Week Eleven

Last Week

- Exam Recap
- Probability Distributions for Count Data
- Count Regression

This Week: GLMs for Count Data & Multicategory data

- Bike Data

Next Week: Multicategory Regression

Multicategory Logit Models

Recall logistic regression for binary model

1. Random component / Probability distribution $[0,1]$

$$Y \sim \text{Binomial}(n, \pi) \text{ or } Y \sim \text{Bernoulli}(\pi)$$

2. Systematic component / linear combination of predictors / functional form of predictors

$$\text{logit}(\pi) = XB \dashrightarrow B_0 + B_1 X_1 + B_2 X_1^2 + B_3 X_3 + B_4 X_1 X_3$$

$$\log\left(\frac{\pi}{1-\pi}\right) = XB$$

3. link function

---

1. Random component for multicategory data?

$$\vec{Y} = (Y_1, Y_2 \ldots, Y_J)$$

$$\vec{Y} \sim \text{Multinomial}(n, \vec{\pi}) \qquad \vec{\pi} = (\pi_1, \pi_2, \ldots \pi_J)$$

$$\forall_j \quad \pi_j \in [0,1]$$

$$\sum_{j=1}^{J} \pi_j = 1$$

2. Functional form $\rightarrow X\vec{B}$

3. Link function?

Baseline-Category Logits

Let J be a baseline, then our baseline logit is $\log\left(\frac{\pi_j}{\pi_J}\right)$ $\rightarrow$

$J = 2$

$\pi_1 + \pi_2 = 1$

$\pi_2 = 1 - \pi_1$

$\log\left(\frac{\pi_1}{\pi_2}\right) = \log\left(\frac{\pi_1}{1-\pi_1}\right)$

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x$$

$$\log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x$$

$\rightarrow$ require $J-1$ equations

AND

$$\log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

$$\log\left(\frac{\pi_1}{\pi_2}\right) = \log\left(\frac{\pi_1 / \pi_J}{\pi_2 / \pi_J}\right) = \log\left(\frac{\pi_1}{\pi_J}\right) - \log\left(\frac{\pi_2}{\pi_J}\right)$$

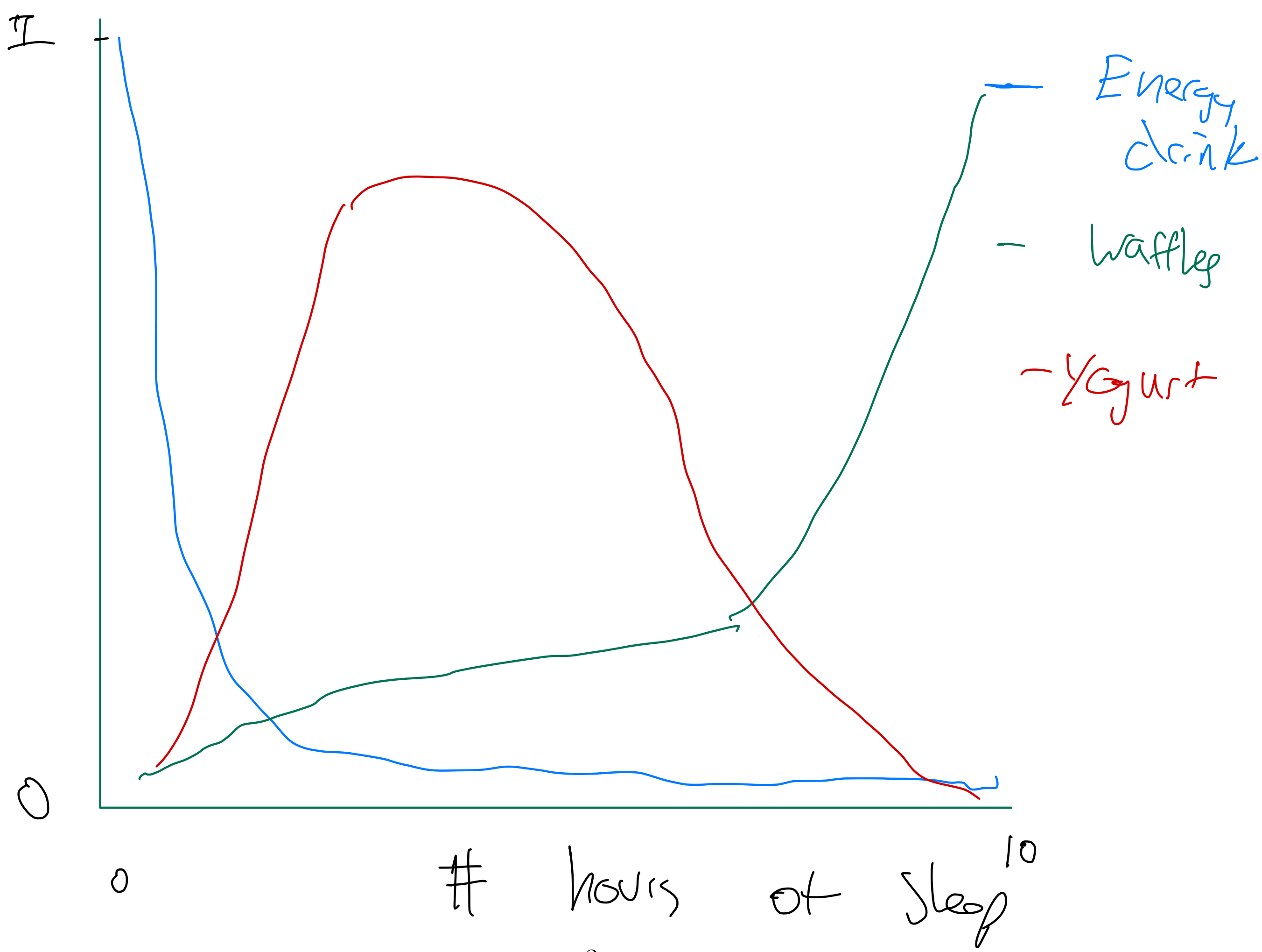
$$= (\alpha_1 + \beta_1 x) - (\alpha_2 + \beta_2 x) \Rightarrow (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x$$

