

Parameter Estimation of a Negative Binomial for the By-product of the Grignard Reaction

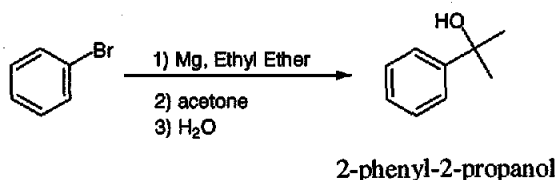
Ling Huang, Sacramento City College
Paul Johnson, University of California San Francisco

Abstract

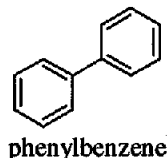
The Grignard reaction is a common undergraduate organic laboratory experiment. At Sacramento City College this reaction is taught within the organic chemistry laboratory to students. It is taught to illustrate an example of organic synthesis. The product produced by way of the reaction is 2-phenyl-2-propanol. Occasionally there is a by-product product. This is phenylbenzene. A negative binomial model can be used since the average by-product is unlikely for the Grignard reaction (i.e., this is the case when no by-product is a common occurrence). The parameters of the negative binomial are estimated using the method of moments. An improved estimate of the sample mean is obtained from use of the bootstrap. The SAS[®] macros listed below are available from the authors at <http://pages.prodigy.net/johnsonp12/grignard.html>. The macros (G1 and G2) require base SAS (SAS, 1990) and SAS/STAT (SAS, 1993) software to run.

Procedure

The Grignard reaction (Bell et al., 1997):



By-product



The students that perform the experiments are typically taking their second semester organic class. They generally have already studied the organometallic chemistry in the lecture and have some ideas of precaution and possible side products. The main focus of the Grignard reaction in this experiment is to prepare phenyl magnesium bromide from phenyl bromide in absolute ethyl ether.

The resulting phenyl magnesium bromide solution is a grayish solution. Reagent grade acetone is added to this solution and 2-phenyl-2-propanol is isolated. The students are then asked to verify the percentage yield and the purity of their product by Gas Chromatography. Most students obtain the desired product in high yield and purity, but a small number of students have low yields and purity. Students are asked to analyze their results and draw conclusions. In a simulation study the following numbers of students obtained by-products. The number of students with the level of by-product (as a percentage of yield) is listed (see Table 1).

Table 1. Frequency distribution of by-product.

% of by-product	ref	# of students
≤ 1	0	372
1 – 2.5	1	16
2.5 – 5	2	12
5 – 7.5	3	5
7.5 – 10	4	3
10 – 12.5	5	3
12.5 – 15	6	1
15 – 17.5	7	2
17.5 – 20	8	1
20 – 22.5	9	0
22.5 – 25	10	2
25 – 27.5	11	1
27.5 – 30	12	1
30 – 32.5	13	0
32.5 – 35	14	0
35 – 37.5	15	0
37 – 40	16	1
≥ 40	17	0

The first SAS subroutine is listed below and is called G1. The program estimates the parameters of a negative binomial model of by-product (Hilborn and Mangel, 1997). The parameters, m and k , are estimated. The data consists of the reference number ('ref') of by-product, c_i , for the i 'th student. The SAS input data consists of this reference number of by-product = 'nb' and the number of students with this level of by-product = 'nh'.

The sample mean number of by-product is calculated (mean), together with its variance (var).

The parameters of the negative binomial are estimated using the method of moments:

$$\hat{m} = \text{mean and } \hat{k} = \frac{\hat{m}^2}{\text{var} - \hat{m}}.$$

A $100(1-\alpha)\%$ confidence interval for m is found:

$$\hat{m} \pm t_{1-\alpha/2, df} * \frac{\text{std}}{\sum nh}, \text{ with } df = (\sum nh) - 1.$$

The estimates for m and k are then used to find probability estimates for the probability that the by-product for the i 'th student, C_i equals c (see Table 2).

$$\Pr\{C_i = c\} = \frac{\Gamma(k+c)}{\Gamma(k)c!} \left(\frac{k}{k+m} \right)^k \left(\frac{m}{m+k} \right)^c.$$

Since c is integer valued we have:

$$\Pr\{C_i = c\} = \frac{\Gamma(k+c)}{\Gamma(k)\Gamma(c+1)} \left(\frac{k}{k+m} \right)^k \left(\frac{m}{m+k} \right)^c.$$

Bootstrap Estimates

The second SAS subroutine is listed below and is called G2. Intervals for the estimates of the mean will be constructed using the nonparametric bootstrap (Efron, 1979). Suppose the parameter of interest is Θ . A random sample, $x_1, x_2, \dots, x_n$, is taken from the population, from which an estimate $\hat{\Theta}$, of Θ , is calculated. The idea behind bootstrapping is that this sample can then be used repeatedly to find out properties (such as bias and variance) of some given estimator of Θ . The original sample of n observed values, each with probability $1/n$, is resampled. The sampling is done with replacement. $\hat{\Theta}_1^*$ is calculated from this new sample. Resampling with replacement again from the original sample generates another sample. $\hat{\Theta}_2^*$ is calculated. The procedure is repeated B times, resulting in B individual estimates of Θ .

