Fundamentals of Statistical Modeling (VT20)

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden

Lab 5

Load the dataset and the `mlci` command

```
. version 14
. use https://raw.githubusercontent.com/anddis/fsm/master/data/lab5_1.dta, clear
. run https://raw.githubusercontent.com/anddis/fsm/master/do/mlci.do
```

Exercise 1

We consider again the oral cancer dataset (see Lab 4). We measure time to death (y) due to oral cancer ($d = 1$) (continuous line in the graphs) or to other causes ($d = 2$) (dashed line in the graphs). Just to get started, we exclude the censored observations ($d = 0$) from our estimation procedure (note the `if d != 0` at the end of the `mlexp` command).

We model the joint distribution $f(y, d)$ through the conditional expansion

$$f(y, d) = f(d|y)f(y).$$

We consider a log-normal distribution for $f(y)$ and a bernoulli distribution for $f(d|y)$. Estimate the model's parameters. Remember to constrain the bounded parameters. How do we interpret $\exp(\gamma_1)$?

```
. local G = "(log(y)-{mu})/exp({theta})"
. local g = "(1/exp({theta})/y)"
. local eta = "invlogit({gamma0}+{gamma1}*y)"
. local f1 = "normalden(`G´)*`g´"
. local f2 = "`eta´^(d==1)*(1-`eta´)^(d==2)"
. mlexp (log(`f1´*`f2´)) if d != 0
initial:       log likelihood = -677.80543
alternative:   log likelihood = -849.52367
rescale:       log likelihood = -669.37878
rescale eq:    log likelihood = -583.30772
Iteration 0:   log likelihood = -583.30772
Iteration 1:   log likelihood = -576.92245
Iteration 2:   log likelihood = -576.89028
Iteration 3:   log likelihood = -576.89028

Maximum likelihood estimation

Log likelihood = -576.89028                    Number of obs     =        184
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/mu	.9087088	.0902927	10.06	0.000	.7317384	1.085679
/theta	.2027691	.0521286	3.89	0.000	.1005989	.3049393
/gamma0	.9899309	.2317795	4.27	0.000	.5356514	1.44421
/gamma1	-.2094189	.0433561	-4.83	0.000	-.2943954	-.1244425

```
. mlci exp /theta
  1.22479   95% CI: 1.105833, 1.356543

. mlci invlogit /gamma0
```

```
      .7290743    95% CI: .6308002, .8091058
. mlci exp /gamma1
      .8110554    95% CI: .7449819,  .882989
```

Plot the estimated distributions $\hat{f}(y)$ and $\hat{f}(d|y)$. Interpret the plots.

```
. gen fhat_y = normalden((log(y) - _b[/mu]) / exp(_b[/theta]))*(1 / y / exp(_b[/theta]))
. tw  (line fhat_y y, sort), name(p1, replace) legend(rows(1))
.
. gen fhat_dy = invlogit(_b[/gamma0]+_b[/gamma1]*y) if d == 1
(171 missing values generated)
. replace fhat_dy = 1-invlogit(_b[/gamma0]+_b[/gamma1]*y) if d == 2
(88 real changes made)
. tw (line fhat_dy y if d==1, sort lc(navy)) ///
> (line fhat_dy y if d==2, sort lc(navy) lp(-)), ///
> name(p2, replace)  legend(off) ylabel(0(0.2)1)
. graph combine p2 p1, name(c1, replace)
. graph export c1.png, replace
(file c1.png written in PNG format)
```
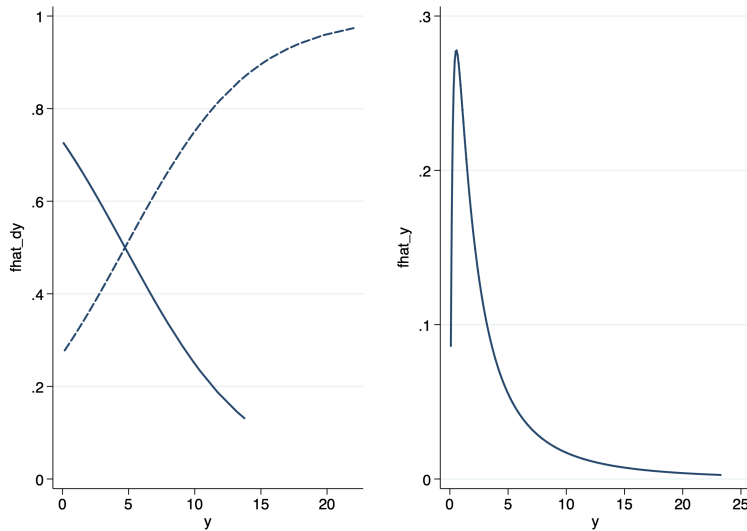

Exercise 2

Some of the times (y) are actually right-censored. Estimate the model's parameters by modifying the likelihood accordingly to take this into account (last equation on slide 137). Remember to constrain the bounded parameters.

