

# GMM\_STAN

Graham Gumbert

2024-12-20

## Read in data

```
data <- read.table(paste0(params$cloud_path, params$beta_vals_file),
                    header = TRUE, sep = "\t")

site_ids <- data[[1]]
tumor_data <- data[,-1]

num_tumors <- ncol(tumor_data)
num_sites <- nrow(tumor_data)

cat("Number of sites:", num_sites, "\n")

## Number of sites: 1000

cat("Number of tumors:", num_tumors, "\n")

## Number of tumors: 213
```

## Initialize Stan model

```
sm <- stan_model("gmm.stan")
```

## Fit Stan model on beta-val distributions and save results in results\_list

Initialize tumor\_params dataframe and fill it out with the Stan model outputs from results\_list.

\_lower is 2.5% and \_upper is 97.5% which give us our 95% confidence interval for the peak location

```
tumor_params <- data.frame(
  Tumor  = names(results_list),
  psi_1_mean   = NA,
  psi_1_lower  = NA,
```

```

psi_1_upper   = NA,
psi_2_mean    = NA,
psi_2_lower   = NA,
psi_2_upper   = NA,
psi_3_mean    = NA,
psi_3_lower   = NA,
psi_3_upper   = NA,

X_mean        = NA,
X_lower       = NA,
X_upper       = NA,

N_mean        = NA,
N_lower       = NA,
N_upper       = NA,

phi_mean      = NA,
phi_lower     = NA,
phi_upper     = NA,

stringsAsFactors = FALSE
)

for (i in seq_along(results_list)) {
  tumor_id      = names(results_list)[i]
  fit           = results_list[[i]]

  if (is.null(fit)) {
    next
  }

  posterior_samples <- rstan::extract(fit)

  ci <- function(x) {
    c(
      lower = quantile(x, probs = 0.025, na.rm = TRUE),
      upper = quantile(x, probs = 0.975, na.rm = TRUE)
    )
  }

  psi_1_vals      <- posterior_samples$psi[, 1]
  psi_2_vals      <- posterior_samples$psi[, 2]
  psi_3_vals      <- posterior_samples$psi[, 3]

  psi_1_ci        <- ci(psi_1_vals)
  psi_2_ci        <- ci(psi_2_vals)
  psi_3_ci        <- ci(psi_3_vals)

  X_vals          <- posterior_samples$X
  X_ci            <- ci(X_vals)

  N_vals          <- posterior_samples$N
  N_ci            <- ci(N_vals)
}

