Submitting results to HepData

Marcin Chrząszcz
mchrzasz@cern.ch

University of Zurich[UZH]

Zurich meeting, CERN
September 27, 2016

The challenge

⇛ The $B \to K^* \mu\mu$ analysis has produced over 80 tables with results and correlation tables...

⇛ Theorists need all the correlation tables to make the global fit.

⇛ Reading the numbers form our tex files from all the theories groups is not really a nice way to do it:

⇛ Many thanks for Alex Grecu for help understanding how the HepData works!

Example:

⇒ The scripts I show here are not $100\ \%$ plug and play.
⇒ Each table in latex is a bit different and needs some special modification.
⇒ But to modify the scripts and apply them to your case should not be much work.

Example:

Table 4: *CP*-averaged angular observables evaluated by the unbinned maximum likelihood fit. The first uncertainties are statistical and the second systematic.

	$0.10 < q^2 < 0.98\,\text{GeV}^2/c^4$	$1.1 < q^2 < 2.5\,\text{GeV}^2/c^4$	$2.5 < q^2 < 4.0\,\text{GeV}^2/c^4$
F_L	$0.263^{+0.045}_{-0.044} \pm 0.017$	$0.660^{+0.083}_{-0.077} \pm 0.022$	$0.876^{+0.109}_{-0.097} \pm 0.017$
S_3	$-0.036^{+0.063}_{-0.063} \pm 0.005$	$-0.077^{+0.087}_{-0.105} \pm 0.005$	$0.035^{+0.098}_{-0.089} \pm 0.007$
S_4	$0.082^{+0.068}_{-0.069} \pm 0.009$	$-0.077^{+0.111}_{-0.113} \pm 0.005$	$-0.234^{+0.127}_{-0.144} \pm 0.006$
S_5	$0.170^{+0.059}_{-0.058} \pm 0.018$	$0.137^{+0.099}_{-0.094} \pm 0.009$	$-0.022^{+0.110}_{-0.103} \pm 0.008$
A_{FB}	$-0.003^{+0.058}_{-0.057} \pm 0.009$	$-0.191^{+0.068}_{-0.080} \pm 0.012$	$-0.118^{+0.082}_{-0.090} \pm 0.007$
S_7	$0.015^{+0.059}_{-0.059} \pm 0.006$	$-0.219^{+0.094}_{-0.104} \pm 0.004$	$0.068^{+0.120}_{-0.112} \pm 0.005$
S_8	$0.079^{+0.076}_{-0.075} \pm 0.007$	$-0.098^{+0.108}_{-0.123} \pm 0.005$	$0.030^{+0.129}_{-0.131} \pm 0.006$
S_9	$-0.083^{+0.058}_{-0.057} \pm 0.004$	$-0.119^{+0.087}_{-0.104} \pm 0.005$	$-0.092^{+0.105}_{-0.125} \pm 0.007$

	$4.0 < q^2 < 6.0\,\text{GeV}^2/c^4$	$6.0 < q^2 < 8.0\,\text{GeV}^2/c^4$	$11.0 < q^2 < 12.5\,\text{GeV}^2/c^4$
F_L	$0.611^{+0.052}_{-0.053} \pm 0.017$	$0.579^{+0.046}_{-0.046} \pm 0.015$	$0.493^{+0.049}_{-0.047} \pm 0.013$
S_3	$0.035^{+0.069}_{-0.068} \pm 0.007$	$-0.042^{+0.058}_{-0.059} \pm 0.011$	$-0.189^{+0.054}_{-0.058} \pm 0.005$
S_4	$-0.219^{+0.086}_{-0.084} \pm 0.008$	$-0.296^{+0.063}_{-0.067} \pm 0.011$	$-0.283^{+0.084}_{-0.095} \pm 0.009$
S_5	$-0.146^{+0.077}_{-0.078} \pm 0.011$	$-0.249^{+0.059}_{-0.060} \pm 0.012$	$-0.327^{+0.076}_{-0.079} \pm 0.009$
A_{FB}	$0.025^{+0.051}_{-0.052} \pm 0.004$	$0.152^{+0.041}_{-0.040} \pm 0.008$	$0.318^{+0.044}_{-0.040} \pm 0.009$
S_7	$-0.016^{+0.081}_{-0.080} \pm 0.004$	$-0.047^{+0.068}_{-0.066} \pm 0.003$	$-0.141^{+0.072}_{-0.074} \pm 0.005$
S_8	$0.167^{+0.094}_{-0.091} \pm 0.004$	$-0.085^{+0.072}_{-0.070} \pm 0.006$	$-0.007^{+0.070}_{-0.072} \pm 0.005$
S_9	$-0.032^{+0.071}_{-0.071} \pm 0.004$	$-0.024^{+0.059}_{-0.060} \pm 0.005$	$-0.004^{+0.070}_{-0.073} \pm 0.006$

