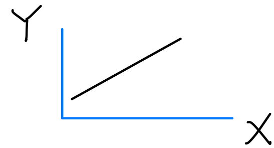
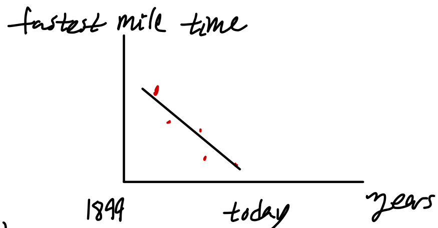
Limited Dependent Variables

2022 - 02 - 15

Regression w/ Y that can only take on certain values

Linear Regression $Y = \boxed{\alpha + \beta X} + U$





NONLINEAR MODELS!

Predict Y

how many?

Count Data : $Y \in \{0, 1, 2, \ldots\}$

Poisson Regression

patents, # kids ...

Binary Outcomes : $Y \in \{0, 1\}$

Probit / Logit

employed / unemployed ; win / lose

Multinomial : $Y \in \{1, 2, 3, \ldots, J\}$

random utility models <u>RUMs</u>

structural economic model of choice!

categorical

Selection : ① "economic model of choice"

Heckman selection model

② Use probit as an ingredient

GRE score

sample selection / missing data

$$Y = \alpha + \boxed{\beta} X + U$$

score on econometrics exam

What if we only observe Y for a subset of people?

* Marno Verbeek Modern Econometrics *

Poisson Regression

$Y \in \{0, 1, 2, 3, \dots\}$

?

$$Y_i \mid \underline{X}_i \sim \boxed{\text{Poisson}}\Big(\underbrace{\boxed{\exp\{\underline{X}_i'\beta\}}}_{e^{(\sim)}}^{\mu_i}\Big) \in \{0, 1, 2, \dots\}$$

When/why would this work?

① Old-fashioned : THE MODEL IS TRUE!

* ② What if the model is NOT true?

How to learn β? Maximum Likelihood!

$$Y_1, \dots, Y_n \sim iid\ f(y; \theta)$$

$$L(\theta) = \prod_{i=1}^{n} f(y_i; \theta) \longleftarrow$$

find θ to max this!

max like est

If model is TRUE then MLE is the best you can do.

$\hat{\theta}_{MLE} \xrightarrow{P} \theta_{true}$, $\hat{\theta}_{MLE}$ has lowest Asymp. Variance

$\boxed{\text{But what if it's NOT true?}}$

more true

$\boxed{\text{TRUE}}$ $Y_i | X_i \sim iid \ \boxed{\cancel{\text{Poisson}}}(\exp\{\underline{X}_i' \beta\})$ $E(Y_i|X_i) = \exp(\underline{X}_i' \beta)$

$\boxed{\text{correct } E(Y|X)}$ NOT Poisson, but $Y_i | X_i \sim iid$ & $E(Y_i|X_i) = \exp(\underline{X}_i'\beta)$

$\boxed{\text{TOTALLY NOT TRUE}}$ $Y_i | X_i \sim iid$?

$\boxed{\text{Does MLE give us anything meaningful?}}$

less true

$\boxed{Y_i \sim \text{Poisson}(\theta_i)}$ Too many params!

Poisson Regression: MODEL for how θ_i varies w/ i

$$\theta_i = \exp\{\alpha + \beta \underline{X}_i\}$$

depending on my X_i different θ_i

TRUE dist of Y

$$KL(p_0 ; f_\theta) = \boxed{\mathbb{E}\left[\log p_0(Y)\right]} - \boxed{\mathbb{E}\left[\log f(Y;\theta)\right]}$$

Doesn't depend on our model f_θ

$\underbrace{\hphantom{xxxxxxxxxxxx}}$

unknown const

this we can calculate! we know f

$\mathbb{E} \approx \frac{1}{n}\Sigma$

plug-in for θ

$$\mathbb{E}(\log f(Y;\theta)) = \int_{-\infty}^{\infty} \boxed{\log f(Y;\theta)}\,\boxed{p_0(y)}\,dy$$

$$\approx \boxed{\frac{1}{n}\sum_{i=1}^{n} \log f(Y_i;\theta)}$$

LLN

$\mathbb{E}(\boxed{Y})$; $Y_1, \dots, Y_N$ iid

$\approx \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

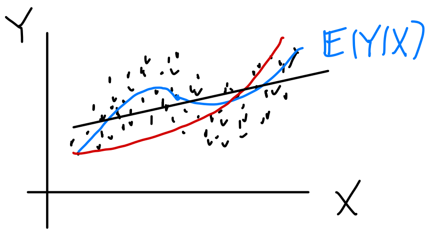
2022-02-16

Info matrix equality (1.4)

AVAR calc for index models (3.6)

Pseudo-R^2 (3.7)

LESS IMPORTANT!



