

Оценивание констант сополимеризации в регрессионной модели с ошибкой Берксона

А. Ю. ТИМОФЕЕВА

Новосибирский государственный технический университет, Новосибирск, Россия
Контактный автор: Тимофеева Анастасия Ю., e-mail: a.timofeeva@corp.nstu.ru

Поступила 10 марта 2020 г., доработана 24 марта 2020 г., принята в печать 10 апреля 2020 г.

Рассматривается проблема оценки относительной активности мономеров на основе дифференциального уравнения сополимеризации. Обосновано включение в модель погрешности измерения входного признака в виде ошибки Берксона. Предложен алгоритм одновременного оценивания констант сополимеризации и дисперсий ошибок с помощью метода максимального правдоподобия. На примере сополимеризации виниловых эфиров произведено сравнение разных методов оценивания констант сополимеризации. Показано, что метод на основе симметричных уравнений дает некорректные результаты. Результаты оценивания с помощью предложенного алгоритма наиболее близки к оценкам, полученным по нелинейному методу наименьших квадратов.

Ключевые слова: сополимеризация, отношение активности мономеров, регрессия, ошибка Берксона, метод максимального правдоподобия.

Цитирование: Тимофеева А.Ю. Оценивание констант сополимеризации в регрессионной модели с ошибкой Берксона. Вычислительные технологии. 2020; 25(3): 18–28.

Введение и постановка задачи

В химической промышленности для определения условий совместной полимеризации двух мономеров с целью синтеза высокомолекулярных соединений заданного состава и строения используется уравнение бинарной сополимеризации в дифференциальной форме [1]. При малой глубине превращения (менее 5 %) обычно исходят из допущения, что при незначительном изменении в ходе реакции концентрации мономеров и состава мономерной смеси выполняется равенство “мгновенных” составов. Тогда дифференциальное уравнение сополимеризации можно представить в следующем виде:

$$\frac{[m_1]}{[m_2]} = \frac{[M_1] r_1 [M_1] + [M_2]}{[M_2] r_2 [M_2] + [M_1]}, \quad (1)$$

где $[M_1]$, $[M_2]$ — мольные концентрации мономеров первого и второго типов M_1 , M_2 ; $[m_1]$, $[m_2]$ — содержание мономерных звеньев M_1 , M_2 в сополимере, выражаемое в тех же единицах, что и концентрации мономеров $[M_1]$, $[M_2]$ в исходной смеси (в мольных долях, мол. % или в молях), r_1 , r_2 — константы сополимеризации, описывающие относительные активности мономеров. Для оценки этих констант стандартно используются графические и аналитические методы, в том числе метод наименьших квадратов (МНК).

Однако эти методы не учитывают погрешности измерения входных переменных, поэтому в течение последних трех десятилетий многие авторы [2–4] отмечают статистическую некорректность большинства методов вычисления констант сополимеризации. Предложенные ими методы оценивания параметров уравнения сополимеризации, учитывающие погрешности измерения входных переменных, имеют существенное ограничение: необходимо иметь дополнительную информацию о величине (дисперсии) погрешностей. Это несколько осложняет их применение на практике. Поэтому далее предлагается новый подход, не требующий привлечения дополнительной информации.

1. Методы оценивания констант сополимеризации

Для простоты вычислений большинство аналитических методов предполагают представление уравнения (1) в виде линейного по параметрам соотношения. Такой форме соответствует наиболее часто используемое уравнение Файнмана — Росса

$$y = r_2 + r_1 x, \quad (2)$$

где $y = \frac{[M_1]}{[M_2]} \left(\frac{[m_2]}{[m_1]} - 1 \right)$, $x = - \left(\frac{[M_1]}{[M_2]} \right)^2 \frac{[m_2]}{[m_1]}$. С помощью МНК оцениваются константа и угловой коэффициент прямой линии (2).