```

```

phi_vals           <- posterior_samples$phi
phi_ci            <- ci(phi_vals)

tumor_params[i, "psi_1_mean"]   <- mean(psi_1_vals, na.rm = TRUE)
tumor_params[i, "psi_1_lower"]  <- psi_1_ci["lower.2.5%"]
tumor_params[i, "psi_1_upper"]  <- psi_1_ci["upper.97.5%"]

tumor_params[i, "psi_2_mean"]   <- mean(psi_2_vals, na.rm = TRUE)
tumor_params[i, "psi_2_lower"]  <- psi_2_ci["lower.2.5%"]
tumor_params[i, "psi_2_upper"]  <- psi_2_ci["upper.97.5%"]

tumor_params[i, "psi_3_mean"]   <- mean(psi_3_vals, na.rm = TRUE)
tumor_params[i, "psi_3_lower"]  <- psi_3_ci["lower.2.5%"]
tumor_params[i, "psi_3_upper"]  <- psi_3_ci["upper.97.5%"]

tumor_params[i, "X_mean"]      <- mean(X_vals, na.rm = TRUE)
tumor_params[i, "X_lower"]     <- X_ci["lower.2.5%"]
tumor_params[i, "X_upper"]     <- X_ci["upper.97.5%"]

tumor_params[i, "N_mean"]      <- mean(N_vals, na.rm = TRUE)
tumor_params[i, "N_lower"]     <- N_ci["lower.2.5%"]
tumor_params[i, "N_upper"]     <- N_ci["upper.97.5%"]

tumor_params[i, "phi_mean"]    <- mean(phi_vals, na.rm = TRUE)
tumor_params[i, "phi_lower"]   <- phi_ci["lower.2.5%"]
tumor_params[i, "phi_upper"]   <- phi_ci["upper.97.5%"]
}

tumor_params

##                                     Tumor psi_1_mean psi_1_lower psi_1_upper psi_2_mean
## 1 TARGET.30.PAIFXV.01A  0.2268148   0.1522082   0.3060589  0.4462592
## 2 TARGET.30.PAIPGU.01A  0.3652404   0.3212801   0.4097767  0.5410444
## 3 TARGET.30.PAISNS.01A  0.2707769   0.2363973   0.3066119  0.4423093
##   psi_2_lower psi_2_upper psi_3_mean psi_3_lower psi_3_upper   X_mean
## 1  0.2902106  0.5902015  0.32692604  0.24595836  0.4207772  0.3535191
## 2  0.4833598  0.5958564  0.09371517  0.06773424  0.1221992  0.2413374
## 3  0.3951547  0.4889455  0.28691384  0.25228559  0.3227009  0.2381459
##   X_lower   X_upper   N_mean   N_lower   N_upper   phi_mean phi_lower phi_upper
## 1 0.3246744  0.3883232 21.00041  14.95252  26.83360  0.6172410  0.5239815  0.7494995
## 2 0.2242560  0.2604991 17.74018  14.58599  20.87008  0.3298567  0.2975677  0.3680254
## 3 0.2263952  0.2503323 22.31617  18.98138  25.60670  0.3235468  0.3014617  0.3472387

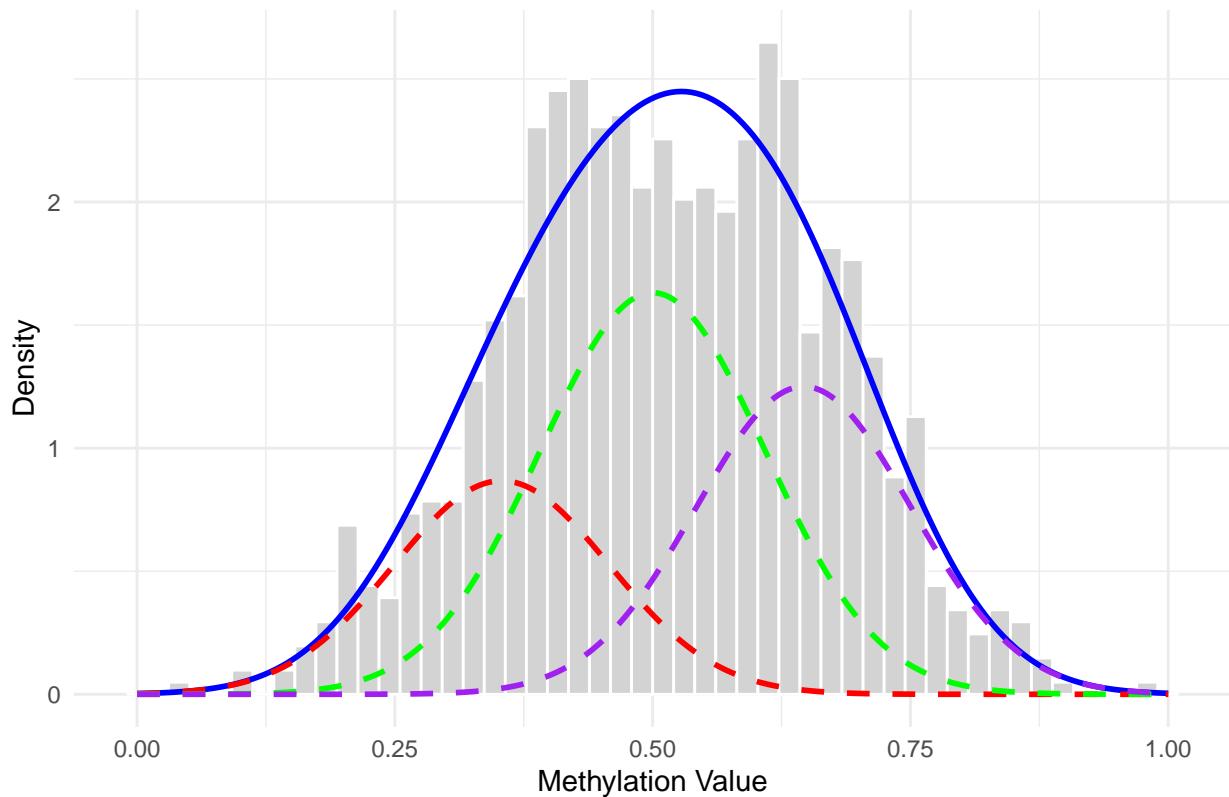
write.csv(tumor_params, params$output_file, row.names = FALSE)

```

Plots tumor beta value distributions with GMM output plotted with it.

Also plots the chain mixing and pairwise correlations of the GMM Stan model outputs.

