NASUG 2001

Decomposition of normal mixture by maximum likelihood: `denormix` package

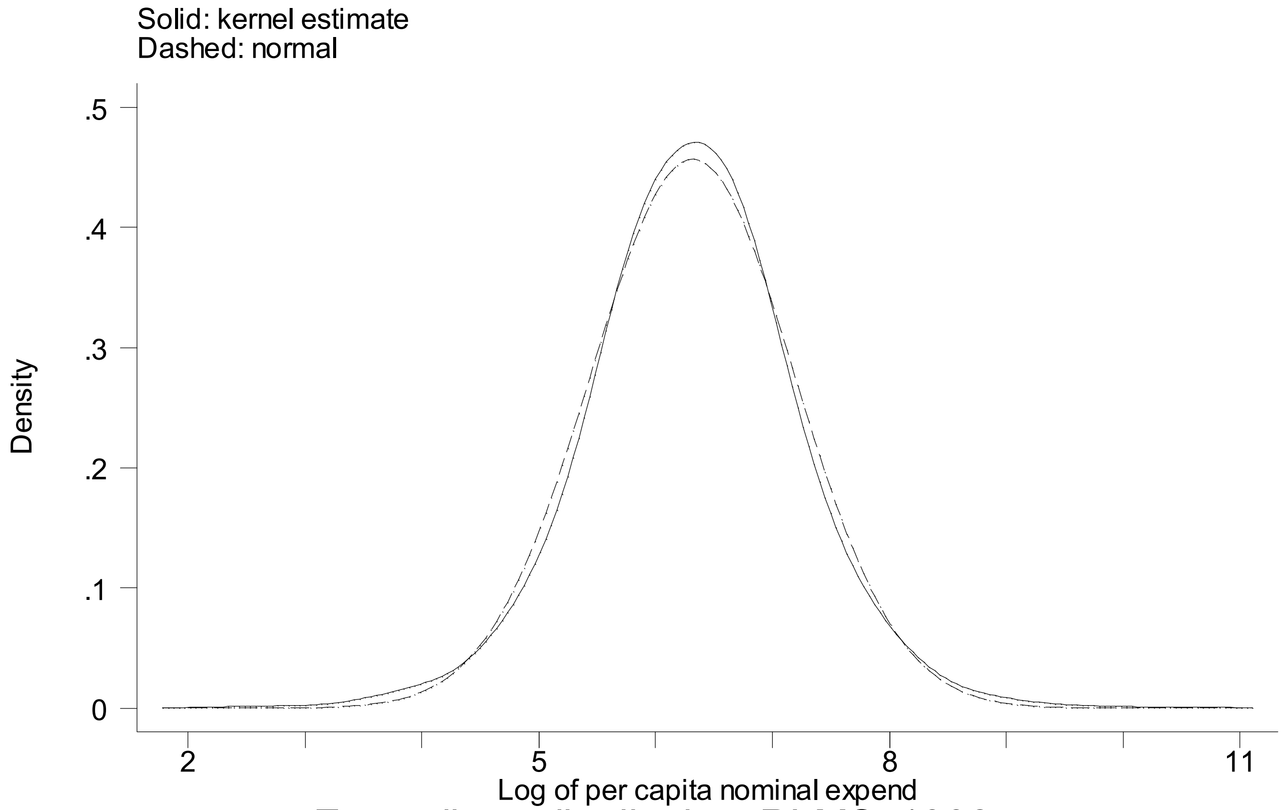
Stanislav Kolenikov

University of North Carolina at Chapel Hill

Center for Economic and Financial Research, Moscow, Russia

Financial support from EERC (Eurasia Foundation) and EPA is gratefully acknowledged.

