

안정성 연구에서의 사용기간에 관한 비모수적 추론

김태규* · 박상규*** · 하명호**

*한남대학교 비즈니스통계학과

**중앙대학교 수학과통계학부

Nonparametric Inference of Shelf-life in Drug's Stability Study

Tai-Kyoo Kim* · Sang-Gue Park*** · Myung-Ho Ha**

*Department of Business Statistics, Hannam University

**Department of Statistics, Chung-Ang University

Key Words : Batch-to-batch variation, Bootstrap, Shelf-life, Stability test

Abstract

The shelf-life of pharmaceutical products is the time that the average product characteristic remains within an approved specification after manufacture. Since the true shelf-life of a drug product is typically unknown, it has to be estimated based on assay results of the drug characteristic from a stability study usually conducted during the process of drug development. The nonparametric statistical methods of assessing the shelf-life of drug are considered with the current FDA regulations. Some simulation studies of nonparametric methods are also presented with the discussion.

1. 서론

안정성 시험(stability test)의 목적은 온도, 습도 및 빛과 같은 다양한 환경 요인 하에서 약물 또는 의약품의 품질이 경시적으로 어떻게 변하는가에 대한 증거를 제공하고, 약물의 재평가 기간(re-test period) 또는 의약품의 사용기간과 권장 저장 조건을 정하는 것이다(식품의약품안전청 고시, 제2007-14호). 즉, 의약품의 안정성시험은 실온 조건에서의 장기보존시험이나 가혹 조건에서의 가속 시험 등을 통해 의약품의 열화(degradation) 정도를 파악한 후 의약품의 규격과 비교하여 의약품의 사용기간(저장기간, shelf life)을 추론하는 과정을 포함한다. 이를 위해 식품의약품안전청에서는 로트의 선정, 보존조건, 시험기간, 측정시기, 시험항목 뿐만 아니라 통계적 시험 설계도 제시하고 있다. 시판되는 모든 의약품은 약사법 제26조, 34조 및 같은 법 시행규칙 제27조 제1항 제3호의 규정에 따라 의약품의 사용기간을 설정하기 위해 적당한 안정성 시험을 수행해야 한다.

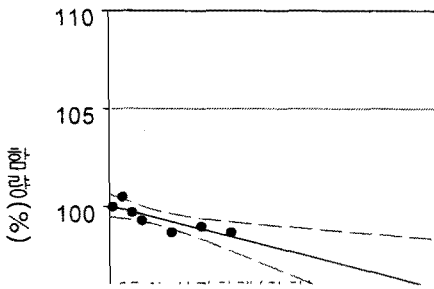
국제적으로는 ICH(International Conference on Harmonization) 등의 기관에서 의약품의 안정성평가에 대한 대체적인 기준을 발표하였고, 우리나라뿐만 아니라 미국이나 유럽 등에서도 이를 반영하고 있다.

본 연구에서는 현재 대다수 국가의 식품의약품안전청이나 ICH 등에서 사용하고 있는 의약품 사용기간 산정 방식에 대한 통계적 추론 방법을 알아보고, 비교적 작은 수의 실험을 통해 수행되는 안정성 시험에서 비모수적 추론의 필요성을 논의한 후 bootstrap 기법을 제안하여 보다 정확한 사용기간의 추론방법을 제안하고자 한다. 2장에서는 현재 사용기간 추론 방법을 소개하고, 3장에서는 bootstrap 비모수 기법을 소개하고 4장에서는 모의실험을 통해 기존의 방식과 제안된 방법의 효율성을 비교하고자 한다.

2. 사용기간의 통계적 추론

우리나라에서 의약품 사용기간을 추론하는 방법은 식품의약품안전청 고시 제2007-14호에 의해 선형회귀

분석을 통한 외삽법을 기준으로 하고 있다. 즉, 시간에 따라 수치가 감소(증가)하는 시험항목(예, 함량)의 경우 95.0% 신뢰도로 하한(상한)의 단측검정을 수행하여 사용기간 등을 외삽한다. 예를 들자면, <그림 1>의 경우 95% 신뢰하한과 규격이 만나는 26개월이 최대 연장 가능한 사용기간으로 결정된다.



