Lecture 3: Hierarchical modeling

Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes

hlopes@ChicagoBooth.edu

Outline

Notation

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

- y_{it}: response of individual i at time t;
- x_{it}: regressors for individual i at time t;
- x_{it} is a k-dimensional vector;
- ϵ_{it}: error terms.

- Matrix notation

$$X_i = \begin{pmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{pmatrix} \quad y_i = \begin{pmatrix} y'_{i1} \\ \vdots \\ y'_{iT} \end{pmatrix} \quad \epsilon_i = \begin{pmatrix} \epsilon'_{i1} \\ \vdots \\ \epsilon'_{iT} \end{pmatrix}$$

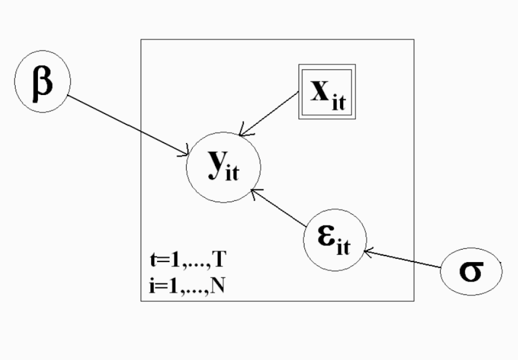
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

Pooled model

For $i = 1, \ldots, N$

$$y_i \sim N(X_i\beta, \sigma^2 I_T)$$

with $\beta = (\alpha, \tilde{\beta})$ and $X_i = (1_T, \tilde{X}_i)$.

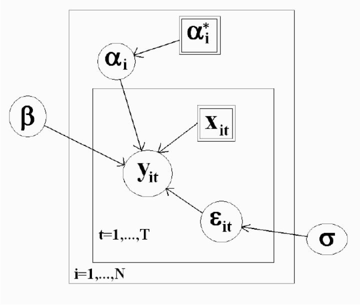


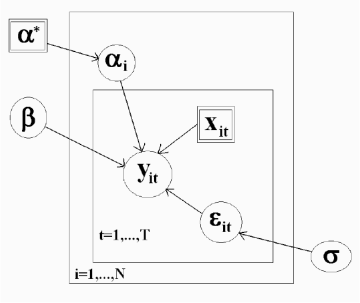
Individual effects model

For $i = 1, \ldots, N$

$$y_i \sim N(1_T \alpha_i + \tilde{X}_i \tilde{\beta}, \sigma^2 I_T)$$

with $\beta_i = (\alpha_i, , \tilde{\beta})$.



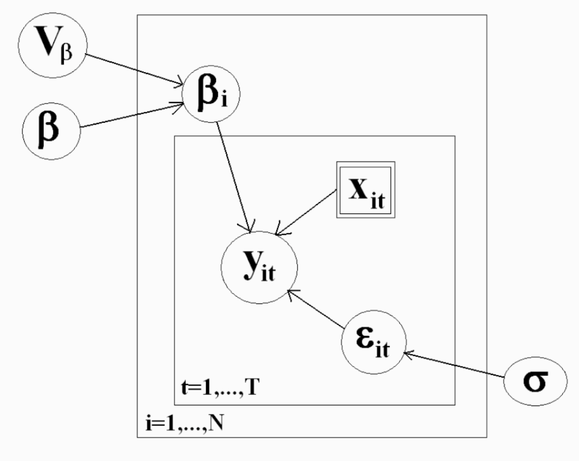


Random coefficients model

For $i = 1, \ldots, N$

$$y_i \sim N(X_i\beta_i, \sigma^2 I_T)$$

with $\beta_i = (\alpha_i, , \tilde{\beta}_i)$.