```
. tab d
```

d	Freq.	Percent	Cum.
Censored	83	31.09	31.09
Oral ca. death	96	35.96	67.04
Other death	88	32.96	100.00
Total	267	100.00	

```
. local G = "(log(y)-{mu})/exp({theta})"
. local g = "(1/exp({theta})/y)"
. local eta = "invlogit({gamma0}+{gamma1}*y)"
. local f1 = "normalden(`G´)*`g´"
. local S = "1-normal(`G´)"
. local f2 = "`eta´^(d==1)*(1-`eta´)^(d==2)"
. mlexp ((d!=0)*log(`f1´*`f2´)+(d==0)*log(`S´))

initial:       log likelihood = -1018.5702
alternative:   log likelihood = -1000.4936
rescale:       log likelihood = -931.72631
```

2

```
rescale eq:     log likelihood = -705.99993
Iteration 0:    log likelihood = -705.99993
Iteration 1:    log likelihood = -689.39801
Iteration 2:    log likelihood = -688.25879
Iteration 3:    log likelihood = -688.25814
Iteration 4:    log likelihood = -688.25814

Maximum likelihood estimation

Log likelihood = -688.25814                     Number of obs      =        267
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
/mu	1.615544	.1030555	15.68	0.000	1.413559	1.817529
/theta	.4444992	.0548368	8.11	0.000	.3370211	.5519773
/gamma0	.9899305	.2317795	4.27	0.000	.535651	1.44421
/gamma1	-.2094188	.0433561	-4.83	0.000	-.2943953	-.1244424

```
. mlci exp /theta
 1.559709   95% CI: 1.400769, 1.736684

. mlci invlogit /gamma0
 .7290742   95% CI: .6308002, .8091057

. mlci exp /gamma1
 .8110555   95% CI:  .744982, .8829891
```

Plot the estimated distributions $\hat{f}(y)$ and $\hat{f}(d|y)$.

```
. gen fhat_y2 = normalden((log(y) - _b[/mu]) / exp(_b[/theta]))*(1 / y / exp(_b[/theta]))
. tw  (line fhat_y2 y, sort), name(p3, replace) legend(rows(1))
.
. gen fhat_dy2 = invlogit(_b[/gamma0]+_b[/gamma1]*y) if d == 1
(171 missing values generated)
. replace fhat_dy2 = 1-invlogit(_b[/gamma0]+_b[/gamma1]*y) if d == 2
(88 real changes made)
. tw (line fhat_dy2 y if d==1, sort lc(navy)) ///
> (line fhat_dy2 y if d==2, sort lc(navy) lp(-)), ///
> name(p4, replace) legend(rows(1)) legend(off) ylabel(0(0.2)1)
. graph combine p4 p3, name(c2, replace)
. graph export c2.png, replace
(file c2.png written in PNG format)
```
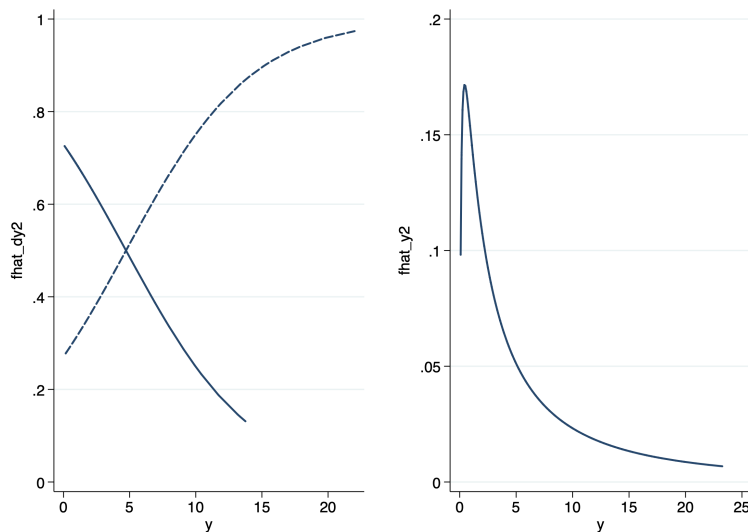

