Beta Regression

Jaerin Kim

Beta Distribution

- Beta distribution is very flexible for $y \in [0, 1]$
- The distribution contains most of what we can think of as "single-peaked", or "unimodal"

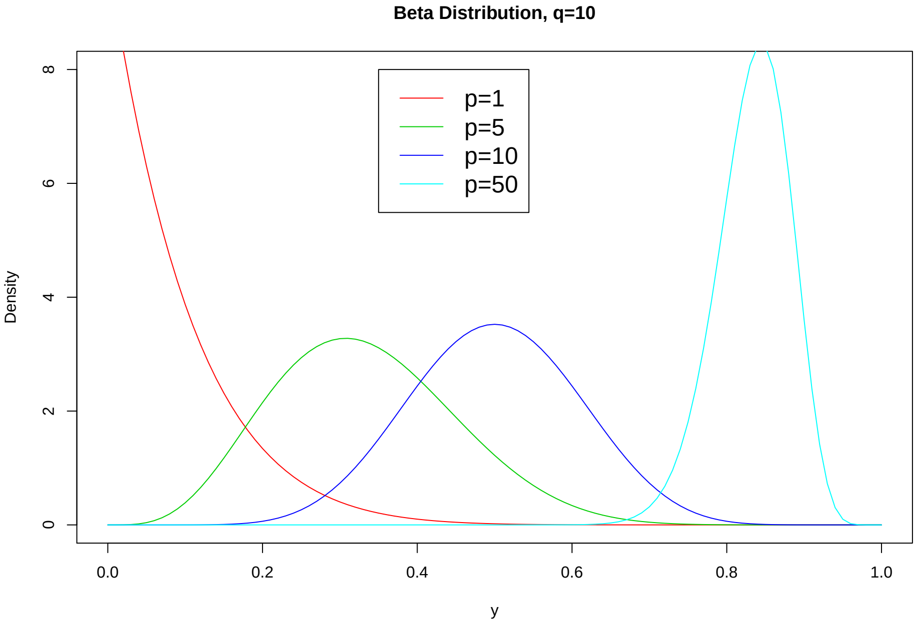
Beta Distribution

- The beta distribution as we know it is parametrized as

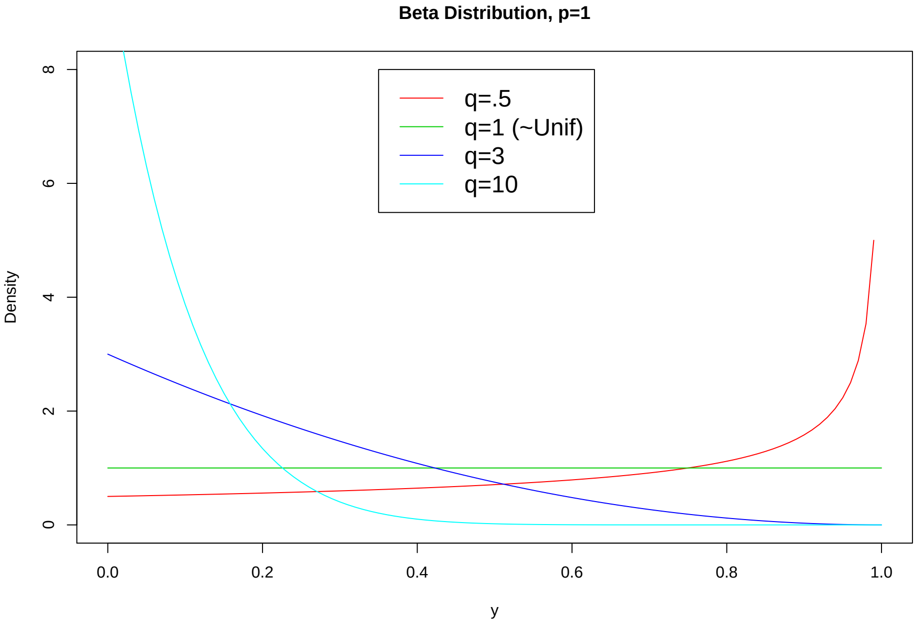
$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)+\Gamma(q)} y^{p-1}(1-y)^{q-1},\ 0<y<1,\ p,q>0$$

- Different values of p and q result in various shapes in distribution

Beta Distribution



Beta Distribution



Beta Distribution

- This will be useful for dependent variables in $[0, 1]$, e.g. proportion.
- But the parameters don't look very intuitive.
- Let's follow Ferrari and Cribari-Neto (2004) and alter parameterization.

