Secure Systems Design and Technology Development. Veliko Tarnovo, Bulgaria
Proceedings of the International Scientific Conference, 2026. Volume III, 19-22

On Some Properties of Posterior Predictive Distributions

Ralitsa Angelova-Slavova
Department of Communication and Information Systems
"Vasil Levski" National Military University
Veliko Tarnovo, Bulgaria
rlslavova@nvu.bg

Dimiter Tsvetkov
Department of Communication and Information Systems
"Vasil Levski" National Military University
Veliko Tarnovo, Bulgaria
dptsvetkov@nvu.bg

***Abstract* – This article considers the method of posterior predictive distributions as the one of the specific tools for checking the correspondence between data and the statistical model. The generated examples demonstrate that in the role of posterior predictive testing statistics (as a discrepancy measure), dispersion measures should be taken, because the central tendency measures do not give the expected results.**

Keywords – Posterior predictive distributions, Markov chain Monte Carlo (MCMC), Metropolis–Hastings algorithm, Data fitting methods.

I. INTRODUCTION

Let a random sample $x_{(j)} \sim f(x|\theta)$ with a size J be given from some distribution $f(x|\theta)$. The estimation for the parameters θ can be performed using the Bayesian method, which examines the posterior distribution of θ, given prior distribution $f(\theta) = f(\theta|\tau)$ with some known hyper-parameters τ [1], [2], [3] and [4]. The likelihood function has the form

$$f\ (\boldsymbol{x}|\theta) = \prod_j f\left(x_{(j)}|\theta\right)$$

and $f(\boldsymbol{x}|\theta)f(\theta)$ is the joint distribution of the data $\boldsymbol{x} = \left(x_{(j)}\right)$ and the parameters θ. For the posterior distribution of θ we have

$$f(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)f(\theta) = \left(\prod_j f\left(x_{(j)}|\theta\right)\right) f(\theta)$$

In this paper, by $f(\cdot)$ we will denote densities of given distributions, which will distinguish through the structure of the arguments.

II. MATHERIALS AND METHODS

This paper brings more theoretical content, which requires using of proper software. Here we use a specific Markov chain Monte Carlo (MCMC) technique, known as Metropolis – Hastings algorithm. Where necessary, numbers are rounded to the third digit after the decimal point. The form of $f(\theta|\boldsymbol{x})$ usually is not easy to find. Its relief however can be achieved by mentioned MCMC method after generating a sufficient number of observations [5], [6], [7] and [8]. The popular Metropolis – Hastings algorithm with independent choice offers the following option. For the working distribution $q(\theta|\theta')$ we set the prior $f(\theta)$ and the target distribution $\pi(\theta)$ is of course the posterior one

$$\pi(\theta) = f(\theta|\boldsymbol{x}) \propto \left(\prod_j f\left(x_{(j)}|\theta\right)\right) f(\theta)$$

The algorithm scheme looks like this. Start by choosing an initial observation $\theta_{(0)}$. Then for a known $\theta_{(n-1)}$ do the following.

1. Select a candidate $\theta_* \sim f(\theta)$ and calculate

$$\boldsymbol{a} = \min\left(1, \frac{\prod_j f\left(x_{(j)}|\theta_*\right)}{\prod_j f\left(x_{(j)}|\theta_{(n-1)}\right)}\right)$$

2. Generate $u \sim U(0,1)$. If $0 < u < \boldsymbol{a}$ the candidate is accepted and put $\theta_{(n)} = \theta_*$. Otherwise the candidate is rejected and put $\theta_{(n)} = \theta_{(n-1)}$.

In this way we obtain a random sample of observations $\left(\theta_{(n)}\right)$ from the posterior distribution $f(\theta|\boldsymbol{x})$. Their local dependence can be mitigated by including in the sample only the members with index $n = ki$, for some fixed $k \in \mathbb{N}$ [9].