Extra 1

Plot the estimated cumulative incidence functions $\hat{F}(y, d = 1)$ and $\hat{F}(y, d = 2)$ (see slide 140) and overlay them to their nonparametric counterparts obtained using Stata's `stcrreg` command.

$$\hat{F}(y,1) = \int_0^y \hat{f}_{Y,D}(u, d=1)du = \int_0^y \hat{f}_Y(u)\hat{f}_{D|Y}(d=1|u)du$$

and

$$\hat{F}(y,2) = \int_0^y \hat{f}_{Y,D}(u, d=2)du = \int_0^y \hat{f}_Y(u)f_{D|Y}(d=2|u)du$$

```
. sum y, meanonly
. range y1 r(min) r(max)
. gen f1 = normalden((log(y1)-_b[/mu])/exp(_b[/theta]))/exp(_b[/theta])/y1 * invlogit(_b[/gamma0]+_b[/gamma1]*y1)
. gen f2 = normalden((log(y1)-_b[/mu])/exp(_b[/theta]))/exp(_b[/theta])/y1 * (1-invlogit(_b[/gamma0]+_b[/gamma1]*y1))
. sort y1
. gen F1 = sum(f1)*(r(max)-r(min))/(_N-1)
. gen F2 = sum(f2)*(r(max)-r(min))/(_N-1)
.
. stset y, fail(d==1)

      failure event:  d == 1
 obs. time interval:  (0, y]
  exit on or before:  failure
───────────────────────────────────────────────────────────────
          267  total observations
            0  exclusions
───────────────────────────────────────────────────────────────
          267  observations remaining, representing
           96  failures in single-record/single-failure data
    1,620.864  total analysis time at risk and under observation
                                     at risk from t =          0
                          earliest observed entry t =          0
                               last observed exit t =     23.258
. stcrreg, compete(d==2)

          failure _d:  d == 1
     analysis time _t:  y
Competing-risks regression                  No. of obs       =        267
                                            No. of subjects  =        267
Failure event  : d == 1                     No. failed       =         96
Competing event: d == 2                     No. competing    =         88
                                            No. censored     =         83

                                            Wald chi2(0)     =       0.00
Log pseudolikelihood = -508.69926           Prob > chi2      =          .

───────────────────────────────────────────────────────────────
                         Robust
         _t │       SHR    Std. Err.      z    P>|z|    [95% Conf. Interval]
────────────┼──────────────────────────────────────────────────

. predict cif1, basecif
. stset y, fail(d==2)

      failure event:  d == 2
 obs. time interval:  (0, y]
  exit on or before:  failure
───────────────────────────────────────────────────────────────
          267  total observations
            0  exclusions
───────────────────────────────────────────────────────────────
          267  observations remaining, representing
           88  failures in single-record/single-failure data
    1,620.864  total analysis time at risk and under observation
                                     at risk from t =          0
                          earliest observed entry t =          0
                               last observed exit t =     23.258
. stcrreg, compete(d==1)

          failure _d:  d == 2
     analysis time _t:  y
Competing-risks regression                  No. of obs       =        267
                                            No. of subjects  =        267
```

```
Failure event  : d == 2                      No. failed      =        88
Competing event: d == 1                      No. competing   =        96
                                             No. censored    =        83
                                             Wald chi2(0)    =      0.00
Log pseudolikelihood = -442.87519            Prob > chi2     =         .
```

_t	SHR	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]

```
. predict cif2, basecif

.
. tw (line cif1 cif2 y, sort connect(J J) lc(orange orange) lp(l -)) ///
> (line F1 F2 y1, sort lc(navy navy) lp(l -)), legend(off) name(p5, replace)
. graph export p5.png, replace
(file p5.png written in PNG format)
```
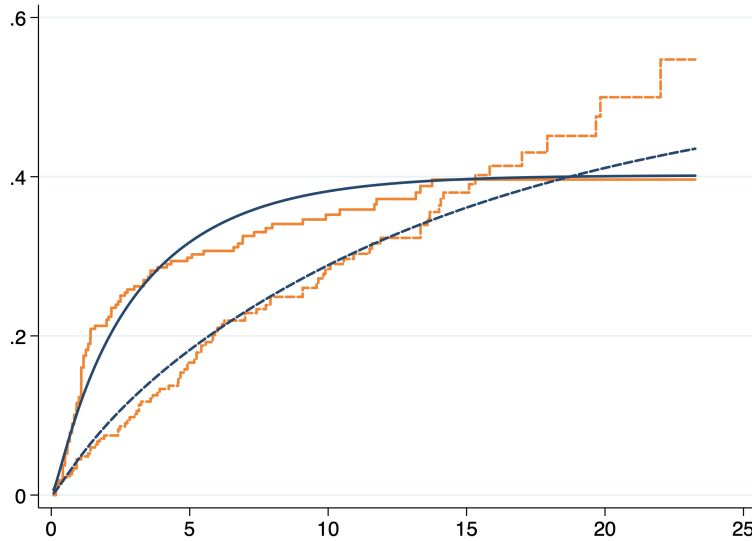

