

Prior and posterior predictive checking

- Bayesian Data Analysis, 3rd ed, Chapter 6
- Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman (2018). Visualization in Bayesian workflow. Journal of the Royal Statistical Society Series A, accepted for publication as discussion paper. [arXiv preprint arXiv:1709.01449](#).
- Graphical posterior predictive checks using the bayesplot package
<http://mc-stan.org/bayesplot/articles/graphical-ppcs.html>
- `demo demos_rstan/ppc/poisson-ppc.Rmd`
- Workflow with prior and posterior predictive checking
https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Replicates vs. future observation

- Predictive \tilde{y} is the next not yet observed possible observation. y^{rep} refers to replicating the whole experiment (with same values of x) and obtaining as many replicated observations as in the original data.

Posterior predictive checking

- Data y
- Parameters θ
- Replicated data y^{rep}
 - assume that the data has been generated by a process which can be well described by the model M with parameters θ
 - replicated data could be observed if the experiment were repeated
 - replace “true” data generating process by the model

$$p(y^{\text{rep}}|y, M) = \int p(y^{\text{rep}}|\theta, M)p(\theta|y, M)d\theta$$

Posterior predictive checking

- Data y
- Parameters θ
- Replicated data y^{rep}
 - assume that the data has been generated by a process which can be well described by the model M with parameters θ
 - replicated data could be observed if the experiment were repeated
 - replace “true” data generating process by the model

$$p(y^{\text{rep}}|y, M) = \int p(y^{\text{rep}}|\theta, M)p(\theta|y, M)d\theta$$

- Graphical checking: compare distribution plots

Posterior predictive checking

- Data y
- Parameters θ
- Replicated data y^{rep}
 - assume that the data has been generated by a process which can be well described by the model M with parameters θ
 - replicated data could be observed if the experiment were repeated
 - replace “true” data generating process by the model

$$p(y^{\text{rep}}|y, M) = \int p(y^{\text{rep}}|\theta, M)p(\theta|y, M)d\theta$$

- Graphical checking: compare distribution plots
- Test quantity (or discrepancy measure) $T(y, \theta)$
 - summary quantity used to compare the observed data and replicates from the predictive distribution

Prior predictive checking

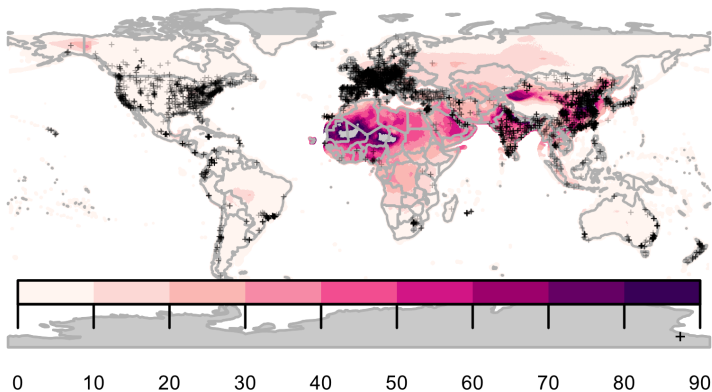
- Similar to posterior predictive checking but without data
- Draws from the prior predictive distribution are compared to external information

Example of prior predictive checking

- Gabry et al (2017). Visualization in Bayesian workflow.
 - Estimation of human exposure to air pollution from particulate matter measuring less than 2.5 microns in diameter ($PM_{2.5}$)
 - A recent report estimated that $PM_{2.5}$ is responsible for three million deaths worldwide each year (Shaddick et al, 2017)

Example of prior predictive checking

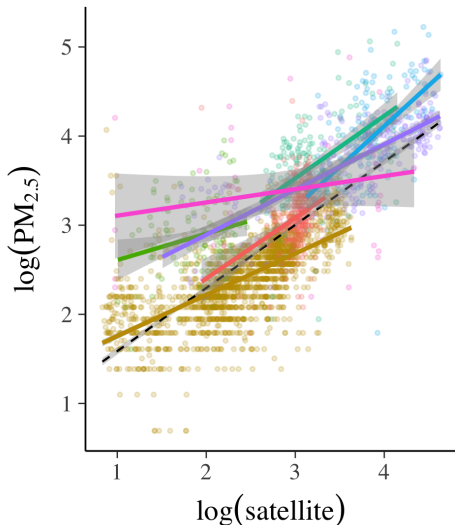
- Gabry et al (2017). Visualization in Bayesian workflow.