<그림 1> 사용기간 결정 방식

의약품 사용기간을 추론하는 통계적 방법은 다음과 같다. 한 배치에서 생산된 의약품 n 개를 n 기간에 랜덤 배치한 후, 각 기간에서 함량 등 의약품의 열화 정도를 파악할 수 있는 지표를 계량화하여 다음과 같은 선형회귀모형에 적합 시킨다.

$$y_j = \alpha + \beta x_j + \epsilon_j, \quad (j = 1, 2, \dots, n), \quad (1)$$

여기서 x_j 는 j -번째 관찰시점을, y_j 는 j -번째 기간에 관측된 의약품의 열화(degradation)량을, ϵ_j 는 평균이 0이고 분산이 σ^2 인 정규분포를 따르는 실험오차를, 그리고 α, β 는 고정모수(fixed parameter)를 의미한다.

최소제곱법을 이용해서 모형 (1)을 추론하면 추정량 $\hat{\alpha}, \hat{\beta}$ 는 다음과 같다.

$$\hat{\beta} = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2} \equiv \frac{S_{xy}}{S_{xx}},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

여기서 $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$ 이다.

시점 t 에서 열화량 y_t 의 $(1-\gamma)100\%$ 신뢰구간의 하한은 다음과 같이 추정된다.

$$\hat{\alpha} + \hat{\beta}t - t(1-\gamma, n-2) \sqrt{MSE \left(\frac{1}{n} + \frac{(t-\bar{x})^2}{S_{xx}} \right)}, \quad (2)$$

여기서 $t(1-\gamma, n-2)$ 는 자유도 $n-2$ 인 t -분포의 $(1-\gamma)$ 백분위수이고, MSE 는 회귀분석의 분산분석표에서 얻어지는 잔차평균제곱(Error Mean Square)을 의미한다. 만약 η 가 의약품의 규격 한계라 한다면 다음을 만족하는 t_s 을 의약품의 사용기간이라 한다.

$$t_s = \inf \left\{ t : \eta \geq \hat{\alpha} + \hat{\beta}t - t(1-\gamma, n-2) \sqrt{MSE \left(\frac{1}{n} + \frac{(t-\bar{x})^2}{S_{xx}} \right)} \right\} \quad (3)$$

일반적으로 의약품은 배치 단위로 생산되며 각 배치들은 어느 정도 변동성을 가지고 있다. 식품의약품안전청에서도 이를 감안하여 최소한 3개 이상의 배치에서 안정성 시험을 하도록 권하고 있다. 만약 k 개의 배치에서 의약품 표본을 선택하여 의약품의 열화량을 파악하려 한다면 다음과 같은 선형회귀모형을 고려할 수 있다.

$$y_{ij} = \alpha_i + \beta_j x_{ij} + \epsilon_{ij}, \quad (i = 1, 2, \dots, k; j = 1, 2, \dots, n), \quad (4)$$

여기서 $x_{ij}, y_{ij}, \epsilon_{ij}$ 는 각각 i -번째 배치에서의 j -번째 관찰시점, 관측된 열화량 그리고 실험오차를 의미한다. 그리고 α_i, β_i 는 각 배치의 특성을 나타내는 고정모수를 의미한다. 만약 배치 간에 변동성을 무시할 수 있다면, 모형 (4)을 $y_{ij} = \alpha + \beta x_{ij} + \epsilon_{ij}$ 로 놓고 각 배치에서 얻은 결과를 하나로 합쳐서 식 (5)와 같은 결과를 얻을 수 있다.

$$t_s = \inf \left\{ t : \eta \geq \hat{\alpha} + \hat{\beta}t - t(1-\gamma, nk-2) \sqrt{MSE \left(\frac{1}{nk} + \frac{(t-\bar{x})^2}{S_{xx}} \right)} \right\} \quad (5)$$

식 (5)을 만족하는 t_s 을 신뢰도 $(1-\gamma) \times 100\%$ 에서의 의약품의 사용기간이라 할 수 있다.

