

SNP tagging conditions

Stas

14 January, 2019

Let Z_o be the observed Z scores for the joint 3-SNP model, with $Z_o \sim N(Z_M, \Sigma)$, $Z_M = \Sigma Z_J$. We are interested in evaluating plausibility of situations in which we would be forced to conclude that SNP 3 is the causal one. Two situations are possible: SNPs 1 and 2 are causal with $Z_J = z_J = (\zeta_1, \zeta_2, 0)'$, or SNP 3 is causal with $Z_J = \tilde{z}_J := (0, 0, \tilde{\zeta})'$. Likelihood of SNP 3 being causal is larger than the likelihood of SNPs 1 and 2 being causal when

$$2(\tilde{\zeta}z_3^o - \zeta_1 z_1^o - \zeta_2 z_2^o) + \zeta_1^2 + \zeta_2^2 - \tilde{\zeta}^2 + 2\zeta_1\zeta_2 r_{12} > 0. \quad (1)$$

SNPs 1 and 2 are causal

We assume $\zeta_1 = \zeta_2 := \zeta$ and $\tilde{\zeta} = \zeta(r_1 + r_2)$ and calculate the probability of the condition (1) above being satisfied as

$$P(r_1, r_2, r_{12}, \zeta) := 1 - \Phi\left(\frac{1}{2}|\zeta|\sqrt{2(1 + r_{12}) - (r_1 + r_2)^2}\right). \quad (2)$$

Additionally conditions

$$\begin{aligned} -2r_1 r_2 r_{12} + r_1^2 + r_2^2 + r_{12}^2 &\leq 1 \\ r_1 + r_2 &< \sqrt{2(1 + r_{12})}. \end{aligned} \quad (3)$$

must be satisfied.