Satellite estimates and ground monitor locations

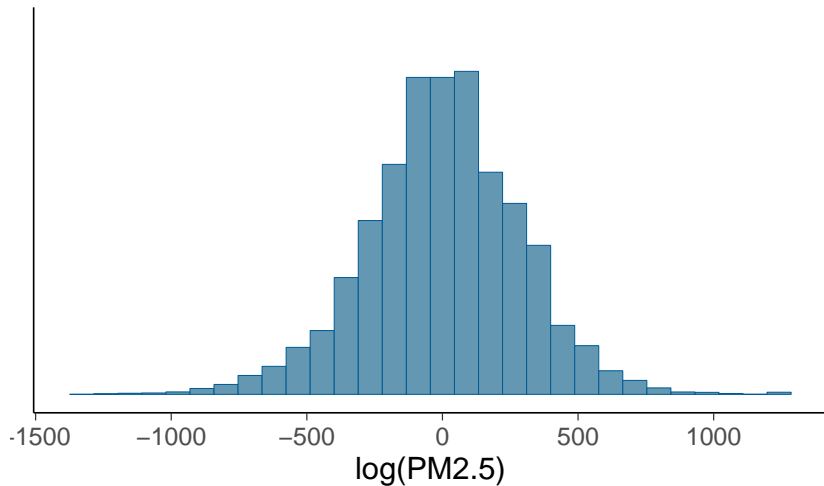
Example of prior predictive checking

- Gabry et al (2017). Visualization in Bayesian workflow.