Extra 2

We now model the joint distribution $f(y, d, x)$ through conditional expansion.

$$f(y, d, x) = f(d|y, x)f(y|x)f(x),$$

The variable x is tumor grade at diagnosis: low ($x = 0$, blue in the graphs) or high ($x = 1$, red in the graphs). We consider a log-normal distribution for $f(y|x)$ and a bernoulli distribution for both $f(d|y, x)$ and $f(x)$.

Estimate the model's parameters. Remember to constrain the bounded parameters. How do we interpret the model's parameters?

```
. local G = "(log(y)-({beta0}+{beta1}*x))/exp({theta})"
. local g = "(1/exp({theta})/y)"
. local eta = "invlogit({gamma0}+{gamma1}*y+{gamma2}*x)"
. local psi = "invlogit({delta})"
. local f1 = "normalden(`G´)*`g´"
. local S = "1-normal(`G´)"
. local f2 = "`eta´^(d==1)*(1-`eta´)^(d==2)"
. local f3 = "`psi´^(x==1)*(1-`psi´)^(x==0)"
. mlexp ((d!=0)*log(`f1´*`f2´)+(d==0)*log(`S´)+log(`f3´))
initial:       log likelihood = -1203.6405
alternative:   log likelihood = -1193.8352
rescale:       log likelihood = -1105.7963
rescale eq:    log likelihood = -876.80228
```

```
Iteration 0:   log likelihood = -876.80228
Iteration 1:   log likelihood = -864.00194
Iteration 2:   log likelihood = -863.66655
Iteration 3:   log likelihood = -863.66597
Iteration 4:   log likelihood = -863.66597

Maximum likelihood estimation

Log likelihood = -863.66597                      Number of obs     =       267
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
/beta0	2.048152	.1475271	13.88	0.000	1.759004	2.3373
/beta1	-.8251894	.1958675	-4.21	0.000	-1.209083	-.4412961
/theta	.4088398	.0546405	7.48	0.000	.3017463	.5159332
/gamma0	.7617058	.3079601	2.47	0.013	.1581151	1.365297
/gamma1	-.2047134	.043271	-4.73	0.000	-.2895229	-.1199038
/gamma2	.3611933	.328396	1.10	0.271	-.2824511	1.004838
/delta	.0974553	.1225434	0.80	0.426	-.1427252	.3376359

```
. mlci exp /beta1
  .438152   95% CI: .2984709, .6432022

. mlci exp /theta
 1.505071   95% CI: 1.352218, 1.675201

.
. mlci exp /gamma1
 .8148808   95% CI: .7486206, .8870057

. mlci exp /gamma2
 1.435041   95% CI: .7539335, 2.731464

.
. mlci invlogit /delta
 .5243446   95% CI: .4643791, .5836161
```

Plot the estimated distributions $\hat{f}(d|y,x)$, $\hat{f}(y|x)$, and $\hat{f}(x)$.

```
. gen fhat_dyx3 = invlogit(_b[/gamma0]+_b[/gamma1]*y+_b[/gamma2]*x) if d == 1
(171 missing values generated)
. replace fhat_dyx3 = 1-invlogit(_b[/gamma0]+_b[/gamma1]*y+_b[/gamma2]*x) if d == 2
(88 real changes made)
. tw (line fhat_dyx3 y if x == 0 & d == 1, sort lc(navy)) ///
> (line fhat_dyx3 y if x == 1 & d == 1, sort lc(maroon)) ///
> (line fhat_dyx3 y if x == 0 & d == 2, sort lc(navy) lp(-)) ///
> (line fhat_dyx3 y if x == 1 & d == 2, sort lc(maroon) lp(-)), ///
> name(p6, replace) legend(rows(1)) legend(off) ylabel(0(0.2)1)

