Ordinal Regression + Compound Models

Categorical Data

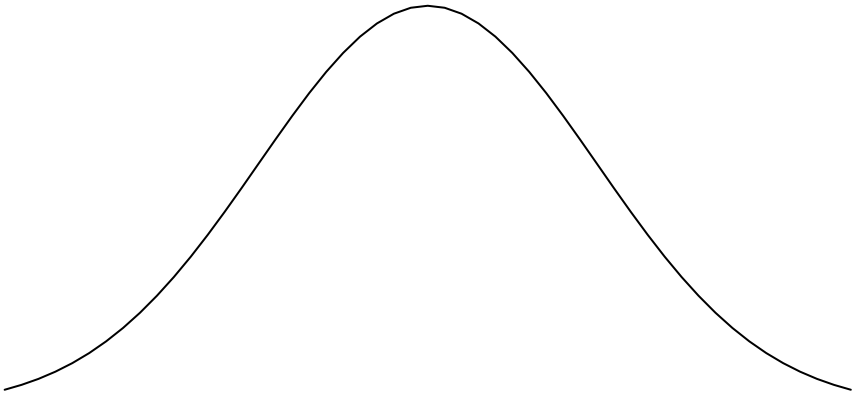
Categorical data comes in two different flavors:

1. Ordinal: Likert scale responses

2. Unordered: favorite winter Montana activity (skiing, sleeping, doing Statistics)

Models for Ordinal Data

Recall the latent formulation for probit model



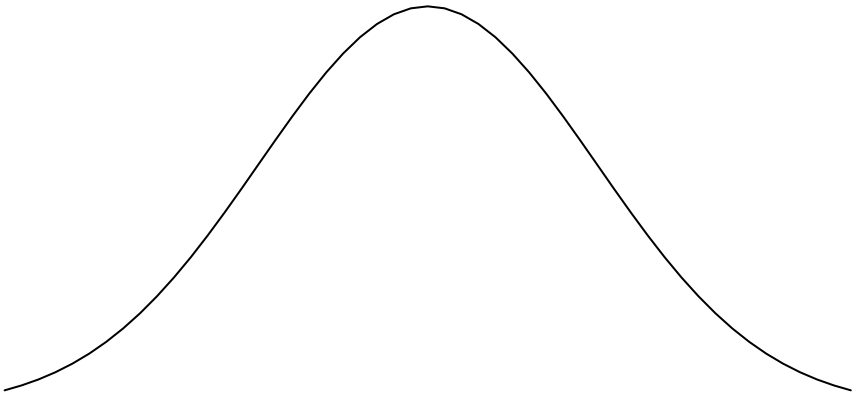
Let y be a vector categorical response

$$y \sim Multinomial(\pi_1, ..., \pi_k) \tag{1}$$
$$\tag{2}$$

Using the latent data formulation, let z be a continuous normal random variable with mean (XB) and variance 1. Then, for $k = 3$,

$$y_i = \begin{cases} 1 \text{ if } \ z_i < c_{1|2} \\ 2 \text{ if } \ c_{2|3} \leq z_i \geq c_{1|2} \\ 3 \text{ if } \ z_i > c_{2|3} \end{cases}$$

Then given $z_i \sim N(X_i\beta, 1)$, $\underline{\pi_i}$ can be calculated by integrating the area of the distribution in the appropriate cut points.



For instance, suppose $X_i\beta = 2$ and $c_{1|2} = 0, c_{2|3} = 2$, then

$\pi_1 =$ `pnorm(0, mean = 2)` $= (0.0227501)$

$\pi_2 =$ `pnorm(2, mean = 2) - pnorm(0, mean = 2)` $= (0.4772499)$

$\pi_3 =$ `1 - pnorm(2, mean = 2)` $= (0.5)$

Vegetation Coverage Class Data Plant coverage is an important ecological indicator. However, estimating plant coverage can be difficult and is often summarized in an ordinal manner.

```
class <- shoshveg  %>%
          mutate(class = case_when(
            ABILAS <= .5 ~ 1,
            ABILAS > .5 & ABILAS <= 3  ~ 2,
            TRUE ~3)) %>%
          mutate(class = factor(class)) %>%
          dplyr::select(class) %>% pull()

plant_cover <- tibble(class = class, elevation = scale(shoshsite$elevation))
```

```
ord <- polr(class ~ elevation, data = plant_cover, method = 'probit')
summary(ord)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = class ~ elevation, data = plant_cover, method = "probit")
##
## Coefficients:
##            Value Std. Error t value
## elevation 0.3129     0.1034   3.026
##
## Intercepts:
##     Value  Std. Error t value
## 1|2 0.2226 0.1057     2.1050
## 2|3 1.4634 0.1520     9.6258
##
## Residual Deviance: 256.0104
## AIC: 262.0104
```

```
bayes_ord <- stan_polr(class ~ elevation, data = plant_cover, method = 'probit',
          prior = R2(0.25, 'mean'), refresh = 0)
print(bayes_ord, digits = 2)
```

```
## stan_polr
##  family:       ordered [probit]
##  formula:      class ~ elevation
##  observations: 150
## ------
##           Median MAD_SD
## elevation 0.30   0.10
##
## Cutpoints:
##     Median MAD_SD
## 1|2 0.21   0.10
## 2|3 1.44   0.15
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```