Motivation



Expenditure distribution, RLMS, 1998

Income distribution analysis:

- The distribution is a mixture of lognormal components (?)
- Transition: changes in labor demand $\Rightarrow$ discrete mixture

Likelihood function

$$f(x\,|\,\theta) = \sum_{k=1}^{K} \phi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(x-\mu_k)^2}{2\sigma_k^2} \right]$$

$$\theta = \left(K, \mu_1, \sigma_1^2, \phi_{1,} \mu_2, \sigma_2^2, \phi_{2,} \ldots\right), \quad \sum_{k=1}^{K} \phi_k = 1$$

Constant variance version: $\sigma_1^2 = \sigma_2^2 = \ldots$

References

Day, N. E. Estimating the components of a mixture of normal distributions. *Biometrika*, **56** (3), 463–474 (1969)

Hathaway, R. A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions. *Ann. Stat.*, **13** (2), 795–800 (1985)

Basford, K. E., G. J. McLachlan. Likelikood Estimation with Normal Mixture Models. *Appl. Stat.*, **34**, 282–289 (1985)

Kiefer, J., J. Wolfowitz. Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *Ann. Math. Stat.*, **27**, 887–906 (1956)

Difficulties

- Homo/heteroskedastic?

  Heteroskedastic: ML estimate need not exist

  Comparison? Models are non-nested…

- Number of components?

  Likelihood ratio has an unusual distribution (estimation on the boundary?)

  McLachlan, G. J. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Appl. Stat.*, **36**, 318–324 (1987)

  Feng, Z. D., C. E. McCulloch. Using bootstrap likelihood ratios in finite mixture models. *J. of the Royal Stat. Soc.*, B, **58**, 609–617 (1996)

  Information criteria (AIC, SBIC, ICOMP, whatever…)

  Goodness of fit (Pearson χ^2, Kolmogorov-Smirnov?)

  Agha, M., D. S. Branker. Algorithm AS 317. Maximum Likelihood Estimation and Goodness-of-fit Tests for Mixtures of Distributions. *Appl. Stat.*, **46** (3), 399--407 (1997).

Syntax

----
help for **denormix**   Stas Kolenikov, skolenik@unc.edu

Univariate normal mixture decomposition
----

denormix varname [weight] [**if** exp] [**in** range],
[**l**evel(#) **noinit diff**icult **chi2 icomp logl**evel(#)
ncomp(#) **nmax**(#) **restart iter**ate(#) **gtol**erance(#)
constr(**var**) **pathto**(directory) **des**criptive **delog**]

Description
----

denormix performs the decomposition of the varname distribution into the mixture of normals. In other words, it assumes that the true density is the weighted sum of normal densities (with possibly different means and/or variances), and estimates the parameters of such mixture by maximum likelihood.

Tricks

➢ Parameter transformations

To ensure numerical stability and maximization without constraints: $\sigma^2 \to \log(\sigma^2)$; $\phi \to$ multinomial logit(ϕ)

➢ Convergence

`gtol(1e-4)` recommended option
`error 430:` convergence not achieved
`diag0cnt(e(V))` : singularity of covariance matrix
`symeigen e(V)` : singularity of covariance matrix
Restart if no convergence is diagnosed

➢ Global macros

~10 global macros used: loglevel; **`ml`** model statements; etc.

```
if `logleve'<=1 { global NMDq1 qui }
           else { global NMDq1 }
...
global NMD_mod $NMD_mod /m$NMD_n /lnV$NMD_n /lp$NMD_n
...
$NMDq1 ml model lf DeNorMix $NMD_mod ...
```

➢ Dynamic `ml model` definition

Separate denormdo.do run from the main denormix.ado

```
cap program drop DeNorMix
program define DeNorMix
   version 6
   args lf $NMD_par
...
```

Experience with real data

RLMS, 3600+ households; ~10 ths. individuals.

Results: "large errors" for heteroskedastic model; three components for homoskedastic model.

- ## Performance of Stata's **`ml`** optimizer

  Rescaling is great; **`difficult`** is also great.

- ## Computation speed

  Might take several hours / several hundred iterations at my Pentium II 333 MHz. Suggestion: specify **`iterate(200)`** so as not to waste time.

- ## Multiple maxima

  Yes, there are. If the number of components is greater than the "optimal" one, then you are bound to find 3-5-… maxima.

- ## Bad identification

  Two or more components may stick together; most of the time diagnosed by the convergence tracker.

- ## Large samples curse

  Sample sizes $1\text{-}3 \cdot 10^3$: χ^2 statistic is U-shaped wrt K.

  Sample size 10^4: χ^2 statistic is ≥ 50.

