ojo, Oluwatobi Blessing, et. al. “Bayesian generalized linear mixed modeling of Tuberculosis using informative priors.” *PLOS One*, 3 Mar 2017, <https://doi.org/10.1371/journal.pone.0172580>.