Beta Distribution

$$\mu = \frac{p}{p+q} \text{ and } \phi = p + q$$

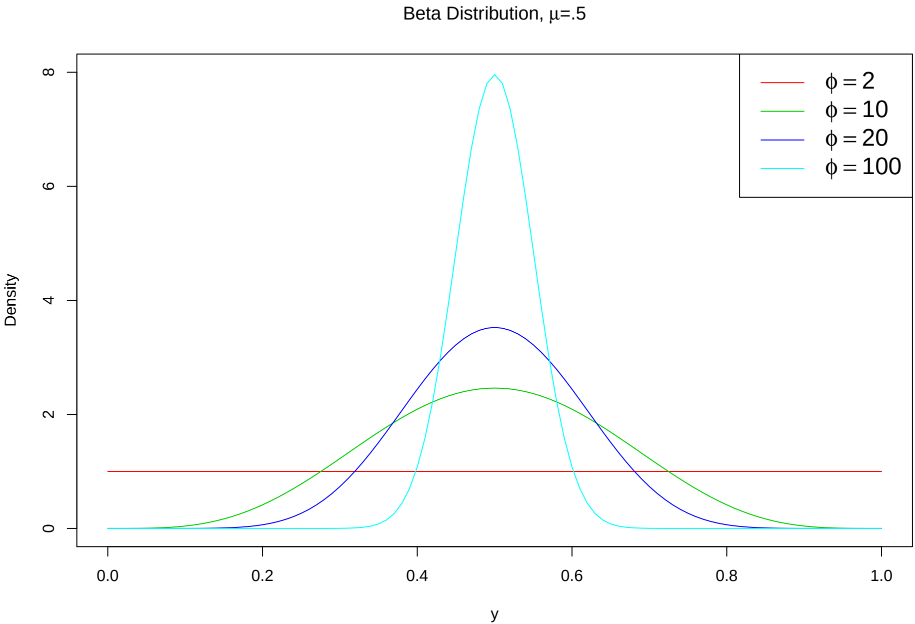
Then,

$$f(y;p,q) = f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi) + \Gamma((1-\mu)\phi} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

where

$$0 < \mu < 1,\ 0 < y < 1, \phi > 0$$

Beta Distribution



Beta Distribution

- Notice that the distribution is centered around $\mu = .5$
- Actually,

$$\mathbb{E}(y) = \mu \text{ and } \mathbb{V}(y) = \frac{\mu(1-\mu)}{1+\phi}$$

- For this reason, ϕ is called "precision parameter"

Beta Regression

- But how do we move on to beta regression?
- We assume that the dependent variable is distributed beta
- Independent variables determine parameters for y_i

Beta Regression

- In short,

$$y_i \sim B(\mu_i, \phi_i)$$

where μ_i and ϕ_i are determined by regressors(x_{ij}) and their coefficients (β_j)

- The problem is, that we are constrained by

$$0 < \mu < 1 \text{ and } \phi > 0$$

Link Function

- To solve this problem, we need a "link function"
  - Some function $(0, 1) \mapsto \mathbb{R}$ (for μ)
  - Some function $(0, \infty) \mapsto \mathbb{R}$ (for ϕ)

Link Function

An example - logit: $(0, 1) \mapsto \mathbb{R}$

$$logit(x) = \log \frac{x}{1 - x}$$

- where we find that, as $x \to 1$, $logit(x) \to \infty$
- and as $x \to 0$, $logit(x) \to -\infty$

An example - log: $(0, \infty) \mapsto \mathbb{R}$

$$\log x$$

- where we find that, as $x \to \infty$, $\log x \to \infty$
- and as $x \to 0$, $\log x \to -\infty$

Link Function

- Where do we plug these into?
- Let g_1 be a link function for μ and g_2 a link function for ϕ. Then,

$$g_1(\mu) = x_i^T \beta = \eta_{1i}$$
$$g_2(\phi) = z_i^T \gamma = \eta_{2i}$$

where the vectors z_i and x_i need not be mutually exclusive.