만약 k 개의 배치간의 변동성을 무시할 수 없다면 각각의 배치에서 k 개의 선형회귀모형을 추정한 후 식 (2)와 (3)을 통해 k 개의 의약품 사용기간 $t_{s(i)}$ 을 구한 후 이 중 최소값을 의약품의 사용기간으로 결정한다. 이는 사용기간을 결정하는 매우 보수적인 방법으로 현재 식품의약품안전청에서 제안하고 있는 방법이기도 하다. 보다 자세한 통계적 추론에 관한 논의는 최근 발표된 김태규와 박상규(2008)을 참고하기 바란다.

3. 사용기간의 비모수적 추론

의약품 사용기간의 추정에 사용되는 선형회귀분석

방법은 일반적으로 이상치에 취약하고 이를 이용해 만들어진 신뢰구간에 큰 영향을 미치는 것으로 알려져 있다. 특히 시험 크기가 작을 때 더욱 문제가 되는 것으로 알려져 있다. 의약품 안정성 연구는 비교적 작은 수의 표본을 이용해서 이루어지므로 오차항의 정규성을 담보하기 어려워 이에 대한 통계적 대안이 꼭 필요하다.

일반적인 선형모형이론에서 정규성 가정이 어려운 경우 최소제곱법에 의한 통계적 추론을 지양하고 Theil(1950)의 비모수적 방법을 많이 사용한다. Theil의 방법은 정규성을 벗어나는 경우에도 간단하게 적용할 수 있으면서 강건한 추정량을 제공하는 것으로 알려져 있다. 모형 (1)에서 α, β 를 비모수적으로 추정하는 방법은 다음과 같다.

관찰시점 $x_{il} > x_{ij}$ ($i = 1, \dots, k; j, l = 1, \dots, n$)에 대하여, 다음과 같은 통계량을 정의한다.

$$S_{ijl} = \frac{y_{il} - y_{ij}}{x_{il} - x_{ij}}, \quad (i = 1, \dots, k; j, l = 1, \dots, n).$$

이렇게 정의된 S_{ijl} 를 순서화하여 이들의 중앙값을 기울기 β 의 추정량 $\hat{\beta}_T$ 이라 한다. 추정량은 $\hat{\beta}_T$ 을 이용해서 $d_{ij} = y_{ij} - \hat{\beta}_T x_{ij}$ 를 계산한 후 이들의 중앙값을 절편 α 의 추정량 $\hat{\alpha}_T$ 라 한다. 이렇게 얻은 추정량을 이용해서 $E(y) = \alpha + \beta t$ 의 95% 신뢰하한을 추정하기 위해 bootstrap 방법을 사용한다. 즉, 원 자료를 대체(replacement)를 허용하고 랜덤하게 선택하여 얻은 (y_{ij}^*, x_{ij}^*) , ($i = 1, \dots, k; j = 1, \dots, n$)을 bootstrap 표본이라고 하고 이를 통해 얻은 Theil 추정량을 $(\hat{\alpha}^*, \hat{\beta}^*)$ 이라 하자. 이러한 과정을 B 번 반복한 후 이렇게 얻은 추정량 $(\hat{\alpha}_1^*, \hat{\beta}_1^*), \dots, (\hat{\alpha}_B^*, \hat{\beta}_B^*)$ 을 통해 B 개의 추정량을 계산한 후 이를 순서화하여 95% 신뢰구간 하한치를 얻고 이를 bootstrap 추정량이라 한다.

