# La régression Bêta

## Une alternative intéressante pour modéliser des proportions

Maxime Lacroix

28 septembre 2018

# Mise en contexte

# Contexte

- Régression sur une variable réponse tenue entre (0,1)
- Par exemple, un taux ou une proportion
- Régression linéaire “classique” à éviter

# Première solution : transformation logit

## Première transformation possible

$$\tilde{y} = log(\frac{y}{1-y}) \quad (1)$$

- Avantage :
  - Les données ne sont plus bornées, la régression linéaire est envisageable
- Désavantages :
  - Interprétation différente
  - Fort potentiel d'hétéroscédasticité
  - Les données sont souvent asymtrique $\rightarrow$ problèmes pour les tests d'hypothèses et les intervalles de confiance.

# Solution : Régression Bêta

# Brève présentation

- Présentée pour la première fois en 2004 par Ferrari et Cribari-Neto
- Intérêt majeur :
  - La densité bêta prend différentes formes dépendemment des paramètres
  - Densité généralement hétéroscédastique
  - Interprétation semblable à la régression logistique

METTRE DES GRAPHIQUES QUI PROUVENT LE POINT 2

# Présentation mathématique

### Densité d'une loin bêta

$$f_Y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \tag{2}$$

De l'équation 2, Ferrari et Cribari-Neto ont proposé une nouvelle paramétrisation, en posant :

- $\mu = \frac{\alpha}{\alpha+\beta}$
- $\phi = \alpha + \beta$

# Nouvelle paramétrisation

On obtient :

**Densité sous la nouvelle paramétrisation**

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3}$$

```
##      speed            dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```