Link Function

- But what we need is estimates for μ and ϕ. That is, we need inverse functions.

$$g_1(\mu) = x_i^T \beta = \eta_{1i}$$

$$g_1(\mu) = x_i^T \beta = \eta_{1i}$$

$$\mu = g_1^{-1}(\eta_{1i}) = g_1^{-1}(x_i^T \beta)$$

and

$$g_2(\phi) = z_i^T \gamma = \eta_{2i}$$

$$\phi = g_2^{-1}(\eta_{2i}) = g_2^{-1}(z_i^T \gamma)$$

Link Function

An example - $g_1^{-1} = logit^{-1}$

$$g_1(\mu_i) = logit(\mu_i) = \log \frac{\mu_i}{1-\mu_i} = x_i^T \beta = \eta_{1i}$$

$$\frac{\mu_i}{1-\mu_i} = e^{\eta_{1i}}$$

$$\mu_i = (1-\mu_i) e^{\eta_{1i}}$$

$$\mu_i = \frac{e^{\eta_{1i}}}{1+e^{\eta_{1i}}} = g_1^{-1}(\eta_{1i}) = g_1^{-1}(x_i^T \beta)$$

Link Function

An example - $g_2^{-1} = log^{-1}$

$$g_2(\phi_i) = \log \phi_i = \gamma_i^T \beta = \eta_{2i}$$

$$\phi_i = e^{\eta_{2i}} = g_2^{-1}(\eta_{2i}) = g_2^{-1}(z_i^T \gamma)$$

- Now we can replace ϕ_i and μ_i with regressors x_i and z_i and regression parameters β and γ
- Let's move on to the likelihood function

Likelihood Function

- Recall that beta distribution,

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi) + \Gamma((1-\mu)\phi} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

- and $y_i \sim B(\mu_i, \phi_i) = B(g^{-1}(x_i^T\beta), g_2^{-1}(z_i^T\gamma))$

Likelihood Function

- Then, the log-likelihood function is

$$\mathcal{L}(\beta, \gamma) = \sum_{i=1}^{n} [\log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1-\mu_i)\phi_i) +$$

$$(\mu_i \phi_i - 1) \log y_i + ((1-\mu_i)\phi_i - 1) \log (1 - y_i)]$$

where $\mu_i = g^{-1}(x_i^T \beta)$ and $\phi_i = g_2^{-1}(z_i^T \gamma)$

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

- Suppose that we have only one regressor. Then, the log-likelihood function will be like

```
simpleb<-function(xi,yi,bg){
  phii<-exp(bg[3]+xi*bg[4])
  mui<-exp(bg[1]+xi*bg[2])/(1+exp(bg[1]+xi*bg[2]))
  return(sum(log(gamma(phii))-log(gamma(mui*phii))-
               log(gamma((1-mui)*phii))+(mui*phii-1)*log(yi)+
               ((1-mui)*phii-1)*log(1-yi)
  ))}
```

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

- An easier alternative is *betareg()* function from **betareg** package.
    - Suppose that $y_i \sim B(logit(x_i), e^{x_i})$. We will see how *betareg()* performs.

```
set.seed(0)
library(betareg)
xi<-runif(10000)
yi<-c()
mui<-function(x)exp(x)/(1+exp(x))
phii<-function(x)exp(x)
for (i in xi){yi<-c(yi,rbeta(1,mui(i)*phii(i),(1-mui(i))*phii(
test<-betareg(yi~xi|xi)
```

```
## Warning in betareg.fit(X, Y, Z, weights, offset, link, link
## control): no valid starting value for precision parameter f
## instead
```

```
betac<-test$coefficients$mean
gammac<-test$coefficients$precision
```

