Moments of K

Alessandro Colombi

27/1/2023

```
suppressWarnings(suppressPackageStartupMessages(library(GDFMM)))
suppressWarnings(suppressPackageStartupMessages(library(ACutils)))
suppressWarnings(suppressPackageStartupMessages(library(tidyverse)))
suppressWarnings(suppressPackageStartupMessages(library(RColorBrewer)))
suppressWarnings(suppressPackageStartupMessages(library(salso)))
suppressWarnings(suppressPackageStartupMessages(library(wesanderson)))
suppressWarnings(suppressPackageStartupMessages(library(mcclust.ext)))
```

Moments of the Number of Clusters K a priori

One level

I first consider the case where $d = 1$ and $n = 100$

```
n_j = 100
Npoints1 = 30
Npoints2 = 10
gamma_grid  = seq(0.01,2,length.out = Npoints1)
Lambda_grid = seq(1,10,length.out = Npoints2)
Exp_val = matrix(nrow = Npoints1,ncol = Npoints2)
Var = matrix(nrow = Npoints1,ncol = Npoints2)

for(i in 1:Npoints1 ){
  for(j in 1:Npoints2){
    Moments = Expected_prior(n_j = n_j, gamma = gamma_grid[i], type = "distinct",
                             prior = "Poisson", lambda = Lambda_grid[j])
    Exp_val[i,j] = Moments$Mean
    Var[i,j] = Moments$Variance
  }
}



Lambda_names = unlist(  lapply( as.list(as.character(round(Lambda_grid, digits = 1))),
                                FUN = function(x){ paste0( "Lambda", "=",x)  }
                              )
                      )

gamma_names = unlist(  lapply( as.list(as.character(round(gamma_grid, digits = 2))),
                                FUN = function(x){ paste0( "gamma", "=",x)  }
                              )
                      )
```

```r
rownames(Exp_val) <- rownames(Var) <- gamma_names
colnames(Exp_val) <- colnames(Var) <- Lambda_names
```

Guardo la media

```r
#Exp_val
round(Exp_val, digits = 3)
```

```
##            Lambda=1 Lambda=2 Lambda=3 Lambda=4 Lambda=5 Lambda=6 Lambda=7
## gamma=0.01    1.050    1.100    1.149    1.199    1.247    1.296    1.345
## gamma=0.08    1.320    1.630    1.931    2.223    2.509    2.788    3.061
## gamma=0.15    1.500    1.979    2.441    2.887    3.320    3.740    4.149
## gamma=0.22    1.625    2.222    2.796    3.349    3.884    4.404    4.909
## gamma=0.28    1.715    2.397    3.052    3.684    4.296    4.889    5.467
## gamma=0.35    1.780    2.525    3.243    3.935    4.606    5.257    5.891
## gamma=0.42    1.828    2.622    3.387    4.127    4.844    5.542    6.222
## gamma=0.49    1.864    2.696    3.498    4.276    5.032    5.767    6.484
## gamma=0.56    1.892    2.753    3.586    4.395    5.181    5.948    6.697
## gamma=0.63    1.913    2.797    3.655    4.489    5.302    6.095    6.870
## gamma=0.7     1.930    2.833    3.711    4.566    5.401    6.216    7.013
## gamma=0.76    1.943    2.861    3.755    4.629    5.482    6.316    7.133
## gamma=0.83    1.953    2.884    3.792    4.680    5.549    6.400    7.234
## gamma=0.9     1.962    2.902    3.822    4.723    5.606    6.471    7.319
## gamma=0.97    1.968    2.917    3.847    4.759    5.654    6.531    7.392
## gamma=1.04    1.974    2.930    3.868    4.789    5.694    6.583    7.455
## gamma=1.11    1.978    2.940    3.885    4.815    5.729    6.627    7.510
## gamma=1.18    1.981    2.948    3.900    4.837    5.759    6.665    7.557
## gamma=1.25    1.984    2.955    3.913    4.856    5.784    6.699    7.598
## gamma=1.31    1.987    2.961    3.923    4.872    5.807    6.728    7.634
## gamma=1.38    1.989    2.967    3.932    4.886    5.826    6.753    7.666
## gamma=1.45    1.990    2.971    3.940    4.898    5.843    6.775    7.695
## gamma=1.52    1.992    2.974    3.947    4.908    5.858    6.795    7.720
## gamma=1.59    1.993    2.978    3.953    4.917    5.871    6.813    7.742
## gamma=1.66    1.994    2.980    3.958    4.925    5.882    6.828    7.762
## gamma=1.73    1.995    2.983    3.962    4.933    5.893    6.842    7.780
## gamma=1.79    1.995    2.985    3.966    4.939    5.902    6.854    7.796
## gamma=1.86    1.996    2.986    3.969    4.944    5.910    6.866    7.811
## gamma=1.93    1.997    2.988    3.972    4.949    5.917    6.876    7.824
## gamma=2       1.997    2.989    3.975    4.954    5.924    6.885    7.836
##            Lambda=8 Lambda=9 Lambda=10
## gamma=0.01    1.393    1.441     1.488
## gamma=0.08    3.329    3.591     3.848
## gamma=0.15    4.549    4.939     5.320
## gamma=0.22    5.401    5.882     6.351
## gamma=0.28    6.029    6.578     7.115
## gamma=0.35    6.509    7.112     7.702
## gamma=0.42    6.885    7.532     8.165
## gamma=0.49    7.184    7.868     8.538
## gamma=0.56    7.428    8.142     8.842
## gamma=0.63    7.627    8.369     9.094
## gamma=0.7     7.793    8.557     9.306
## gamma=0.76    7.933    8.716     9.485
## gamma=0.83    8.051    8.852     9.637
## gamma=0.9     8.151    8.968     9.769
```

```
## gamma=0.97    8.238    9.067     9.882
## gamma=1.04    8.312    9.154     9.981
## gamma=1.11    8.377    9.230    10.068
## gamma=1.18    8.434    9.296    10.144
## gamma=1.25    8.484    9.355    10.212
## gamma=1.31    8.527    9.406    10.272
## gamma=1.38    8.566    9.452    10.325
## gamma=1.45    8.601    9.493    10.373
## gamma=1.52    8.631    9.530    10.416
## gamma=1.59    8.659    9.563    10.455
## gamma=1.66    8.684    9.593    10.490
## gamma=1.73    8.706    9.620    10.522
## gamma=1.79    8.726    9.644    10.551
## gamma=1.86    8.744    9.667    10.577
## gamma=1.93    8.761    9.687    10.601
## gamma=2       8.776    9.705    10.623
```