Example 1. Let $x_{(j)} \sim Poisson(\lambda)$ be a random sample with size J from a Poisson distribution with parameter $\lambda > 0$. Then

$$f(\boldsymbol{x}|\lambda) = \prod_{j=1}^{J} e^{-\lambda} \frac{\lambda^{x_{(j)}}}{x_{(j)}!} \propto e^{-J\lambda}\lambda^{J\bar{x}}$$

Online ISSN 3044-7771
https://doi.org/10.68302/std2026.vol3.37

About the parameter λ we assume a gamma prior distribution $\lambda \sim Gamma(\alpha, \beta)$, where α (shape) and β (rate) are beforehand chosen hyper-parameters. For the joint density of $\boldsymbol{x}$ and λ we have

$$f(\boldsymbol{x}, \lambda) = \left(\prod_{j=1}^{J} e^{-\lambda} \frac{\lambda^{x_{(j)}}}{x_{(j)}!} \right) \frac{\beta^{\alpha} \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)}$$

A random sample of size $J = 100$ was taken from $Poisson(3)$ distribution. The M-H algorithm here is specified as follows. Assume $\alpha = \overline{\boldsymbol{x}}$ and $\beta = 1$, for which the mean of the prior distribution coincides with the mean of the data. Choose $\lambda_{(0)} \sim Gamma(\overline{\boldsymbol{x}}, 1)$. Let the observation $\lambda_{(n-1)}$ be known. Then for the next observation $\lambda_{(n)}$ do the following.

1. Select a candidate $\lambda_* \sim Gamma(\overline{\boldsymbol{x}}, 1)$ and calculate

$$\boldsymbol{a} = \min\left(1, \frac{e^{-J\lambda_*} \lambda_*^{J\overline{x}}}{e^{-J\lambda_{(n-1)}} \lambda_{(n-1)}^{J\overline{x}}}\right)$$

2. Generate $u \sim U(0,1)$. If $0 < u < \boldsymbol{a}$ the candidate is accepted and put $\lambda_{(n)} = \lambda_*$. Otherwise the candidate is rejected and put $\lambda_{(n)} = \lambda_{(n-1)}$.

Perform $m = 1\,000$ initial iterations and yet $N = 1\,000\,000$ basic iterations, from which we take every 100-th in a row [9], obtaining a random sample of size $10\,000$. On the other hand

$$f(\boldsymbol{x}, \lambda) \propto e^{-(\beta+J)\lambda} \lambda^{(\alpha + x_{(1)} + \cdots + x_{(J)} - 1)}$$

whence the posterior distribution $f(\lambda|\boldsymbol{x})$ appears as the following gamma distribution

$$Gamma\left(\alpha + x_{(1)} + \cdots + x_{(J)}, \beta + J\right)$$

Fig. 1 shows a very good practical fit between the histogram of the MCMC sample and the corresponding theoretical gamma distribution $[ChiSquare(9) = 7.582; p - value = 0.577]$. The following Table 1 contains sample estimated characteristics of the posterior distribution $f(\lambda|\boldsymbol{x})$, including the hyper-parameters $shape$ and $rate$.

TABLE 1. COMPARISON BETWEEN VALUES.

	Shape	Rate	Mean	SD
Theoretical	312.090	101	3.090	0.175
Empirical	315.793	102.247	3.089	0.174

III. RESULTS AND DISCUSSION

Let data observations $\boldsymbol{x}_{obs} = \left(o_{(j)}\right)$ be given, for which a statistical model with a likelihood function $f(\boldsymbol{x}|\theta)$ is assumed with certain parameters θ. Posterior predictive distributions are formed by generating new possible data $\boldsymbol{x}_{rep}$, according to the observed data $\boldsymbol{x}_{obs}$ [10], [11] and [12]. This new possible data is generated by means of the formula

$$\boldsymbol{x}_{rep(n)} \sim f(\boldsymbol{x}|\boldsymbol{x}_{obs}) = \int f(\boldsymbol{x}|\theta) f(\theta|\boldsymbol{x}_{obs}) d\theta$$

where $f(\theta|\boldsymbol{x}_{obs})$ is the posterior distribution of θ. Any observation $\theta_{(n)} \sim f(\theta|\boldsymbol{x}_{obs})$ gives a response $\boldsymbol{x}_{rep(n)}$ by random selection $\boldsymbol{x}_{rep(n)} \sim f\left(\boldsymbol{x}|\theta_{(n)}\right)$.