## Tumor: TARGET.30.PAIFXV.01A – Posterior Mixture Plot

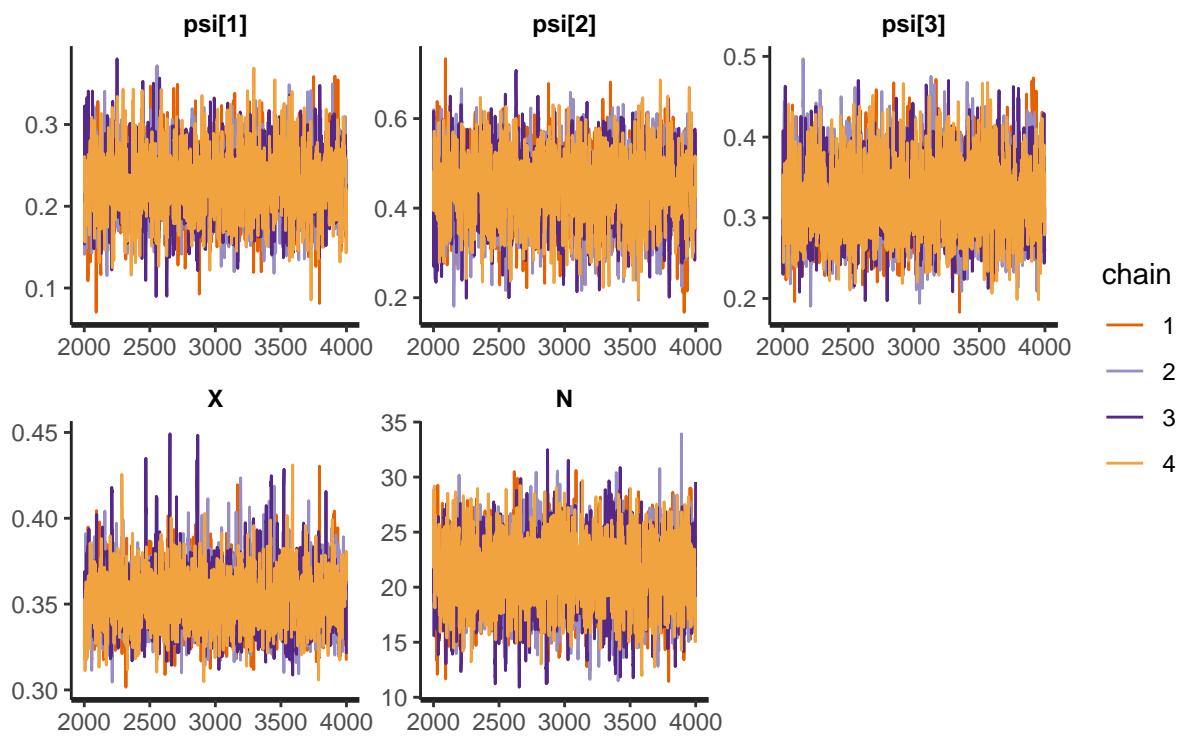


```

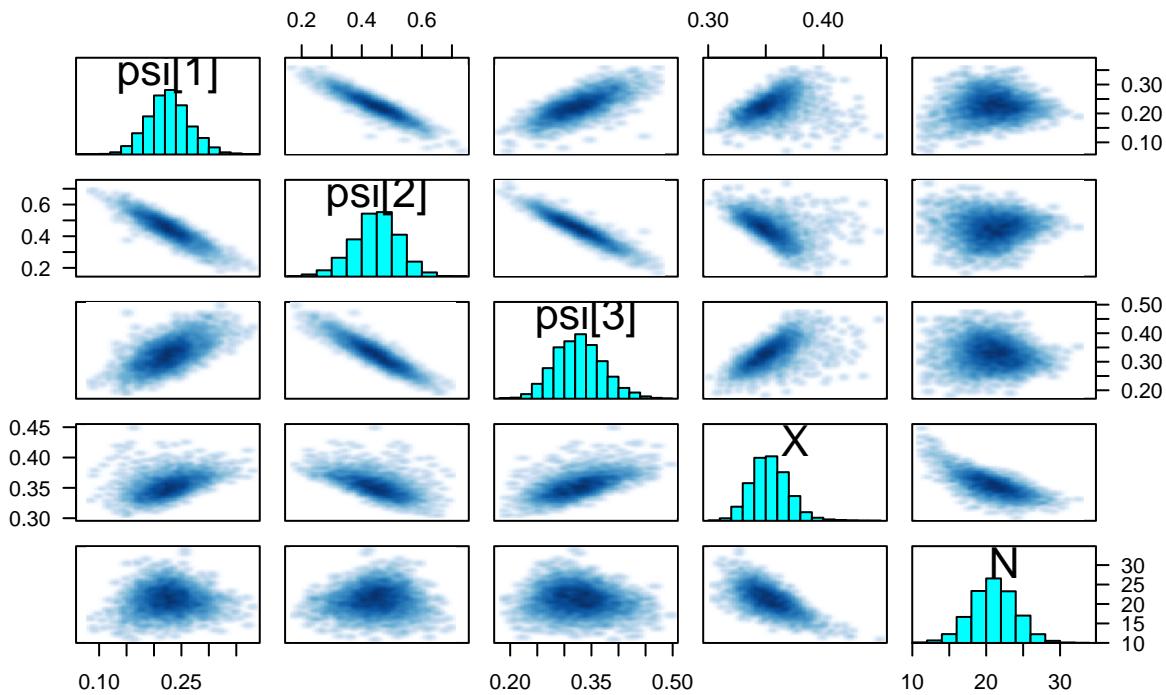
## ### Tumor: TARGET.30.PAIFXV.01A - Posterior Summary (Stan Fit)
## Inference for Stan model: anon_model.
## 4 chains, each with iter=4000; warmup=2000; thin=1;
## post-warmup draws per chain=2000, total post-warmup draws=8000.
##
##           mean   se_mean    sd   2.5%   25%   50%   75% 97.5% n_eff Rhat
## psi[1]    0.23    0.00  0.04   0.15   0.20   0.23   0.25   0.31  1930    1
## psi[2]    0.45    0.00  0.08   0.29   0.40   0.45   0.50   0.59  1827    1
## psi[3]    0.33    0.00  0.04   0.25   0.30   0.33   0.36   0.42  2186    1
## X         0.35    0.00  0.02   0.32   0.34   0.35   0.36   0.39  2126    1
## N        21.00    0.06  3.01  14.95  18.99  20.99  23.00  26.83  2610    1
## phi       0.62    0.00  0.06   0.52   0.58   0.61   0.65   0.75  2073    1
##
## Samples were drawn using NUTS(diag_e) at Fri May  2 12:10:10 2025.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).