일반적으로 bootstrap 추정량은 표본이 한쪽으로 치우쳐있는 경우 편의(bias)되는 경우가 많이 발생된다(Davidson and Hinkley, 1997). 안정성 시험의 경우 김태규와 박상규(2008)의 사례처럼 초기 열화량이 한방향으로 몰려 있는 경우가 많아 편의가 발생할 가능성이 높다. 따라서 일반적인 bootstrap 추정량을 사용하는 것보다 편의를 수정하여 사용하는 것이 더 효율적일 수 있다. 편의 수정(biased-corrected) bootstrap 추정량은 다음과 같은 방식으로 계산된다.

Bootstrap 표본 (y_{ij}^*, x_{ij}^*) , ($i = 1, \dots, k; j = 1, \dots, n$)

에 대하여 편의 추정값은 식 (6)과 같다.

$$bias_{\theta} = \hat{\theta}^*(\cdot) - \hat{\theta} \tag{6}$$

여기서 $\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$ 이고, $\hat{\theta}^*(b)$ 는 b 번째 bootstrap

표본으로부터 구한 추정량이며, $\hat{\theta}$ 은 원 표본 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 으로부터 구한 추정량이다. 이렇게 계산된 편의 추정값이 0과 크게 다를 때 bootstrap 추정량의 값의 정확도는 떨어지게 된다. 따라서 이러한 편의값을 보정한 새로운 추정량의 제안이 필요하게 되는데 편의 수정 bootstrap 추정량은 식 (7)과 같다.

$$\hat{\theta}_{adj} = \hat{\theta}^*(b) - bias_{\theta} \tag{7}$$

이러한 편의 수정 bootstrap 추정량을 구하는 방법은 다음과 같다.

step 1 원 자료 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 로부터 bootstrap 표본 $(x_{i1}^*, y_{i1}^*), (x_{i2}^*, y_{i2}^*), \dots, (x_{in}^*, y_{in}^*)$ 을 생성한다.

step 2 부스트랩 표본 $(x_{i1}^*, y_{i1}^*), (x_{i2}^*, y_{i2}^*), \dots, (x_{in}^*, y_{in}^*)$ 로부터 $(\hat{\alpha}(i)^*, \hat{\beta}(i)^*)$ 를 추정한다.

step 3 이러한 과정을 B 번 반복한다.

step 4 원자료 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 로부터 $(\hat{\alpha}, \hat{\beta})$ 을 추정한다.

step 5 편의 수정 bootstrap 추정량은 다음과 같다.

$$\begin{aligned} \hat{\alpha}_{adj}(b) &= \hat{\alpha}^*(b) - (\hat{\alpha}^*(\cdot) - \hat{\alpha}) \\ \hat{\beta}_{adj}(b) &= \hat{\beta}^*(b) - (\hat{\beta}^*(\cdot) - \hat{\beta}) \end{aligned}$$

step 6 추정량 $(\hat{\alpha}_{adj}^*(1), \hat{\beta}_{adj}^*(1)), \dots, (\hat{\alpha}_{adj}^*(B), \hat{\beta}_{adj}^*(B))$ 을 통해 B 개의 추정량을 계산한 후 이를 순서화하여 95% 신뢰구간 하한치를 얻는다.

4. 모의실험

의약품의 사용기간을 통계적으로 추론하는 3가지 방법을 비교하기 위하여 모의실험을 수행하였다. 모의실

험은 (1)의 모형을 토대로 하여 α 값을 100에서 102, β 값을 -0.2에서 -0.1까지 변화시키면서 $x=0, 3, 6, 9, 12$ 에서 자료를 생성하여 각각 1000회 반복하여 수행하였으며, 오차는 $N(0,0.2)$ 에서 발생시켰다. 생성한 자료를 이용하여 MLE 방법과 bootstrap 방법, 편의 수정 bootstrap 방법으로 각각 $\hat{\alpha}, \hat{\beta}$ 를 추정하고, 추정된 $\hat{\alpha}, \hat{\beta}$ 를 이용하여 95% 신뢰하한을 계산하였다. 모의실험 수행 결과는 <표 1>에 정리를 했고 <표 2>는 이를 요약한 표이다. <표 1>에서 α 와 β 는 실제 모수값을 의미하고 True value(참값)은 실제로 부여한 α 와 β 값에서 Y가 규격 하한과 일치되는 X값이다. MLE, bootstrap, 편의 수정 bootstrap열에 있는 값은 1000번 중에 추정된 95% 신뢰 하한이 참값을 포함한 횟수이다.