Метод Келена и Тьюдоша похож на метод Файнмана и Росса, но учитывает степень разброса экспериментальных данных с помощью фактора α :

$$\alpha = \sqrt{\left(\min_i \frac{Q_i^2}{q_i} \max_i \frac{Q_i^2}{q_i} \right)}, \quad (3)$$

где $Q_i = \frac{[M_1]_i}{[M_2]_i}$, $q_i = \frac{[m_1]_i}{[m_2]_i}$, $[M_1]_i$, $[M_2]_i$, $[m_1]_i$, $[m_2]_i$ — значения концентраций мономеров $[M_1]$, $[M_2]$, $[m_1]$, $[m_2]$ в i -м эксперименте. Тем самым показатель (3) характеризует экстремальные значения соотношений концентраций мономеров в опытных данных. Далее по данным экспериментов вычисляются величины

$$\varsigma_i = \left(1 + \alpha \frac{q_i}{Q_i^2} \right)^{-1}, \quad \pi_i = \frac{Q_i(q_i - 1)}{\alpha q_i + Q_i^2},$$

на основе которых строится уравнение

$$\pi_i = -\frac{r_2}{\alpha} + \left(r_1 + \frac{r_2}{\alpha} \right) \varsigma_i. \quad (4)$$

Если оценить неизвестные параметры $\theta_0 = -\frac{r_2}{\alpha}$, $\theta_1 = r_1 + \frac{r_2}{\alpha}$ с помощью МНК, то через них легко выразить константы сополимеризации.

Если сравнивать два подхода — оценивание уравнения (2) и уравнения (4), то второй имеет тот недостаток, что оценки констант сополимеризации не оцениваются непосредственно, а выражены как некоторые функции оценок параметров θ_0, θ_1 . Поэтому если в первом подходе можно сразу вычислить стандартные ошибки оценок относительных активностей мономеров, то во втором необходимы дополнительные вычисления для

определения погрешностей оценивания. Кроме того, r_2 выражается через фактор α , вычисляемый в соответствии с (3) по экспериментальным данным, что вносит дополнительные погрешности в оценку этой константы сополимеризации, а также усложняет определение ее стандартной ошибки.

В целом применение МНК для оценивания уравнений (2) и (4) имеет ряд негативных последствий. Хорошие свойства оценок МНК (в частности, эффективность) доказываются в предположении бесконечного объема выборки. В реальности предполагается, что выборки достаточно большие. Однако в практических исследованиях синтеза высокомолекулярных соединений производится лишь небольшое число экспериментов в одних и тех же начальных условиях с разным мольным составом мономеров. Например, в работе [5] приведены результаты пяти опытов, в авторефератах диссертаций [6] и [7] представлены данные по трем — шести и трем — пяти экспериментам. Наконец, в исследовании [8] было проведено 10 экспериментов, что считается довольно существенным. Хотя применение МНК допустимо в том случае, если число опытов превышает число оцениваемых параметров, но тем не менее это затрудняет статистическую процедуру оценки качества построенной модели. Кроме того, ограниченный объем данных не позволяет применять многие современные методы статистического анализа, обрабатывающие большие выборки, учитывающие повторные наблюдения и т. п.

Основной проблемой представления уравнений сополимеризации в виде (2) или (4) является то, что зависимые и независимые переменные рассчитываются на основе одних и тех же мольных концентраций. Это приводит к тому, что как в правой, так и в левой частях уравнений фигурируют одни и те же погрешности измерения. Следовательно, кроме того что в принципе независимая переменная наблюдается с ошибкой, эта ошибка еще и коррелирует с ошибкой измерения зависимой переменной.

Несколько ослабляет эту проблему использование нелинейных соотношений. Например, дифференциальное уравнение (1) может быть представлено в виде

$$Y = \frac{r_2 X_2 + X(1 - X)}{r_1(1 - X)^2 + 2X(1 - X) + r_2 X^2}, \quad (5)$$

где $Y = [m_2]$, $X = [M_2]$. При построении уравнения (5) учитываются балансовые соотношения $[m_1] + [m_2] = 1$, $[M_1] + [M_2] = 1$. Для оценивания параметров уравнения (5) используется нелинейный МНК. Недостатком является то, что в отличие от обычного МНК решение задачи оценивания нельзя представить в аналитическом виде, необходимо использовать численные методы оптимизации [9]. В качестве начального приближения могут быть использованы оценки, полученные по МНК описанными выше способами.

При применении МНК предполагается, что переменная с левой стороны уравнения наблюдается с погрешностями, тогда как переменная в правой части считается независимой, контролируемой и наблюдаемой без искажения [10]. О том, что мольные концентрации наблюдаются с некоторыми погрешностями, свидетельствует тот факт [1, с. 19], что значения параметров, оцененных по несимметричным уравнениям, будут различаться, иногда довольно сильно. Это значит, что если оценивать обратную регрессию, т. е. от уравнения (2) переходить к

$$x = -\frac{r_2}{r_1} + \frac{1}{r_1}y, \quad (6)$$

то оценки констант сополимеризации, выраженные через константу и угловой коэффициент уравнения (6), оцененного по МНК, будут другими.

