Inverse Probability Weighting

In backdoor criterion part, we learned that PA(X) always satisfies backdoor criterion. Hence, we can represent postintervention causal effect as such:

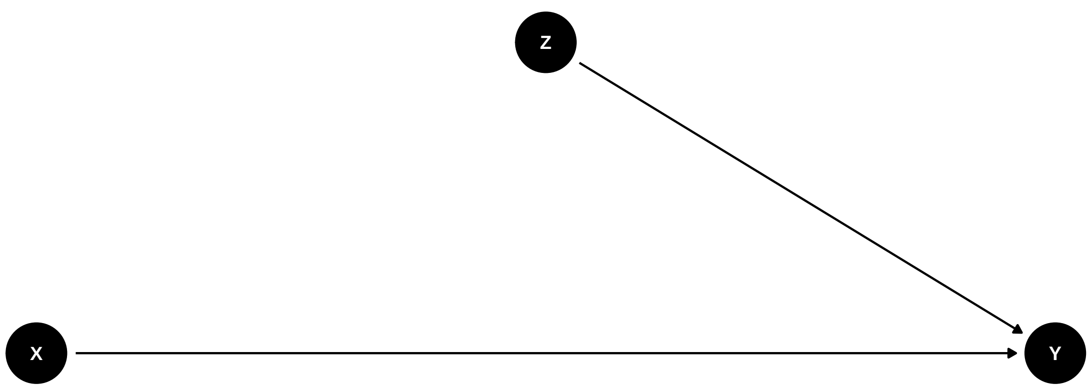
$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, PA = z)P(PA = z)$

Modifying this equation, we can obtain the following:

$P(Y = y \mid do(X = x)) = \sum_z \frac{P(Y=y,X=x,PA=z)}{P(X=x|PA=z)}$

The factor $P(X = x \mid PA = z)$ in the denominator is known as the "propensity score."

Now, assume we have a graphical model with an intervention, do(X=x), on the model.



$P(Y = y \mid do(X = x))$

$= \; P_m(Y = y \mid X = x) \quad \text{(by definition)}$

$= \; \sum_z P_m(Y = y \mid X = x, Z = z)P(Z = z \mid X = x) \quad \text{(Bayes' rule)}$

$= \; \sum_z P_m(Y = y \mid X = x, Z = z)P(Z = z) \quad (X \perp\!\!\!\perp Z)$

$= \; \sum_z P(Y = y \mid X = x, Z = z)P(Z = z) \quad \text{(invariance relations)}$

$= \; \sum_z \frac{P(Y=y|X=x,Z=z)P(X=x|Z=z)P(Z=z)}{P(X=x|Z=z)}$

$= \; \sum_z \frac{P(Y=y,X=x,Z=z)}{P(X=x|Z=z)}$

From this result, we can see that postintervention causal effect can be computed by multiplying the pretreatment distribution of (X,Y,Z) by a factor $1/P(X = x \mid Z = z)$, propensity score. Namely, each case $(Y = y, X = x, Z = z)$in the population should boost its probability by the inverse conditional probability of assignment to a treatment condition given a set of observed covariates. This is the reason why this method is called "inverse probability weighting."

Importing data

```
nhefs <- read_csv("nhefs.csv")

nhefs_uncensored <-
  nhefs %>%
  mutate(cens = ifelse(is.na(wt82), 1, 0)) %>%
  relocate(cens, wt82) %>%
  filter(!is.na(wt82))
```