```

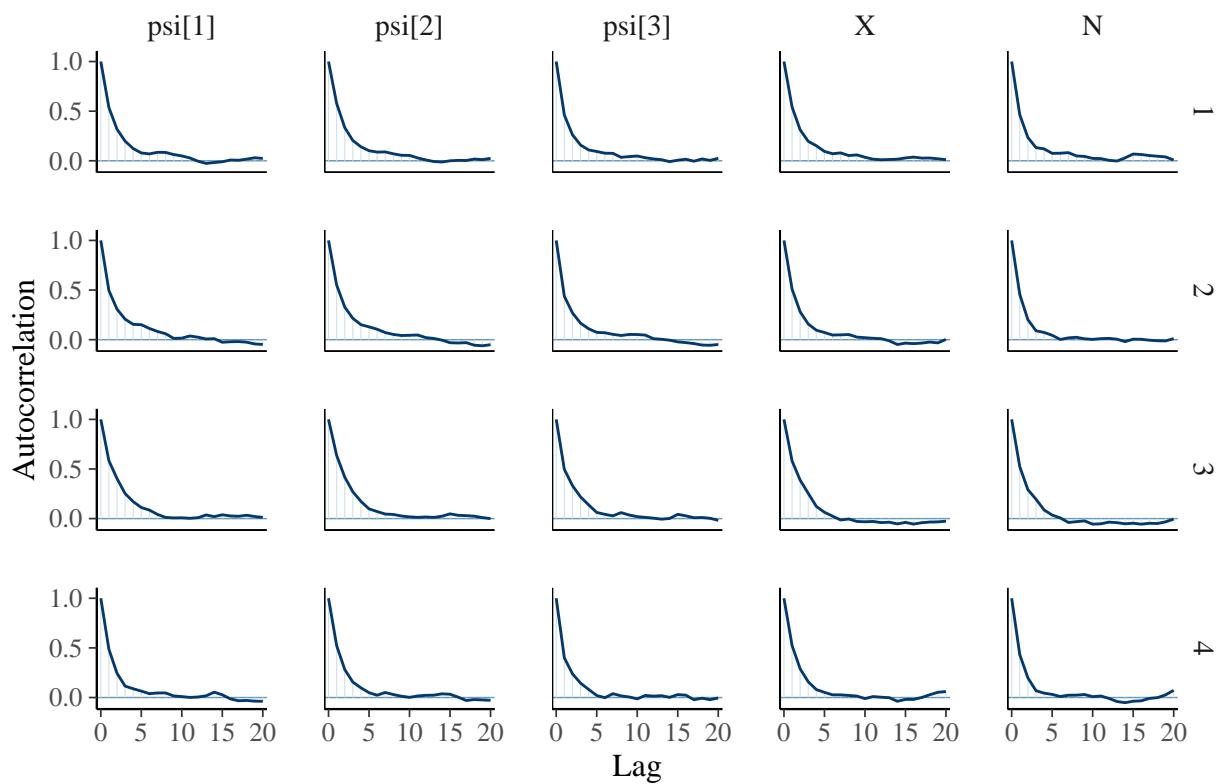
## Traceplot for tumor: TARGET.30.PAIFXV.01A



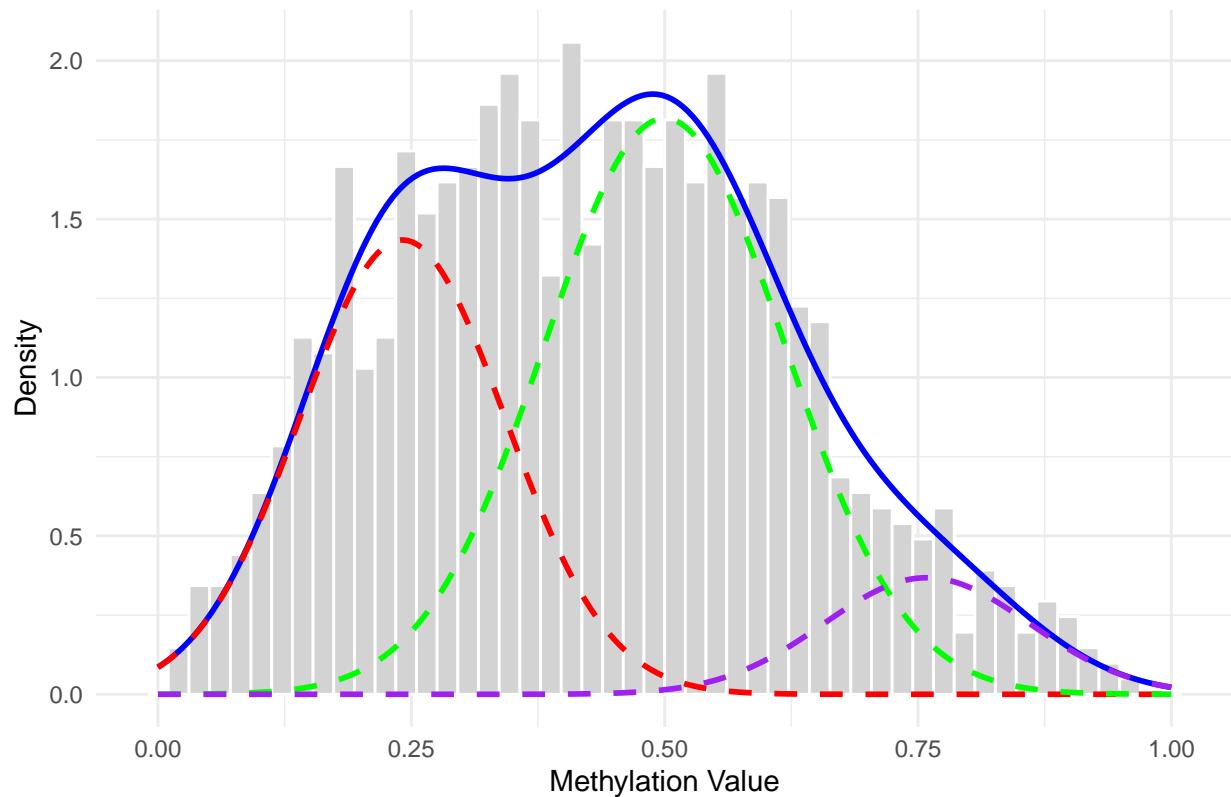
## Pairs Plot for tumor: TARGET.30.PAIFXV.01A