Частично эта проблема ослабляется при использовании аналитического метода расчета по симметричным уравнениям, предложенного в [11]. Авторы представляют уравнение (2) и соответствующее ему уравнение (6) несколько в другой форме:

$$x - \frac{1}{k} = \frac{x}{k}r_1 - r_2, \quad \frac{1}{x} - k = \frac{k}{x}r_2 - r_1, \quad (7)$$

где $k = \frac{[m_1][M_2]}{[m_2][M_1]}$, $x = \frac{[M_1]}{[M_2]}$. При применении первого уравнения более точно оценивается r_1 , а при применении второго — r_2 . В тех случаях, когда трудно сделать обоснованные предположения об ошибке измерения, наиболее рационально считать уравнения (7) равносильными и при расчетах констант сополимеризации учитывать их с равным весом. Для этого в [11] предлагается перейти от (7) к симметричному уравнению

$$\sqrt{kx} - \frac{1}{\sqrt{kx}} = r_1\sqrt{\frac{x}{k}} - r_2\sqrt{\frac{k}{x}}. \quad (8)$$

Использование обычного МНК для оценивания уравнения (8) приводит к простым выражениям для вычисления оценок [11]. Очевидно, что этот способ не решает проблемы коррелированности погрешностей измерения левой и правой частей уравнения, поскольку в обоих случаях фигурируют одни и те же величины. Кроме того, теперь регрессоры очень сильно коррелируют, между ними нелинейная функциональная связь. Это порождает новые проблемы, связанные с мультиколлинеарностью [12].

В настоящее время наибольшее развитие получили методы оценивания констант сополимеризации, рассматривающие уравнения сополимеризации как модель с ошибками в переменных [3, 4]. Эти работы основаны на постановке с классической ошибкой измерения [10]. В таком случае для оценивания модели разработаны специальные методы, в частности, можно использовать метод максимального правдоподобия (ММП). В работе [10] выписана функция правдоподобия и показано, что для оценивания в предположении нормального распределения требуется задавать соотношение дисперсий ошибок. Обычно исходят из соотношения, равного единице, предполагая, что вариация погрешностей одинакова как для входного, так и для выходного признаков. Это, однако, может быть не так.

В целом, далее поставлена под вопрос постановка, в которой уравнение сополимеризации рассматривается как модель с классической ошибкой. При таких предположениях входная концентрация мономеров была бы случайной величиной, варьирующейся около истинного значения. Однако на практике обычно исследователь может зафиксировать исходный состав смеси. В частности, на этом основаны рекомендации по оптимальному составу исходной смеси, обеспечивающему наиболее точное оценивание констант сополимеризации. Такой оптимальный состав получается на основе методов теории планирования эксперимента [13, 14].

2. Алгоритм оценивания констант сополимеризации без привлечения дополнительной информации

Анализ результатов экспериментов по сополимеризации (см., например, [15]) показывает, что исследователь задает соотношение мономеров в исходной смеси, в том числе

для разных экспериментов такое соотношение может быть одинаковым, тогда как соотношение мономеров в сополимере несколько отличается в силу погрешностей. Это наиболее соответствует постановке с ошибкой Берксона [10]. В этом случае мы имеем дело с активным экспериментом, когда исследователь задает значения входного фактора x_i в ходе i -го опыта. Эти величины являются детерминированными. Однако когда происходит реакция сополимеризации, значение исходной концентрации несколько искажается, поэтому на выходной фактор воздействует случайная величина:

$$X_i = x_i + \delta_i. \quad (9)$$

Для того чтобы оценить модель (5) в предположении ошибки Берксона, воспользуемся ММП. В этом случае наблюдаемой является единственная случайная величина η_i — полученная в ходе i -го эксперимента мольная концентрация мономера M_2 в сополимере, имеющая конкретную реализацию y_i . Предположим, что распределение случайных ошибок, возникающих в эксперименте, нормальное с нулевым математическим ожиданием. В соответствии с уравнением сополимеризации (5) и в предположении, что погрешность измерения исходного состава смеси мономеров входит в соотношение (9) аддитивно, получаем следующее представление случайной величины η_i :

$$\eta_i = \zeta_i + \varepsilon_i, \quad (10)$$

$$\zeta_i = \frac{r_2(x_i + \delta_i)^2 + (x_i + \delta_i)(1 - (x_i + \delta_i))}{r_1(1 - (x_i + \delta_i))^2 + 2(x_i + \delta_i)(1 - (x_i + \delta_i)) + r_2(x_i + \delta_i)^2}, \quad (11)$$

