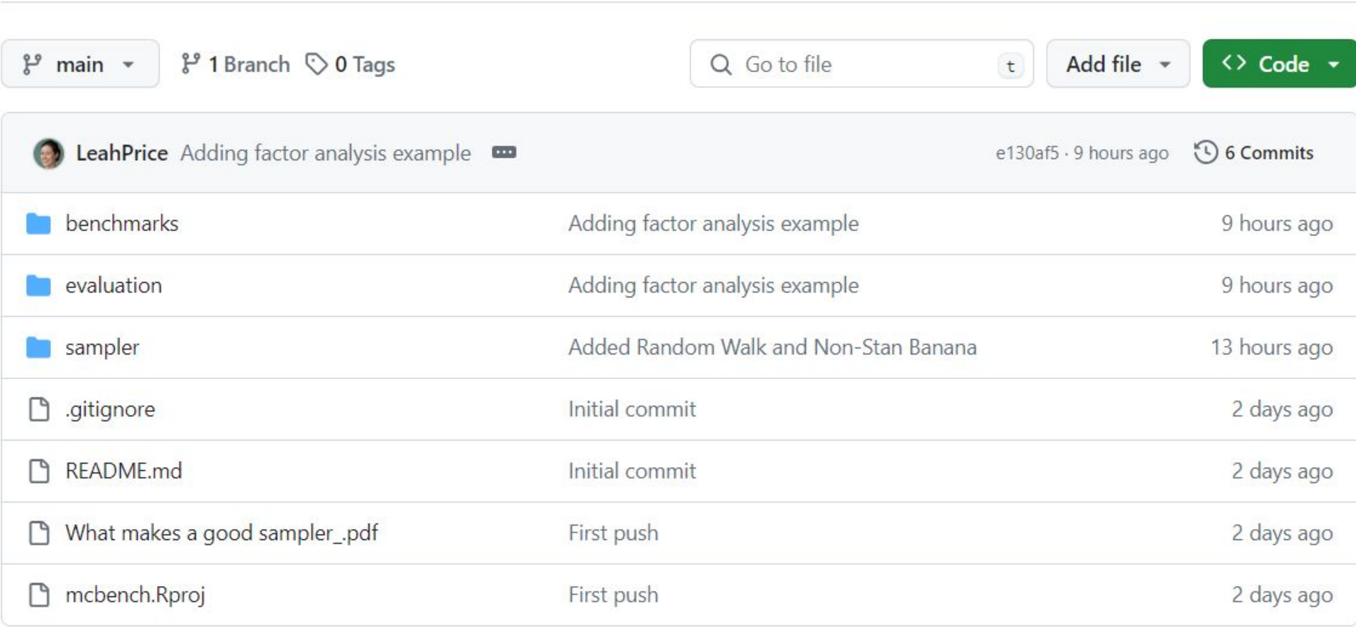
# What makes a good sampler?

A BOB Challenge!

# Github repo for assessing samplers

1. Benchmarks - example test problems
2. Evaluation metrics / plotting



# The Benchmarks

1. R script return: log_prob, grad_log_prob, samples
2. Heavy tails, multi-modality, complex dependencies

Example Ideas:

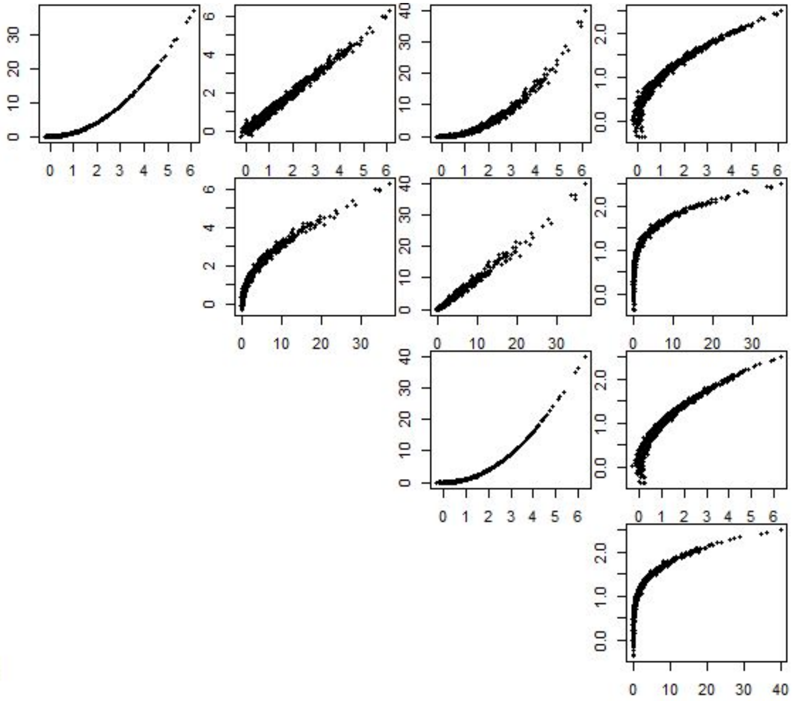
- multi modal
- Stoch Vol
- Anisometric
- Heavy Tail
- Complicated Joints
- Neural Net
- Spectral (Matt Moores)
- Factor Analysis (Leah)
- Twin Moons
- High Dim
- Neal's Funnel

```r
banana_info <- get_banana(mu = 1, a = 1, b = matrix(5,2,2))

banana_info$log_p(x0)
banana_info$grad_log_p(x0)
banana_info$samples(nsamples = 100)
```

# Benchmarks - Starter (5D-Banana)

1. Hybrid Rosenbrock
2. Narrow ridge / curved



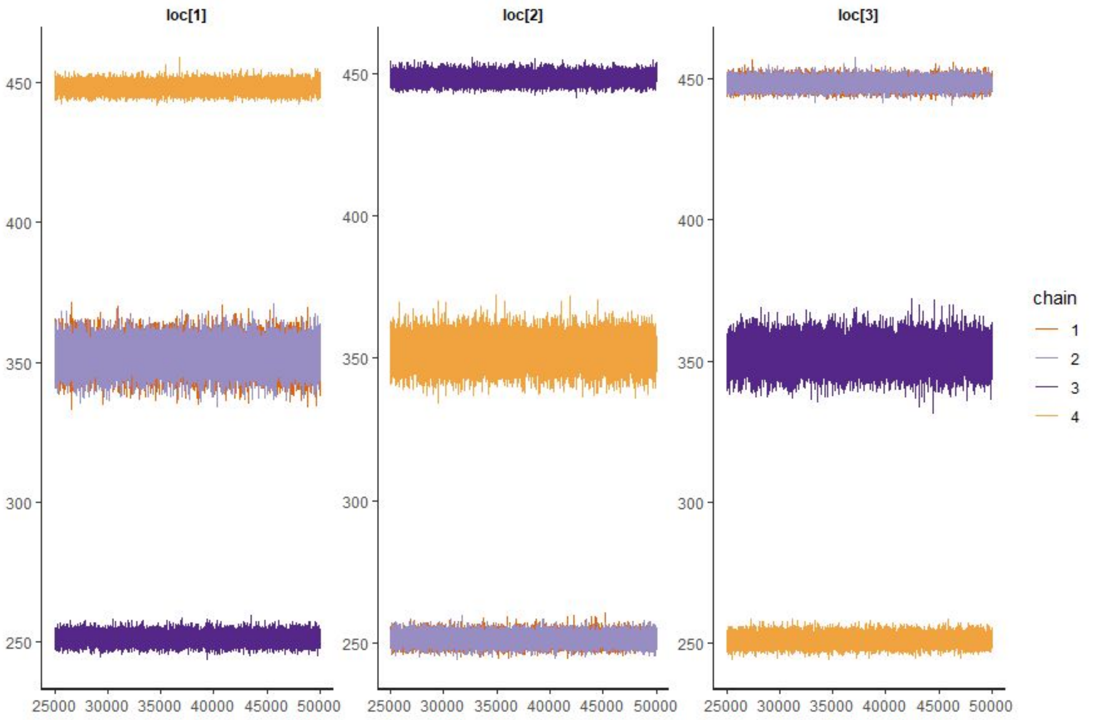
```
data {
  real mu;
  real a;
  int <lower=0> n1;
  int <lower=0> n2;
  matrix[n2,n1] b;
}

parameters {
  vector[n1*n2 + 1] X;
}

model {
  X[n1*n2 + 1] ~ normal(mu, 1/(2*a));
  for ( j in 1:n2 ) {
    X[(j-1)*n1+1] ~ normal(X[n1*n2 + 1]^2, 1/(2*b[j,1]));
    for ( i in 2:n1 ) {
      X[(j-1)*n1+i] ~ normal(X[(j-1)*n1+i-1]^2, 1/(2*b[j,i]));
    }
  }
}
```