NP regression estimate a conditional mean function

$Y = m(X) + U$

$U \equiv Y - \mathbb{E}(Y|X)$

$$\int \log f(y|\theta)\, \boxed{p_0(y)}\, dy \approx \frac{1}{n}\sum_{i=1}^{n} f(y_i|\theta)$$

Are there discrete Y dists for which we could develop regression model?

⟹ Logit / probit / multinomial logit

categorical

⟹ Binomial Regression (related to logit)

KNOW max count N

⟹ Ordered probit / logit model for rankings...

NOT in this class

Binomial Dist (n, p)

$$\mathbb{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$Y \in \{0,1\} \implies \mathbb{E}(Y|X) = \mathbb{P}(Y=1|X)$$

$$\mathrm{Var}(Y|X) = \mathbb{P}(Y=1|X)\underbrace{\boxed{\mathbb{P}(Y=0|X)}}_{1-\mathbb{P}(Y=1|X)}$$

Poisson Regression: can have $\mathbb{E}(Y|X) = \exp(X'\beta)$

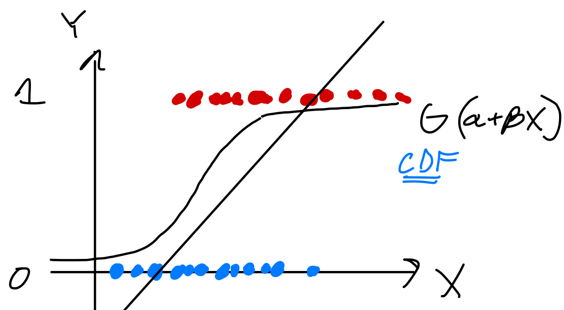
$$\boxed{\text{BUT } \mathrm{Var}(Y|X) \neq \exp(X'\beta)}$$

! Not really Poisson data BUT still getting $\mathbb{E}(Y|X)$ correct

Index Models

① logit has a simple CDF ← mathematically

& is similar to Normal



Look at random utility models $\Rightarrow$ motivation for logit...

② Normal Everyone's favorite dist!

Nice way to build more complicated models of choice from normal dists... ← linear structure

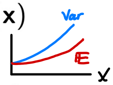
③ t-distribution... (~~rorit~~? robit?) Robust est.

1. Poisson Assumption: $\text{Var}(y|\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$

most real-world count data DON'T satisfy ①

   - holds if Poisson model is correct.

2. Quasi-Poisson Assumption: $\text{Var}(y|\mathbf{x}) = \sigma^2 \mathbb{E}(y|\mathbf{x})$



   - Allows for possibility that $y|\mathbf{x}$ is *not* Poisson
   - Overdispersion: $\sigma^2 > 1 \implies \text{Var}(y|\mathbf{x}) > \mathbb{E}(y|\mathbf{x})$
   - Underdispersion $\sigma^2 < 1 \implies \text{Var}(y|\mathbf{x}) < \mathbb{E}(y|\mathbf{x})$

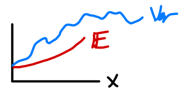
$E(Y|X) = \exp(X'\beta)$

   - If $\sigma^2 = 1$ we're back to the Poisson Assumption.

2' negative binomial regression

model Var(Y|X)

3. No Assumption: $\text{Var}(y|\mathbf{x})$ unspecified



& no ~~attempt~~ to model Var(Y|X)

DIG this out!

$\boxed{\text{How do we KNOW ~~that~~ pseudo-}R^2 \text{ is always well-defined? OLD PS question...}}$

LATENT variable interp. of probit / logit?
YES! ⟹ Chapter 4

Another way to think about Poisson regression...

MLE, What is the Popn. FOC? (max E(log-Like))

$$(2.6) \quad \underline{s}_i \equiv \frac{\partial \ell_i}{\partial \beta} = \underline{X}_i\left[Y_i - \exp(\underline{X}_i'\beta)\right]$$

$$\mathbb{E}(\underline{s}_i) = 0 \quad ; \quad \boxed{\mathbb{E}\left[\underline{X}_i \left\{ Y_i - \exp(\underline{X}_i' \beta) \right\}\right] = 0}$$

$$\mathbb{E}\left[\underline{X}_i (Y_i - \underline{X}_i' \beta)\right] = 0$$

if $\mathbb{E}(Y_i | X_i) = \exp(\underline{X}_i' \beta_0)$

$$\boxed{\mathbb{E}\left[X \left\{ Y - \mathbb{E}(Y|X) \right\}\right] = 0}$$