Two-door vs. back-door

Tetiana Gorbach

Here we compare $var\varphi_{td}$ and $var\varphi_{bd}$ using the fact that when the two-door and the back-door assumptions are fulfilled,

$$var\varphi_{td} - var\varphi_{bd} = \sum_{z,c} p(c)var(Y|z,c)\left((p(z|a^*,c) - p(z|a,c))^2 \sum_{a'} \frac{p(a'|c)}{p(z|a',c)} - \frac{p(z|a^*,c)}{p(a^*|c)} - \frac{p(z|a,c)}{p(a|c)}\right)$$

Additionally, $p(z|a,c) = p(z|a)\ \forall a$ for the considered in the example data distribution.

Functions needed to calculate the difference:

```
# A  -> Z : par.z.a
# (Z, C)  -> Y : par.y.zc
# C -> A : par.a.c

expit <- function(x) {
  return(exp(x) / (1 + exp(x)))
}
diff_two_door_minus_back_door <- function(astar, a, par.c, par.a.c, par.z.a, par.y.zc) {
  # computes  difference between $var \varphi_{td}$ and $var \varphi_{bd}$
  pc <- function(c_value){
     # returns p(C)
     par.c^c_value * (1 - par.c)^(1 - c_value)
  }
  pa_given_c <- function(a_value, c_value) {
    # returns p(A = a_value| C = c_value)
    p1 <- expit(par.a.c * c_value)
    p1^a_value * (1 - p1)^(1 - a_value)
  }
  pz_given_a <- function(z_value, a_value) {
    # returns p(Z = z_value|A = a_value, C=c_value)
    p1 <-  expit(par.z.a * a_value)
    p1^z_value * (1 - p1)^(1 - z_value)
  }
  varYZC <- function(z_value, c_value) {
    # returns var(Y|Z = z_value, C = c_value)
    p <- expit(par.y.zc %*% c(z_value, c_value))
    p * (1 - p)
  }
  summation_pos <- function(z_value, c_value) {
    (pz_given_a(z_value, astar) - pz_given_a(z_value, a))^2 *
      (pa_given_c(1, c_value) / pz_given_a(z_value, 1) +
       pa_given_c(0, c_value) / pz_given_a(z_value, 0))
  }
  summation_neg <- function(z_value, c_value) {
     - pz_given_a(z_value, astar) / pa_given_c(astar, c_value) -
       pz_given_a(z_value, a) / pa_given_c(a, c_value)
  }
```

```
  difference <- 0
  for (c_value in c(0,1)){
    for (z_value in c(0,1)){
      difference <- difference +
       pc(c_value)*varYZC(z_value, c_value) *
        (summation_pos(z_value, c_value) + summation_neg(z_value, c_value) )
    }
  }
    difference
}
```

Calculations:

```
sim.cases <- expand.grid(par.z.a = seq(-4, 4, by = 0.2), zy = seq(-4, 4, by = 1),
                         par.a.c = seq(-4, 4, by = 1), cy = seq(-4, 4, by = 1),
                         par.c = c(0.1, 0.3, 0.6, 0.9))%>%
   dplyr::arrange(par.z.a, par.a.c, zy, cy, par.c)

for (h in 1:nrow(sim.cases)) {
  par.c <- sim.cases$par.c[h]
  par.a.c <- sim.cases$par.a.c[h]
  par.z.a <- sim.cases$par.z.a[h]
  par.y.zc <- c(sim.cases$zy[h], sim.cases$cy[h])
  sim.cases$diff[h] <-  diff_two_door_minus_back_door(astar = 1, a = 0, par.c, par.a.c, par.z.a, par.y.z
}
rm(par.z.a, par.y.zc, par.a.c,  par.c, h)
save(sim.cases, file="twodoor_minus_backdoor_example.rData")
load(file="twodoor_minus_backdoor_example.rData")
```

Plotting:

```
bitmap("example_var_td_minus_var_bd.png", height = 80, width = 150, units = "mm",  res=600, pointsize=10
op <- par(no.readonly=T)
par(mar=c(4,3,0,1))
plot(expit(sim.cases$par.z.a), sim.cases$diff, pch=20, cex = 0.1, xlab = TeX("P(Z(1)=1)"), ylab="")
abline(0,0, lwd=2)
abline(v = (3-2*sqrt(2))/2)
abline(v = (2*sqrt(2)-1)/2)
text(x=(3-2*sqrt(2))/2, par("usr")[3],
     labels = TeX("\\frac{3-2\\sqrt{2}}{2}"),  pos = 1, xpd = TRUE )
text(x=(2*sqrt(2)-1)/2, par("usr")[3],
     labels = TeX("\\frac{2\\sqrt{2}-1}{2}"),  pos = 1, xpd = TRUE)
title(ylab=TeX("$var \\varphi_{td}-var \\varphi_{bd}$"), line=2, cex.lab=1)
par(op)
dev.off()
```

```
## pdf
##   2
```