.
. gen fhat_yx3 = normalden((log(y) - (_b[/beta0]+_b[/beta1]*x)) / exp(_b[/theta]))*(1 / y / exp(_b[/theta]))
. tw  (line fhat_yx3 y if x == 0, sort lc(navy)) ///
> (line fhat_yx3 y if x == 1, sort lc(maroon)), name(p7, replace) legend(off)

.
. gen fhat_x3 = invlogit(_b[/delta]) if x==1
(127 missing values generated)
. replace fhat_x3 = 1-invlogit(_b[/delta]) if x==0
(127 real changes made)

.
. tw (dropline fhat_x3 x if x == 1, lc(maroon) lc(maroon) mc(maroon)) ///
> (dropline fhat_x3 x if x == 0, lc(navy) lc(navy) mc(navy)), ///
> name(p8, replace) legend(off) ylabel(0(0.2)1) xlabel(0 1) xscale(range(-.5 1.5))
. graph combine p6 p7 p8, name(c3, replace)

. graph export c3.png, replace
(file c3.png written in PNG format)
```

fhat_dyx3

y

fhat_yx3

y

fhat_x3

x
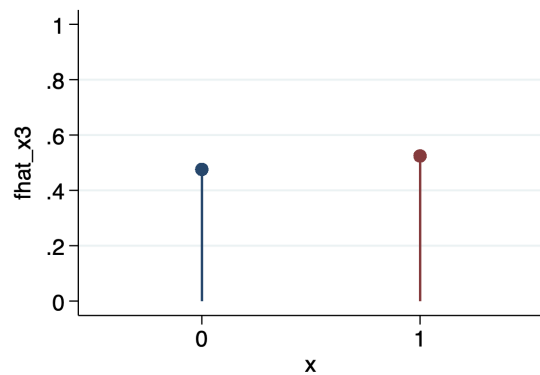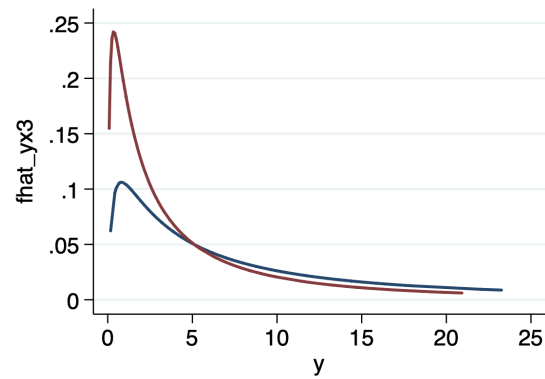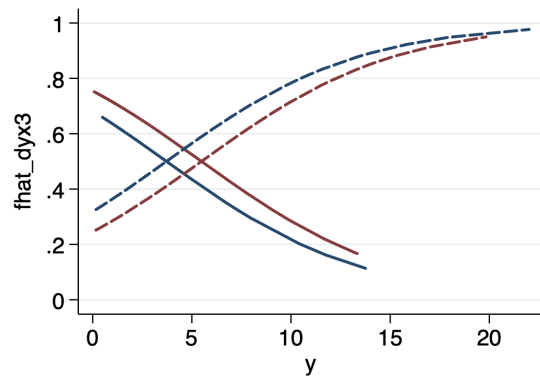
Exercise 3

We recruited 2,784 subjects in Sweden at the time of their first myocardial infarction. We took a blood sample and measured LDL cholesterol (mmol/L) on a first follow-up visit, 1 month after the MI (variable `ldl1`). We then measured LDL cholesterol again 6 months after the MI (second follow-up visit) (variable `ldl2`). Plot the sample histogram of the 2 variables. What can we say about them?