- Valori piccoli di γ hanno un certo impatto, che però svanisce al crescere di γ.
- La soglia da cui in poi γ non impatta più, dipende da Λ, più crece Λ, più valori di γ hanno effetto
- Da quella soglia in poi, $E[K]$ è determinato solo da Λ

```
#Var
round(Var, digits = 3)
```

```
##            Lambda=1 Lambda=2 Lambda=3 Lambda=4 Lambda=5 Lambda=6 Lambda=7
## gamma=0.01    0.050    0.099    0.148    0.196    0.244    0.291    0.337
## gamma=0.08    0.313    0.605    0.879    1.138    1.384    1.619    1.844
## gamma=0.15    0.484    0.922    1.324    1.699    2.050    2.382    2.699
## gamma=0.22    0.601    1.138    1.628    2.080    2.503    2.901    3.279
## gamma=0.28    0.685    1.295    1.850    2.361    2.839    3.288    3.714
## gamma=0.35    0.748    1.414    2.020    2.580    3.103    3.594    4.059
## gamma=0.42    0.795    1.507    2.156    2.757    3.318    3.845    4.343
## gamma=0.49    0.832    1.581    2.267    2.903    3.497    4.056    4.583
## gamma=0.56    0.861    1.641    2.359    3.026    3.650    4.236    4.790
## gamma=0.63    0.884    1.691    2.437    3.131    3.781    4.393    4.970
## gamma=0.7     0.903    1.732    2.502    3.221    3.895    4.530    5.130
## gamma=0.76    0.918    1.767    2.558    3.299    3.996    4.651    5.271
## gamma=0.83    0.930    1.796    2.607    3.368    4.084    4.759    5.397
## gamma=0.9     0.940    1.821    2.649    3.427    4.162    4.855    5.510
## gamma=0.97    0.949    1.843    2.685    3.480    4.231    4.941    5.612
## gamma=1.04    0.956    1.861    2.717    3.527    4.293    5.018    5.705
## gamma=1.11    0.962    1.877    2.745    3.568    4.349    5.088    5.789
## gamma=1.18    0.967    1.891    2.770    3.605    4.399    5.151    5.865
## gamma=1.25    0.971    1.902    2.791    3.638    4.444    5.209    5.935
## gamma=1.31    0.975    1.913    2.811    3.668    4.484    5.261    5.999
## gamma=1.38    0.978    1.922    2.828    3.694    4.521    5.309    6.057
## gamma=1.45    0.981    1.930    2.843    3.718    4.555    5.352    6.111
## gamma=1.52    0.983    1.937    2.857    3.740    4.585    5.392    6.160
## gamma=1.59    0.985    1.943    2.869    3.759    4.613    5.428    6.206
## gamma=1.66    0.987    1.949    2.880    3.777    4.638    5.462    6.248
## gamma=1.73    0.988    1.953    2.890    3.793    4.661    5.493    6.287
## gamma=1.79    0.990    1.958    2.898    3.807    4.682    5.521    6.323
## gamma=1.86    0.991    1.962    2.906    3.821    4.702    5.547    6.357
## gamma=1.93    0.992    1.965    2.914    3.833    4.720    5.572    6.388
## gamma=2       0.993    1.968    2.920    3.844    4.736    5.594    6.417
```

```
##            Lambda=8 Lambda=9 Lambda=10
## gamma=0.01    0.384    0.429     0.474
## gamma=0.08    2.061    2.270     2.472
## gamma=0.15    3.001    3.290     3.569
## gamma=0.22    3.639    3.983     4.314
## gamma=0.28    4.118    4.504     4.873
## gamma=0.35    4.499    4.919     5.320
## gamma=0.42    4.815    5.264     5.692
## gamma=0.49    5.082    5.557     6.008
## gamma=0.56    5.314    5.811     6.284
## gamma=0.63    5.517    6.035     6.527
## gamma=0.7     5.696    6.234     6.744
## gamma=0.76    5.857    6.412     6.938
## gamma=0.83    6.000    6.572     7.114
## gamma=0.9     6.130    6.717     7.274
## gamma=0.97    6.248    6.849     7.420
## gamma=1.04    6.355    6.970     7.553
## gamma=1.11    6.452    7.081     7.676
## gamma=1.18    6.541    7.182     7.789
## gamma=1.25    6.623    7.276     7.894
## gamma=1.31    6.698    7.362     7.990
## gamma=1.38    6.768    7.442     8.080
## gamma=1.45    6.832    7.515     8.164
## gamma=1.52    6.891    7.584     8.242
## gamma=1.59    6.945    7.648     8.314
## gamma=1.66    6.996    7.707     8.382
## gamma=1.73    7.044    7.763     8.446
## gamma=1.79    7.087    7.815     8.505
## gamma=1.86    7.128    7.863     8.561
## gamma=1.93    7.167    7.909     8.613
## gamma=2       7.203    7.951     8.663
```

- Per quanto riguarda la varianza, l'impatto di γ sembra trascurabile rispetto a Λ, soprattutto per $\gamma > 1$.
- In più, la varianza sembra essere sempre simile alla media, cosa che può dato che Λ è distribuita con una Poisson. Potrebbe tuttavia essere una limitazione.

Distribution for Λ fixed

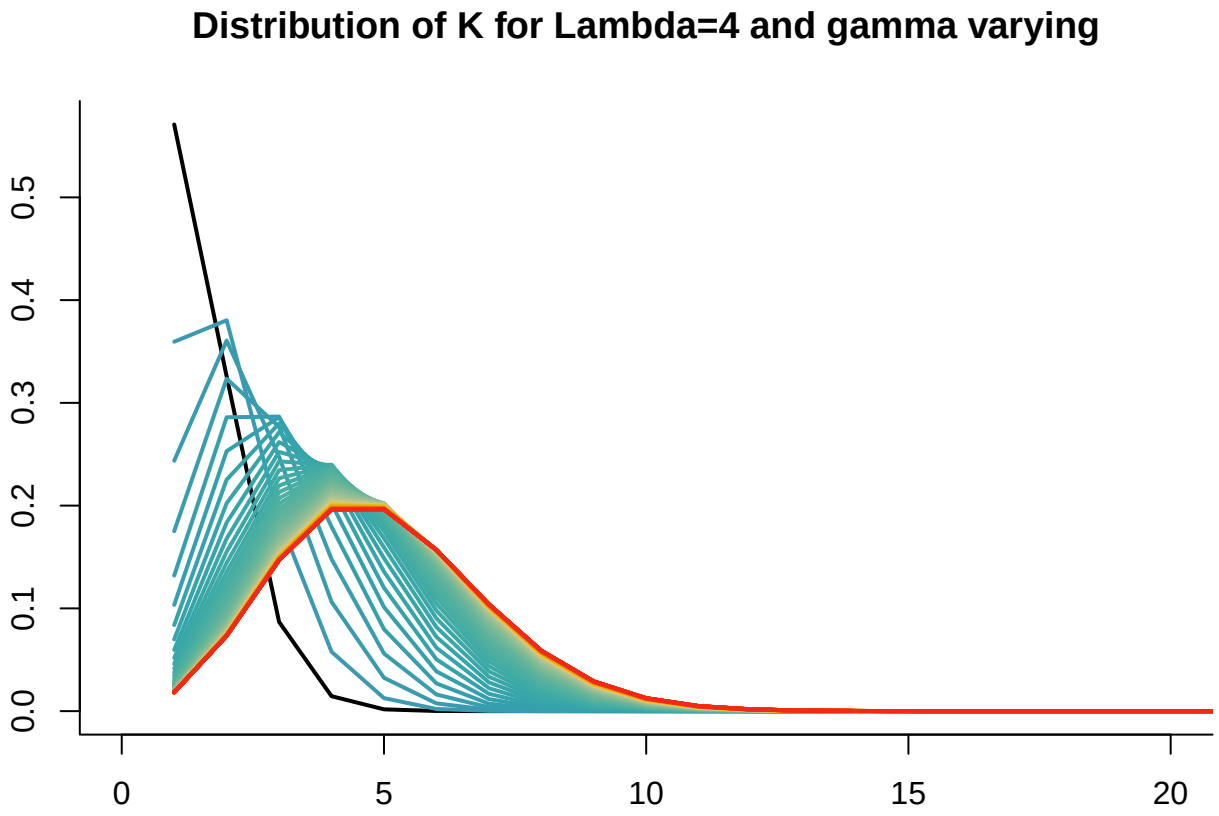
Fisso $\Lambda = 4$ e guardo la distribuzione del numero di cluster al variare di γ

```
Npoints = 100
Lambda = 4
gamma_grid  = seq(0.03,3,length.out = Npoints)
prob_k = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:n_j ){
    prob_k[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                    gamma = gamma_grid[i],
                                    prior = "Poisson",
                                    lambda = Lambda)
    #if(sum(prob_k)>=threshold)
      #break
```

```
  }
}
```