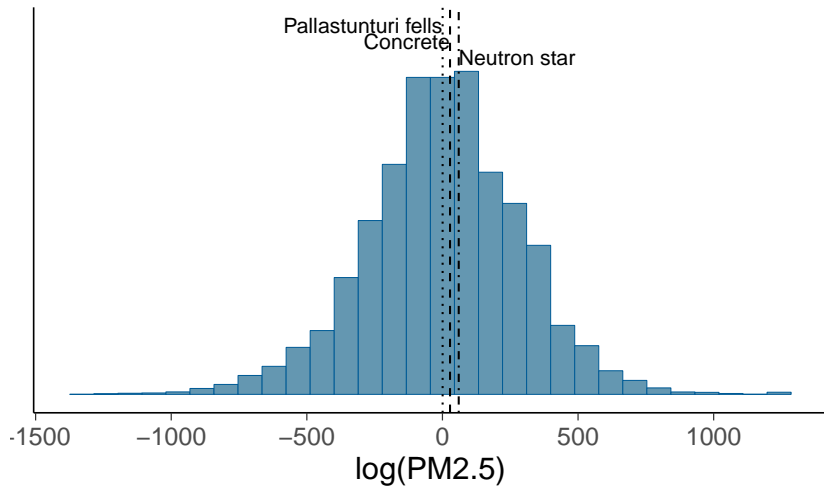
Example of prior predictive checking

Prior predictive distribution with vague prior



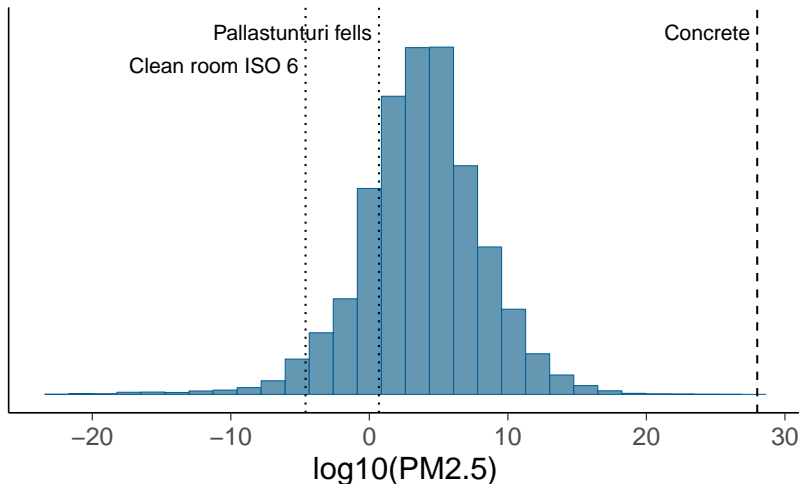
Example of prior predictive checking

Prior predictive distribution with vague prior



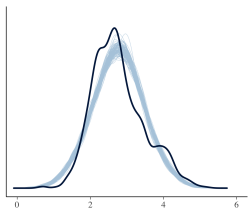
Example of prior predictive checking

Prior predictive distribution with weakly informative

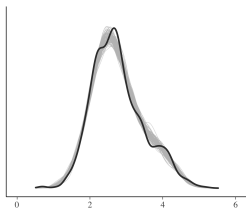


Example of posterior predictive checking

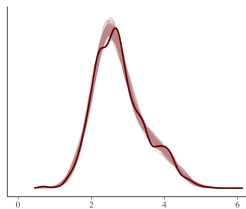
Posterior predictive checking – predictive distributions



(a) Model 1



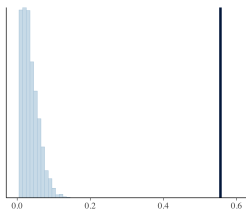
(b) Model 2



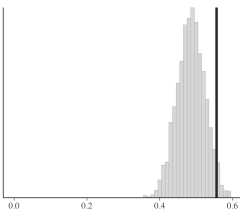
(c) Model 3

Example of posterior predictive checking

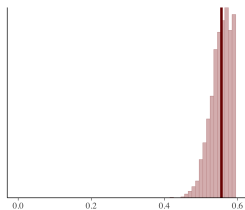
Posterior predictive checking – test statistic (skewness)



(a) Model 1



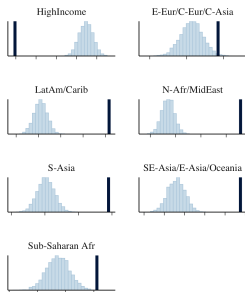
(b) Model 2



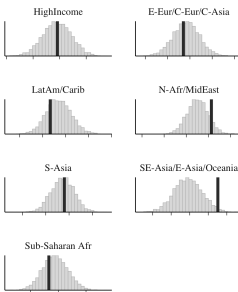
(c) Model 3

Example of posterior predictive checking

Posterior predictive checking – test statistic (median for groups)



(a) Model 1



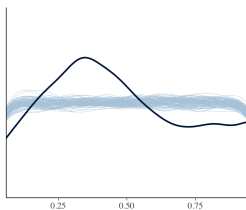
(b) Model 2



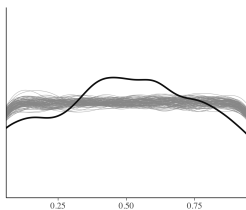
(c) Model 3

Example of posterior predictive checking

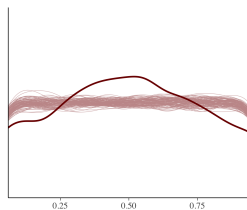
LOO predictive checking – LOO-PIT



(a) Model 1



(b) Model 2



(c) Model 3

Posterior predictive checking

- demo demos_rstan/ppc/poisson-ppc.Rmd

```
data {  
  int<lower=1> N;  
  int<lower=0> y[N];  
}  
parameters {  
  real<lower=0> lambda;  
}  
model {  
  lambda ~ exponential(0.2);  
  y ~ poisson(lambda);  
}  
generated quantities {  
  real log_lik[N];  
  int y_rep[N];  
  for (n in 1:N) {  
    y_rep[n] = poisson_rng(lambda);  
    log_lik[n] = poisson_lpmf(y[n] | lambda);  
  }  
}
```