	$15.0 < q^2 < 17.0\,\text{GeV}^2/c^4$	$17.0 < q^2 < 19.0\,\text{GeV}^2/c^4$
F_L	$0.349^{+0.039}_{-0.039} \pm 0.009$	$0.354^{+0.049}_{-0.048} \pm 0.025$
S_3	$-0.142^{+0.044}_{-0.049} \pm 0.007$	$-0.188^{+0.074}_{-0.084} \pm 0.017$
S_4	$-0.321^{+0.055}_{-0.074} \pm 0.007$	$-0.266^{+0.063}_{-0.072} \pm 0.010$
S_5	$-0.316^{+0.051}_{-0.057} \pm 0.009$	$-0.323^{+0.063}_{-0.072} \pm 0.009$
A_{FB}	$0.411^{+0.041}_{-0.037} \pm 0.008$	$0.305^{+0.049}_{-0.048} \pm 0.013$
S_7	$0.061^{+0.058}_{-0.058} \pm 0.005$	$0.044^{+0.073}_{-0.072} \pm 0.013$
S_8	$0.003^{+0.061}_{-0.061} \pm 0.003$	$0.013^{+0.071}_{-0.070} \pm 0.005$
S_9	$-0.019^{+0.054}_{-0.056} \pm 0.004$	$-0.094^{+0.065}_{-0.067} \pm 0.004$

Example:

RE	**P P --> B0 < K*(892) < K+ PI- > MU+ MU- > X**
SQRT(S)	**7000.0 GeV**
SQRT(S)	**8000.0 GeV**

q^2 IN GEV**2	F_{L}	S_3	S_4	S_5	A_{FB}	S_7	S_8	S_9
0.10 – 0.98	0.263 +0.045,-0.044 (stat) ± 0.017 (sys)	-0.036 ± 0.063 (stat) ± 0.005 (sys)	0.082 +0.068,-0.069 (stat) ± 0.009 (sys)	0.170 +0.059,-0.058 (stat) ± 0.018 (sys)	-0.003 +0.058,-0.057 (stat) ± 0.009 (sys)	0.015 ± 0.059 (stat) ± 0.006 (sys)	0.079 +0.076,-0.075 (stat) ± 0.007 (sys)	-0.083 +0.058,-0.057 (stat) ± 0.004 (sys)
1.1 – 2.5	0.660 +0.083,-0.077 (stat) ± 0.022 (sys)	-0.077 +0.087,-0.105 (stat) ± 0.005 (sys)	-0.077 +0.111,-0.113 (stat) ± 0.005 (sys)	0.137 +0.099,-0.094 (stat) ± 0.009 (sys)	-0.191 +0.068,-0.080 (stat) ± 0.012 (sys)	-0.219 +0.094,-0.104 (stat) ± 0.004 (sys)	-0.098 +0.108,-0.123 (stat) ± 0.005 (sys)	-0.119 +0.087,-0.104 (stat) ± 0.005 (sys)
2.5 – 4.0	0.876 +0.109,-0.097 (stat) ± 0.017 (sys)	0.035 +0.098,-0.089 (stat) ± 0.007 (sys)	-0.234 +0.127,-0.144 (stat) ± 0.006 (sys)	-0.022 +0.110,-0.103 (stat) ± 0.000 (sys)	-0.118 +0.082,-0.090 (stat) ± 0.007 (sys)	0.068 +0.120,-0.112 (stat) ± 0.005 (sys)	0.030 +0.129,-0.131 (stat) ± 0.006 (sys)	-0.092 +0.105,-0.125 (stat) ± 0.007 (sys)
4.0 – 6.0	0.611 +0.052,-0.053 (stat) ± 0.017 (sys)	0.035 +0.069,-0.068 (stat) ± 0.007 (sys)	-0.219 +0.086,-0.084 (stat) ± 0.008 (sys)	-0.146 +0.077,-0.078 (stat) ± 0.011 (sys)	0.025 +0.051,-0.052 (stat) ± 0.004 (sys)	-0.016 +0.081,-0.080 (stat) ± 0.004 (sys)	0.167 +0.094,-0.091 (stat) ± 0.004 (sys)	-0.032 ± 0.071 (stat) ± 0.004 (sys)
6.0 – 8.0	0.579 ± 0.046 (stat) ± 0.015 (sys)	-0.042 +0.058,-0.059 (stat) ± 0.011 (sys)	-0.296 +0.063,-0.067 (stat) ± 0.011 (sys)	-0.249 +0.059,-0.060 (stat) ± 0.012 (sys)	0.152 +0.041,-0.040 (stat) ± 0.008 (sys)	-0.047 +0.068,-0.066 (stat) ± 0.003 (sys)	-0.085 +0.072,-0.070 (stat) ± 0.006 (sys)	-0.024 +0.059,-0.060 (stat) ± 0.005 (sys)
11.0 – 12.5	0.493 +0.049,-0.047 (stat) ± 0.013 (sys)	-0.189 +0.054,-0.058 (stat) ± 0.005 (sys)	-0.283 +0.084,-0.095 (stat) ± 0.009 (sys)	-0.327 +0.076,-0.079 (stat) ± 0.009 (sys)	0.318 +0.044,-0.040 (stat) ± 0.009 (sys)	-0.141 +0.072,-0.074 (stat) ± 0.005 (sys)	-0.007 +0.070,-0.072 (stat) ± 0.005 (sys)	-0.004 +0.070,-0.073 (stat) ± 0.006 (sys)
15.0 – 17.0	0.349 ± 0.039 (stat) ± 0.009 (sys)	-0.142 +0.044,-0.049 (stat) ± 0.007 (sys)	-0.321 +0.055,-0.074 (stat) ± 0.007 (sys)	-0.316 +0.051,-0.057 (stat) ± 0.009 (sys)	0.411 +0.041,-0.037 (stat) ± 0.008 (sys)	0.061 ± 0.058 (stat) ± 0.005 (sys)	0.003 ± 0.061 (stat) ± 0.003 (sys)	-0.019 +0.054,-0.056 (stat) ± 0.004 (sys)
17.0 – 19.0	0.354 +0.049,-0.048 (stat) ± 0.025 (sys)	-0.188 +0.074,-0.084 (stat) ± 0.017 (sys)	-0.266 +0.063,-0.072 (stat) ± 0.010 (sys)	-0.323 +0.063,-0.072 (stat) ± 0.009 (sys)	0.305 +0.049,-0.048 (stat) ± 0.013 (sys)	0.044 +0.073,-0.072 (stat) ± 0.013 (sys)	0.013 +0.071,-0.070 (stat) ± 0.005 (sys)	-0.094 +0.065,-0.067 (stat) ± 0.004 (sys)
	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot	Plot SelectPlot