La curva nera è la prima, quella con γ minore

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k), type = "l", lwd = 2,xlim = c(0,20),lty = 1,
          col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
          main = paste0(  paste0( "Distribution of K for Lambda=4 and gamma varying") ))
```



Come mi aspettavo, fissato Λ ci sono delle differenze iniziali, che diventano poi minime quando la media di avvicina a 4.

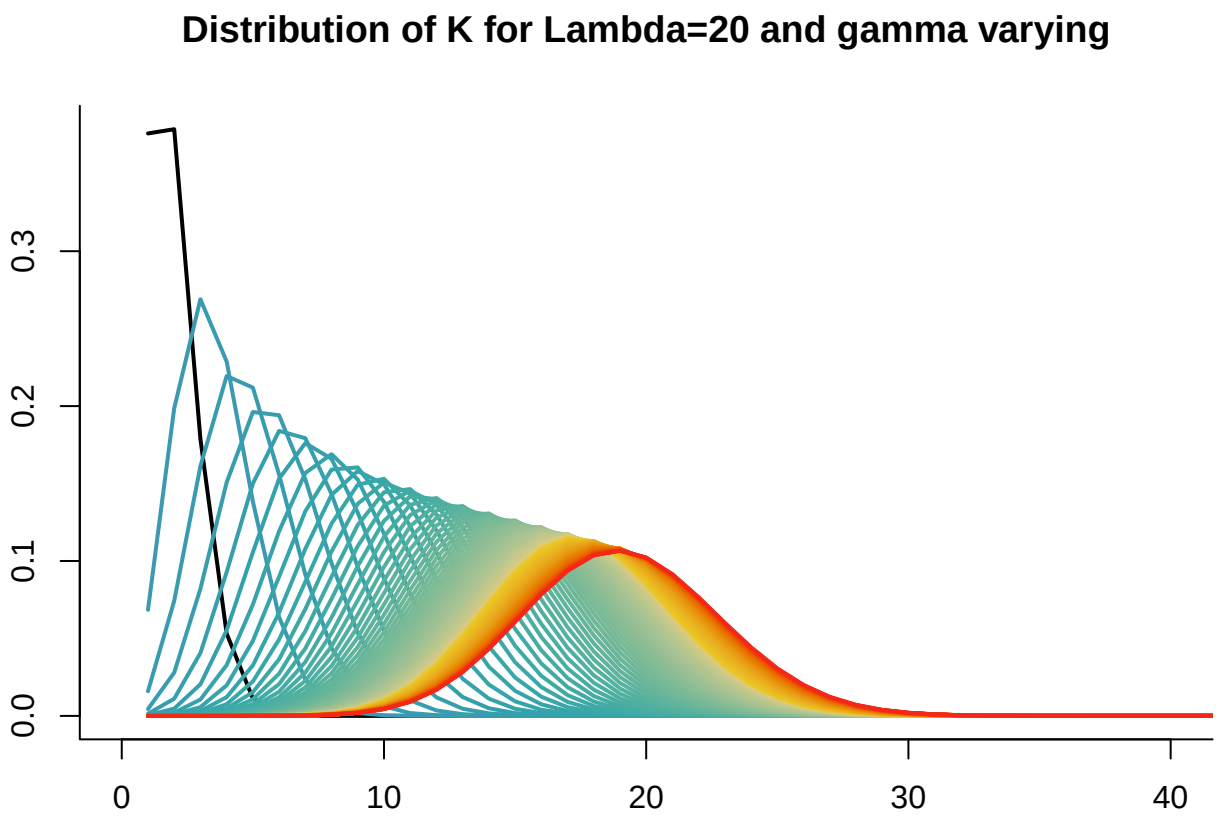
Rifaccio con $\Lambda = 20$

```
Npoints = 100
Lambda = 20
gamma_grid  = seq(0.01,2,length.out = Npoints)
prob_k2 = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:n_j ){
    prob_k2[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                     gamma = gamma_grid[i],
                                     prior = "Poisson",
                                     lambda = Lambda)
```

```
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con γ minore

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k2), type = "l", lwd = 2,xlim = c(0,40), lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for Lambda=20 and gamma varying") ))
```



La variabilità sembra essere sempre simile

Distribution for γ fixed

Ora faccio l'esperimento inverso, tengo fisso γ e faccio variare Λ.

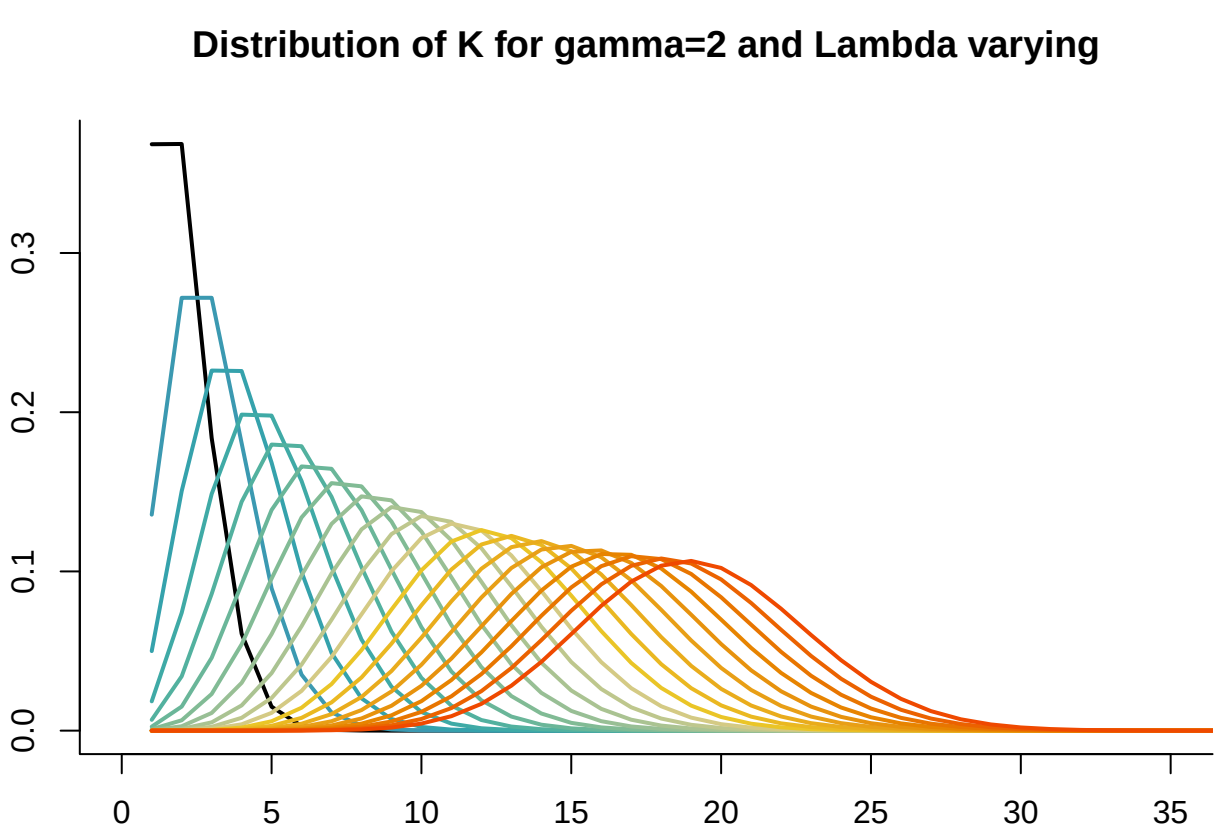
gamma = 2

```
Npoints = 20
gamma = 2
Lambda_grid  = seq(1,20,length.out = Npoints)
prob_k3 = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){
```

```
  #threshold = 0.995
  for(kk in 1:n_j ){
    prob_k3[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                     gamma = gamma,
                                     prior = "Poisson",
                                     lambda = Lambda_grid[i])
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con Λ minore. Ogni curva, corrisponde ad aumentare Λ di 1.

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k3), type = "l", lwd = 2,xlim = c(0,35),lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for gamma=2 and Lambda varying") ))
```



gamma = 0.1

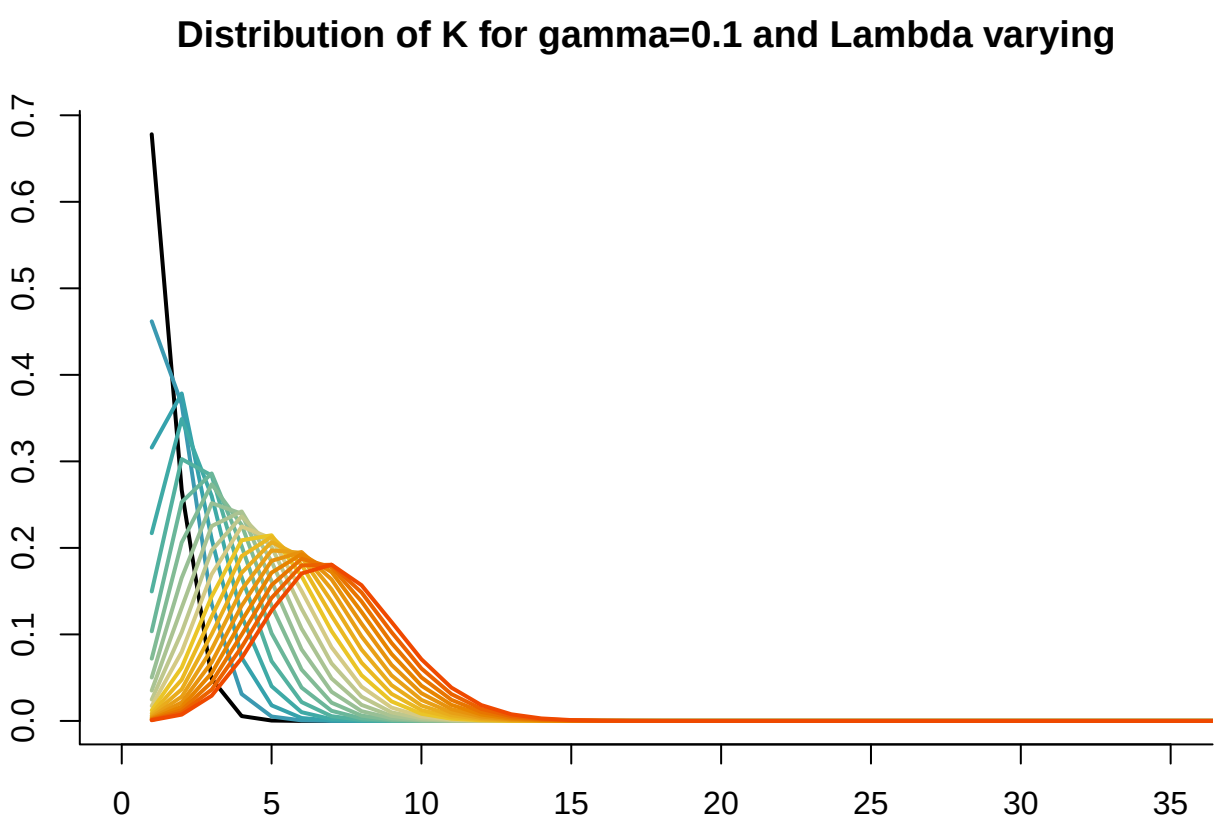
Ripeto lo stesso esperimento abbassando il valore di γ. Ogni curva, corrisponde ad aumentare Λ di 1.