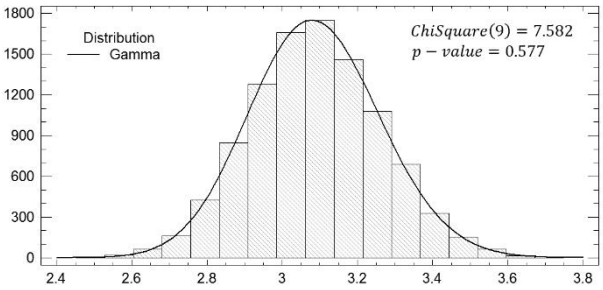

Fig 1. Explanation to Example 1.

Let $g(\boldsymbol{x})$ be some test statistic for the model $f(\boldsymbol{x}|\theta) f(\theta)$. Then $\left(g\left(\boldsymbol{x}_{rep(n)}\right)\right)$ are observations for the random variable $g(\boldsymbol{x})$. The belonging of the particular value $g(\boldsymbol{x}_{obs})$ far from the center of the empirical distribution of $\left(g\left(\boldsymbol{x}_{rep(n)}\right)\right)$ indicates a misfit between the model and the data.

The reader may see [13], [14], [15] and [16], where this approach is rigorously grounded.

Example 2. In Example 1, $\boldsymbol{x}_{obs}$ are actually generated from $Poisson(3)$ distribution. Let the test statistic be the standard deviation $\sigma(\boldsymbol{x})$. We perform $10\,000$ data generations $\left(\boldsymbol{x}_{rep(n)}\right)$, based on the one constructed in Example 1 sample from $f(\lambda|\boldsymbol{x}_{obs})$. The distribution of standard deviations is presented in the following Fig. 2.

For the observed data, we have $\sigma(\boldsymbol{x}_{obs}) = 1.787$, which represents approximately the 61% quantile for the empirical posterior predictive distribution of $\left(\sigma\left(\boldsymbol{x}_{rep(n)}\right)\right)$ observations. This value is close to the center of the distribution and indicates a good enough fit between data and the assumed model, which is an expected result, since the data are actually generated from a Poisson distribution.

Example 3. Go back to Example 1, but this time let data be generated from a uniform distribution between the integers 0 and 9. Again, we will look for a model with a Poisson distribution, and as in Example 2, we will examine the posterior predictive distribution of $\sigma\left(\boldsymbol{x}_{rep}\right)$. We repeat the scheme from the previous two examples with 10000 observations $\left(\boldsymbol{x}_{rep(n)}\right)$. The empirical distribution of the corresponding standard deviations is shown in Fig. 3.

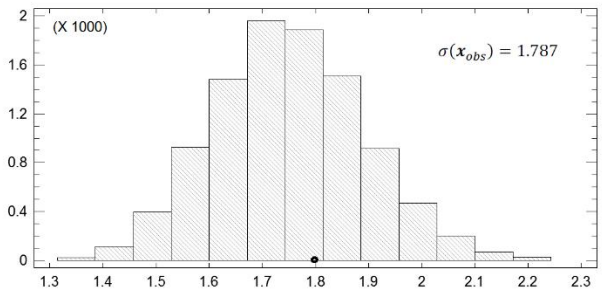

Fig 2. Histogram of $\left(\sigma\left(\boldsymbol{x}_{rep(n)}\right)\right)$ for Example 2.

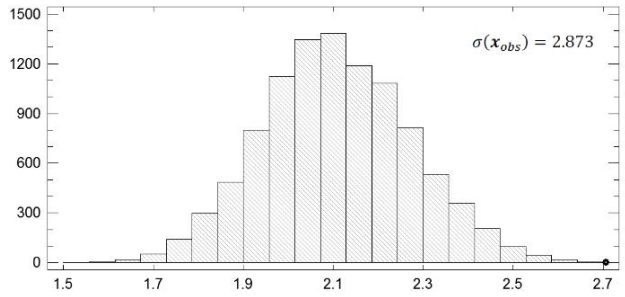

Fig. 3. Histogram of $\left(\sigma(\boldsymbol{x}_{rep(n)})\right)$ for Example 3.

For the observed data, we have $\sigma(\boldsymbol{x}_{obs}) = 2.873$, which lies far in the right tail of the posterior predictive distribution. This value indicates a poor fit between the observed data and the actually assumed model.