### Autocorrelation Plot for Tumor: TARGET.30.PAIFXV.01A

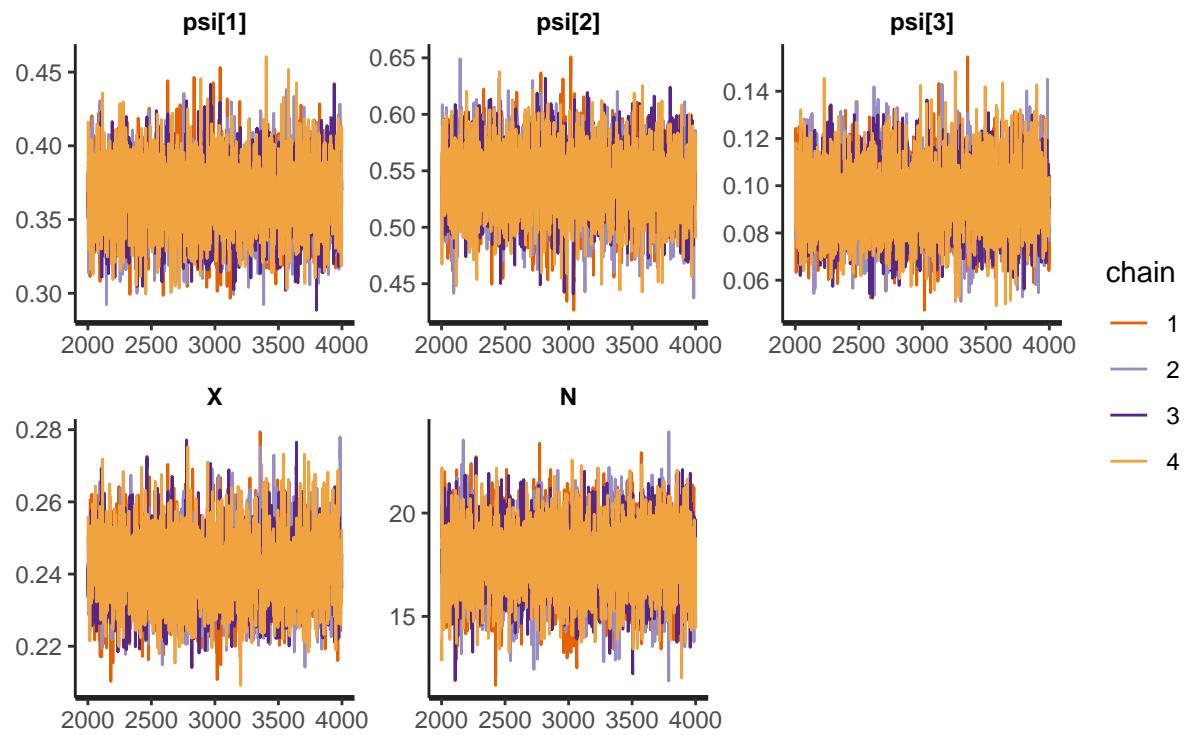


## Tumor: TARGET.30.PAIPGU.01A – Posterior Mixture Plot

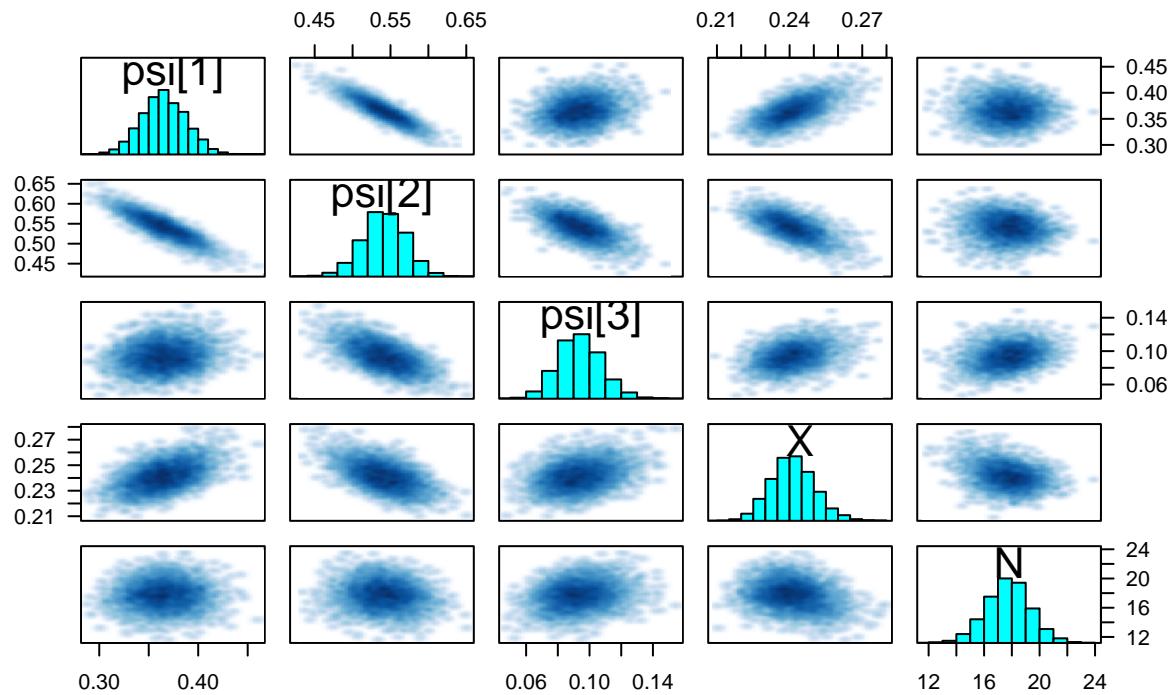


```
## ### Tumor: TARGET.30.PAIPGU.01A - Posterior Summary (Stan Fit)
## Inference for Stan model: anon_model.
## 4 chains, each with iter=4000; warmup=2000; thin=1;
## post-warmup draws per chain=2000, total post-warmup draws=8000.