где случайными величинами являются ошибки δ_i, ε_i ; величина x_i — детерминированная, т. е. рассматривается как заданное число. Обозначим функцию плотности распределения случайной величины η_i в точке y_i как $f_{\eta_i}(y_i) = f_{\eta_i}(y_i, x_i, r_1, r_2, \sigma_\delta^2, \sigma_\varepsilon^2)$. Она зависит от x_i , констант сополимеризации r_1 и r_2 , а также от параметров распределения (дисперсий) ошибок $\sigma_\delta^2, \sigma_\varepsilon^2$. В предположении независимости ошибок в разных экспериментах логарифмическая функция правдоподобия имеет вид

$$\log L = \sum_{i=1}^n \log f_{\eta_i}(y_i; x_i, r_1, r_2, \sigma_\delta^2, \sigma_\varepsilon^2). \quad (12)$$

Задача оценивания неизвестных параметров $r_1, r_2, \sigma_\delta^2, \sigma_\varepsilon^2$ сводится к максимизации функции правдоподобия (12). Для решения этой задачи необходимо получить функцию плотности распределения случайной величины η_i .

В силу независимости случайных ошибок распределение случайной величины (10) представляет собой свертку распределений величин ζ_i, ε_i , которую можно выразить по известной формуле свертки:

$$f_{\eta_i}(y_i) = (f_{\zeta_i} * f_\varepsilon)(y_i) = \int_{-\infty}^{+\infty} f_{\zeta_i}(y_i - t) f_\varepsilon(t) dt, \quad (13)$$

где f_{ζ_i} — функция плотности случайной величины ζ_i ; f_ε — функция плотности нормального распределения с нулевым параметром сдвига и дисперсией, равной σ_ε^2 . Необходимо получить выражение для функции плотности f_{ζ_i} . Рассмотрим случайную величину ζ_i

как функцию от случайной величины δ_i , $\zeta_i = g(\delta_i)$. Для того чтобы восстановить распределение от функции случайной величины, нужно получить обратную функцию g^{-1} . В силу того что выражение в числителе (11) является полиномом, g^{-1} в общем случае неоднозначна. Здесь имеет место полином второй степени, следовательно, будет два корня квадратного уравнения вида

$$\frac{a_1x^2 + b_1x + c_1}{a_2x^2 + b_2x + c_2} - u = 0,$$

где $a_1 = r_2 - 1$, $b_1 = 2x_i(r_2 - 1) + 1$, $c_1 = r_2x_i^2 + x_i(1 - x_i)$, $a_2 = r_2 + r_1 - 2$, $b_2 = 2x_i(r_1 + r_2 - 2) + 2(1 - r_1)$, $c_2 = r_1(1 - x_i)^2 + 2x_i(1 - x_i) + r_2x_i^2$, u — аргумент функции плотности. После приведения к общему знаменателю в числителе получаем квадратное уравнение. Для поиска корней квадратного уравнения находим дискриминант

$$D = (b_1 - b_2u)^2 - 4(a_1 - a_2u)(c_1 - c_2u).$$

Тогда имеем два корня уравнения:

$$v_1(u) = \frac{-(b_1 - b_2u) - \sqrt{D}}{2(a_1 - a_2u)}, \quad v_2(u) = \frac{-(b_1 - b_2u) + \sqrt{D}}{2(a_1 - a_2u)}.$$

Тем самым мы получили обратные функции g^{-1} . Далее функция плотности величины $\zeta_i = g(\delta_i)$ выражается следующим образом:

$$f_{\zeta_i}(u) = f_{\delta}(v_1(u))v_1'(u) + f_{\delta}(v_2(u))v_2'(u), \quad (14)$$

где f_{δ} — функция плотности нормального распределения с нулевым параметром сдвига и дисперсией, равной σ_{δ}^2 ; v_1' , v_2' — первые производные функций v_1 , v_2 .

В качестве аргумента функции (14) выступает $(y_i - t)$. Результаты расчета подставляются в выражение для свертки (13). В результате находим значение функции плотности случайной величины η_i в точке y_i , которое подставляется в логарифмическую функцию правдоподобия (12), и решается задача ее максимизации. Оптимизация производится численными методами, в частности, использовался метод Нелдера — Мида.

