

Version 6: Cluster first

Simulate data

```
I <- 100; K <- 10; S <- 3
set.seed(123)
test.data <- simulateData(I, K, S)
```

Connect ω by cluster

```
##
## Cluster means and variances
##
W <- test.data$w[test.data$z, ]
meds <- apply(W, 2, median)
slevels <- colnames(test.data$w)[order(meds, decreasing=TRUE)]
w2 <- test.data$w %>%
  as_tibble() %>%
  mutate(z=factor(1:K)) %>%
  gather("sample", "omega", -z) %>%
  mutate(sample=factor(sample, levels=slevels))
fig2 <- ggplot(w2, aes(sample, omega, group=z)) +
  geom_point(size=1, aes(color=z)) +
  geom_line(aes(color=z)) +
  ylab(expression(omega)) +
  xlab("") +
  theme(axis.text.x=element_text(angle=45, hjust=1),
        panel.background=element_rect(fill="white", color="black"),
        legend.key=element_rect(fill="white", color="white")) +
  guides(color=guide_legend(title="Variant cluster\n(z)"))
fig2
```

Bayesian model: cluster mutations

```
jags.file <- file.path("...", "code", "w.jags")
inits <- list(".RNG.name" = "base::Wichmann-Hill",
              ".RNG.seed" = 123)
params <- c("z", "w", "ystar")
n.iter = 10000
thin = 7
K <- 10
samps <- runMCMC(test.data,
                 K,
                 jags.file,
```

```

        inits,
        params,
        n.iter,
        thin)

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 300
##   Unobserved stochastic nodes: 431
##   Total graph size: 4208
##
## Initializing model

chains <- ggs(samps)
z.chain <- chains[grepl("z", chains$Parameter), ]
w.chain <- chains[grepl("w", chains$Parameter), ]
mcmc_w <- group_by(w.chain, Parameter) %>%
  summarize(mean=mean(value))
saveRDS(chains, file.path(outdir, "chains.rds"))

```