The bootstrap estimate (Efron, 1979) of bias is calculated by subtracting the simple estimate of Θ from the overall bootstrap estimate obtained by averaging the B estimates, $\hat{\Theta}_i^*$

$$(i = 1, \dots, B), \text{ i.e., } \text{BIAS}_\Theta^* = \bar{\hat{\Theta}}^* - \hat{\Theta}$$

$$= \frac{\sum_{i=1}^B \hat{\Theta}_i^*}{B} - \hat{\Theta}.$$

An improved estimate of Θ is found by subtracting this bias from the simple estimate of Θ :

$$\hat{\Theta}^I = \hat{\Theta} - \left(\bar{\hat{\Theta}}^* - \hat{\Theta} \right) = 2\hat{\Theta} - \bar{\hat{\Theta}}^*.$$

Hence an improved estimate of m is:

$$m^I = 2\hat{m} - m^*.$$

A $100(1-\alpha)\%$ Bootstrap confidence interval for m is found by first sorting the B bootstraps of m . The $B * (\alpha/2)^{\text{th}}$ observation is the lower bound and the $B * (1 - (\alpha/2))^{\text{th}}$ observation is the upper bound.

Results

Mean = 0.4024; Var = 2.5847; Std = 1.6077

Moment Estimators for m and k :

$$\hat{m} = 0.4024 \text{ and } \hat{k} = 0.0742.$$

A 95% confidence interval for m is (0.2482, 0.5566).

Table 2. Probability that the by-product for the i^{th} student = c i.e., $P(C_i = c)$

c	$P(C_i = c)$
0	0.87110
1	0.05457
2	0.02475
3	0.01445
4	0.00937
5	0.00645
6	0.00460
7	0.00337
8	0.00252
9	0.00191
10	0.00146
11	0.00113
12	0.00088
13	0.00069
14	0.00054
15	0.00043
16	0.00034
17	0.00027

$$m^* = 0.4004 \text{ and } m^I = 0.4044.$$

A 95% Bootstrap confidence interval for m is (0.2643, 0.5476). The Bootstrap provides for a tighter confidence interval for the true mean, m .

Program Name: G1
Estimating Parameters of a Negative Binomial Model of
By-product.

```

data cboot1;scan: set bb1 end=last;n+1;
  if not last then goto scan; do j=1 to n;
    seed=floor(&i+1000000000*(sqrt(time())-
    floor(sqrt(time())))); k=ceil(ranuni(seed)*n);
    set bb1 point=k; if _error_ then abort;
    output; end; stop;

data aa1;set cboot1; data aa1;set aa1;ind = 1;

proc sort data = aa1;by nb;

proc means data = aa1 noprint;by nb;var ind;
output out = a sum = nh;

data a;set a;keep nb nh;
data a;set a; totc = nb*nh;

proc means data = a noprint;var nh totc;
output out = b sum = sum_nh sum_totc;

data b&i;set b;m=sum_totc/sum_nh;

%end;

%do j = 2 %to &n_boot;
  proc append base = b1 data = b&j; %end;

%mend boot;
%boot

proc sort data = b1;by m;
proc print data = b1;var m;
Title1 'Print out of Bootstraps for Confidence
Interval for m';

proc means data = b1 noprint; var m;
output out = c mean = mean_m;

proc print data = c;var mean_m;
Title1 'Mean of the m Bootstraps';

data bb1;set bb1;ind = 1;
proc sort data = bb1;by nb;

proc means data = bb1 noprint;by nb;var ind;
output out = bb2 sum = nh;

data bb2;set bb2;keep nb nh;
data bb2;set bb2; totc = nb*nh;

proc means data = bb2 noprint;var nh totc;
output out = bb3 sum = sum_nh sum_totc;

data bb3;set bb3;mhat=sum_totc/sum_nh;
proc print data = bb3;var mhat;
Title1 'Simple Estimate for m (mhat)';

```

```

data c;set c;dr = 'a'; data bb3;set bb3;dr = 'a';

data d;merge bb3 c;by dr;

data d;set d;m_i = 2*mhat-mean_m;
m_boot = mean_m;

proc print data = d; var mhat m_boot m_i;
Title1 'Simple Estimate, Bootstrap and Improved
Estimate for m';

%mend negbin;

%negbin

run;

```

References

- Bell, C. E., Clark, A. K., Taber, D. F., and Rodig, O. R. (1997). *Organic Chemistry Laboratory: Standard & Microscale Experiments*. New York: Harcourt Brace College Publishers.
- Efron, B. (1979). "Bootstrap methods: Another look at the Jackknife," *Annals of Statistics*, 7, 1-26.
- Hilborn, R., and Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton, New Jersey: Princeton University Press.
- SAS Institute Inc. (1990). *SAS Procedures Guide: Basics, Version 6, 3rd ed.*, Cary, NC: SAS Institute Inc.
- SAS Institute, Inc. (1993). *SAS/STAT User's Guide, Version 6, 4th ed., Volume 1 and Volume 2*, Cary, NC: SAS Institute Inc.

Acknowledgments

SAS and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Contact Information

Ling Huang,
Sacramento City College, Science and Allied Health,
3835 Freeport Blvd., Sacramento CA 95822-1386
E-Mail: HuangL@scc.losrios.edu

Paul Johnson,
P.O. Box 4146, Davis CA 95617-4146
E-Mail: P.Johnson@prodigy.net
Fax: (530) 280-4246