Разработанный алгоритм реализован в виде пользовательской функции в среде статистического программирования R. Производные $v_1'(u)$, $v_2'(u)$ взяты аналитически с помощью функции `D{stats}`. Здесь они не представлены в силу громоздкости выражений. Для расчета функции плотности нормального распределения использовалась стандартная функция `dnorm`. Для расчета интеграла применялись численные методы, в частности самый простой — метод левых прямоугольников. Фиксировалось число значений переменной интегрирования t . Оно составляло 1000. Бесконечный интервал $(-\infty, +\infty)$ заменен на отрезок $[-5\sigma_{\varepsilon}, 5\sigma_{\varepsilon}]$ в силу того, что для нормального распределения в этом отрезке 99.99994% значений случайной величины.

Дополнительно в программной реализации учтены вырожденные случаи: $\sigma_{\delta} = 0$ или $\sigma_{\varepsilon} = 0$. В первом случае $f_{\eta_i}(y_i)$ является плотностью нормального распределения, во втором случае $f_{\eta_i}(y_i) \equiv f_{\zeta_i}(y_i)$.

3. Результаты оценивания констант сополимеризации

Продemonстрируем результаты практического применения известных методов определения констант сополимеризации, а также предложенного алгоритма ММП-оценивания в предположении модели с ошибкой Берксона на примере сополимеризации виниловых эфиров [16, 17]. Для иллюстрации выбрана реакция сополимеризации винилхлорида и винилфенилового эфира. В работе [16] приведены исходные данные для расчета, насчитывающие семь экспериментов. В табл. 1 для i -го эксперимента представлено содержание винилхлорида в исходной смеси $[M_1]_i$ и в сополимере $[m_1]_i$, мол. %, содержание винилфенилового эфира в исходной смеси $[M_2]_i$ и сополимере $[m_2]_i$, мол. %.

Как было отмечено выше, исследователь может задавать одно и то же соотношение, например 40 и 60 %, и получать несколько различающиеся результаты, что характерно для схемы активного эксперимента с ошибками в переменных. При этом ошибка входного фактора, скорее, должна рассматриваться в постановке с ошибкой Берксона.

Оценки \hat{r}_1, \hat{r}_2 констант сополимеризации r_1, r_2 , полученные известными аналитическими методами, обобщены в табл. 2. При оценивании наклонной прямой Файнмана и Росса (2), симметричного уравнения (8) дополнительно рассчитаны стандартные ошибки оценок коэффициентов сополимеризации, которые указаны в скобках.

Процедура оценивания с применением предложенного алгоритма представляет собой нелинейную оптимизацию с помощью численных методов, поэтому результат может зависеть от начального приближения неизвестных параметров. В качестве начальных приближений использовались оценки констант сополимеризации, найденные известными методами и представленные в табл. 2. Это позволило дополнительно проверить стабильность работы алгоритма при разных начальных приближениях.

Т а б л и ц а 1. Результаты экспериментов / Table. 1. The results of experiments

i	$[M_1]_i$	$[m_1]_i$	$[M_2]_i$	$[m_2]_i$
1	0.2	0.4332	0.8	0.5668
2	0.4	0.5855	0.6	0.4145
3	0.4	0.5889	0.6	0.4111
4	0.6	0.7312	0.4	0.2688
5	0.6	0.7171	0.4	0.2829
6	0.8	0.8833	0.2	0.1167
7	0.8	0.8593	0.2	0.1407

Т а б л и ц а 2. Результаты оценивания констант сополимеризации известными методами / Table. 2. The results for the estimation of the copolymerization reactivity ratios by known methods

Метод	\hat{r}_1	\hat{r}_2
Файнмана и Росса	1.514 (0.126)	0.263 (0.171)
Келена и Тьюдоша	1.474	0.237
Файнмана и Росса (обратная регрессия)	1.566	0.316
Нелинейный МНК	1.365	0.203
По симметричным уравнениям	1.478 (0.100)	0.226 (0.057)

Т а б л и ц а 3. Результаты оценивания с помощью предложенного алгоритма
 Table. 3. The results for the estimation using the proposed algorithm