$\rightarrow$ odds ratio

Baseline-Category Logits

We want estimate $\vec{\pi}(x)$

$$\pi_i(x) = \frac{\exp(\alpha_i + \beta_i x)}{\sum_{j=1}^{J} \exp(\alpha_j + \beta_j x)} \rightarrow \alpha_J + \beta_J = 0$$



Ordinal Data

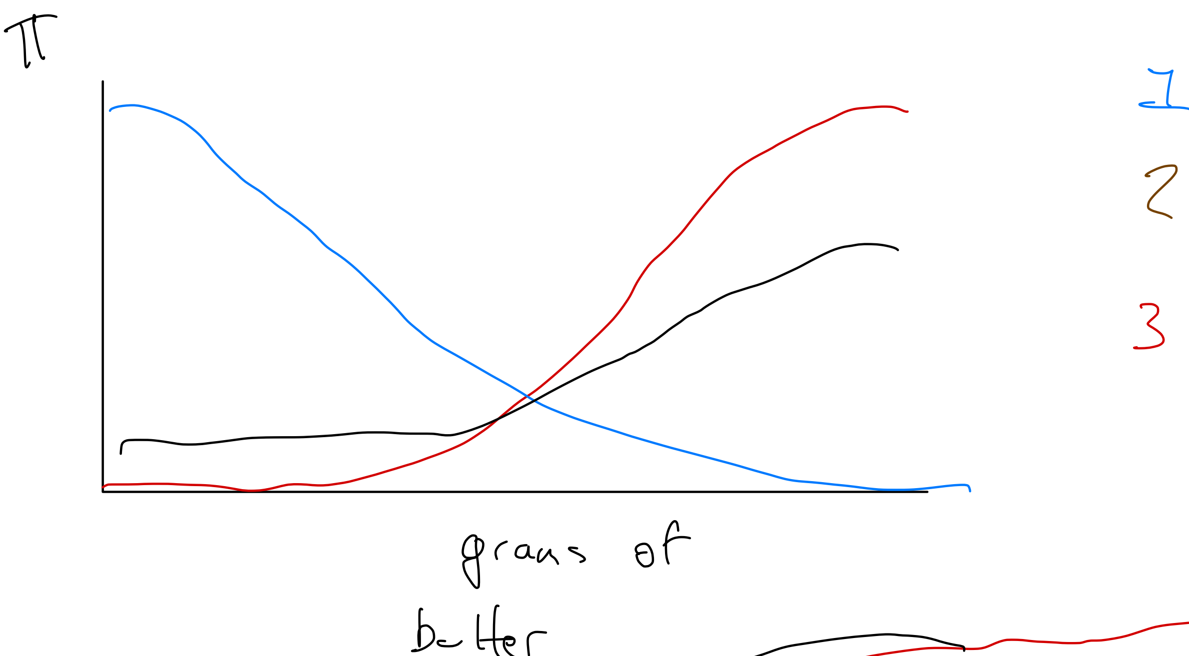
$$\Pr[Y \le j] = \pi_1 + \pi_2 + \cdots \pi_j$$