```
Npoints = 20
gamma = 0.1
Lambda_grid  = seq(1,20,length.out = Npoints)
prob_k3 = matrix(nrow = Npoints, ncol = n_j)
```

```
for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:n_j ){
    prob_k3[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                     gamma = gamma,
                                     prior = "Poisson",
                                     lambda = Lambda_grid[i])
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con Λ minore

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k3), type = "l", lwd = 2,xlim = c(0,35), lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for gamma=0.1 and Lambda varying") ))
```



Al crescere di Λ le curve si spostano ma molto più lentamente.

CONCLUSIONI

- Λ è una sorta di upper bound per il numero di cluster, $E[K] < \Lambda$.
- γ aiuta a ridurre $E[K]$ al di sotto della soglia Λ. Partendo da valori di γ piccoli:
  - Se γ cresce, cresce anche $E[K]$ ma rispettando il vincolo $E[K] < \Lambda$. Quindi da un certo valore di γ in poi, anche aumentando γ non cambia niente.

- **Non possiamo controllare sia media che varianza**. Possiamo fissare la media al valore che vogliamo ma non stringere significativamente la varianza attorno a quel valore, quindi non possiamo essere troppo informativi sul numero di cluster che vogliamo.

Two levels

I repeat the experiment when $d = 2$. Now, the process parameters are $\gamma_1, \gamma_2, \Lambda$.

I consider an unbalanced number of points in the two levels, $n = (100, 50)$.

Grid for fixed Lambda

Fisso $\Lambda = 6$

```
n_j = c(100,50)
Lambda = 6

Npoints = 20
gamma1_grid  = seq(0.01,1,length.out = Npoints)
gamma2_grid  = seq(0.01,1,length.out = Npoints)

#Lambda_grid = seq(2,10,length.out = Npoints)

Exp_val = matrix(nrow = Npoints,ncol = Npoints)
Var     = matrix(nrow = Npoints,ncol = Npoints)

for(i in 1:Npoints){
  for(j in 1:Npoints){
    Moments = Expected_prior(n_j = n_j, gamma = c(gamma1_grid[i],gamma2_grid[j]),
                             type = "distinct",
                             prior = "Poisson", lambda = Lambda )
    Exp_val[i,j] = Moments$Mean
    Var[i,j]     = Moments$Variance
  }
}



gamma1_names = unlist(  lapply( as.list(as.character(round(gamma1_grid, digits = 2))),
                                FUN = function(x){ paste0( "gamma1", "=",x)  }
                              )
                      )
gamma2_names = unlist(  lapply( as.list(as.character(round(gamma2_grid, digits = 2))),
                                FUN = function(x){ paste0( "gamma2", "=",x)  }
                              )
                      )

rownames(Exp_val) <- rownames(Var) <- gamma1_names
colnames(Exp_val) <- colnames(Var) <- gamma2_names
```

Guardo la media

```
#Exp_val
round(Exp_val, digits = 3)
```

```
##             gamma2=0.01 gamma2=0.06 gamma2=0.11 gamma2=0.17 gamma2=0.22
```

```
## gamma1=0.01       2.298       3.147       3.744       4.194       4.546
## gamma1=0.06       3.284       3.955       4.426       4.781       5.060
## gamma1=0.11       3.970       4.516       4.901       5.190       5.417
## gamma1=0.17       4.478       4.933       5.253       5.493       5.682
## gamma1=0.22       4.871       5.254       5.524       5.727       5.886
## gamma1=0.27       5.181       5.509       5.738       5.912       6.047
## gamma1=0.32       5.431       5.713       5.912       6.061       6.178
## gamma1=0.37       5.636       5.881       6.053       6.183       6.285
## gamma1=0.43       5.806       6.021       6.171       6.284       6.373
## gamma1=0.48       5.948       6.137       6.270       6.369       6.447
## gamma1=0.53       6.069       6.236       6.353       6.441       6.510
## gamma1=0.58       6.171       6.320       6.424       6.502       6.564
## gamma1=0.64       6.259       6.392       6.485       6.555       6.610
## gamma1=0.69       6.335       6.454       6.538       6.600       6.650
## gamma1=0.74       6.401       6.508       6.583       6.640       6.684
## gamma1=0.79       6.458       6.555       6.623       6.674       6.714
## gamma1=0.84       6.509       6.597       6.658       6.704       6.741
## gamma1=0.9        6.553       6.633       6.689       6.731       6.764
## gamma1=0.95       6.592       6.665       6.716       6.754       6.784
## gamma1=1          6.627       6.694       6.740       6.775       6.803
##              gamma2=0.27 gamma2=0.32 gamma2=0.37 gamma2=0.43 gamma2=0.48
## gamma1=0.01       4.830       5.064       5.259       5.424       5.566
## gamma1=0.06       5.284       5.469       5.623       5.754       5.865
## gamma1=0.11       5.599       5.750       5.876       5.982       6.073
## gamma1=0.17       5.834       5.959       6.063       6.152       6.227
## gamma1=0.22       6.014       6.119       6.208       6.282       6.346
## gamma1=0.27       6.157       6.247       6.322       6.386       6.440
## gamma1=0.32       6.272       6.350       6.414       6.469       6.516
## gamma1=0.37       6.366       6.434       6.490       6.538       6.579
## gamma1=0.43       6.445       6.504       6.553       6.595       6.631
## gamma1=0.48       6.510       6.562       6.606       6.642       6.674
## gamma1=0.53       6.566       6.612       6.650       6.683       6.711
## gamma1=0.58       6.613       6.654       6.688       6.717       6.742
## gamma1=0.64       6.654       6.691       6.721       6.747       6.769
## gamma1=0.69       6.689       6.722       6.749       6.772       6.792
## gamma1=0.74       6.720       6.749       6.774       6.795       6.813
## gamma1=0.79       6.746       6.773       6.795       6.814       6.830
## gamma1=0.84       6.770       6.794       6.814       6.831       6.846
## gamma1=0.9        6.790       6.812       6.831       6.846       6.859
## gamma1=0.95       6.809       6.829       6.845       6.859       6.872
## gamma1=1          6.825       6.843       6.858       6.871       6.882
##              gamma2=0.53 gamma2=0.58 gamma2=0.64 gamma2=0.69 gamma2=0.74
## gamma1=0.01       5.688       5.794       5.888       5.970       6.043
## gamma1=0.06       5.962       6.046       6.120       6.185       6.242
## gamma1=0.11       6.152       6.220       6.281       6.334       6.381
## gamma1=0.17       6.293       6.350       6.400       6.444       6.483
## gamma1=0.22       6.402       6.450       6.492       6.529       6.562
## gamma1=0.27       6.488       6.529       6.565       6.597       6.625
## gamma1=0.32       6.557       6.593       6.624       6.651       6.675
## gamma1=0.37       6.614       6.645       6.672       6.696       6.717
## gamma1=0.43       6.661       6.688       6.712       6.733       6.752
## gamma1=0.48       6.701       6.725       6.746       6.764       6.780
## gamma1=0.53       6.735       6.756       6.774       6.791       6.805
## gamma1=0.58       6.764       6.782       6.799       6.813       6.826
```