Метод поиска начального приближения	\hat{r}_1	\hat{r}_2	$\hat{\sigma}_\delta$	$\hat{\sigma}_\varepsilon$	$\max \log L$
Файнмана и Росса	1.367	0.204	$2.33 \cdot 10^{-5}$	0.011	-24.819
Келена и Тюдоса	1.356	0.200	$2.72 \cdot 10^{-5}$	0.011	-23.747
Файнмана и Росса (обратная регрессия)	1.356	0.200	$3.09 \cdot 10^{-5}$	0.012	-23.277
Нелинейный МНК	1.354	0.204	$1.96 \cdot 10^{-5}$	0.013	-26.859
По симметричным уравнениям	1.366	0.203	$2.21 \cdot 10^{-5}$	0.011	-25.170

В качестве начального приближения дисперсии ошибок использовалось $\sigma_\delta^2 = 0$. Тем самым априори предполагалось, что входной фактор наблюдается без погрешностей. Начальное приближение для σ_ε^2 оценивалось как средний квадрат отклонений наблюдаемых значений выходного признака в уравнении (5) от рассчитанных по соотношению (5) при заданном начальном приближении констант сополимеризации. Максимальное число итераций метода Нелдера — Мида установлено равным 4000. В табл. 3 представлены все полученные результаты, включая оценки стандартных отклонений ошибок $\hat{\sigma}_\delta, \hat{\sigma}_\varepsilon$. Видно, что они достаточно близки, различия для оценок констант сополимеризации наблюдаются только во втором и третьем знаках после десятичной точки. Кроме оценок в табл. 3 представлены полученные оптимальные значения функции правдоподобия. Это позволяет выбрать наилучший результат. Он достигается при использовании начального приближения констант, оцененных методом Файнмана и Росса по обратной регрессии.

Выводы

Стандартные ошибки оценок констант сополимеризации, полученные методом Файнмана и Росса, достаточно велики (см. табл. 2). Особенно велико значение стандартного отклонения \hat{r}_2 , которое составляет 0.171 при значении самого параметра, равном 0.263. В частности, при проверке гипотезы о значимости параметра статистический вывод заключается в том, что r_2 не отличен от нуля, что несколько обескураживает.

Если оценивать уравнение Файнмана и Росса (2) как обратную регрессию (6), то получим совсем другие значения оценок констант сополимеризации (см. табл. 2). Особенно сильное отличие наблюдается для \hat{r}_2 . Это связано с тем, что константа в уравнениях (2) и (6) оценивается наименее точно.

Оценивание уравнения (5) с помощью нелинейного МНК дает опять же отличные от всех результаты (см. табл. 2). Оценки констант сополимеризации в этом случае получаются самыми низкими. Именно этот подход применен в работе [16]. Там же вычислены погрешности оценивания: для параметра r_1 20 %, а для параметра r_2 40 %.

Аналитический метод расчета по симметричным уравнениям дает меньшие стандартные ошибки по сравнению с методом Файнмана и Росса. При этом оценки констант сополимеризации близки к результатам, полученным по методу Келена и Тюдоса. Однако эти оценки сильно отличаются от полученных на основе предложенного алгоритма (см. табл. 3). Это свидетельствует о некорректности оценок по симметричным уравнениям. В целом полученные по ММП значения достаточно близки к оценкам, найденным с помощью нелинейного МНК. Это вполне закономерно, поскольку оценка стандартного отклонения ошибки входного фактора очень мала и имеет порядок 10^{-5} .