Example 4. Let $x_{(j)} \sim Geometric(p)$ be a random sample with size J from the (zero based) geometric distribution with parameter $0 < p < 1$. Then

$$f(\boldsymbol{x}|p) = \prod_{j=1}^{J}(1-p)^{x_{(j)}}p = p^J(1-p)^{x_{(1)}+\cdots+x_{(J)}} = p^J(1-p)^{J\overline{x}}$$

The parameter p is assumed to have beta prior distribution $p \sim Beta(\alpha, \beta)$, where the hyper-parameters α (shape) and β (shape) are chosen in advance. For the joint density of $\boldsymbol{x}$ and p we have

$$f(\boldsymbol{x}, p) \propto p^{\alpha+J-1}(1-p)^{\beta+J\overline{x}-1} \sim \mathrm{B}(\alpha + J, \beta + J\overline{\boldsymbol{x}})$$

For this Example 4 is realized a random sample of size $J = 100$ from distribution $Geometric(0.1)$. The M-H algorithm here is specified as follows. Assume $\alpha = 1$ and $\beta = 1 + \overline{\boldsymbol{x}}$, under which the mean of the prior distribution coincides with the MLE estimate of p from the observed data.

Choose $p_{(0)} \sim Beta(1, 1 + \overline{\boldsymbol{x}})$ and suppose the observation $p_{(n-1)}$ is known. Then for the next one $p_{(n)}$ do the following.

1. Select a candidate $\lambda_* \sim Gamma(\overline{\boldsymbol{x}}, 1)$ and calculate

$$\boldsymbol{a} = \min\left(1, \frac{p_*^{\alpha+J-1}(1-p_*)^{\beta+J\overline{x}-1}}{p_{(n-1)}^{\alpha+J-1}\left(1-p_{(n-1)}\right)^{\beta+J\overline{x}-1}}\right)$$

2. Generate $u \sim U(0,1)$. If $0 < u < \boldsymbol{a}$ the candidate is accepted and set $p_{(n)} = p_*$. Otherwise, the candidate is rejected and set $p_{(n)} = p_{(n-1)}$.

Now perform $m = 1000$ initial iterations and yet more $N = 1\,000\,000$ basic iterations, from which take every 100-th in turn. In this way we get a good random sample from $f(\lambda|\boldsymbol{x})$ with a size $10\,000$. The posterior distribution $f(p|\boldsymbol{x})$ is beta $\mathrm{B}(\alpha + J, \beta + J\overline{\boldsymbol{x}})$.

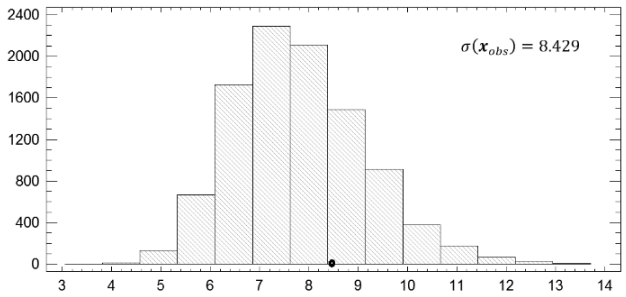

Fig. 4. Histogram of $\left(\sigma(\boldsymbol{x}_{rep(n)})\right)$ for Example 4.

Let, as in Example 2, the test statistic be the standard deviation $\sigma(\boldsymbol{x})$ and let perform $10\,000$ data generations. $\left(\boldsymbol{x}_{rep(n)}\right)$. The distribution of standard deviations is given in Fig. 4.

For the observed data, we have $\sigma(\boldsymbol{x}_{obs}) = 8.429$, which represents approximately the 71% quantile for the empirical posterior predictive distribution This value is close to the center of the distribution and indicates a good fit between the observed data and the model, which is an expected result since the data are actually generated from a geometric distribution.

Example 5. Consider again Example 4, but this time the data is generated from a uniform distribution between the integers 0 to 9. Again, we will look for a model with a geometric distribution, and as in example 4, we will examine the posterior predictive distribution of $\sigma(\boldsymbol{x}_{rep})$. We repeat the scheme from the previous examples with $10\,000$ generations $\left(\boldsymbol{x}_{rep(n)}\right)$. The following Fig. 5 shows the empirical predictive distribution of the standard deviations.