##
##           mean   se_mean    sd   2.5%   25%   50%   75% 97.5% n_eff Rhat
## psi[1]    0.37     0.00  0.02   0.32   0.35   0.36   0.38   0.41  4249     1
## psi[2]    0.54     0.00  0.03   0.48   0.52   0.54   0.56   0.60  3712     1
## psi[3]    0.09     0.00  0.01   0.07   0.08   0.09   0.10   0.12  5082     1
## X        0.24     0.00  0.01   0.22   0.24   0.24   0.25   0.26  3925     1
## N       17.74     0.02 1.59 14.59 16.69 17.77 18.78 20.87  5059     1
## phi      0.33     0.00  0.02   0.30   0.32   0.33   0.34   0.37  3888     1
##
## Samples were drawn using NUTS(diag_e) at Fri May 2 12:10:22 2025.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

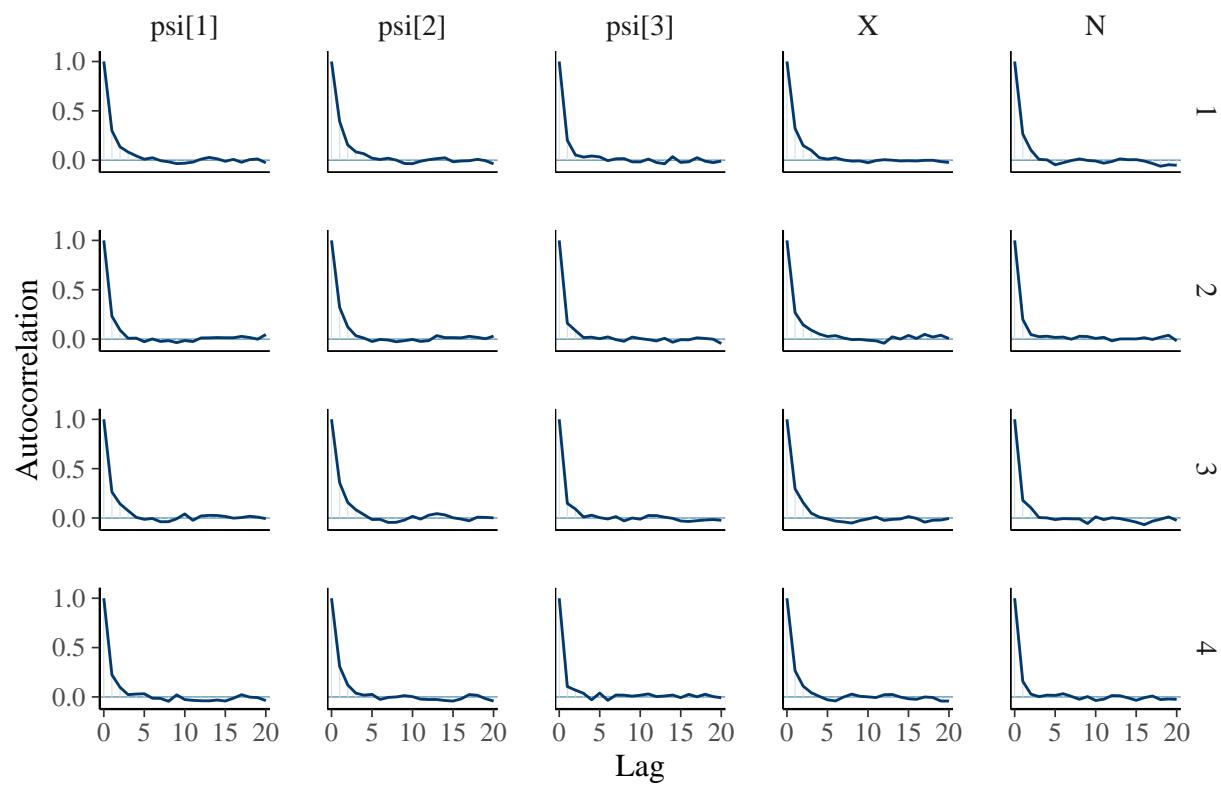
## Traceplot for tumor: TARGET.30.PAIPGU.01A



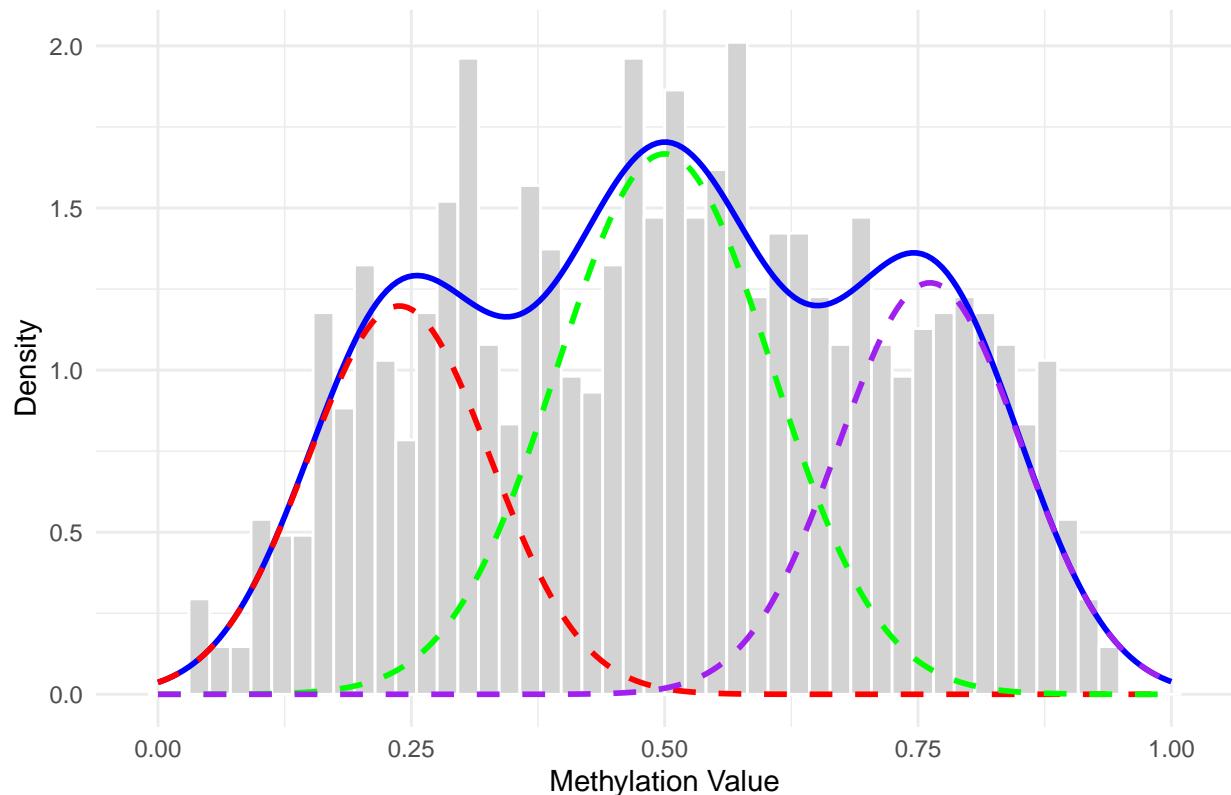
## Pairs Plot for tumor: TARGET.30.PAIPGU.01A



### Autocorrelation Plot for Tumor: TARGET.30.PAIPGU.01A



### Tumor: TARGET.30.PAISNS.01A – Posterior Mixture Plot

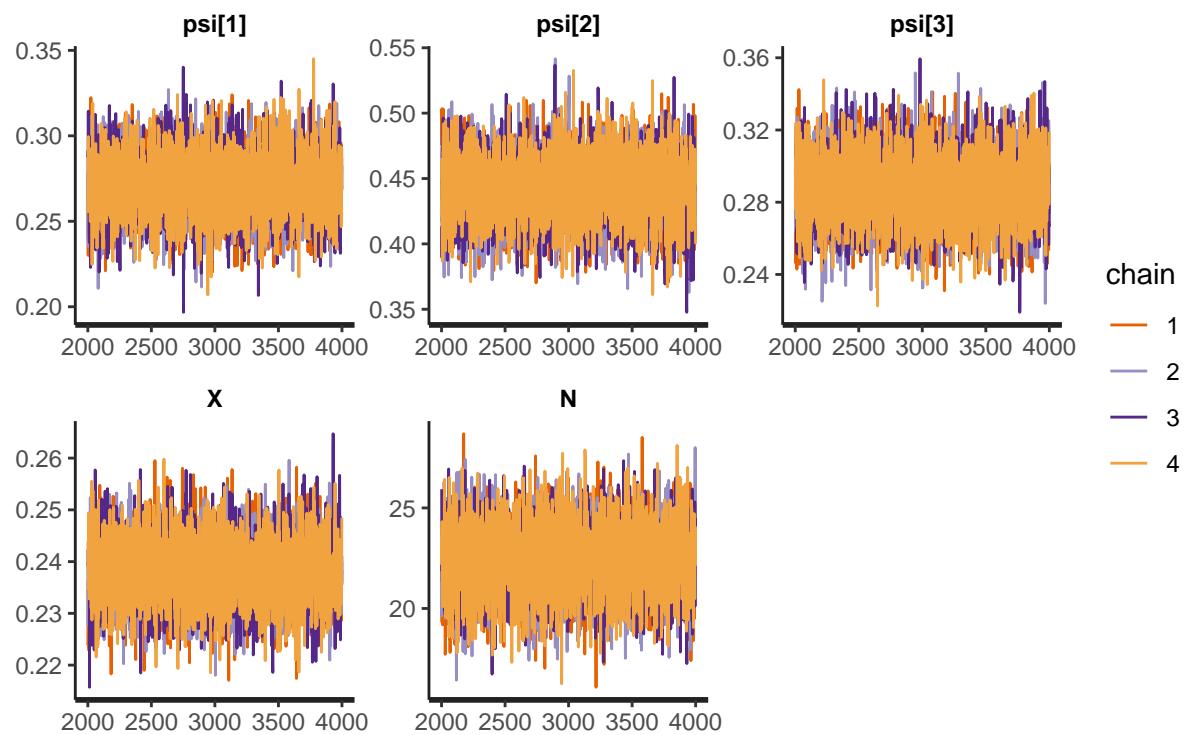