Pooled Model

Individual level:

$$y_i \sim N(X_i\beta, \sigma^2 I_T)$$

Combining individuals

$$y \sim N(X\beta, \sigma^2 I_{TN})$$

Likelihood function

$$p(y|\beta,\sigma^2) \propto \sigma^{-TN} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}$$

Prior distribution

$$\beta \sim N(\beta_0, V_0) \quad \text{and} \quad \sigma^2 \sim IG(\nu_0/2, \nu_0 s_0^2/2)$$

Full conditionals

$$\beta|\sigma^2 \sim N(\beta_1, V_1)$$

where

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + \sigma^{-2} X'X \\ V_1^{-1}\beta_1 &= V_0^{-1}\beta_0 + \sigma^{-2} X'y \end{aligned}$$

$$\sigma^2|\beta \sim IG(\nu_1/2, \nu_1 s_1^2/2)$$

where

$$\begin{aligned} \nu_1 &= \nu_0 + TN \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (y - X\beta)'(y - X\beta) \end{aligned}$$

Individual effects model: non-hierarchical prior

Individual level

$$y_i \sim N(1_T \alpha_i + \tilde{X}_i \tilde{\beta}, \sigma^2 I_T)$$

Combining individuals

$$y \sim N(X^* \beta^*, \sigma^2 I_{TN})$$

where $\beta^* = (\alpha_1, \ldots, \alpha_N, \tilde{\beta}')$ and

$$X^* = \begin{pmatrix} 1_T & 0 & 0 & \cdots & 0 & \tilde{X}_1 \\ 0 & 1_T & 0 & \cdots & 0 & \tilde{X}_2 \\ 0 & 0 & 1_T & \cdots & 0 & \tilde{X}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1_T & \tilde{X}_1 \end{pmatrix}$$

Prior distribution

$$\beta^* \sim N(\beta_0^*, V_0^*) \quad \text{and} \quad \sigma^2 \sim IG(\nu_0/2, \nu_0 s_0^2/2)$$

Full conditionals

$$\beta^*|\sigma^2 \sim N(\beta_1, V_1)$$

where

$$\begin{aligned} V_1^{-1} &= V_0^{*-1} + \sigma^{-2} X^{*\prime} X^* \\ V_1^{-1}\beta_1 &= V_0^{*-1}\beta_0^* + \sigma^{-2} X^{*\prime} y \end{aligned}$$

$$\sigma^2|\beta^* \sim IG(\nu_1/2, \nu_1 s_1^2/2)$$

where

$$\begin{aligned} \nu_1 &= \nu_0 + TN \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (y - X^*\beta^*)'(y - X^*\beta^*) \end{aligned}$$

Individual effects model: hierarchical prior

Individual level

$$y_i \sim N(1_T \alpha_i + \tilde{X}_i \tilde{\beta}, \sigma^2 I_T)$$

Prior distributions (1st level)

$$\begin{aligned} \alpha_i &\sim N(\mu_\alpha, V_\alpha) \qquad i = 1, \ldots, N \\ \tilde{\beta} &\sim N(\beta_0, V_\beta) \\ \sigma^2 &\sim IG(\nu_0/2, \nu_0 s_0^2/2) \end{aligned}$$

Prior distributions (2nd level)

$$\begin{aligned} \mu_\alpha &\sim N(\mu_0, V_\mu) \\ V_\alpha &\sim IG(\nu_0/2, \nu_0 \sigma_0^2/2) \end{aligned}$$

Full conditionals

$$\alpha_i|\tilde{\beta}, \sigma^2 \sim N(\tilde{\mu}_{i\alpha}, \tilde{V}_\alpha) \qquad i = 1, \ldots, N$$

where

$$\begin{aligned} \tilde{V}_\alpha^{-1} &= V_\alpha^{-1} + \sigma^{-2} T \\ \tilde{V}_\alpha^{-1} \tilde{\mu}_{i\alpha} &= V_\alpha^{-1} \mu_\alpha + \sigma^{-2}(y_i - \tilde{X}_i \tilde{\beta}) \end{aligned}$$

$$\tilde{\beta}|\alpha_1, \ldots, \alpha_N, \sigma^2 \sim N(\beta_1, \tilde{V}_\beta)$$

where

$$\begin{aligned} \tilde{V}_\beta^{-1} &= V_\beta^{-1} + \sigma^{-2} \sum_{i=1}^{N} \tilde{X}_i' \tilde{X}_i \\ \tilde{V}_\beta^{-1} \beta_1 &= V_\beta^{-1} \beta_0 + \sigma^{-2} \sum_{i=1}^{N} \tilde{X}_i' (y_i - 1_T \alpha_i) \end{aligned}$$