Estimation of ip weights via a logistic model

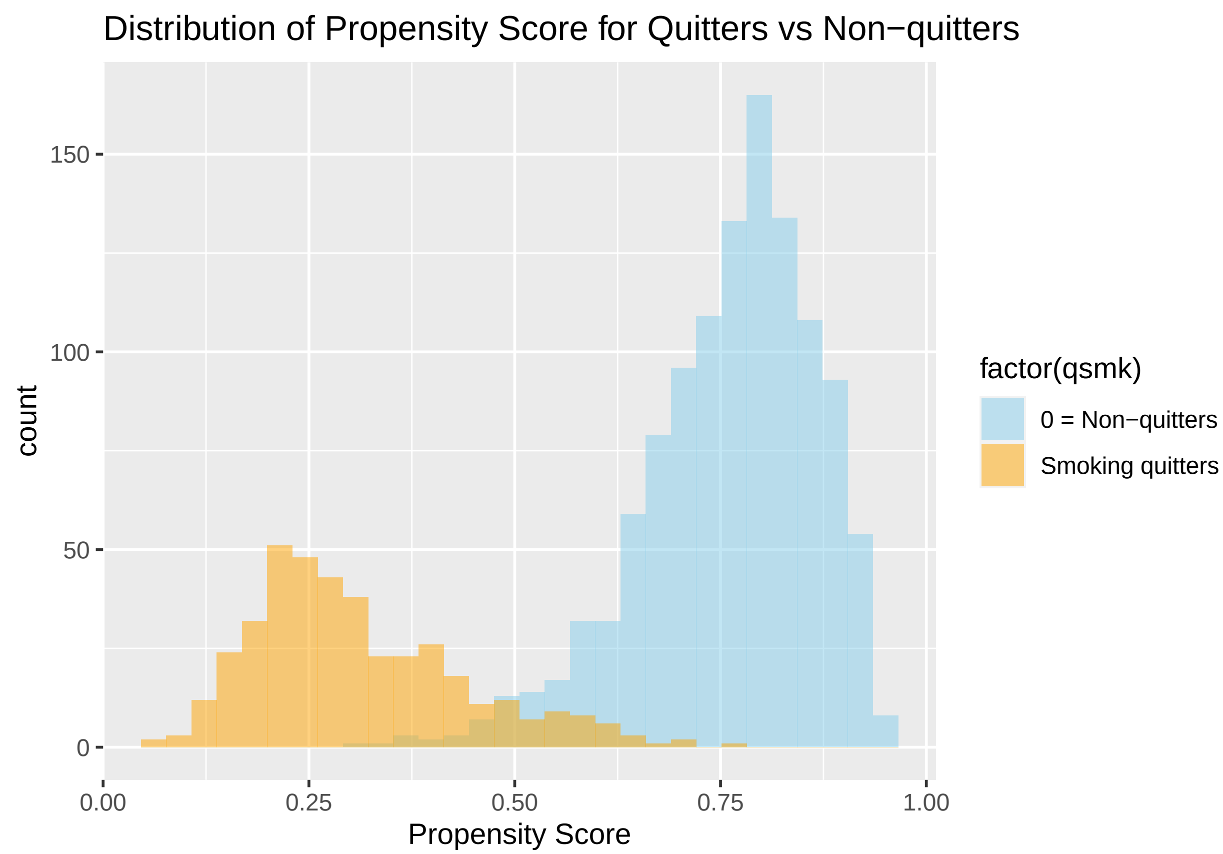
```
propensity_model <- glm(
  qsmk ~ sex + race + age + I(age ^ 2) +
    as.factor(education) + smokeintensity +
    I(smokeintensity ^ 2) + smokeyrs + I(smokeyrs ^ 2) +
    as.factor(exercise) + as.factor(active) + wt71 + I(wt71 ^ 2),
  family = binomial(),
  data = nhefs_uncensored
)
```

Computing propensity score. Note that Pr[A=0|L] = 1-Pr[A=1|L]

```
p.qsmk.obs <-
  ifelse(nhefs_uncensored$qsmk == 0,
         1 - predict(propensity_model, type = "response"),
         predict(propensity_model, type = "response"))

nhefs_uncensored <-
  nhefs_uncensored %>%
  mutate(w = 1/p.qsmk.obs,
         ps = p.qsmk.obs) #computing propensity score

nhefs_uncensored %>%
  ggplot(aes(ps, group = qsmk, fill = factor(qsmk))) +
  geom_histogram(alpha = 0.5, position = "identity") +
  scale_fill_manual(values = c("skyblue", "orange"),
                    labels = c("0 = Non-quitters", "Smoking quitters")) +
  labs(x = "Propensity Score",
       title = "Distribution of Propensity Score for Quitters vs Non-quitters")
```



Getting coefficients of the marginal structural model

```
msm.w <- geeglm(
  wt82_71 ~ qsmk,
  data = nhefs_uncensored,
  weights = w,
  id = seqn,
  corstr = "independence"
)

beta <- coef(msm.w)
SE <- coef(summary(msm.w))[, 2]
lcl <- beta - qnorm(0.975) * SE
ucl <- beta + qnorm(0.975) * SE
cbind(beta, lcl, ucl)
```

```
##                 beta      lcl      ucl
## (Intercept) 1.779978 1.339514 2.220442
## qsmk        3.440535 2.410587 4.470484
```

Reference

Pearl, J., Glymour, M., & Jewell, N. P. (2019). Causal inference in statistics a primer. Wiley.

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.