Список литературы

- [1] **Дерябина Г.И.** Сополимеризация. Самара: Самарский ун-т; 2013: 48.
 - [2] **Polic L., Duever T.A., Penlidis A.** Case studies and literature review on the estimation of copolymerization reactivity ratios. *Journal of Polymer Science. Part A: Polymer Chemistry*. 1998; 36(7):813–822.
 - [3] **Kazemi N., Duever T. A., Penlidis A.** A powerful estimation scheme with the error-in-variables-model for nonlinear cases: Reactivity ratio estimation examples. *Computers & Chemical Engineering*. 2013; 48(7):200–208.
 - [4] **Zwanzig S.** On the criteria for experimental design in nonlinear error-in-variables models. *Advances in Stochastic Simulation Methods. Statistics for Industry and Technology*. N. Balakrishnan, V.B. Melas, S. Ermakov. (Eds). Boston: Springer; 2000:153–163.
 - [5] **Даванков А.Б., Зубакова Л.Б., Гуров А.А.** Определение констант сополимеризации 2-метил-5-винилпиридина с триэтиленгликольдиметакрилатом. *Высокомолекулярные соединения*. 1964; 6(2):237–240.
 - [6] **Юлусов В.В.** Закономерности образования сополимеров из мономеров разной активности в условиях полимеризации с обратимой передачей цепи: Дис. ... канд. хим. наук. М.: МГУ им. М.В. Ломоносова; 2014: 147.
 - [7] **Поздняков А.С.** Полифункциональные (со)полимеры 1-винил-1,2,4-триазола и нанокompозиты на их основе: Дис. ... канд. хим. наук. Иркутск: ИрИХ СО РАН; 2011: 175.
 - [8] **Van der Meer R., Linssen H.N., German A.L.** Improved methods of estimating monomer reactivity ratios in copolymerization by considering experimental errors in both variables. *Journal of Polymer Science. Part A: Polymer Chemistry*. 1978; 16(11):2915–2930.
 - [9] **Калиткин Н.Н.** Численные методы. СПб.: БХВ-Петербург; 2011: 592.
 - [10] **Denisov V.I., Timofeeva A.Y., Khailenko E.A.** Estimating parameters of polynomial models with errors in variables and no additional information. *Journal of Applied and Industrial Mathematics*. 2016; 10(3):322–332.
 - [11] **Езриелев А.И., Брохина Э.Л., Роскин Е.С.** Аналитический метод вычисления констант сополимеризации. *Высокомолекулярные соединения*. 1969; (A)11(8):1670–1680.
 - [12] **Тимофеев В.С., Фаддеенков А.В., Щеколдин В.Ю.** Эконометрика. М.: ЮРАЙТ; 2016: 328.
 - [13] **Hautus F.L.M., Linssen H.N., German A.L.** Dependence of computed copolymer reactivity ratios on the calculation method. II. Effects of experimental design and error structure. *Journal of Polymer Science. Part A: Polymer Chemistry*. 1984; 22(12):3661–3671.
 - [14] **Kazemi N., Lessard B.H., Marić M., Duever T.A., Penlidis A.** Reactivity ratio estimation in radical copolymerization: from preliminary estimates to optimal design of experiments. *Industrial & Engineering Chemistry Research*. 2014; 53(18):7305–7312.
 - [15] **Дерябина Г.И., Фролов Ю.Л., Калабина А.В.** Реакционная способность винилариловых эфиров в сополимеризации с хлористым винилом и винилиденхлоридом. *Высокомолекулярные соединения*. 1971; (A)13(5):1162–1169.
 - [16] **Дерябина Г.И., Гершенгорн Г.И.** Расчет констант сополимеризации систем винил(винилиден) хлорид – винилфениловый эфир. *Высокомолекулярные соединения*. 1969; 11(5):941–945.
 - [17] **Смирнов А.И., Дерябина Г.И., Калабина А.В., Владимирова Т.Л.** Особенности сополимеризации простых виниловых эфиров с электроноакцепторными мономерами. *Высокомолекулярные соединения*. 1980; 22(3):173–177.
-

MATHEMATICAL MODELLING

DOI:10.25743/ICT.2020.25.3.003

Estimation of copolymerization reactivity ratios in the Berkson error regression model

TIMOFEEVA ANASTASIIA YU.

Novosibirsk State Technical University, 630073, Novosibirsk, Russia

Corresponding author: Timofeeva Anastasiia Yu., e-mail: a.timofeeva@corp.nstu.ru

*Received March 10, 2020, revised March 24, 2020, accepted April 10, 2020***Abstract**

Purpose. The purpose of this paper is to study methods for estimating copolymerization reactivity ratios based on the differential composition equation.

Methodology. Most estimation methods reduce the differential composition equation to a linear form. They are based on the least squares method and do not take into account the measurement error in the input variable. Therefore they lead to statistically incorrect results. When analyzing the problem on the basis of the error-in-variables model in the classical case, additional information is required to determine the magnitude of the errors in measuring the concentration of monomers in the mixture and in the copolymer. Inclusion of the measurement error in the input variable into the model as the Berkson error is more consistent with the actual conditions of the experiments. It allows simultaneous estimating both the reactivity ratios and the variances of measurement errors using the maximum likelihood method.

Results. The algorithm have been developed for estimating reactivity ratios with no additional information. The empirical study of estimation methods has been carried out using the example of copolymerization of vinyl esters.

Findings. It is shown that the method based on symmetric equations gives incorrect results. Estimation results using the proposed algorithm are closest to the estimates obtained by the non-linear least squares method.

Keywords: copolymerization, monomer reactivity ratio, regression, Berkson error, maximum likelihood method.