Correlation tables

- ⇛ The main problem was the correlation tables...
- ⇛ We had around 80 of them...

	F_{L}	S_3	S_4	S_5	A_{FB}	S_7	S_8	S_9
F_{L}	1.00	0.06	0.00	0.03	0.04	−0.02	0.07	0.08
S_3		1.00	0.01	0.10	−0.00	−0.07	−0.01	−0.03
S_4			1.00	0.08	0.11	−0.00	0.07	0.02
S_5				1.00	0.05	−0.01	0.00	0.04
A_{FB}					1.00	0.03	−0.07	0.02
S_7						1.00	0.01	0.11
S_8							1.00	0.02
S_9								1.00

Correlation tables

- ⇛ The main problem was the correlation tables...
- ⇛ We had around 80 of them...

q^2 = M**2(<MU+ MU->)	0.1-0.98 GeV^2							
RE	P P --> B0 < K*(892) < K+ PI- > MU+ MU- > X							
SQRT(S)	7000.0 GeV							
SQRT(S)	8000.0 GeV							
CORR	F_L	S_3	S_4	S_5	A_{FB}	S_7	S_8	S_9
F_L	1.00	0.06	0.00	0.03	0.04	-0.02	0.07	0.08
S_3	0.06	1.00	0.01	0.10	0.00	-0.07	-0.01	-0.03
S_4	0.00	0.01	1.00	0.08	0.11	0.00	0.07	0.02
S_5	0.03	0.10	0.08	1.00	0.05	-0.01	0.00	0.04
A_{FB}	0.04	0.00	0.11	0.05	1.00	0.03	-0.07	0.02
S_7	-0.02	-0.07	0.00	-0.01	0.03	1.00	0.01	0.11
S_8	0.07	-0.01	0.07	0.00	-0.07	0.01	1.00	0.02
S_9	0.08	-0.03	0.02	0.04	0.02	0.11	0.02	1.00