# Benchmarks - Matt Moores (synthetic Raman spectrum)

1. Multi-Modal
2. Raman spectroscopy identifies molecules, like DNA, through laser light scattering.




**Matt** 12:16 PM

https://CRAN.R-project.org/package=serrsBayes

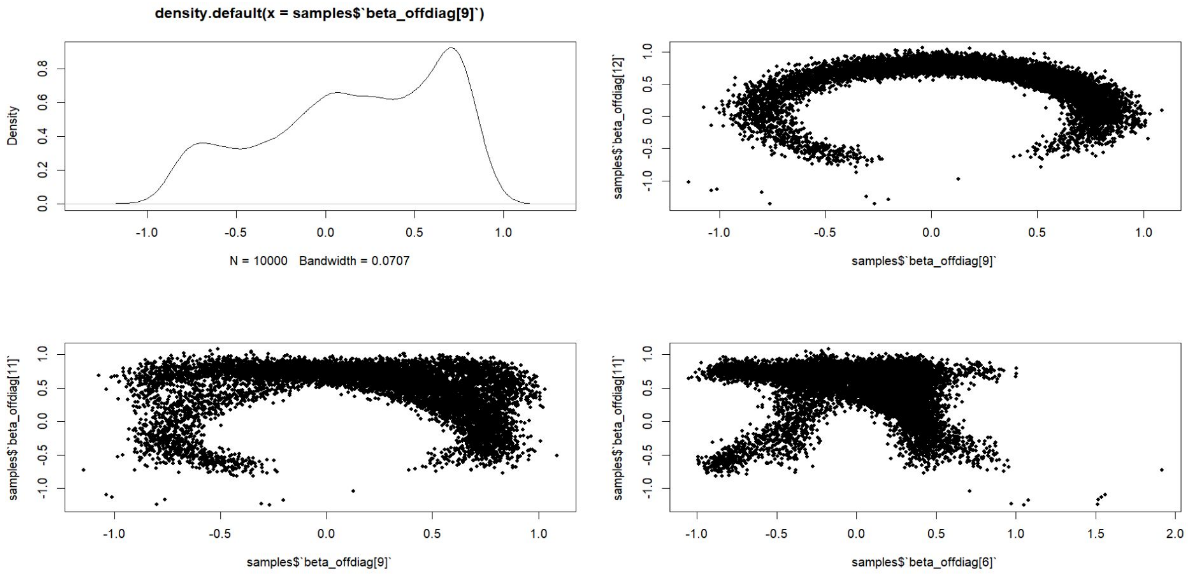
**cran.r-project.org**

**serrsBayes: Bayesian Modelling of Raman Spectroscopy**

Sequential Monte Carlo (SMC) algorithms for fitting a generalised additive mixed model (GAMM) to surface-enhanced resonance Raman spectroscopy (SERRS), using

