

YOUR QUESTIONS ANSWERED

For frequentist approach, the sample size is quite important to decide the reading power using PROC POWER. What about the Bayesian approach? How do you decide the sample size for hypothesis tests?

Here are some suggested articles about Bayesian power.

Wang, Fei and Alan E. Gelfand. "A Simulation-Based Approach to Bayesian Sample Size Determination for Performance Under a Given Model and for Separating Models." *Statistical Science* 17, no. 2 (2002): 193-208.

Spiegelhalter, David J., Keith R. Abrams and Jonathan P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Vol. 13. John Wiley & Sons, 2004.

De Santis, Fulvio and Stefania Gubbiotti. "Sample Size Requirements for Calibrated Approximate Credible Intervals for Proportions in Clinical Trials." *International Journal of Environmental Research and Public Health* 18, no. 2 (2021): 595.

How can Bayesian statistics be combined with statistical process control?

For more information, please read [this paper](#).

Are there any rules of thumb we can apply when looking at the effective sample sizes? Any tips on judging what is an adequate ESS and does it vary depending on the outcome (continuous vs. categorical)? In the [SAS examples](#), ESS for SLR was the same as nmc but was much lower for the multiple logistic example (nmc=20,000 vs ess=[401.8-1054.7]), which was determined to be "adequate."

I prefer using the AutoCorrelation Time in that table. This value is the length of the saved chain divided by the ESS. I put a cut value of this diagnostic. For example, in my work I have used 3 as a bound. Individual areas of research have different viewpoints of bounds here.

Are informed priors ever considered biased?

Priors from subject-matter experts can easily be considered biased. Some of these biases can be incorporated into the model and we need to be on the lookout for all of them during the analysis. Sensitivity analyses can help here.

For someone who is starting their data science career, where should I start with Bayesian statistics?

Start with basic model types like simple linear regression. Then move onward to the more elaborate types of analyses. As you learn how to develop the Bayesian viewpoint for that model type, then move forward.

Is Bayesian analysis often used in causal inference analysis? Are there any caveats in this regard?

[You can find more information in this paper.](#)

As someone who is only starting to get into SAS® code, is there a path to use Bayesian statistics through a SAS GUI such as VA reporting objects?

PROC MCMC, being a very flexible procedure with fairly open-ended modeling capabilities, although it is not that easy to make a generic catch-all GUI for Bayesian analysis. However, you could create your own SAS Studio task to build simpler models that you use frequently. [This site](#) can help you learn to develop a SAS Studio task.

What other applications are there? Forecasting sales/revenue in a startup business or new industry seems a natural, given little prior data. Does that make sense? Other areas?

The number of applications for Bayesian are always expanding. Don't focus only on areas where the amount of data is restricted. That is not a requirement. As Bayesian gains popularity, we are seeing it expand into many other areas. I have started seeing its entrance into time series and areas of machine learning.

How are priors adjusted in a temporal sense? For instance, if a prior is based on one identical event versus multiple identical events.

To allow for temporal aspects, you could use one of the multivariate distributions that are available within PROC MCMC or custom write your own likelihood using GENERAL or DGENERAL. Included as part of PROC MCMC are matrix capabilities similar to that of SAS/IML® software. These could also be helpful.

Suppose I have a data set but I don't have the explicit probability distribution of it. Is there a way to construct an empirical distribution that approximates the "true" one? Could you point out a reference or some keywords so I can dig deeper?

If you have a distribution that is not considered "typical" or "known," you can build the likelihood to the distribution of interest using the GENERAL or DGENERAL statement. (GENERAL is for new continuous distributions and DGENERAL is for new discrete distributions.)

Could you please give me performance information of SAS MCMC procedure? Also, could you comment on MCMC convergence speed on sparse data?

Here is a [GitHub link](#) that may help with the understanding of how the performance of MCMC was enhanced as updates were made.

[And a link about sparse data.](#)

You said you prefer frequentist approach sometimes, but Bayesian other times. You've mentioned large sample sizes being more for frequentist. Are there other issues that affect your choice?

Correction to my comment: Larger sample sizes cause the results of the Bayesian to mimic the results of the frequentist due to the data taking over in the mixing with the priors. A large enough sample will

wash out any prior. To answer the preference question, to me the question at hand is the determination of which would be useful in practice. Bayesian analysis can answer different styles of questions than the frequentist thanks to the perspective on the parameters. Also, with Bayesian not needing to deal with asymptotes, when sample sizes get to a point where this breaks down within frequentist approaches the Bayesian is happy to continue without issue. The main use of Bayesian is the inclusion of information other than the data.