```
## gamma1=0.64       6.788       6.805       6.820       6.833       6.844
## gamma1=0.69       6.810       6.824       6.838       6.849       6.860
## gamma1=0.74       6.828       6.842       6.853       6.864       6.873
## gamma1=0.79       6.844       6.856       6.867       6.877       6.885
## gamma1=0.84       6.858       6.870       6.879       6.888       6.895
## gamma1=0.9        6.871       6.881       6.890       6.898       6.905
## gamma1=0.95       6.882       6.891       6.899       6.906       6.913
## gamma1=1          6.892       6.900       6.908       6.914       6.920
##             gamma2=0.79 gamma2=0.84 gamma2=0.9 gamma2=0.95 gamma2=1
## gamma1=0.01       6.108       6.166      6.218       6.266    6.309
## gamma1=0.06       6.294       6.340      6.381       6.418    6.452
## gamma1=0.11       6.422       6.460      6.494       6.524    6.552
## gamma1=0.17       6.518       6.549      6.577       6.603    6.626
## gamma1=0.22       6.592       6.618      6.642       6.663    6.683
## gamma1=0.27       6.650       6.673      6.693       6.711    6.728
## gamma1=0.32       6.697       6.717      6.734       6.750    6.764
## gamma1=0.37       6.736       6.753      6.768       6.782    6.794
## gamma1=0.43       6.768       6.783      6.796       6.808    6.819
## gamma1=0.48       6.795       6.808      6.820       6.831    6.840
## gamma1=0.53       6.818       6.830      6.840       6.849    6.858
## gamma1=0.58       6.837       6.848      6.857       6.865    6.873
## gamma1=0.64       6.854       6.864      6.872       6.879    6.886
## gamma1=0.69       6.869       6.877      6.885       6.891    6.897
## gamma1=0.74       6.881       6.889      6.896       6.902    6.907
## gamma1=0.79       6.893       6.899      6.905       6.911    6.916
## gamma1=0.84       6.902       6.908      6.914       6.919    6.923
## gamma1=0.9        6.911       6.916      6.921       6.926    6.930
## gamma1=0.95       6.918       6.924      6.928       6.932    6.936
## gamma1=1          6.925       6.930      6.934       6.938    6.941
```

- L'impatto di γ_1 e γ_2 è un'estensione di quello che abbiamo visto nel caso $d = 1$.
- $E[K]$ cresce più velocemente con γ_1 che con γ_2, ovvero con il livello più numeroso.
- Un'importante differenza è che ora l'upper bound per $E[K]$ sembra essere $\Lambda + 1$ e non solo Λ

Guardo anche la varianza

```
#Var
round(Var, digits = 3)
```

```
##             gamma2=0.01 gamma2=0.06 gamma2=0.11 gamma2=0.17 gamma2=0.22
## gamma1=0.01       0.604       1.405       1.930       2.314       2.614
## gamma1=0.06       1.552       2.213       2.645       2.962       3.211
## gamma1=0.11       2.184       2.752       3.123       3.397       3.612
## gamma1=0.17       2.651       3.151       3.478       3.721       3.912
## gamma1=0.22       3.018       3.464       3.759       3.977       4.149
## gamma1=0.27       3.316       3.721       3.988       4.186       4.342
## gamma1=0.32       3.567       3.935       4.179       4.361       4.505
## gamma1=0.37       3.780       4.119       4.343       4.511       4.643
## gamma1=0.43       3.965       4.277       4.485       4.640       4.762
## gamma1=0.48       4.127       4.416       4.609       4.753       4.866
## gamma1=0.53       4.270       4.539       4.718       4.852       4.958
## gamma1=0.58       4.398       4.648       4.815       4.940       5.039
## gamma1=0.64       4.512       4.746       4.902       5.019       5.111
## gamma1=0.69       4.615       4.833       4.980       5.089       5.175
## gamma1=0.74       4.708       4.913       5.050       5.152       5.233
## gamma1=0.79       4.792       4.984       5.113       5.210       5.286
```

```
## gamma1=0.84       4.869       5.050       5.171       5.262       5.333
## gamma1=0.9        4.939       5.109       5.223       5.309       5.376
## gamma1=0.95       5.003       5.164       5.271       5.352       5.415
## gamma1=1          5.062       5.213       5.315       5.391       5.451
##             gamma2=0.27 gamma2=0.32 gamma2=0.37 gamma2=0.43 gamma2=0.48
## gamma1=0.01       2.859       3.065       3.241       3.395       3.532
## gamma1=0.06       3.414       3.585       3.732       3.861       3.974
## gamma1=0.11       3.788       3.937       4.065       4.176       4.275
## gamma1=0.17       4.068       4.201       4.314       4.413       4.501
## gamma1=0.22       4.290       4.409       4.511       4.601       4.680
## gamma1=0.27       4.471       4.579       4.672       4.754       4.825
## gamma1=0.32       4.622       4.722       4.807       4.881       4.946
## gamma1=0.37       4.751       4.843       4.921       4.989       5.049
## gamma1=0.43       4.862       4.947       5.019       5.082       5.137
## gamma1=0.48       4.959       5.037       5.104       5.162       5.213
## gamma1=0.53       5.044       5.117       5.179       5.233       5.280
## gamma1=0.58       5.119       5.187       5.245       5.295       5.339
## gamma1=0.64       5.186       5.249       5.303       5.350       5.391
## gamma1=0.69       5.246       5.305       5.355       5.399       5.437
## gamma1=0.74       5.299       5.355       5.402       5.443       5.478
## gamma1=0.79       5.348       5.400       5.444       5.482       5.516
## gamma1=0.84       5.391       5.440       5.482       5.518       5.549
## gamma1=0.9        5.431       5.477       5.516       5.550       5.579
## gamma1=0.95       5.467       5.510       5.547       5.579       5.607
## gamma1=1          5.500       5.541       5.576       5.606       5.632
##             gamma2=0.53 gamma2=0.58 gamma2=0.64 gamma2=0.69 gamma2=0.74
## gamma1=0.01       3.653       3.763       3.863       3.954       4.037
## gamma1=0.06       4.076       4.168       4.251       4.326       4.396
## gamma1=0.11       4.364       4.443       4.515       4.581       4.641
## gamma1=0.17       4.579       4.650       4.713       4.771       4.824
## gamma1=0.22       4.750       4.813       4.870       4.922       4.969
## gamma1=0.27       4.889       4.946       4.997       5.044       5.087
## gamma1=0.32       5.004       5.056       5.103       5.146       5.184
## gamma1=0.37       5.102       5.150       5.193       5.231       5.267
## gamma1=0.43       5.186       5.230       5.269       5.305       5.337
## gamma1=0.48       5.259       5.299       5.335       5.368       5.398
## gamma1=0.53       5.322       5.359       5.393       5.423       5.451
## gamma1=0.58       5.378       5.412       5.443       5.471       5.497
## gamma1=0.64       5.427       5.459       5.488       5.514       5.538
## gamma1=0.69       5.471       5.501       5.528       5.552       5.574
## gamma1=0.74       5.510       5.538       5.563       5.585       5.606
## gamma1=0.79       5.545       5.571       5.595       5.616       5.635
## gamma1=0.84       5.577       5.601       5.623       5.643       5.661
## gamma1=0.9        5.605       5.628       5.649       5.667       5.684
## gamma1=0.95       5.631       5.653       5.672       5.689       5.705
## gamma1=1          5.655       5.675       5.693       5.710       5.724
##             gamma2=0.79 gamma2=0.84 gamma2=0.9 gamma2=0.95 gamma2=1
## gamma1=0.01       4.114       4.185      4.252       4.313    4.370
## gamma1=0.06       4.460       4.519      4.574       4.625    4.673
## gamma1=0.11       4.696       4.747      4.794       4.838    4.879
## gamma1=0.17       4.873       4.918      4.960       4.998    5.034
## gamma1=0.22       5.013       5.053      5.090       5.124    5.156
## gamma1=0.27       5.126       5.162      5.195       5.226    5.254
## gamma1=0.32       5.220       5.253      5.283       5.310    5.336
```

```
## gamma1=0.37       5.299       5.329      5.356       5.381    5.405
## gamma1=0.43       5.367       5.394      5.419       5.442    5.463
## gamma1=0.48       5.425       5.450      5.473       5.494    5.513
## gamma1=0.53       5.476       5.499      5.520       5.539    5.557
## gamma1=0.58       5.520       5.541      5.561       5.578    5.595
## gamma1=0.64       5.559       5.579      5.597       5.613    5.629
## gamma1=0.69       5.594       5.612      5.629       5.644    5.658
## gamma1=0.74       5.624       5.641      5.657       5.671    5.684
## gamma1=0.79       5.652       5.668      5.682       5.696    5.708
## gamma1=0.84       5.677       5.692      5.705       5.718    5.729
## gamma1=0.9        5.699       5.713      5.726       5.737    5.748
## gamma1=0.95       5.719       5.732      5.744       5.755    5.765
## gamma1=1          5.738       5.750      5.761       5.771    5.781
```

Grid for fixed $\gamma_1 = \gamma_2$