We see that probability (2) depends on the quantity $w := (r_1 + r_2)^2$, squared sum of correlation of causal SNPs with SNP3, correlation r_{12} between causal SNPs, and absolute value of effect size ζ . We plot P for varying values w , r_{12} and ζ (note that for $r_{12} = -1$, there are no combinations of correlations r_1 and r_2 simultaneously satisfying both conditions (3)):

```
r12 <- c(-1, -0.5, 0, 0.5, 1)
zeta <- c(0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1)

foo <- function(r12, a = 0.01) {
  r <- seq(-1, 1, by = a)
  dat <- data.table(expand.grid(r, r))
  colnames(dat) <- c('r1', 'r2')
  cnd1 <- -2 * r12 * dat$r1 * dat$r2 + dat$r1^2 + dat$r2^2 + r12^2 <= 1
  cnd2 <- round(dat$r1 + dat$r2, 2) < sqrt(2 * (1 + r12))
  dat <- dat[cnd1 & cnd2, ]
  dat[, rsum := r1 + r2]
  dat
}

dat <- lapply(r12, FUN = function(r) {
  dat <- foo(r)
  if(nrow(dat) == 0) return(NA)
  else
    P(r, r12, rsum, zeta)
})
```

```

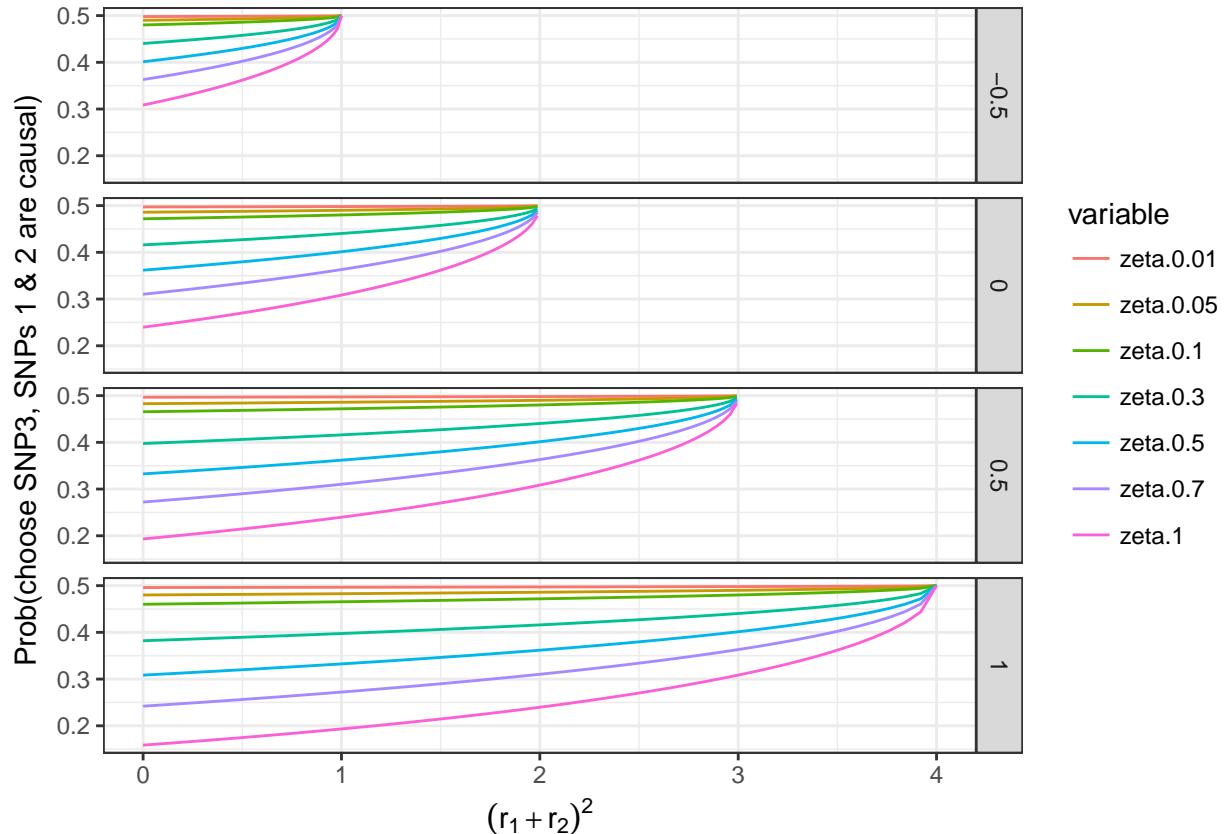
val.mat <- lapply(zeta, FUN = function(z) {
  v <- 0.5 * z * sqrt(2 * (1 + r) - dat$rsum^2)
  1 - pnorm(v)
}) %>% do.call(cbind, .)
colnames(val.mat) <- paste0("zeta=", zeta)

dat <- data.frame(dat, val.mat)
dat <- melt(dat, id = c("r1", "r2", "rsum"))
dat$r12 <- r
dat
}) %>% do.call(rbind, .)

dat <- dat[!is.na(dat$r12), ]

ggplot(dat, aes(x = rsum^2, y = value)) + geom_line(aes(colour = variable)) +
  facet_grid(r12 ~ .) +
  labs(x = expression((r[1] + r[2])^"2"), y = "Prob(choose SNP3, SNPs 1 & 2 are causal)") + theme_bw()