Full conditionals

$$\sigma^2|\tilde{\beta},\alpha_1,\ldots,\alpha_N \sim IG(\nu_1/2, \nu_1 s_1^2/2)$$

where $\nu_1 = \nu_0 + TN$ and

$$\nu_1 s_1^2 = \nu_0 s_0^2 + \sum_{i=1}^{N}(y_i - 1_T\alpha_i - \tilde{X}_i\tilde{\beta})'(y_i - 1_T\alpha_i - \tilde{X}_i\tilde{\beta})$$

$$\mu_\alpha|\alpha_1,\ldots,\alpha_N,\sigma^2 \sim N(\mu_1, \tilde{V}_\mu)$$

where

$$\tilde{V}_\mu^{-1} = V_\mu^{-1} + \sigma^{-2}N$$

$$\tilde{V}_\mu^{-1}\mu_1 = V_\mu^{-1}\mu_0 + \sigma^{-2}\sum_{i=1}^{N}\alpha_i$$

$$V_\alpha|\alpha_1,\ldots,\alpha_N \sim IG(\eta_1/2, \eta_1\sigma_1^2/2)$$

where $\eta_1 = \eta_0 + N$ and

$$\eta_1\sigma_1^2 = \eta_0\sigma_0^2 + \sum_{i=1}^{N}(\alpha_i - \mu_\alpha)^2$$

Random coefficients model

Individual level

$$y_i \sim N(X_i\beta_i, \sigma^2 I_T)$$

Prior distribution (1st level)

$$\begin{aligned} \beta_i &\sim N(\beta_0, V_0) \qquad i = 1, \ldots, N \\ \sigma^2 &\sim IG(\eta_0/2, \eta_0 s_0^2/2) \end{aligned}$$

Prior distribution (2nd level)

$$\begin{aligned} \beta_0 &\sim N(\beta_{00}, V_\beta) \\ V_0^{-1} &\sim W(\nu_0, V_{00}^{-1}) \end{aligned}$$

Full conditionals

$$\beta_i|\sigma^2 \sim N(\beta_{1i}, V_{1i}) \qquad i = 1, \ldots, N$$

where

$$\begin{aligned} V_{1i}^{-1} &= V_0^{-1} + \sigma^{-2} X_i' X_i \\ V_{1i}^{-1}\beta_{1i} &= V_0 \beta_0^{-1} + \sigma^{-2} X_i' y_i \end{aligned}$$

$$\sigma^2|\beta_1, \ldots, \beta_N \sim IG(\eta_1/2, \eta_1 s_1^2/2)$$

where

$$\begin{aligned} \eta_1 &= \eta_0 + TN \\ \eta_1 s_1^2 &= \eta_0 s_0^2 + \sum_{i=1}^{N} (y_i - X_i\beta_i)'(y_i - X_i\beta_i) \end{aligned}$$

Full conditionals

$$\beta_0 | \beta_1, \ldots, \beta_N \sim N(\beta_{01}, \tilde{V}_\beta)$$

where

$$\begin{aligned} \tilde{V}_\beta^{-1} &= V_\beta^{-1} + N V_0 \\ \tilde{V}_\beta^{-1} \beta_{01} &= V_\beta^{-1} \beta_{00} + V_0^{-1} \sum_{i=1}^{N} \beta_i \end{aligned}$$

$$V_0^{-1} | \beta_0, \beta_1, \ldots, \beta_N \sim W(\nu_1, V_{11}^{-1})$$

where

$$\begin{aligned} \nu_1 &= \nu_0 + N \\ V_{11}^{-1} &= V_{00}^{-1} + \sum_{i=1}^{N} (\beta_i - \beta_0)(\beta_i - \beta_0)' \end{aligned}$$