# Benchmarks - Leah South (Factor Analysis)

1. Complex dependencies
2. Data from monthly exchange of 6 currency rates relative to the pound
3. Model Comparison



# Metrics (comparing samplers)

Metrics:
Bayes LOO
LPPD
KSD
Scoring
C2ST

ess_metrics.R

```r
# Multivariate ESS
get_multiESS <- function(samples){
  return(mcmcse::multiESS(samples))
}

# Min ESS
get_minESS <- function(samples){
  return(min(mcmcse::ess(samples)))
}

# log_p ESS
get_logpESS <- function(samples, log_p){
  lp_samp <- apply(samples, MARGIN = 1, log_p)
  return(mcmcse::ess(lp_samp))
}

```

## Energy Score (thanks Adam Bretherton)

```r
energy_score <- function(Pdist, Qdist, beta=1){
    # ES is a strictly proper scoring rule, that is ES(P, Q) = ES(Q, Q) iff
    # P = Q
    if (is.vector(Pdist)){
        n <- length(Pdist)
        m <- length(Qdist)
    } else {
        n <- dim(Pdist)[1]
        m <- dim(Qdist)[1]
    }
    # Generate permutation vector
    perm <- sample(c(1:n))
    escore <- 0
    for (i in c(1:m)){
        if (is.vector(Pdist)){
            # Single param
            escore = escore + score(Pdist, Qdist[i], perm, beta)
        } else {
            # Multiple params
            escore = escore + score(Pdist, Qdist[i,], perm, beta)
        }
    }
    return(escore/m)
}
```

# Metrics (Assessing samples)

**Matt** 7:13 AM

some metrics for RStan and PyStan https://github.com/betanalpha/mcmc_diagnostics

> **GitHub**
>
> **GitHub - betanalpha/mcmc_diagnostics: Markov chain Monte Carlo general, and Hamiltonian Monte Carlo specific, diagnostics for Stan**
>
> Markov chain Monte Carlo general, and Hamiltonian Monte Carlo specific, diagnostics for Stan - GitHub - betanalpha/mcmc_diagnostics: Markov chain Monte Carlo general, and Hamiltonian Monte Carlo sp... (154 kB)

**Adam Bretherton** 12:41 PM

Could also consider Scoring rules for comparing to the "true" distribution, instead of relying on MSE/Bias.
https://sites.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf

👍 1

**Kenyon Ng** 1 day ago

Kinetic and configurational temperature diagnostics (Appendix I): https://arxiv.org/pdf/2002.02405.pdf

👀 1 👍 2

1 reply

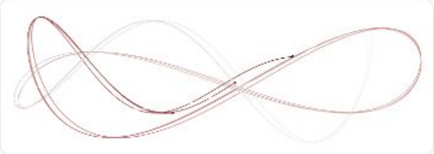
**Kenyon Ng** 1 day ago

Kinetic temperature estimator: Eq 40, Configurational temperature estimator: Eq 43 (edited)

**Kenyon Ng** 18 hours ago

https://mc-stan.org/users/interfaces/shinystan

> **stan-dev.github.io**
>
> ShinyStan (153 kB)