```
n_j = c(100,50)

Npoints1 = 30
Npoints2 = 10
Lambda_grid  = seq(1,10,length.out = Npoints2)
gamma_grid  = seq(0.01,1,length.out = Npoints1)

#Lambda_grid = seq(2,10,length.out = Npoints)

Exp_val = matrix(nrow = Npoints1,ncol = Npoints2)
Var     = matrix(nrow = Npoints1,ncol = Npoints2)

for(i in 1:Npoints1){
  for(j in 1:Npoints2){
    Moments = Expected_prior(n_j = n_j, gamma = c(gamma_grid[i],gamma_grid[i]),
                             type = "distinct",
                             prior = "Poisson", lambda = Lambda_grid[j] )
    Exp_val[i,j] = Moments$Mean
    Var[i,j]     = Moments$Variance
  }
}



Lambda_names = unlist(  lapply( as.list(as.character(round(Lambda_grid, digits = 1))),
                                FUN = function(x){ paste0( "Lambda", "=",x)  }
                              )
                      )
gamma2_names = unlist(  lapply( as.list(as.character(round(gamma_grid, digits = 2))),
                                FUN = function(x){ paste0( "gamma1=gamma2", "=",x)  }
                              )
                      )

rownames(Exp_val) <- rownames(Var) <- gamma2_names
colnames(Exp_val) <- colnames(Var) <- Lambda_names
```

Guardo la media

```r
#Exp_val
round(Exp_val, digits = 3)
```

```
##                     Lambda=1 Lambda=2 Lambda=3 Lambda=4 Lambda=5 Lambda=6
## gamma1=gamma2=0.01    1.426    1.698    1.893    2.048    2.180    2.298
## gamma1=gamma2=0.04    1.581    2.045    2.448    2.816    3.162    3.494
## gamma1=gamma2=0.08    1.689    2.286    2.830    3.341    3.829    4.301
## gamma1=gamma2=0.11    1.767    2.457    3.101    3.714    4.304    4.875
## gamma1=gamma2=0.15    1.823    2.581    3.299    3.987    4.652    5.298
## gamma1=gamma2=0.18    1.864    2.674    3.446    4.191    4.914    5.617
## gamma1=gamma2=0.21    1.895    2.743    3.558    4.347    5.115    5.863
## gamma1=gamma2=0.25    1.918    2.796    3.644    4.468    5.271    6.056
## gamma1=gamma2=0.28    1.935    2.836    3.711    4.562    5.395    6.209
## gamma1=gamma2=0.32    1.949    2.868    3.763    4.637    5.493    6.332
## gamma1=gamma2=0.35    1.959    2.893    3.804    4.697    5.573    6.432
## gamma1=gamma2=0.39    1.967    2.912    3.838    4.746    5.638    6.514
## gamma1=gamma2=0.42    1.974    2.928    3.864    4.785    5.691    6.582
## gamma1=gamma2=0.45    1.979    2.940    3.886    4.817    5.735    6.638
## gamma1=gamma2=0.49    1.983    2.950    3.904    4.844    5.771    6.686
## gamma1=gamma2=0.52    1.986    2.958    3.919    4.866    5.802    6.725
## gamma1=gamma2=0.56    1.988    2.965    3.931    4.885    5.828    6.759
## gamma1=gamma2=0.59    1.990    2.971    3.941    4.900    5.849    6.788
## gamma1=gamma2=0.62    1.992    2.975    3.949    4.913    5.868    6.813
## gamma1=gamma2=0.66    1.993    2.979    3.956    4.924    5.884    6.834
## gamma1=gamma2=0.69    1.994    2.982    3.962    4.934    5.897    6.852
## gamma1=gamma2=0.73    1.995    2.985    3.967    4.942    5.909    6.868
## gamma1=gamma2=0.76    1.996    2.987    3.971    4.949    5.919    6.882
## gamma1=gamma2=0.8     1.997    2.989    3.975    4.955    5.928    6.894
## gamma1=gamma2=0.83    1.997    2.990    3.978    4.960    5.936    6.904
## gamma1=gamma2=0.86    1.998    2.991    3.981    4.964    5.942    6.914
## gamma1=gamma2=0.9     1.998    2.993    3.983    4.968    5.948    6.922
## gamma1=gamma2=0.93    1.998    2.994    3.985    4.972    5.953    6.929
## gamma1=gamma2=0.97    1.998    2.994    3.987    4.975    5.958    6.936
## gamma1=gamma2=1       1.999    2.995    3.988    4.977    5.962    6.941
##                     Lambda=7 Lambda=8 Lambda=9 Lambda=10
## gamma1=gamma2=0.01    2.408    2.512    2.612     2.709
## gamma1=gamma2=0.04    3.814    4.126    4.431     4.730
## gamma1=gamma2=0.08    4.759    5.206    5.642     6.070
## gamma1=gamma2=0.11    5.431    5.974    6.505     7.025
## gamma1=gamma2=0.15    5.927    6.542    7.144     7.734
## gamma1=gamma2=0.18    6.304    6.975    7.632     8.277
## gamma1=gamma2=0.21    6.595    7.311    8.014     8.703
## gamma1=gamma2=0.25    6.824    7.577    8.316     9.042
## gamma1=gamma2=0.28    7.008    7.791    8.561     9.317
## gamma1=gamma2=0.32    7.156    7.965    8.760     9.542
## gamma1=gamma2=0.35    7.277    8.107    8.924     9.728
## gamma1=gamma2=0.39    7.377    8.225    9.061     9.884
## gamma1=gamma2=0.42    7.460    8.324    9.176    10.015
## gamma1=gamma2=0.45    7.529    8.407    9.273    10.127
## gamma1=gamma2=0.49    7.588    8.477    9.356    10.222
## gamma1=gamma2=0.52    7.637    8.537    9.426    10.304
## gamma1=gamma2=0.56    7.680    8.589    9.487    10.375
## gamma1=gamma2=0.59    7.716    8.633    9.540    10.436
## gamma1=gamma2=0.62    7.747    8.672    9.586    10.490
```

```
## gamma1=gamma2=0.66    7.774    8.705    9.626    10.537
## gamma1=gamma2=0.69    7.798    8.734    9.661    10.579
## gamma1=gamma2=0.73    7.818    8.760    9.692    10.616
## gamma1=gamma2=0.76    7.836    8.782    9.719    10.648
## gamma1=gamma2=0.8     7.852    8.802    9.744    10.677
## gamma1=gamma2=0.83    7.866    8.820    9.765    10.703
## gamma1=gamma2=0.86    7.878    8.835    9.785    10.726
## gamma1=gamma2=0.9     7.889    8.849    9.802    10.747
## gamma1=gamma2=0.93    7.899    8.861    9.817    10.766
## gamma1=gamma2=0.97    7.907    8.873    9.831    10.783
## gamma1=gamma2=1       7.915    8.883    9.844    10.798
```

Il comportamento è simile a prima ma lo spazio di manovra sembra molto più limitato, i valori convergono molto più velocemente all'upper bound dato da $E[K]$.

Guardo la varianza