Further development (?)

- prediction: densities, cdfs, discriminant analysis
- EM-type algorithm: update means & variances

  — update proportions

  Peters, B. C. Jr., H. F. Walker. An Iterative Procedure for Obtaining Maximum-Likelihood Estimates of the Parameters for a Mixture of Normal Distributions. *SIAM J. of Appl. Math.*, **35** (2), 362–378 (1978).

  Xu, L., M. I. Jordan. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Computation*, **8**, 129–151 (1996)
- graphical output
- penalized likelihood for difference in variances
- **`constr(mean)`** was thought of initially

VISIT MY STATA SITE

net from http://www.komkon.org/~tacik/stata

Large sample

weighted noinit: 9176 observations

Results with strata <1% are discarded (most of the runs with 4-5 components)

#	LL, +11000	Chi2	df	p	AIC, -23000	SBIC, -23000	sigma	m1	share 1	m2	share 2	m3	share 3	m4	share 4	m5	share 5
1 component																	
20	-684.61	175.18	11	0.00	387.46	.	.865	6.343									
2 components																	
9	-618.34	124.08	9	0.00	244.67	273.17	.826	6.370	.989	3.914	.011						
10	-633.65	125.11	9	0.00	275.29	303.79	.838	6.326	.9936	8.968	.0064						
Model 1 identified: 5																	
3 components																	
18	-532.21	76.867	7	0.00	76.42	119.17	.756	6.340	.958	8.282	.023	4.159	.018				
Model 2.2 identified: 3																	
Model 2.1 identified: 2																	
5 components																	
8	-515.97	69.32	3	0.00	51.95	123.19	.684	6.294	.8762	3.022	.0022	9.766	.0023	4.652	.0354	7.562	.0840
Model 3.1 identified: 5																	
Model 2.2 identified: 3																	

Smaller sample

3619 observations (1 outlier)

	initial	improve	Iteration 0	Last iteration	AIC	ICOMP	Chi2()	Prob	Freq	Components
1				(1)	3564.37	3565.68	65.55 (12)	0.000	always	Mode
2	-2727.33	-2093.25	-2000.10	-1730.96 (5/9)	3461.93	3473.32	43.27 (10)	0.000	5	Mode + R hump (<1%)
2	-2727.33	-1896.25	-1896.25	-1761.27 (6/9)	3522.53	3534.15	43.80 (10)	0.000	4	Mode + L hump (1.1%)
2									3	Mode^2
3	-2789.54	-1810.20	-1810.20	-1698.39 (6/7)	3396.78	3410.52	17.40 (8)	**0.026**	5	Mode + L hump (2.0%) + R hump (0.5%)
3	-2789.54	-1799.46	-1729.60	-1710.85 (8)	3421.71	3439.75	33.88 (8)	0.000	1	Mode + R hump (1.5%)+ outlier
3				-1730.96 -1761.27					4	Mode + hump^2 or Mode^2 + hump
4				-1698.39					6	Mode^2 + L hump (2.0%) + R hump (0.5%)
4	-2814.77	-2482.82 -2170.17	-1830.13 -2066.80	-1669.55 (6/24)	3339.09	3367.64	12.00 (6)	**0.062**	3	Mode + L hump (2.9%) + R hump (1.4%) + outlier
4	-2826.54	-2043.54	-1777.81	-1698.22 (8)	3396.44	3482.75	17.42 (6)	0.008	1	Mode + L hump (2.1%) + LL hump (<.1%) + R hump (1.1%)
5	-2826.54	-2052.48 -1997.30 -1752.56	-1794.78 –1707.12 -1697.35	-1668.41 (7/8)	3336.82	3361.98	11.04 (4)	**0.026**	3	Mode + L hump (3.6%) + LL hump (0.1%) + R hump (1.4%) + outlier
5	-2826.54	-1912.52 -1838.38	-1900.05 -1687.89	-1667.17 (8/16)	3334.34	3361.70	9.22 (4)	**0.056**	2	Mode + R hump (3.1%) + L hump (3.2%) + RR hump (0.3%) + outlier
5				-1669.55					4	Mode^2 + L hump (2.9%) + R hump (1.4%) + outlier