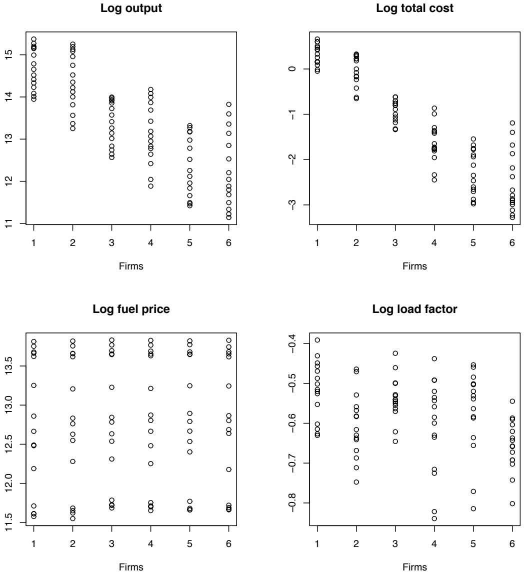
US Airline companies

Annualized data on 6 firms for 1970-1984 (15 years)

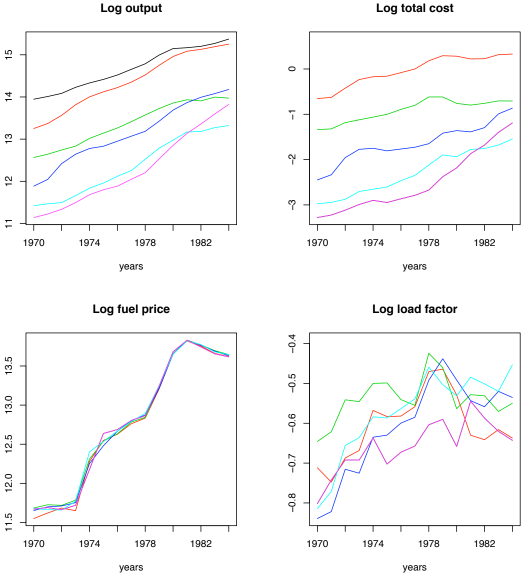
Variables:

- Airline index
- Year
- Output (revenue passenger miles)
- Total cost (thousands of dollars)
- Fuel price
- Load factor (average capacity used)

Cross-sectional data



Time-series data



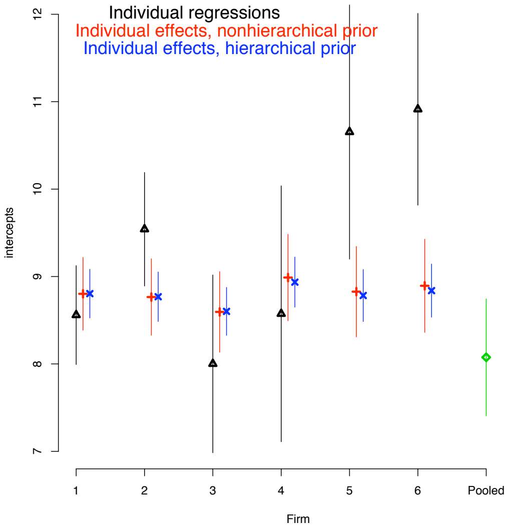
Individual and pooled regressions

Dependent variable: output
Covariates: total cost, fuel price and load factor

	Estimate				Standard Error			
Firm	b0	b1	b2	b3	b0	b1	b2	b3
1	8.56	1.17	0.39	-1.46	0.28	0.10	0.02	0.25
2	9.54	1.46	0.31	-1.52	0.32	0.08	0.03	0.14
3	8.00	0.72	0.45	-0.42	0.51	0.15	0.04	0.36
4	8.57	0.94	0.46	-0.38	0.73	0.08	0.04	0.26
5	10.65	1.06	0.30	-0.61	0.73	0.08	0.04	0.17
6	10.91	0.97	0.30	0.09	0.55	0.03	0.03	0.24
Pooled	8.08	0.88	0.45	-0.89	0.33	0.01	0.02	0.19

b0: Intercept
b1: Total cost
b2: Fuel price
b3: Load factor

Comparing the intercepts



WinBUGS

The BUGS (Bayesian inference Using Gibbs Sampling) project is concerned with flexible software for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods.