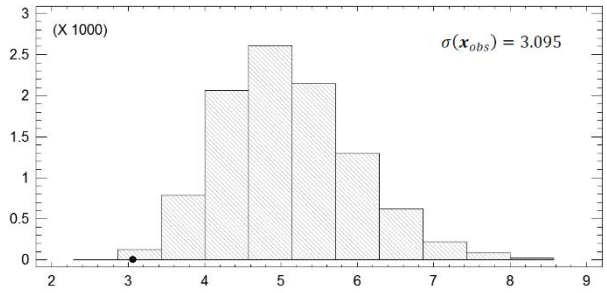

Fig. 5. Histogram of $\left(\sigma(\boldsymbol{x}_{rep(n)})\right)$ for Example 5.

For the observed data, we have $\sigma(\boldsymbol{x}_{obs}) = 3.095$, which is approximately the 0.15% quantile for the empirical posterior predictive distribution and lies far from the center of the distribution. This fact confirms the misfit between the observed data and the assumed model.

The above examples raise a further discussion about the extent to which the size of the empirical sample affects the results. The question of finding a wider variety of test statistics that highlight the lack of fit between the data and the model also remains under further discussion.

IV. Conclusions

Posterior predictive distributions provide a unique method for testing the fit between the observed data and the corresponding statistical model. Some measure of dispersion should be chosen as a test statistic, because the measures of central tendency are not sufficiently consistent with the method, described above. More information about this approach, the reader can find for example in [13], [14], [15] and [16].

One may conclude that unimodal distributions are similar in their central areas, and the differences appear in the tails of their distributions.

V. Acknowledgments

The authors are very grateful to the department colleagues for the support of this work, which was partially developed in the framework with the national project BG05M2OP001-2.016-0003 "Modernization of the National Military University "Vasil Levski" and Sofia University "St. Kliment Ohridski".

REFERENCES

[1] M. Aitkin, Statistical Inference: An Integrated Bayesian/Likelihood Approach (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/EBK1420093438

[2] I. Ntzoufras, Bayesian modeling using WinBUGS, Hoboken, N. J: Wiley, 2009.

[3] P. Congdon, Bayesian Models for Categorical Data, New York: Wiley, 2005.

[4] J. Albert, Bayesian Computation with R, 2nd ed., New York: Springer, 2009. https://doi.org/10.1007/978-0-387-92298-0

[5] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys., vol. 21, pp. 1087-1091, 1953. https://doi.org/10.1063/1.1699114

[6] C. P. Robert and G. Casella, Monte Carlo statistical methods, New York: Springer, 2004.

[7] C. P. Robert and G. Casella, Introducing Monte Carlo methods with R, New York: Springer, 2010.

[8] R. Angelova-Slavova, Convergence and Applications of the Metropolis-Hasting algorithm, Plovdiv: Astarta, 2020.

[9] W. Link and M. Eaton, M. "On thinning of chains in MCMC," Methods in Ecology and Evolution, vol. 3, pp. 112–115, 2012. https://doi.org/10.1111/j.2041-210X.2011.00131.x

[10] A. Gelman, X.-L. Meng and H. Stern, "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies," Statistica Sinica, vol. 6(4), pp. 733-807, 1996. [Online]. Oxford Reference Online, https://sites.stat.columbia.edu/gelman/research/published/A6n41.pdf, [Accessed December 06, 2025].

[11] M. A. Tanner and W. H. Wong, "The Calculation of Posterior Distributions by Data Augmentation," Journal of the American Statistical Association, vol. 82(398), pp. 528-540, 1987.

[12] M. M. Barbieri and J. O. Berger, "Optimal predictive model selection," Ann. Statist., vol. 32(3), pp. 870-897, 2004. https://doi.org/10.1214/009053604000000238

[13] X.-L. Meng, "Posterior predictive p-values", The Annals of Statistics, vol. 22(3), pp. 1142-1160, 1994.

[14] T. Ando, Bayesian Model Selection and Statistical Modeling, Taylor and Francis Group, 2010.

[15] J.-P. Fox, Bayesian Item Response Modeling. Theory and Applications, New-York: Springer, 2010. https://doi.org/10.1007/978-1-4419-0742-4

[16] C. A. Stone and Z. Xiaowen, Bayesian Analysis of Item Response Theory Models Using SAS, Cary, NC: SAS Institute Inc., 2015.