```
#Var
round(Var, digits = 3)
```

```
##                    Lambda=1 Lambda=2 Lambda=3 Lambda=4 Lambda=5 Lambda=6
## gamma1=gamma2=0.01    0.296    0.386    0.437    0.487    0.543    0.604
## gamma1=gamma2=0.04    0.472    0.769    1.026    1.276    1.524    1.772
## gamma1=gamma2=0.08    0.597    1.041    1.438    1.817    2.186    2.544
## gamma1=gamma2=0.11    0.688    1.240    1.739    2.210    2.663    3.098
## gamma1=gamma2=0.15    0.756    1.389    1.965    2.508    3.024    3.519
## gamma1=gamma2=0.18    0.807    1.503    2.140    2.740    3.309    3.852
## gamma1=gamma2=0.21    0.846    1.591    2.279    2.925    3.539    4.123
## gamma1=gamma2=0.25    0.876    1.661    2.390    3.076    3.727    4.348
## gamma1=gamma2=0.28    0.899    1.718    2.481    3.201    3.885    4.537
## gamma1=gamma2=0.32    0.918    1.763    2.555    3.305    4.018    4.698
## gamma1=gamma2=0.35    0.932    1.800    2.617    3.393    4.132    4.837
## gamma1=gamma2=0.39    0.944    1.830    2.669    3.467    4.229    4.957
## gamma1=gamma2=0.42    0.954    1.855    2.712    3.531    4.313    5.062
## gamma1=gamma2=0.45    0.961    1.876    2.749    3.585    4.386    5.153
## gamma1=gamma2=0.49    0.968    1.893    2.781    3.632    4.449    5.234
## gamma1=gamma2=0.52    0.973    1.908    2.807    3.672    4.504    5.305
## gamma1=gamma2=0.56    0.977    1.920    2.830    3.708    4.553    5.368
## gamma1=gamma2=0.59    0.980    1.931    2.850    3.739    4.596    5.424
## gamma1=gamma2=0.62    0.983    1.940    2.867    3.766    4.634    5.473
## gamma1=gamma2=0.66    0.986    1.947    2.882    3.789    4.668    5.518
## gamma1=gamma2=0.69    0.988    1.954    2.895    3.810    4.698    5.558
## gamma1=gamma2=0.73    0.990    1.959    2.906    3.829    4.725    5.594
## gamma1=gamma2=0.76    0.991    1.964    2.916    3.845    4.748    5.626
## gamma1=gamma2=0.8     0.992    1.968    2.925    3.859    4.770    5.655
## gamma1=gamma2=0.83    0.993    1.972    2.933    3.872    4.789    5.681
## gamma1=gamma2=0.86    0.994    1.975    2.939    3.884    4.806    5.705
## gamma1=gamma2=0.9     0.995    1.978    2.945    3.894    4.822    5.727
## gamma1=gamma2=0.93    0.996    1.980    2.951    3.903    4.836    5.746
## gamma1=gamma2=0.97    0.996    1.983    2.955    3.911    4.848    5.764
## gamma1=gamma2=1       0.997    1.984    2.959    3.919    4.860    5.781
##                    Lambda=7 Lambda=8 Lambda=9 Lambda=10
## gamma1=gamma2=0.01    0.671    0.740    0.811     0.883
## gamma1=gamma2=0.04    2.017    2.258    2.495     2.727
## gamma1=gamma2=0.08    2.892    3.230    3.559     3.878
## gamma1=gamma2=0.11    3.518    3.923    4.314     4.692
```

```
## gamma1=gamma2=0.15    3.994    4.450    4.890     5.315
## gamma1=gamma2=0.18    4.372    4.871    5.351     5.814
## gamma1=gamma2=0.21    4.682    5.218    5.732     6.228
## gamma1=gamma2=0.25    4.941    5.509    6.054     6.579
## gamma1=gamma2=0.28    5.160    5.757    6.330     6.881
## gamma1=gamma2=0.32    5.349    5.972    6.570     7.145
## gamma1=gamma2=0.35    5.512    6.159    6.780     7.377
## gamma1=gamma2=0.39    5.655    6.324    6.966     7.583
## gamma1=gamma2=0.42    5.780    6.469    7.130     7.766
## gamma1=gamma2=0.45    5.890    6.597    7.277     7.931
## gamma1=gamma2=0.49    5.988    6.712    7.409     8.079
## gamma1=gamma2=0.52    6.074    6.815    7.527     8.213
## gamma1=gamma2=0.56    6.152    6.907    7.634     8.334
## gamma1=gamma2=0.59    6.221    6.990    7.731     8.444
## gamma1=gamma2=0.62    6.283    7.065    7.818     8.544
## gamma1=gamma2=0.66    6.339    7.132    7.898     8.636
## gamma1=gamma2=0.69    6.390    7.194    7.970     8.720
## gamma1=gamma2=0.73    6.435    7.250    8.037     8.797
## gamma1=gamma2=0.76    6.477    7.300    8.097     8.867
## gamma1=gamma2=0.8     6.514    7.347    8.153     8.932
## gamma1=gamma2=0.83    6.548    7.389    8.204     8.992
## gamma1=gamma2=0.86    6.579    7.428    8.251     9.048
## gamma1=gamma2=0.9     6.608    7.464    8.294     9.099
## gamma1=gamma2=0.93    6.634    7.497    8.334     9.146
## gamma1=gamma2=0.97    6.657    7.527    8.371     9.191
## gamma1=gamma2=1       6.679    7.555    8.405     9.231
```

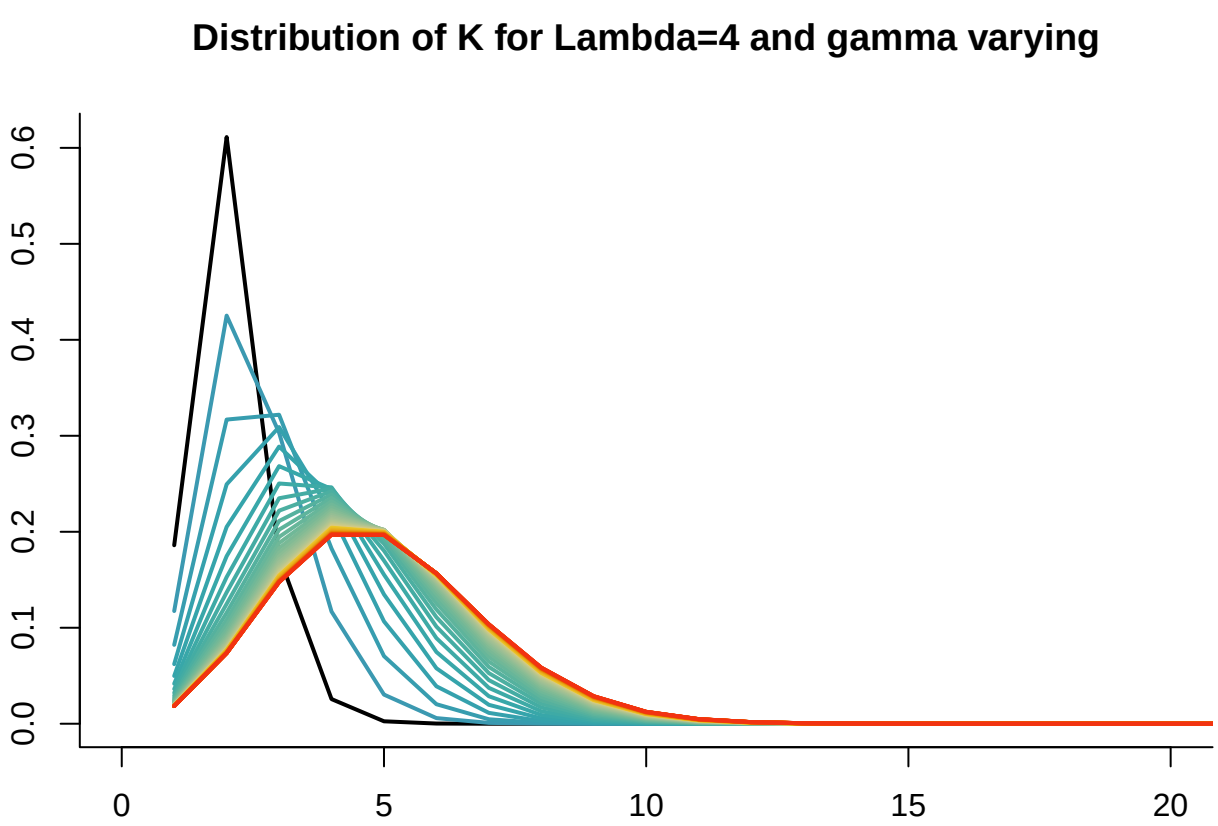
Distribution for Λ fixed and equal γs

Fisso $\Lambda = 4$ e guardo la distribuzione del numero di cluster al variare di $\gamma_1 = \gamma_2$

```
Npoints = 50
Lambda = 4
gamma_grid  = seq(0.01,1,length.out = Npoints)
prob_k = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:max(n_j) ){
    prob_k[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                    gamma = c(gamma_grid[i],gamma_grid[i]),
                                    prior = "Poisson",
                                    lambda = Lambda)
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con γ minore