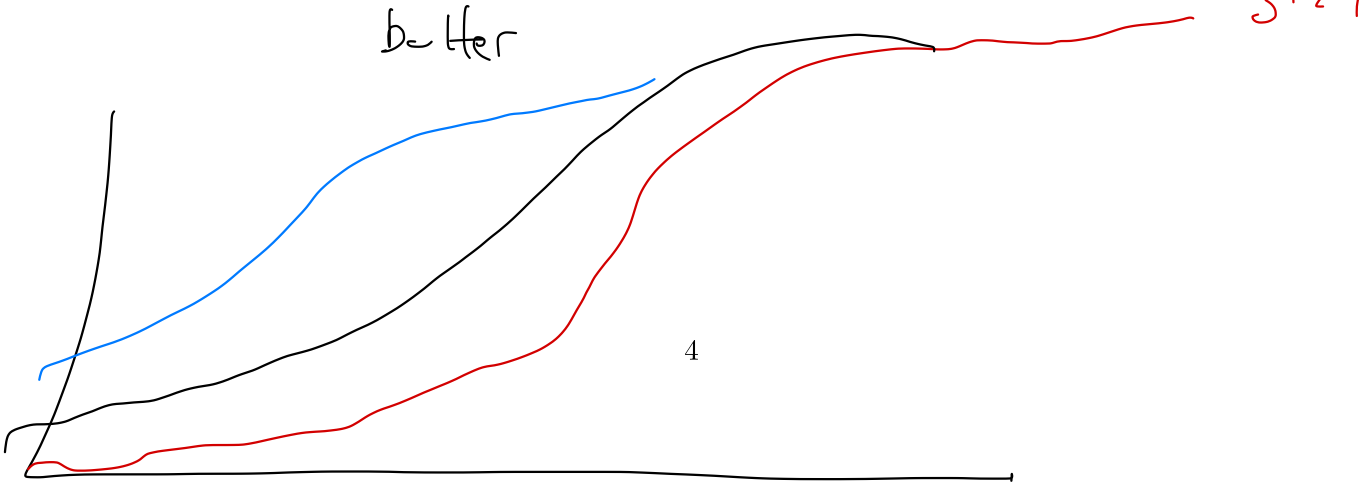
5

cumulative probabilities

1. Multinomial

$$\operatorname{logit}(P[Y \le j]) = \alpha_j + \beta x$$



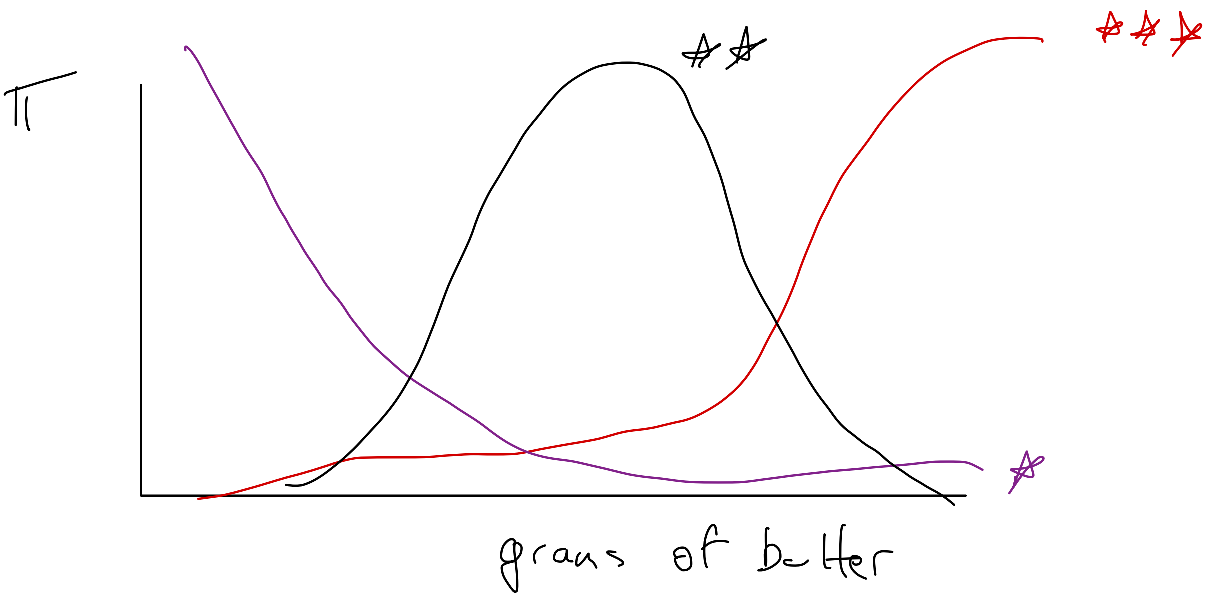


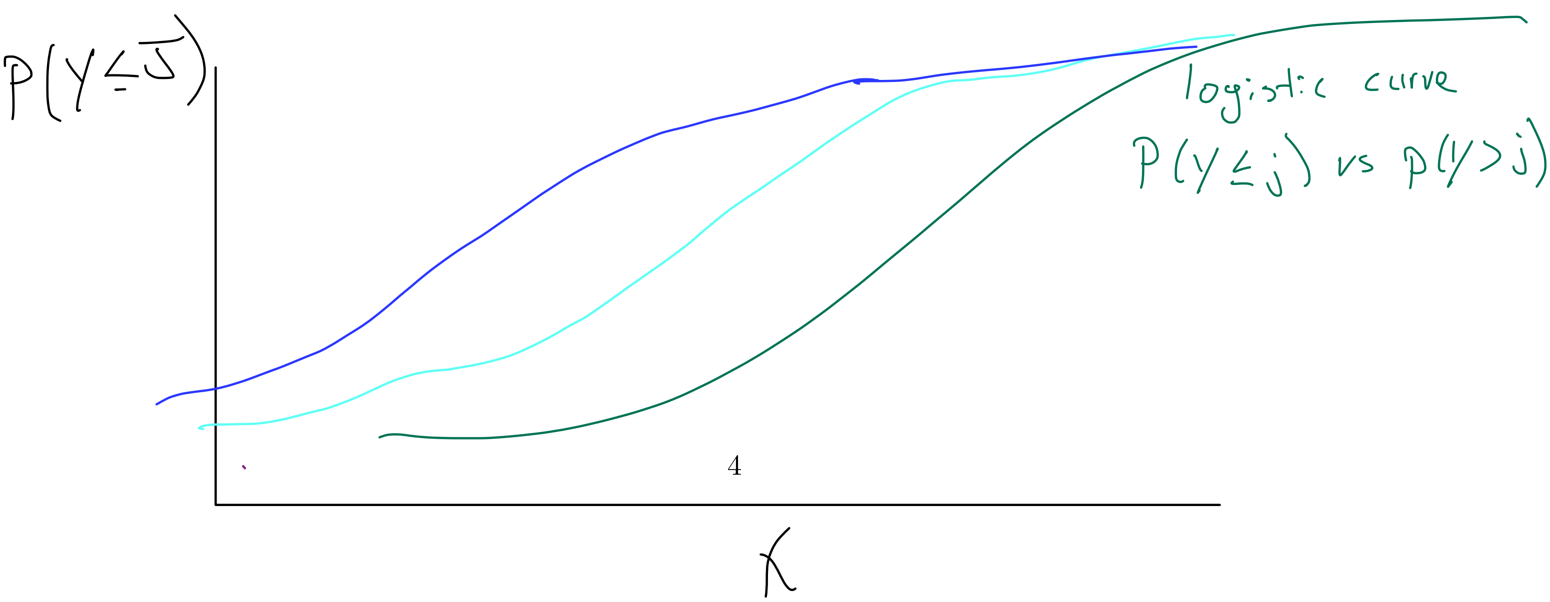
Ordinal Data

5