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

- What did we get?

```
## $mean
##  (Intercept)            xi
## -0.002001234  1.047367021
##
## $precision
##  (Intercept)            xi
## -0.003598346  1.027131001
```

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

- So the model says,

$$\hat{\mu} = \frac{e^{0+1.05x}}{1 + e^{0+1.05x}}$$

$$\frac{\partial \hat{\mu}}{\partial x} = \frac{\partial[1-(1+e^{1.05x})^{-1}]}{\partial x} = 1.05e^{1.05x}(1+e^{1.05x})^{-2} > 0$$

- Then, the expectation of y increases as x increases

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

- Also, the model says,

$$\hat{\phi}_i = e^{0+1.03x_i}$$

$$\frac{\partial \hat{\phi}}{\partial x} = 1.03e^{1.03x} > 0$$

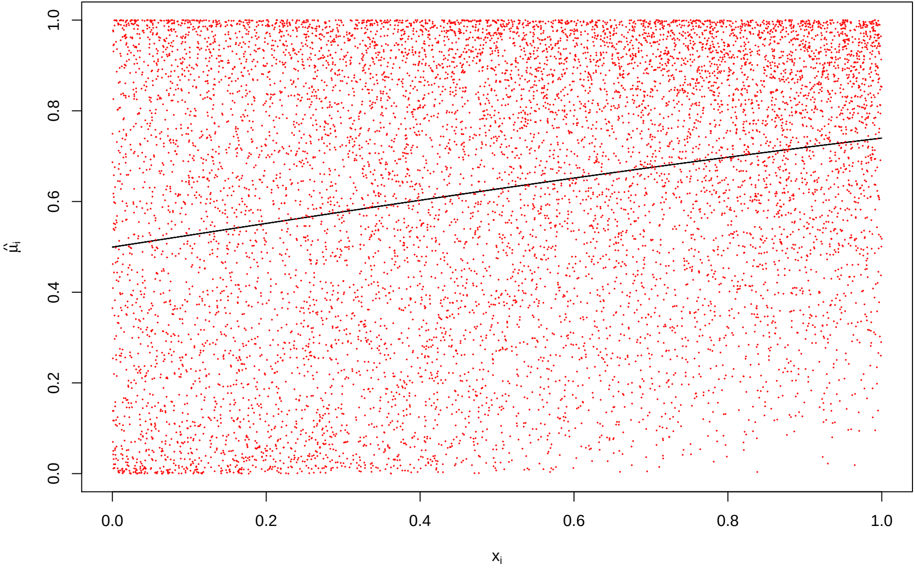
- This means that, as x increases, the precision parameter increases
  - Put another way, increased x will result in decreased dispersion in y.

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$

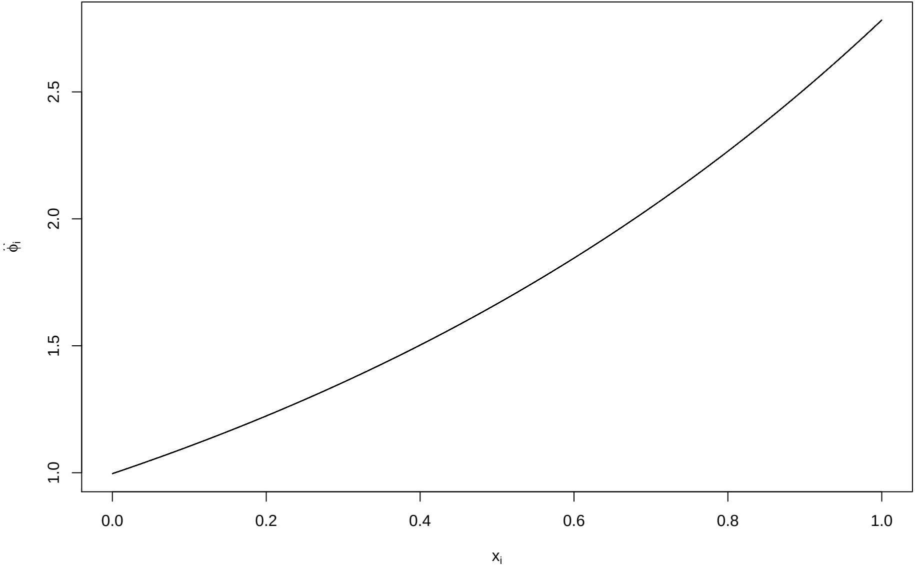
- Let's see how $\hat{\mu}_i$ and $\hat{\phi}_i$ behave as x_i changes

```
xi1<-cbind(rep(1,length(xi)),xi)
muihat<-exp(xi1%*%betac)/(1+exp(xi1%*%betac))
phiihat<-exp(xi1%*%gammac)
```

Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$



Example: $g_1(\mu_i) = logit(\mu_i)$ and $g_2(\phi_i) = \log \phi_i$



Application: Union Density and Subjective Class

- Do people living in a country with higher union density identify with working class more?
- They surely can, since union membership might make them conscious of belonging to the working class
- Yet, it is still possible that people are more conscious of their class when there is no union to protect them

Application: Union Density and Subjective Class

- OECD data on union density (percent)
- Eurobarometer data on subjective identification with working class (proportion)
- European countries in 2014
- Codes for preparing the data omitted

Application: Union Density and Subjective Class

- Data: **ud2014adm**
- y: Subjective Identification with Working Class, **sclass**
- x: Union Density **obsValue** Unit: Country

```
## Warning: Unknown or uninitialised column: 'sclass'.
```

```
## Running beta regression with betareg()
## betareg(y~x|z) where x is regressor(s) for mu
## and z is regressor(s) for phi
scl<-betareg(ud2014admp$sclass~ud2014admp$obsValue|ud2014admp$
## By default, link functions are
## g[1](x)=logit(x) and
## g[2](x)=log(x)
```

Application: Union Density and Subjective Class

- Again, $\hat{\mu_i} = g^{-1}(x_i^T \hat{\beta})$ and $\hat{\phi}_i = g_2^{-1}(z_i^T \hat{\gamma})$
- Since $g_1(x) = logit(x)$ and $g_2(x) = \log x$,

$$\hat{y}_i \sim B(g^{-1}(x_i^T \hat{\beta}), g_2^{-1}(z_i^T \hat{\gamma})$$

- And we just got $\hat{\beta}$ and $\hat{\gamma}$ from *betareg()*

Application: Union Density and Subjective Class

```r
scl$coefficients
```

```
## $mean
##         (Intercept) ud2014admp$obsValue
##         -0.63784174         -0.01268848
##
## $precision
##         (Intercept) ud2014admp$obsValue
##          1.53632092          0.04347422
```

Application: Union Density and Subjective Class

$$\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta}) = \frac{e^{x_i^T \hat{\beta}}}{1 + e^{x_i^T \hat{\beta}}} = \frac{e^{-.638 - .013 x_i}}{1 + e^{-.638 - .013 x_i}}$$

and

$$\hat{\phi}_i = g_2^{-1}(z_i^T \hat{\gamma}) = e^{1.54 + .043 x_i}$$

- where we estimate that y_i is distributed beta.

Application: Union Density and Subjective Class

- Germany has union density of 17.74%. How many people would think that they belong to working class?

```
print(Guden<-scl$model[3,2])
```

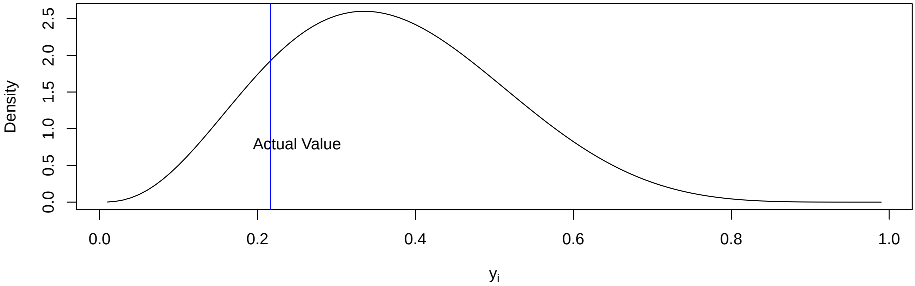
```
## [1] 17.7378
```

```
Gxbeta<-coef(scl)[1]+coef(scl)[2]*Guden
Gzgamma<-coef(scl)[3]+coef(scl)[4]*Guden
Gmui<-exp(Gxbeta)/exp(1+Gxbeta)
Gphii<-exp(Gzgamma)
```