The project began in 1989 in the MRC Biostatistics Unit and led initially to the 'Classic' BUGS program, and then onto the WinBUGS software developed jointly with the Imperial College School of Medicine at St Mary's, London.

Development now also includes the OpenBUGS project in the University of Helsinki, Finland.

There are now many versions of BUGS, which can be confusing.

Simple hierarchical model

Data

School	y_i	σ_i
A	28.39	14.9
B	7.94	10.2
C	-2.75	16.3
D	6.82	11.0
E	-0.64	9.4
F	0.63	11.4
G	18.01	10.4
H	12.16	17.6

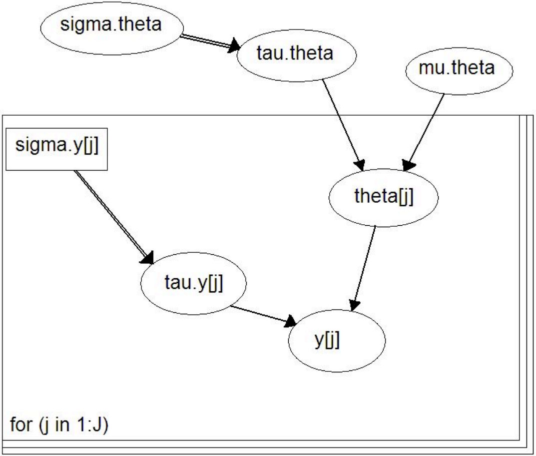
Hierarchical model: For $i = 1, \ldots, I = 8$

$$y_i \sim N(\theta_i, \sigma_i^2)$$
$$\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$$

Hyperprior

$$\mu_\theta \sim N(0, 10^6)$$
$$\sigma_\theta \sim U(0, 100)$$

Graphical model



Bugs code

```
model;
{
   for( j in 1 : J ) {
      theta[j] ~ dnorm(mu.theta,tau.theta)
   }
   for( j in 1 : J ) {
      y[j] ~ dnorm(theta[j],tau.y[j])
   }
   mu.theta ~ dnorm( 0.0,1.0E-6)
   tau.theta <- pow(sigma.theta, -2)
   for( j in 1 : J ) {
      tau.y[j] <- pow(sigma.y[j], -2)
   }
   sigma.theta ~ dunif(0,100)
}
```

Bugs from R - R2WinBUGS

```
install.packages("R2WinBUGS")
library("R2WinBUGS")
data(schools)
J       <- nrow(schools)
y       <- schools$estimate
sigma.y <- schools$sd
data    <- list("J","y","sigma.y")
inits   <- function(){
  list(theta=rnorm(J,0,100),mu.theta=rnorm(1,0,100),
       sigma.theta=runif(1,0,100))
}
schools.sim = bugs(data,inits,
  model.file="hierarchicalmodel.bug",
  parameters=c("theta","mu.theta","sigma.theta"),
  n.chains=3,n.iter=2000,n.burnin=1000,n.thin=1,
  bugs.directory="c:/Program Files/WinBUGS14/",
  codaPkg=FALSE)
print(schools.sim)
plot(schools.sim)
```

Nonlinear growth curve

Carlin and Gelfand (1991) present a nonconjugate Bayesian analysis of the following data set from Ratkowsky (1983):

Dugong (sea cows)	1	2	3	...	26	27
Age (X)	1.00	1.50	1.50	...	29.0	31.50
Length (Y)	1.80	1.85	1.87	...	2.27	2.57

Carlin and Gelfand (1991) model this data using a nonlinear growth curve with no inflection point and an asymptote as x_i tends to infinity:

$$
\begin{aligned}
y_i &\sim N(\mu_i, \tau^{-1}) \\
\mu_i &= \alpha - \beta\gamma^{x_i}
\end{aligned}
$$

for $i = 1, \ldots, 27$, $\alpha, \beta > 1$ and $0 < \gamma < 1$.
Standard noninformative priors are adopted for α, β and τ, and a uniform prior on (0,1) is assumed for γ.