$$Pr\left[Y \le j\right] = \pi_1 + \pi_2 + \dots \pi_j$$

cumulative probabilities

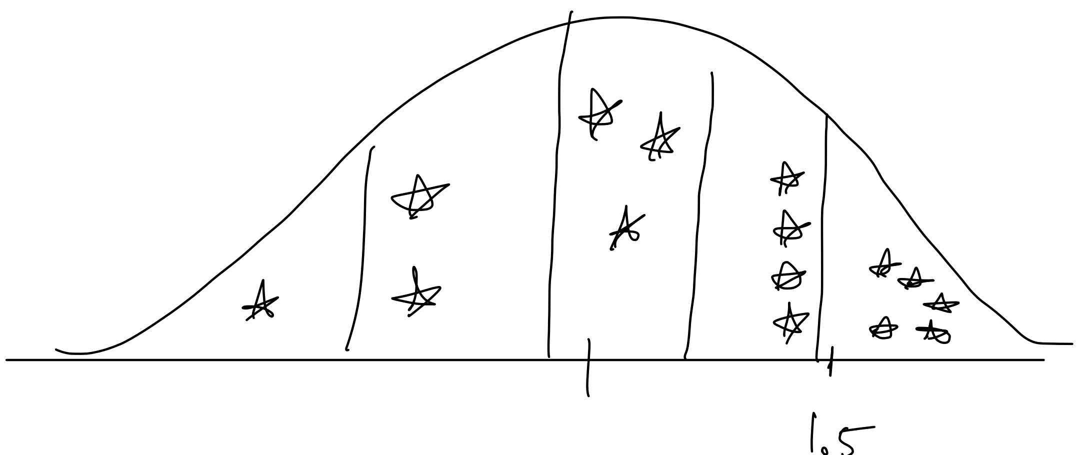




Latent variable interpretation

↳ unmeasured (hidden) continuous variable

↳ Normal continuous



Probit → CDF of normal