Citation: Timofeeva A.Yu. Estimation of copolymerization reactivity ratios in the Berkson error regression model. Computational Technologies. 2020; 25(3):18–28. (In Russ.)

References

1. Deryabina G.I. Sopolimerizatsiya [Copolymerization]. Samara: Samarskiy Univ.; 2013: 48. (In Russ.)
2. Polic L., Duever T.A., Penlidis A. Case studies and literature review on the estimation of copolymerization reactivity ratios. Journal of Polymer Science. Part A: Polymer Chemistry. 1998; 36(7):813–822.
3. Kazemi N., Duever T. A., Penlidis A. A powerful estimation scheme with the error-in-variables-model for nonlinear cases: Reactivity ratio estimation examples. Computers & Chemical Engineering. 2013; 48(7):200–208.
4. On the criteria for experimental design in nonlinear error-in-variables models. Advances in Stochastic Simulation Methods. Statistics for Industry and Technology. N. Balakrishnan, V.B. Melas, S. Ermakov. (Eds). Boston: Springer; 2000:153–163.
5. Davankov A.B., Zubakova L.B., Gurov A.A. Determination of the monomer reactivity ratios for copolymerization of 2-methyl-5-vinylpyridine and triethyleneglycoldimethyl acrylate. Polymer Science. 1964; 6(2):237–240. (In Russ.)

6. Yulusov V.V. Zakonomernosti obrazovaniya sopolimerov iz monomerov raznoy aktivnosti v usloviyakh polimerizatsii s obratimoy peredachey tsepi [Patterns of the formation of copolymers from monomers of various activity under polymerization conditions with reversible chain transfer]. Dis. ... kand. khim. nauk. Moskva: MGU im. M.V. Lomonosova; 2014: 147. (In Russ.)
7. Pozdnyakov A.S. Polifunktsional'nye (so)polimery 1-vinil-1,2,4-triazola i nanokompozity na ikh osnove [Multifunctional (co)polymers of 1-vinyl-1,2,4-triazole and nanocomposites based on them]. Dis. ... kand. khim. nauk. Irkutsk: IrIKh SO RAN; 2011: 175. (In Russ.)
8. Van der Meer R., Linssen H.N., German A.L. Improved methods of estimating monomer reactivity ratios in copolymerization by considering experimental errors in both variables. *Journal of Polymer Science. Part A: Polymer Chemistry*. 1978; 16(11):2915–2930.
9. Kalitkin N.N. Chislennyye metody [Numerical methods]. SPb.: BKhV-Peterburg; 2011: 592. (In Russ.)
10. Denisov V.I., Timofeeva A.Y., Khailenko E.A. Estimating parameters of polynomial models with errors in variables and no additional information. *Journal of Applied and Industrial Mathematics*. 2016; 10(3):322–332.
11. Ezrielev A.I., Brokhina E.L., Roskin E.S. Analytical method for determination of copolymerization ratios. *Polymer Science. Series A*. 1969; 11(8):1670–1680. (In Russ.)
12. Timofeev V.S., Faddeenkov A.V., Shchekoldin V.Yu. *Ekonometrika [Econometrics]*. Moscow: YuRAYT; 2016: 328. (In Russ.)
13. Hautus F.L.M., Linssen H.N., German A.L. Dependence of computed copolymer reactivity ratios on the calculation method. II. Effects of experimental design and error structure. *Journal of Polymer Science. Part A: Polymer Chemistry*. 1984; 22(12):3661–3671.
14. Kazemi N., Lessard B.H., Marić M., Duever T.A., Penlidis A. Reactivity ratio estimation in radical copolymerization: from preliminary estimates to optimal design of experiments. *Industrial & Engineering Chemistry Research*. 2014; 53(18):7305–7312.
15. Deryabina G.I., Frolov Yu.L., Kalabina A.V. The reaction capacity of vinylidene ethers during copolymerization with vinyl chloride and vinylidene chloride. *Polymer Science. Series A*. 1971; 13(5): 1162–1169. (In Russ.)
16. Deryabina G.I., Gershengorn G.I. Calculation of copolymerization constants in systems vinyl(vinylidene) chloride-vinylphenyl ether. *Polymer Science. Series A*. 1969; 11(5):941–945. (In Russ.)
17. Smirnov A.I., Deryabina G.I., Kalabina A.V., Vladimirova T.L. Characteristics of copolymerization of vinyl ethers with electron-withdrawing monomers. *Polymer Science. Series B*. 1980; 22(3):173–177. (In Russ.)