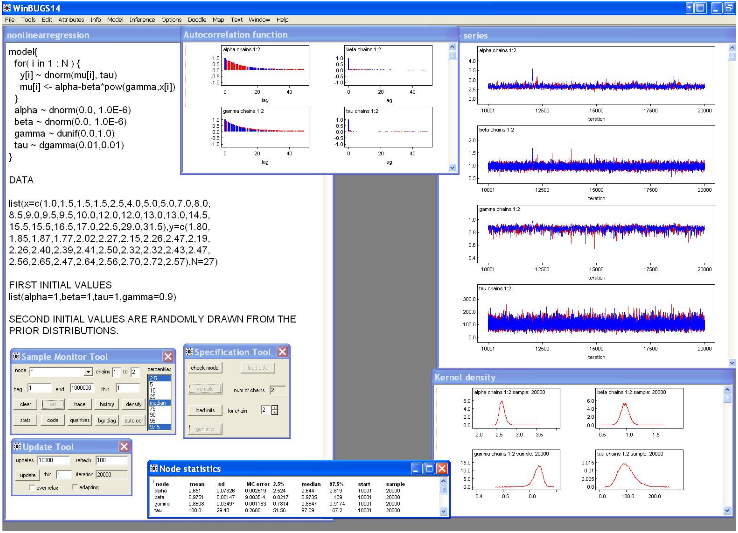
WinBugs code

```
model{
  for( i in 1 : N ) {
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- alpha - beta * pow(gamma,x[i])
  }
  alpha ~ dnorm(0.0, 1.0E-6)
  beta ~ dnorm(0.0, 1.0E-6)
  gamma ~ dunif(0.0, 1.0)
  tau ~ dgamma(0.001, 0.001)
}
```

R code

```
x = c(1.0,1.5,1.5,1.5,2.5,4.0,5.0,5.0,7.0,8.0,8.5,
      9.0,9.5,9.5,10.0,12.0,12.0,13.0,13.0,14.5,
      15.5,15.5,16.5,17.0,22.5,29.0,31.5)
y = c(1.80,1.85,1.87,1.77,2.02,2.27,2.15,2.26,2.47,
      2.19,2.26,2.40,2.39,2.41,2.50,2.32,2.32,2.43,
      2.47,2.56,2.65,2.47,2.64,2.56,2.70,2.72,2.57)
N      <- length(x)
data   <- list("x","y","N")
inits <- function(){
  list(alpha=1,beta=1,tau=1,gamma=0.9)
}
nonlinear.sim = bugs(data,inits,
     model.file="nonlinearmodel.bug",
     parameters=c("alpha","beta","tau","gamma"),
     n.chains=1,n.iter=10000,n.burnin=5000,n.thin=1,
     bugs.directory="c:/Program Files/WinBUGS14/",
     codaPkg=FALSE)
```

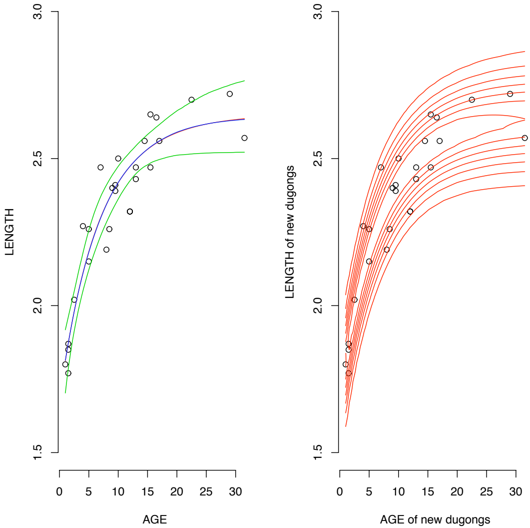
Winbugs output



Posterior summary

	α	β	τ	γ
1st Qu.	2.600	0.9244	85.06	0.8418
Median	2.643	0.9739	104.70	0.8646
Mean	2.652	0.9755	107.80	0.8607
3rd Qu.	2.694	1.0240	127.20	0.8845

$$E(y|x) = \alpha - \beta\gamma^x$$



$p(y_{new} | x_{new}, x, y)$