Application: Union Density and Subjective Class

- Note that $B(p, q) = Beta(\mu\phi, (1-\mu)\phi)$

```
plot(1:99/100,dbeta(1:99/100,Gmui*Gphii,(1-Gmui)*Gphii),"l",
     xlab=expression(y[i]),ylab="Density")
abline(v=scl$model[3,1],col=4)
text(.25,.8,"Actual Value")
```



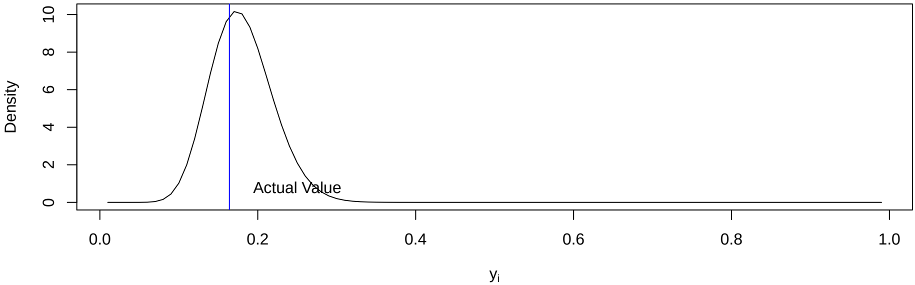
Application: Union Density and Subjective Class

- How about Denmark?

```
Dxb<-t(c(1,scl$model[16,2]))%*%scl$coefficients$mean
Dmu<-exp(Dxb)/(1+exp(Dxb))
Dphi<-exp(t(c(1,scl$model[16,2]))%*%scl$coefficients$precision
```

Application: Union Density and Subjective Class

```
plot(1:99/100,dbeta(1:99/100,Dmu*Dphi,(1-Dmu)*Dphi),"l",
     xlab=expression(y[i]),ylab="Density")
abline(v=scl$model[16,1],col=4)
text(.25,.8,"Actual Value")
```



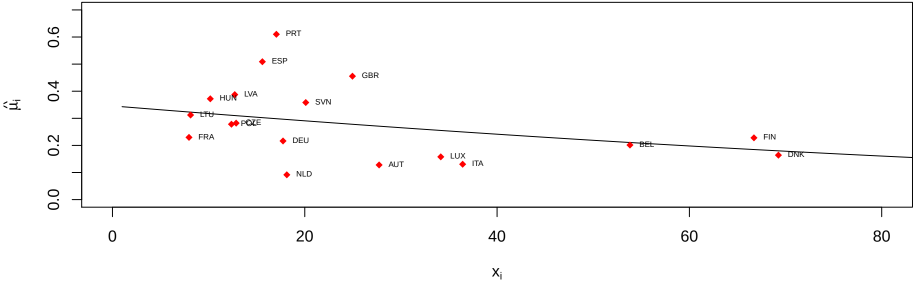
Application: Union Density and Subjective Class

- Let's generalize

```
xi1<-cbind(rep(1,length(1:100)),1:100)
betac<-scl$coefficients$mean
gammac<-scl$coefficients$precision
muihat<-exp(xi1%*%betac)/(1+exp(xi1%*%betac))
phiihat<-exp(xi1%*%gammac)
```