```



SNP3 is causal

We assume $\zeta_1 = \tilde{\zeta}r_1$ and $\zeta_2 = \tilde{\zeta}r_2$ and probability of event (1) becomes

$$P(r_1, r_2, r_{12}, \zeta) = 1 - \Phi \left(-\frac{1}{2} |\zeta| \sqrt{1 - r_1^2 - r_2^2 + 2r_1 r_2 r_{12}} \right)$$

with an additional condition

$$-2r_1 r_2 r_{12} + r_1^2 + r_2^2 < 1.$$

and first of (3) have to be satisfied.

```
r12.vec <- c(-1, -0.5, 0, 0.5, 1)
zeta <- c(0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1)

foo <- function(x, a = 0.01) {
  r <- seq(-1, 1, by = a)
  dat <- data.table(expand.grid(r, r))
  colnames(dat) <- c('r1', 'r2')
  cnd1 <- -2 * x * dat$r1 * dat$r2 + dat$r1^2 + dat$r2^2 + x^2 <= 1
  cnd2 <- -2 * x * dat$r1 * dat$r2 + dat$r1^2 + dat$r2^2 <= 1
  dat <- dat[cnd1 & cnd2, ]
  dat
}

dat <- lapply(r12.vec, FUN = function(r12) {
  dat <- foo(r12)
  if(nrow(dat) == 0) return(NA)

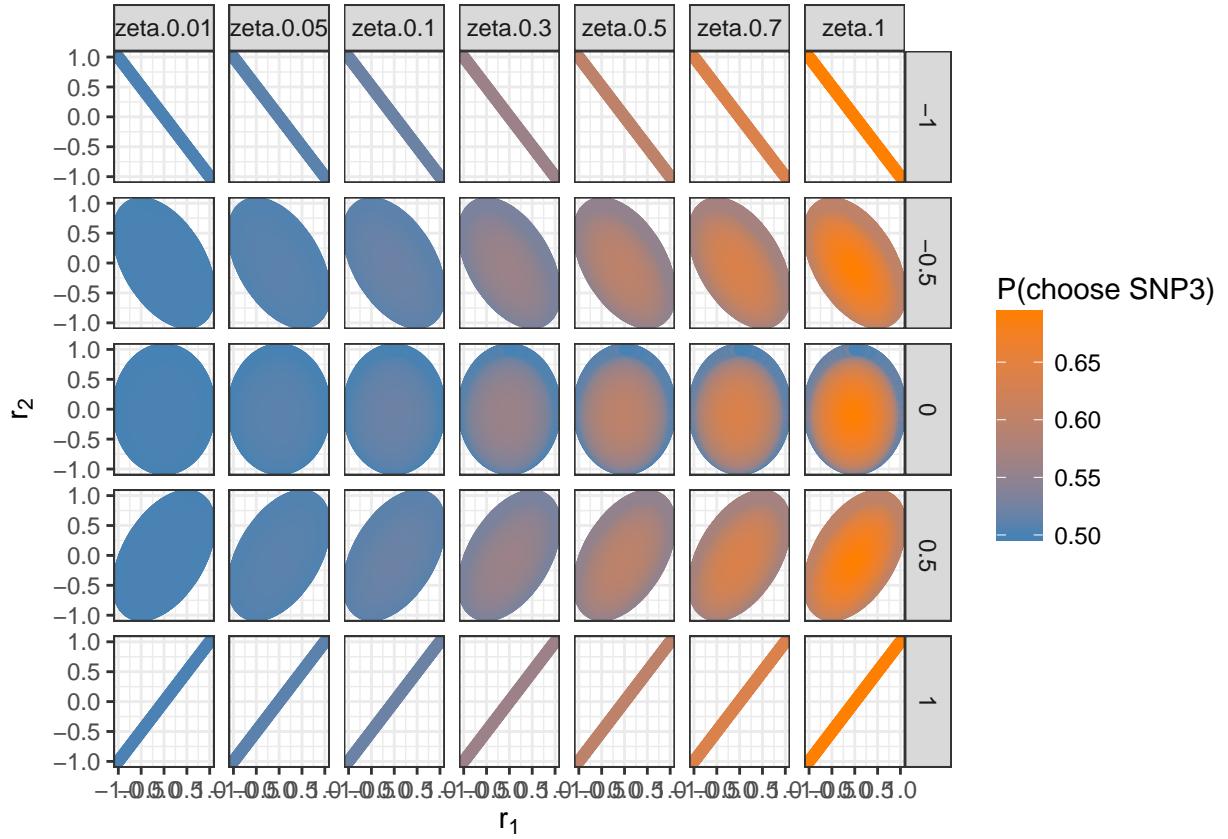
  val.mat <- lapply(zeta, FUN = function(z) {
    prob <- 1 - pnorm(-0.5 * z * sqrt(round(1 - dat$r1^2 - dat$r2^2 + 2 * dat$r1 * dat$r2 * r12, 4)))
  }) %>% do.call(cbind, .)
  colnames(val.mat) <- paste0("zeta=", zeta)

  dat <- data.frame(dat, val.mat)
  dat <- melt(dat, id = c("r1", "r2"))
  dat$r12 <- r12
  dat
}) %>% do.call(rbind, .)

dat <- dat[!is.na(dat$r12), ]

ggplot(dat, aes(x = r1, y = r2)) + geom_point(aes(colour = value)) +
  facet_grid(r12 ~ variable) +
  labs(x = expression(r[1]), y = expression(r[2]), y = "Prob(choose SNP3, SNP is causal)") +
  scale_colour_gradient("P(choose SNP3)", low = "steelblue", high = "darkorange1") +
  theme_bw()

## Warning: The plyr::rename operation has created duplicates for the
## following name(s): (`y`)
```



Appendix

Suppose SNPs 1 and 2 are causal and $r_{12} = 0$. Then the above conditions (3) reduce to

$$r_1^2 + r_2^2 \leq 1$$

$$r_1 + r_2 \leq \sqrt{2}.$$

```

r <- seq(-1, 1, by = 0.01)
dat <- expand.grid(r1 = r, r2 = r)

plot(c(-2, 2), c(-2, 2), type = 'n', xlab = 'x2 + z2', ylab = 'x + z')
rect(-2, -2, 2, sqrt(2), col = alpha('firebrick1', 0.5), border = NA)
rect(-2, -2, 1, 2, col = alpha('dodgerblue', 0.5), border = NA)
abline(h = sqrt(2), lty = 2, col = 'firebrick1', lwd = 2)
abline(v = 1, lty = 2, col = 'dodgerblue', lwd = 2)

points(dat$r1^2 + dat$r2^2, dat$r1 + dat$r2, col = "purple")
lines(r^2 + r^2, r + r, col = "seagreen3", type = 'l', lwd = 3)

```