```
. use https://raw.githubusercontent.com/anddis/fsm/master/data/lab5_2.dta, clear
.
. su ldl1 ldl2, detail
                             ldl1
-----------------------------------------------------------------
      Percentiles      Smallest
 1%         .833            .123
 5%        1.164            .273
10%        1.372            .322       Obs              2,784
25%       1.6945            .412       Sum of Wgt.      2,784

50%       2.1195                       Mean          2.195147
                        Largest        Std. Dev.     .7253906
75%        2.581           5.667
90%        3.073           5.803       Variance      .5261915
95%        3.512           5.829       Skewness      .9833921
99%        4.441           6.274       Kurtosis      5.316994
                             ldl2
-----------------------------------------------------------------
      Percentiles      Smallest
 1%         .977            .077
 5%        1.431            .337
10%        1.634            .408       Obs              2,784
25%        2.028            .434       Sum of Wgt.      2,784

50%        2.509                       Mean          2.670523
                        Largest        Std. Dev.     .9806975
75%        3.129           7.839
90%        3.886           7.925       Variance      .9617677
```
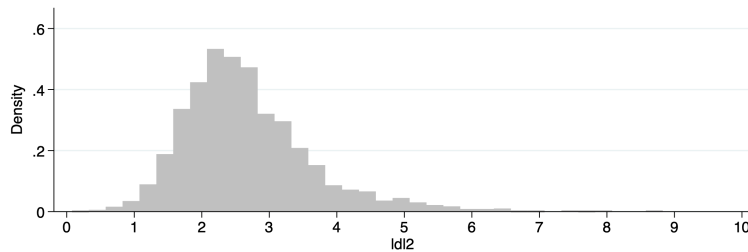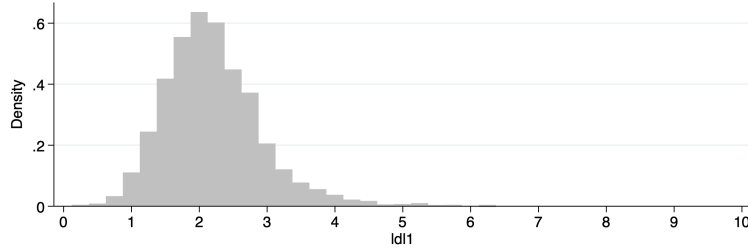
```
     95%          4.525           8.611         Skewness        1.33667
     99%          5.884           8.823         Kurtosis        6.399398

.
. hist ldl1, xlabel(0/10) width(.25) name(s1, replace)
(bin=25, start=.123, width=.25)

. hist ldl2, xlabel(0/10) width(.25) name(s2, replace)
(bin=35, start=.077, width=.25)

. graph combine s1 s2, name(s0, replace) ycommon cols(1)

. graph export s0.png, replace
(file s0.png written in PNG format)
```


We model the marginal distributions of the 2 variables: $f(ldl_1), f(ldl_2)$. We consider skew-normal models. This means that $Z = G(Y)$ follows a standard skew-normal distribution.

$$G(y) = (y - \mu)/\sigma$$

$$g(y) = 1/\sigma$$

$$f_Z(z) = 2F_N(\alpha z)f_N(z)$$

where $F_N(z)$ and $f_N(z)$ are the standard normal CDF and PDF, respectively.

Estimate the 2 models' parameters and plot the densities over the sample histograms. Are the data suggesting that the skewness parameter α is different from 0?

```
. local G = "(ldl1-{mu})/exp({theta})"

. local g = "1 / exp({theta})"

. local f1 = "normal({alpha}*`G´)"

. local f2 = "normalden(`G´)*`g´"

. mlexp (log(2*`f1´*`f2´))
initial:        log likelihood =  -9998.109
alternative:    log likelihood = -4801.7168
rescale:        log likelihood = -4801.7168
rescale eq:     log likelihood = -3154.1605
Iteration 0:    log likelihood = -3154.1605
Iteration 1:    log likelihood = -2945.0724
Iteration 2:    log likelihood = -2924.1425
Iteration 3:    log likelihood = -2920.4731
Iteration 4:    log likelihood = -2920.4395
Iteration 5:    log likelihood = -2920.4395