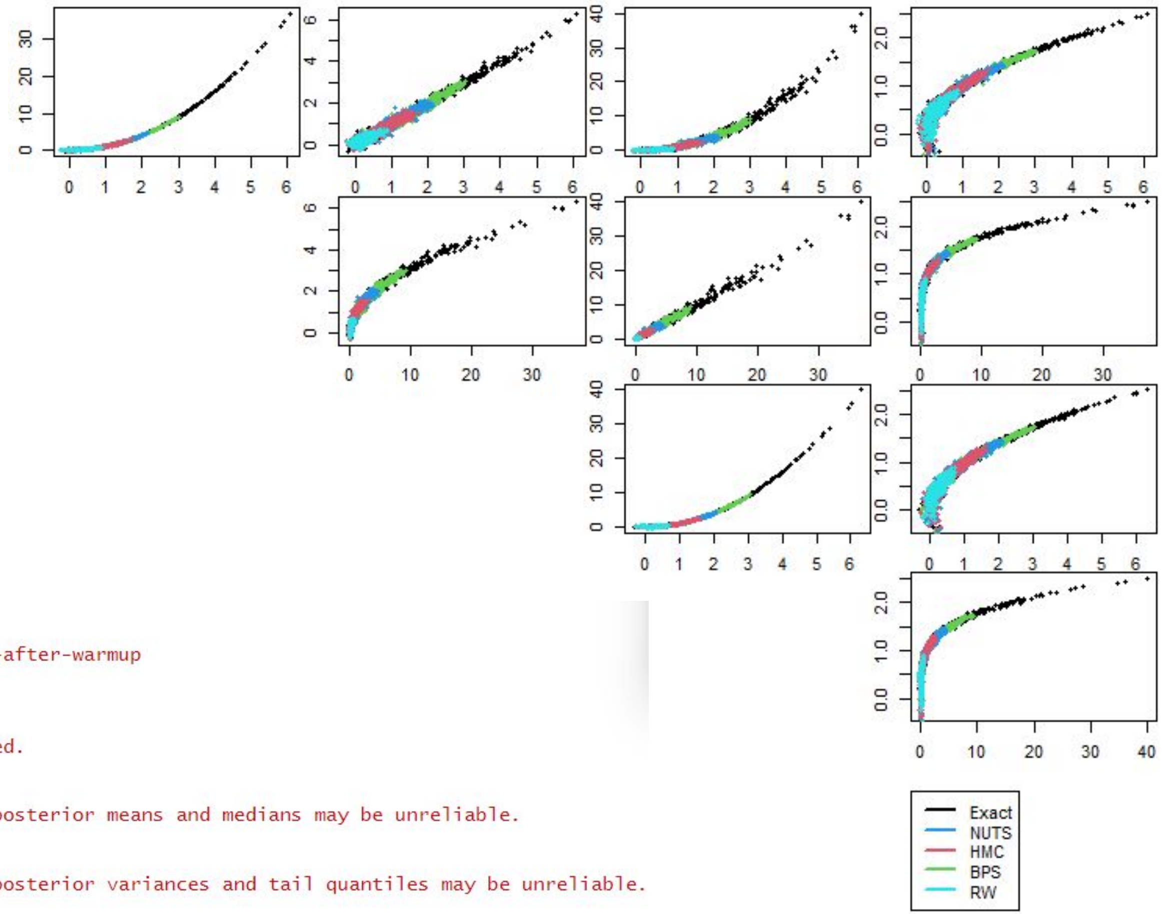


1 reply

**Kenyon Ng** 1 hour ago

one line of R code `shinystan::launch_shinystan(my_stanfit)` to produce ESS, standard R hat etc.. and also some HMC specific diagnostics: showing diverged samples, BFMI, marginal energy distribution (Fig 34 of Betancourt 2017 https://arxiv.org/abs/1701.02434 , all in a nice interactive webpage!

# Plots - comparing some samplers




```
Warning messages:
1: There were 274 divergent transitions after warmup. See
https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
to find out why this is a problem and how to eliminate them.
2: Examine the pairs() plot to diagnose sampling problems

3: The largest R-hat is 1.82, indicating chains have not mixed.
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#r-hat
4: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and medians may be unreliable.
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#bulk-ess
5: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quantiles may be unreliable.
Running the chains for more iterations may help. See
https://mc-stan.org/misc/warnings.html#tail-ess
```

# Conclusion

Looking for challenging test problems?

→ Find stan code on the Github!

Have an application that current samplers fail at?

→ Consider adding it to the Github!

GitHub @ matt-sutton/mcbench