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k), type = "l", lwd = 2,xlim = c(0,20),lty = 1,
          col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
          main = paste0(  paste0( "Distribution of K for Lambda=4 and gamma varying") ))
```

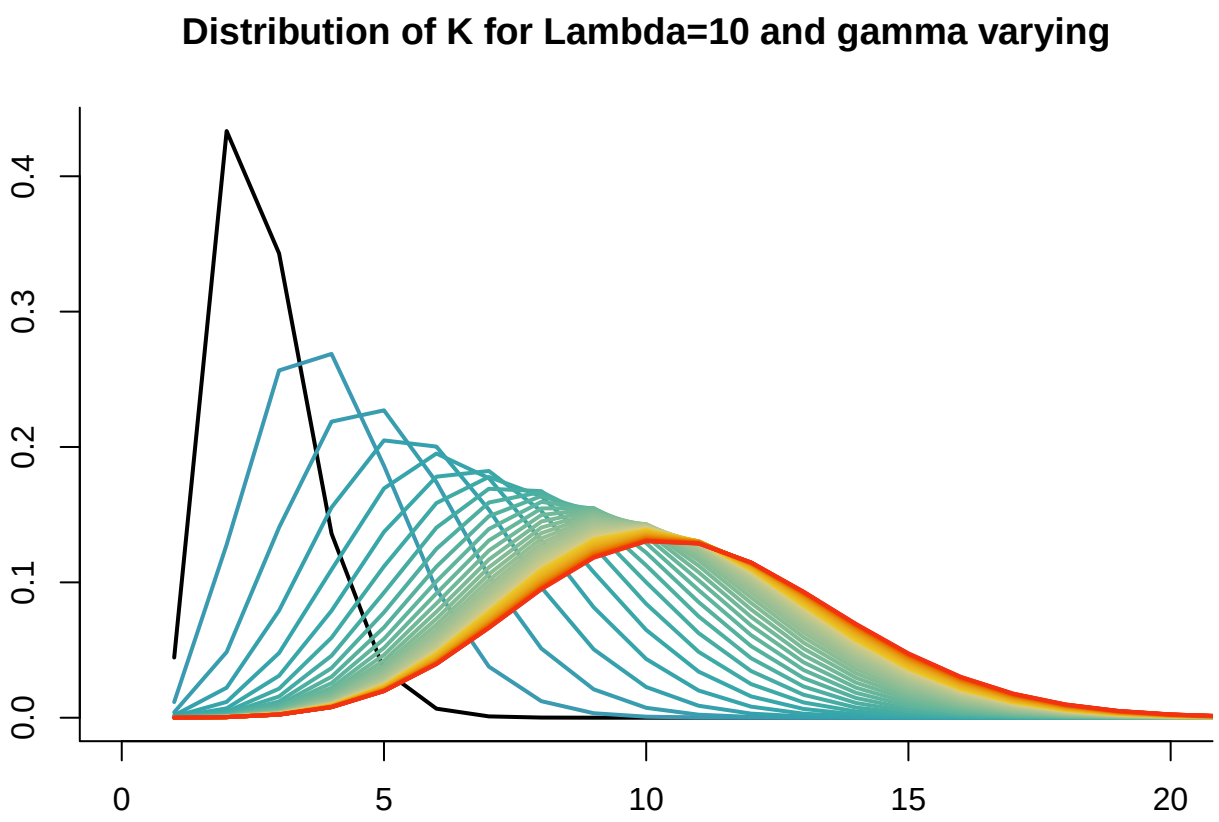


Ripeto per $\Lambda = 10$

```
Npoints = 50
Lambda = 10
gamma_grid  = seq(0.01,1,length.out = Npoints)
prob_k = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:max(n_j) ){
    prob_k[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                    gamma = c(gamma_grid[i],gamma_grid[i]),
                                    prior = "Poisson",
                                    lambda = Lambda)
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con γ minore

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k), type = "l", lwd = 2,xlim = c(0,20),lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for Lambda=10 and gamma varying") ))
```



Distribution for γ fixed

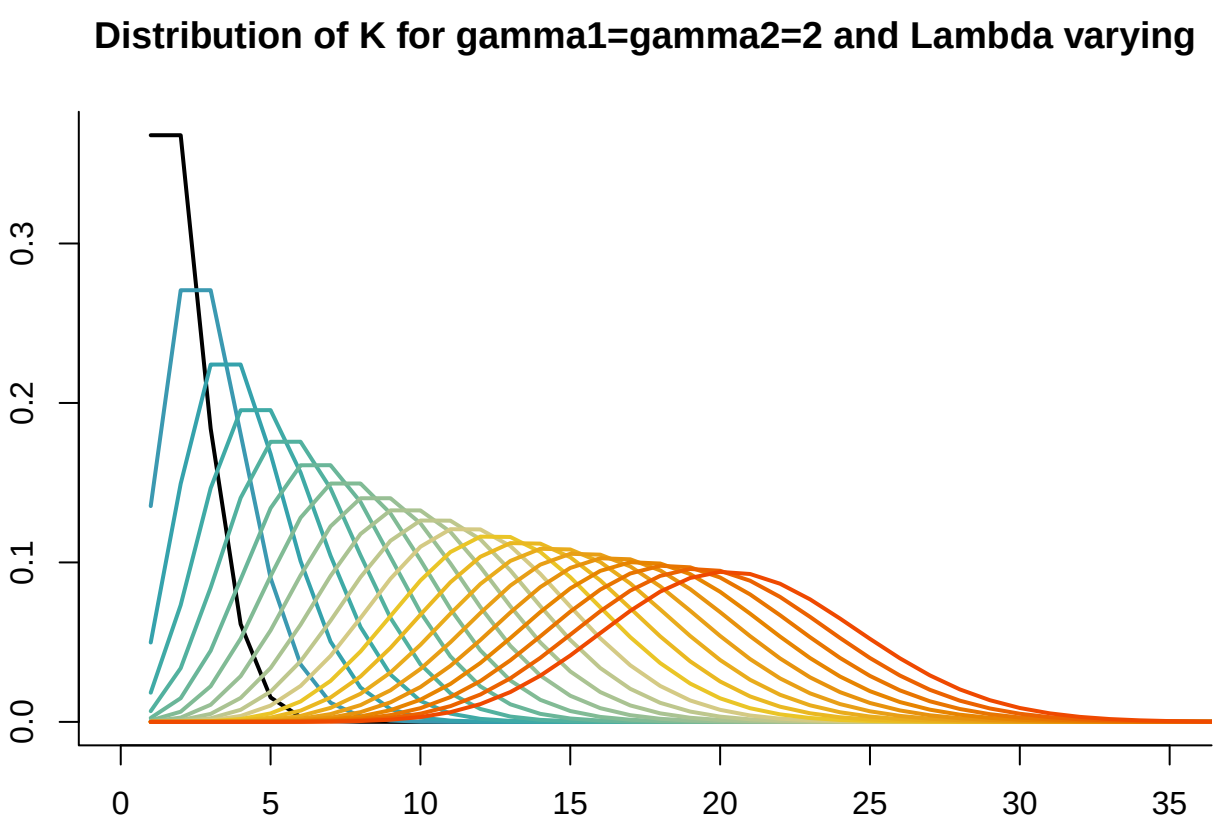
Ora faccio l'esperimento inverso, tengo fisso γ e faccio variare Λ.

gamma1 = gamma2 = 2

```
Npoints = 20
gamma1 <- gamma2 <- 2
Lambda_grid  = seq(1,20,length.out = Npoints)
prob_k3 = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:max(n_j) ){
    prob_k3[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                     gamma = c(gamma1,gamma2),
                                     prior = "Poisson",
                                     lambda = Lambda_grid[i])
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con Λ minore. Ogni curva, corrisponde ad aumentare Λ di 1.

```r
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k3), type = "l", lwd = 2,xlim = c(0,35),lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for gamma1=gamma2=2 and Lambda varying") ))
```



gamma1 = gamma2 = 0.1

Ripeto lo stesso esperimento abbassando il valore di γ_1 e γ_2. Ogni curva, corrisponde ad aumentare Λ di 1.

```r
Npoints = 20
gamma1 <- gamma2 <- 0.1
Lambda_grid  = seq(1,20,length.out = Npoints)
prob_k4 = matrix(nrow = Npoints, ncol = n_j)

for(i in 1:Npoints){

  #threshold = 0.995
  for(kk in 1:max(n_j) ){
    prob_k4[i,kk] = p_distinct_prior(k = kk, n_j = n_j,
                                     gamma = c(gamma1,gamma2),
                                     prior = "Poisson",
                                     lambda = Lambda_grid[i])
    #if(sum(prob_k)>=threshold)
      #break
  }
}
```

La curva nera è la prima, quella con Λ minore. Ogni curva, corrisponde ad aumentare Λ di 1.

```
par(mar = c(2,2,4,1), bty = "l")
matplot( t(prob_k4), type = "l", lwd = 2,xlim = c(0,35),lty = 1,
         col = c("black",hcl.colors(n=Npoints, palette = "Zissou 1")),
         main = paste0(  paste0( "Distribution of K for gamma1=gamma1=0.1 and Lambda varying") ))
```