Maximum likelihood estimation

Log likelihood = -2920.4395                     Number of obs      =       2,784
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]

```
        /alpha │    2.611656    .1571988     16.61   0.000     2.303552    2.91976
          /mu  │    1.406456    .0213084     66.00   0.000     1.364692    1.44822
        /theta │    .0690262    .0198467      3.48   0.001     .0301275    .107925
```

```
. gen fhat_ldl1 = 2*normal(_b[/alpha]*(ldl1-_b[/mu])/exp(_b[/theta]))*normalden((ldl1-_b[/mu])/exp(_b[/theta]))/exp(_b[/
> theta])

. tw (hist ldl1, width(.25)) (line fhat_ldl1 ldl1, sort), name(s1, replace) xlabel(0/10)  legend(off)

. 
. local G = "(ldl2-{mu})/exp({theta})"

. local g = "1 / exp({theta})"

. local f1 = "normal({alpha}*`G´)"

. local f2 = "normalden(`G´)*`g´"

. mlexp (log(2*`f1´*`f2´))
initial:        log likelihood = -13823.944
alternative:    log likelihood = -5798.7897
rescale:        log likelihood = -5072.3055
rescale eq:     log likelihood = -4031.7673
Iteration 0:    log likelihood = -4031.7673
Iteration 1:    log likelihood = -3868.3234
Iteration 2:    log likelihood = -3860.2447  (not concave)
Iteration 3:    log likelihood = -3675.2321
Iteration 4:    log likelihood = -3651.3767
Iteration 5:    log likelihood = -3649.0576
Iteration 6:    log likelihood =  -3649.036
Iteration 7:    log likelihood =  -3649.036

Maximum likelihood estimation

Log likelihood =  -3649.036                      Number of obs     =      2,784
```

```
             │      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
        /alpha │    3.524436    .1928855     18.27   0.000     3.146387    3.902484
          /mu  │    1.561897    .0223357     69.93   0.000     1.518119    1.605674
        /theta │    .3920594    .0175325     22.36   0.000     .3576964    .4264223
```

```
. gen fhat_ldl2 = 2*normal(_b[/alpha]*(ldl2-_b[/mu])/exp(_b[/theta]))*normalden((ldl2-_b[/mu])/exp(_b[/theta]))/exp(_b[/
> theta])

. tw (hist ldl2, width(.25)) (line fhat_ldl2 ldl2, sort), name(s2, replace) xlabel(0/10) legend(off)

. 
. graph combine s1 s2, name(p6, replace) ycommon cols(1)

. graph export p6.png, replace
(file p6.png written in PNG format)
```
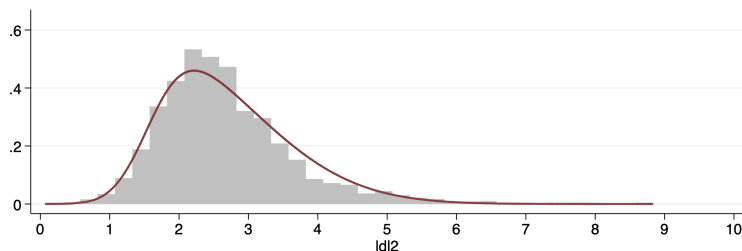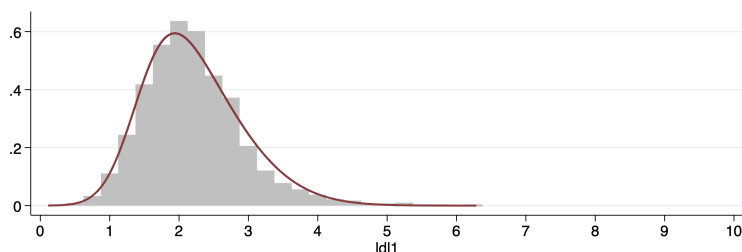
Exercise 4