HepData format

⇛ So HepData format is completely different then tex.

```
*dataset:
*location: Appendix C
*dscomment: Likelihood correlation matrix $0.1 < q^2 < 0.98~{\rm GeV}^2/c^4$.
*qual: $q^2$ = M**2(<MU+ MU->) IN GEV**2: 0.1 TO 0.98
*reackey: P P --> B0 + X
*obskey: CORR
*qual: RE : P P --> B0 < K*(892) < K+ PI- > MU+ MU- > X
*qual: SQRT(S) IN GEV : 7000.0
*qual: SQRT(S) IN GEV : 8000.0
*yheader: $F_{\rm L}$ : $S_3$ : $S_4$ : $S_5$ : $A_{\rm FB}$ : $S_7$ : $S_8$ : $S_9$
*xheader: CORR
*data: x : y : y : y : y : y : y : y : y
 $F_{\rm L}$; 1.00; 0.06; 0.00; 0.03; 0.04; -0.02; 0.07; 0.08;
 $S_3$; 0.06; 1.00; 0.01; 0.10; 0.00; -0.07; -0.01; -0.03;
 $S_4$; 0.00; 0.01; 1.00; 0.08; 0.11; 0.00; 0.07; 0.02;
 $S_5$; 0.03; 0.10; 0.08; 1.00; 0.05; -0.01; 0.00; 0.04;
 $A_{\rm FB}$; 0.04; 0.00; 0.11; 0.05; 1.00; 0.03; -0.07; 0.02;
 $S_7$; -0.02; -0.07; 0.00; -0.01; 0.03; 1.00; 0.01; 0.11;
 $S_8$; 0.07; -0.01; 0.07; 0.00; -0.07; 0.01; 1.00; 0.02;
 $S_9$; 0.08; -0.03; 0.02; 0.04; 0.02; 0.11; 0.02; 1.00;
*dataend:
```

My scripts

⇛ Written in python.

⇛ Will briefly go through the some of it's functions.

```
class Table:
    def __init__(self, header, columns, matrix):
        self.header = header
        self.columns = columns
        self.matrix = matrix
        self.matrix_fix = []
        self.table = ''

    def get(self, row, column):
        if row > column:
            return self.get(column, row)
        return self.matrix[row][column - row]

    def get_row(self, row):
        return self.matrix[row]

    def __str__(self):
        return self.__unicode__()

    def fixmatrix(self):
        print self.matrix
        self.matrix_fix = self.matrix
        for i in range(0, len(self.matrix)):
            print self.matrix[i]
            for j in range(0, len(self.matrix[i])):
                if (self.matrix[i][j] == ' ' or self.matrix[i][j] == '    ' or self.matrix[i][j] == '    ' or self.matrix[i][j] == '
                    self.matrix_fix[i][j] = get_ride_of_latex(self.matrix[j - 1][i + 1])
                    #print(i,j, self.matrix[j-1][i+1])
                else:
                    #print(i,j, get_ride_of_latex(self.matrix[i][j]))
                    self.matrix_fix[i][j] = get_ride_of_latex(self.matrix[i][j])
```

My scripts

⇛ Written in python.

⇛ Will briefly go through the some of it's functions.

```
def get_string(self):
    s = '*data:  x : '
    for i in range(1, len(self.matrix_fix[0]) - 1):
        s += "y :"
    s += " y "
    self.table += s + '\n'

    for i in range(0, len(self.matrix_fix)):
        s = ''
        s += self.columns[i + 1] + " : "
        for j in range(1, len(self.matrix_fix[i])):
            #print('a',self.matrix_fix[j-1][i+1])
            s += get_ride_of_latex(self.matrix_fix[j - 1][i + 1]) + ' : '
        self.table += s + '\n'
        # self.table+= \

    print("TEST: ", self.matrix_fix)
def __unicode__(self):
    dupa=":".join(self.columns)
    dupa=dupa.replace(': ', '', 1)
    #print(dupa)
```

My scripts

⇛ Written in python.

⇛ Will briefly go through the some of it's functions.

```
        print("TEST: ", self.matrix_fix)
    def __unicode__(self):
        dupa=":".join(self.columns)
        dupa=dupa.replace(': ', '', 1)
        #print(dupa)


        return """*dataset:
*location: Appendix G
*dscomment: %s
*yheader: %s

*data CHANGEME
%s
*dataend:
""" % (self.header, dupa, self.table)


class TableBuilder:
    def __init__(self):
        self.header = None
        self.columns = None
        self.matrix = []
        self.matrix_fix = []

    def add_row(self, row):
        self.matrix.append(row)

    def build(self):
        return Table(self.header, self.columns, self.matrix)

    def parse_table_header(self, line):
```

My scripts

- ⇛ Written in python.
- ⇛ Will briefly go through the some of it's functions.

```
def is_hline(self, line):
    return line.startswith(r'\hline')

def is_table_header(self, line):
    return line.startswith(r'\subsection')

def is_table_begin(self, line):
    return line.startswith(r'\begin{tabular}')

def is_table_end(self, line):
    return line.startswith(r'\end{tabular}')

def build(self):
    self.state = ParserStateEnum.DONE
    self.table = self.builder.build()
```

My scripts

⇛ Written in python.