```

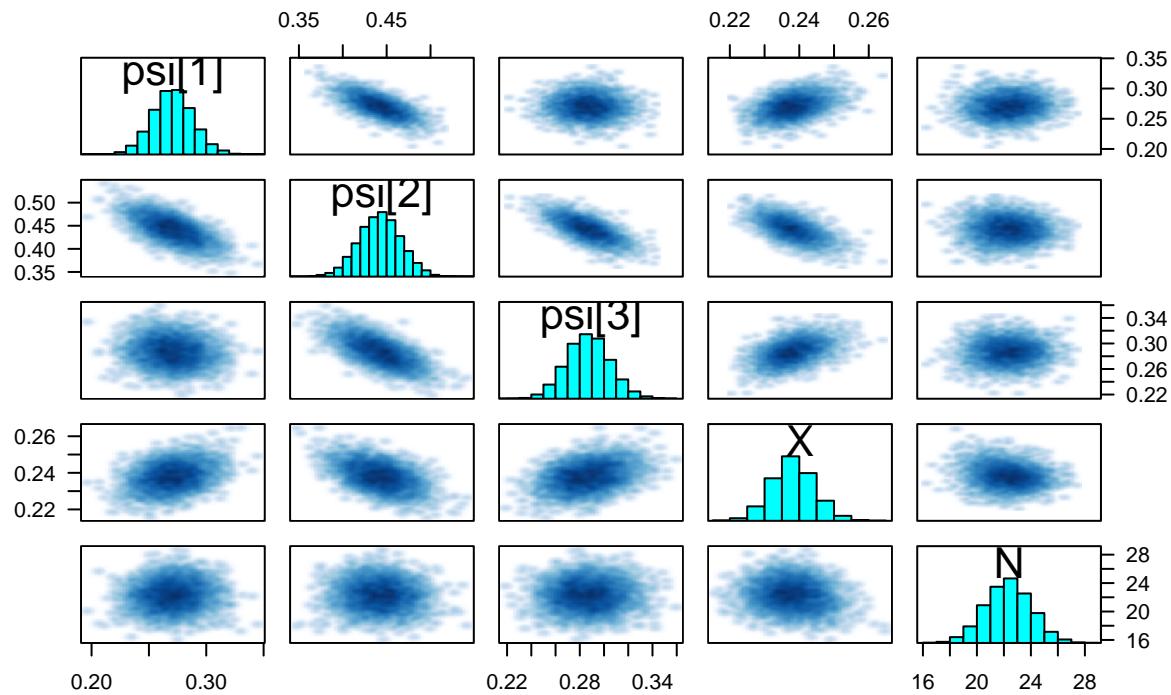
## ### Tumor: TARGET.30.PAISNS.01A - Posterior Summary (Stan Fit)
## Inference for Stan model: anon_model.
## 4 chains, each with iter=4000; warmup=2000; thin=1;
## post-warmup draws per chain=2000, total post-warmup draws=8000.
##
##           mean   se_mean    sd   2.5%   25%   50%   75% 97.5% n_eff Rhat
## psi[1]    0.27     0.00  0.02   0.24   0.26   0.27   0.28   0.31  5174    1
## psi[2]    0.44     0.00  0.02   0.40   0.43   0.44   0.46   0.49  4199    1
## psi[3]    0.29     0.00  0.02   0.25   0.27   0.29   0.30   0.32  5826    1
## X        0.24     0.00  0.01   0.23   0.23   0.24   0.24   0.25  4533    1
## N       22.32     0.02  1.70 18.98 21.16 22.33 23.47 25.61  5760    1
## phi      0.32     0.00  0.01   0.30   0.32   0.32   0.33   0.35  4525    1
##
## Samples were drawn using NUTS(diag_e) at Fri May 2 12:10:33 2025.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).

```

## Traceplot for tumor: TARGET.30.PAISNS.01A



## Pairs Plot for tumor: TARGET.30.PAISNS.01A



### Autocorrelation Plot for Tumor: TARGET.30.PAISNS.01A