We consider the joint distribution of ldl_1 and ldl_2

$$f(ldl_1, ldl_2) = f(ldl_2|ldl_1)f(ldl_1)$$

We assume that $f(ldl_1)$ and $f(ldl_2|ldl_1)$ are skew-normal. Allow all parameters of $f(ldl_2|ldl_1)$ to depend on ldl_1. Estimate the model's parameters. Remember to constrain the bounded parameters.

```
. local f1 = "normal({alpha1}*(ldl1-{beta1})/exp({theta1}))"
. local f2 = "normalden((ldl1-{beta1})/exp({theta1}))/exp({theta1})"
.
. local f3 = "normal({alpha2: _cons ldl1}*(ldl2-{beta2: _cons ldl1})/exp({theta2: _cons ldl1}))"
. local f4 = "normalden((ldl2-{beta2:})/exp({theta2:}))/exp({theta2:})"
.
. // The code above is equivalent to the (probably more familar) code below, but it´s more compact
. // and easier to read. The output of the code above is easier to read, too.
. // local f3 = "normal(({alpha20}+{alpha21}*ldl1)*(ldl2-({beta20}+{beta21}*ldl1))/exp({theta20}+{theta21}*ldl1))"
. // local f4 = "normalden((ldl2-({beta20}+{beta21}*ldl1))/exp({theta20}+{theta21}*ldl1))/exp({theta20}+{theta21}*ldl1)"
.
. mlexp(log(2*`f1´*`f2´) + log(2*`f3´*`f4´))
initial:       log likelihood = -23822.053
alternative:   log likelihood = -10600.506
rescale:       log likelihood = -10018.378
rescale eq:    log likelihood = -7185.9278
Iteration 0:   log likelihood = -7185.9278  (not concave)
Iteration 1:   log likelihood = -6701.0692
Iteration 2:   log likelihood = -6388.4497  (not concave)
Iteration 3:   log likelihood = -6266.5071
Iteration 4:   log likelihood = -6179.0375
Iteration 5:   log likelihood = -6171.9849
Iteration 6:   log likelihood = -6171.8148
Iteration 7:   log likelihood = -6171.8146

Maximum likelihood estimation

Log likelihood = -6171.8146                    Number of obs     =      2,784
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
alpha1						
_cons	2.611652	.1571985	16.61	0.000	2.303549	2.919756
beta1						
_cons	1.406457	.0213085	66.00	0.000	1.364693	1.44822
theta1						
_cons	.0690259	.0198467	3.48	0.001	.0301271	.1079246
alpha2						
ldl1	-1.154094	.1203596	-9.59	0.000	-1.389995	-.9181938
_cons	5.912853	.3713762	15.92	0.000	5.184969	6.640737
beta2						
ldl1	.6225221	.0290767	21.41	0.000	.5655327	.6795114
_cons	.3965575	.0557818	7.11	0.000	.2872271	.5058879
theta2						
ldl1	.0917335	.0204176	4.49	0.000	.0517159	.1317512
_cons	.0259962	.0464537	0.56	0.576	-.0650512	.1170437

```
. mlexp, coeflegend // Display the legend for the model´s coefficients, so that we know how to reference them
Maximum likelihood estimation

Log likelihood = -6171.8146                    Number of obs     =      2,784
```

	Coef.	Legend
alpha1		
_cons	2.611652	_b[alpha1:_cons]
beta1		

	_cons	1.406457	_b[beta1:_cons]
theta1			
	_cons	.0690259	_b[theta1:_cons]
alpha2			
	ldl1	-1.154094	_b[alpha2:ldl1]
	_cons	5.912853	_b[alpha2:_cons]
beta2			
	ldl1	.6225221	_b[beta2:ldl1]
	_cons	.3965575	_b[beta2:_cons]
theta2			
	ldl1	.0917335	_b[theta2:ldl1]
	_cons	.0259962	_b[theta2:_cons]

Draw a scatterplot of ldl_2 versus ldl_1. Do the results above agree with the plot? Make a qualitative assessment.

Plot the estimated conditional denisity $\hat{f}(ldl_2|ldl_1)$ for ldl_1 values of 2, 3, and 5 mmol/L. Again, make a qualitative assessment of the plot.

```
. tw (scatter ldl2 ldl1, msize(tiny) msym(Oh)), name(p7, replace)
. graph export p7.png, replace
(file p7.png written in PNG format)

.
. gen fhat_y2 = 2 * normal((_b[alpha2:_cons]+_b[alpha2:ldl1]*2)*(ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*2))/exp((_b[theta2
> :_cons]+_b[theta2:ldl1]*2))) * ///
> normalden((ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*2))/exp((_b[theta2:_cons]+_b[theta2:ldl1]*2)))/exp((_b[theta2:_cons]+_
> b[theta2:ldl1]*2))

.
. gen fhat_y3 = 2 * normal((_b[alpha2:_cons]+_b[alpha2:ldl1]*3)*(ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*3))/exp((_b[theta2
> :_cons]+_b[theta2:ldl1]*3))) * ///
> normalden((ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*3))/exp((_b[theta2:_cons]+_b[theta2:ldl1]*3)))/exp((_b[theta2:_cons]+_
> b[theta2:ldl1]*3))

.
. gen fhat_y5 = 2 * normal((_b[alpha2:_cons]+_b[alpha2:ldl1]*5)*(ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*5))/exp((_b[theta2
> :_cons]+_b[theta2:ldl1]*5))) * ///
> normalden((ldl2-(_b[beta2:_cons]+_b[beta2:ldl1]*5))/exp((_b[theta2:_cons]+_b[theta2:ldl1]*5)))/exp((_b[theta2:_cons]+_
> b[theta2:ldl1]*5))

.
. tw (line fhat_y2 fhat_y3 fhat_y5 ldl2, sort lc(blue purple red)), name(p8, replace) legend(rows(1))
. graph export p8.png, replace
(file p8.png written in PNG format)
```