⇛ Will briefly go through the some of it's functions.

```
*dataset:
*location: Appendix G
*dscomment: None
*yheader:   $A_{3}$ : $A_{4}$ : $A_{5}$ : $A_{6s}$ : $A_{7}$ : $A_{8}$ : $A_{9}$

*data CHANGEME
*data:  x : y :y :y :y :y :y : y
  $A_{3}$  :  1.00   :   -0.12   :   -0.18   :   0.00   :   0.01   :   0.01   :   -0.05  :
 $A_{4}$  :   -0.12   :   1.00   :   0.26   :   -0.14   :   0.02   :   -0.08   :   0.03  :
 $A_{5}$  :   -0.18   :   0.26   :  1.00   :   -0.13   :   -0.09   :   0.02   :   0.07  :
 $A_{6s}$  :   0.00   :   -0.14   :   -0.13   :   1.00   :   0.0   :   0.01   :   -0.01  :
 $A_{7}$  :   0.01   :   0.02   :   -0.09   :   0.0   :   1.00   :   0.14   :   -0.15  :
 $A_{8}$  :   0.01   :   -0.08   :   0.02   :   0.01   :   0.14   :   1.00   :   -0.07  :
 $A_{9}$ :   -0.05   :   0.03  :   0.07  :   -0.01  :   -0.15  :   -0.07  :   1.00  :

*dataend:
```

User case

⇛ There will be things that need to be changed for each table:

```
def get_ride_of_latex(line):
    line2 = line
    line2 = line2.replace(r"\\rm", r"\rm")
    line2 = line2.replace(r"\\", "")
    line2 = line2.replace(r'$', '')
    line2 = line2.replace(r'\ ', '')
    line2 = line2.replace(r'\ ', '')
    #line2 = line2.replace(r'\gevgevcccc', 'GeV/c^4')
    return line2
```

Error encoding

⇛ If you need to encode errors in HepData format:

```
*yheader: $F_{\rm L}$ : $S_{3}$ : $S_{4}$ : $S_{5}$ : $A_{\rm FB}$ : $S_{7}$ : $S_{8}$ : $S_{9}$
*data: x : y : y : y : y : y : y : y : y
0.10 TO 0.98 ;    0.263 +0.045,-0.044(DSYS=0.017) ;       -0.036 +0.063,-0.063(DSYS=0.005) ;
+0.058,-0.057(DSYS=0.009) ;       0.015 +0.059,-0.059(DSYS=0.006) ;       0.079 +0.076,-0.075(DSYS
1.1 TO 2.5;       0.660 +0.083,-0.077(DSYS=0.022) ;       -0.077 +0.087,-0.105(DSYS=0.005) ;
+0.068,-0.080(DSYS=0.012) ;       -0.219 +0.094,-0.104(DSYS=0.004) ;      -0.098 +0.108,-0.123(DSY
2.5 TO 4.0  ;     0.876 +0.109,-0.097(DSYS=0.017)  ;      0.035 +0.098,-0.089(DSYS=0.007)  ;
+0.082,-0.090(DSYS=0.007)  ;     0.068 +0.120,-0.112(DSYS=0.005)  ;      0.030 +0.129,-0.131(DSYS=
4.0 TO 6.0  ;     0.611 +0.052,-0.053(DSYS=0.017) ;       0.035 +0.069,-0.068(DSYS=0.007) ;
+0.051,-0.052(DSYS=0.004) ;       -0.016 +0.081,-0.080(DSYS=0.004) ;      0.167 +0.094,-0.091(DSYS
6.0 TO 8.0 ;      0.579 +0.046,-0.046(DSYS=0.015) ;       -0.042 +0.058,-0.059(DSYS=0.011) ;
```

Summary

⇛ HepData is a tricky format that unfortunately is not similar to other ones :(

⇛ I am afraid that everyone has to adjust those scripts for they tables.

⇛ If you want to put the results to HepData do it with the paper submission!

⇛ If you already produced some results and want to copy them from tex files to Hepdata you can start with my scripts:
`http://www.physik.uzh.ch/~mchrzasz/HepData/KstarMuMu/`

Backup