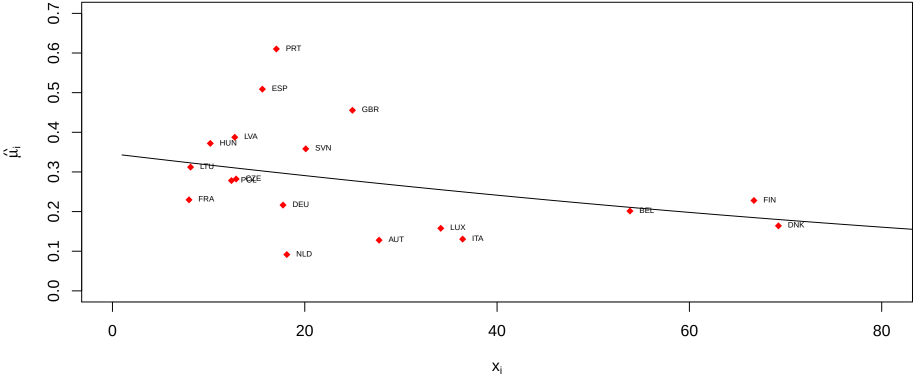
Application: Union Density and Subjective Class

```
plot(1:100,muihat,"l",ylim=c(0,.7),xlim=c(0,80),xlab=expressio
points(ud2014admp$obsValue,ud2014admp$sclass,col=2,pch=18)
text(ud2014admp$obsValue[!is.na(ud2014admp$sclass)],ud2014admp
```



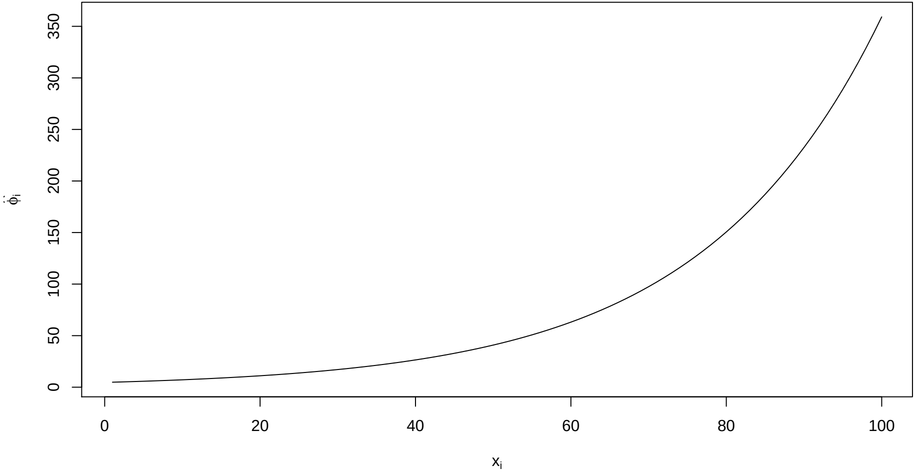
Application: Union Density and Subjective Class

- We have heteroskedasticity here
  - Dispersion is higher for lower values of y_i
- Does ϕ_i reflect that?



Application: Union Density and Subjective Class

```
plot(1:100,phiihat,"l",
     xlab=expression(x[i]),ylab=expression(hat(phi[i])))
```



Application: Union Density and Subjective Class

- In sum, higher union density leads to lower working class identification
- And dispersion of working class identification decreases as union density increases
- Change in union density by

```
scl$model[16,2]-scl$model[3,2]
```

```
## [1] 51.52154
```

leads to decrease in working class identification by

```
Gmui-Dmu
```

```
##              [,1]
## [1,] 0.1879222
```

Application: Union Density and Subjective Class

- Change in union density by

```
scl$model[16,2]-scl$model[3,2]
```

```
## [1] 51.52154
```

leads to increase in the precision parameter by

```
Dphi-Gphii
```

```
##          [,1]
## [1,] 84.32998
```

Application: Union Density and Subjective Class

- Countries with lower union density seems to have higher identification with working class
  - And the dispersion is higher when union density is lower
- My interpretation is that when union density is high, workers are more likely to get nation-wide benefits, making them better off. Identities other than class might become more important to them in this case.

Conclusion

- A flexible model suited for proportional data
  - There are proportional data everywhere in political science
  - Turnout, vote share, seat share, issue salience, Gini index, etc.
- Heteroskedasticity? No problem
  - ϕ_i means that you can capture each i's dispersion
- Still, its interpretation requires some algebra