<표 2>를 보면 크게 MLE에 비해 bootstrap 추정량이 개선되고 있음을 볼 수 있다. 즉, bootstrap 추정량은 MLE에 비해 주어진 유의수준인 95%를 잘 유지하고 있는 것을 볼 수 있고, 편의수정 bootstrap 추정량은 MLE와 비슷한 유의수준을 유지하고 있지만 표준편차의 크기가 크게 작아 어떤 상황에서도 일관된 추정량을 제시하고 있는 것을 알 수 있다. 특히 <표 1>을 보면 α 의 값이 100에 가깝고, β 의 값이 -0.20에 가까운 경우 MLE와 bootstrap 추정량은 참값을 포함한 비율이 주어진 신뢰수준인 95%에 크게 못미치는 것을 볼 수 있지만 편의수정 bootstrap 추정량은 어떤 상황 하에서도 일관된 추정량을 제시하고 있다는 것이 큰 장점이라 할 수 있다.

<표 1> 95% 포함확률 모의실험 결과

α	β	True value	MLE	bootstrap	편의 수정 bootstrap
100.0	-0.20	25.00	538	599	968
100.0	-0.18	27.78	840	843	967
100.0	-0.16	31.25	973	947	975
100.0	-0.14	35.71	995	969	976
100.0	-0.12	41.67	1000	975	981
100.0	-0.10	50.00	1000	984	989
100.2	-0.20	26.00	687	729	956
100.2	-0.18	28.89	917	908	963
100.2	-0.16	32.50	987	958	967
100.2	-0.14	37.14	999	977	969
100.2	-0.12	43.33	1000	975	982
100.2	-0.10	52.00	1000	974	986
100.4	-0.20	27.00	797	789	965
100.4	-0.18	30.00	967	941	970
100.4	-0.16	33.75	990	967	968

100.4	-0.14	38.57	1000	972	976
100.4	-0.12	45.00	1000	977	984
100.4	-0.10	54.00	1000	980	987
100.6	-0.20	28.00	869	863	959
100.6	-0.18	31.11	967	952	968
100.6	-0.16	35.00	992	965	969
100.6	-0.14	40.00	1000	968	986
100.6	-0.12	46.67	1000	978	983
100.6	-0.10	56.00	1000	973	991
100.8	-0.20	29.00	930	910	961
100.8	-0.18	32.22	983	960	969
100.8	-0.16	36.25	995	977	974
100.8	-0.14	41.43	1000	977	983
100.8	-0.12	48.33	1000	974	979
100.8	-0.10	58.00	1000	975	985
101.0	-0.20	30.00	957	936	970
101.0	-0.18	33.33	994	978	969
101.0	-0.16	37.50	999	969	975
101.0	-0.14	42.86	1000	973	978
101.0	-0.12	50.00	1000	975	982
101.0	-0.10	60.00	1000	969	990
101.2	-0.20	31.00	970	945	954
101.2	-0.18	34.44	993	961	958
101.2	-0.16	38.75	999	965	969
101.2	-0.14	44.29	1000	978	974
101.2	-0.12	51.67	1000	970	977
101.2	-0.10	62.00	1000	981	989
101.4	-0.20	32.00	979	951	965
101.4	-0.18	35.56	998	971	972
101.4	-0.16	40.00	1000	973	982
101.4	-0.14	45.71	1000	973	978
101.4	-0.12	53.33	1000	972	976
101.4	-0.10	64.00	1000	965	987
101.6	-0.20	33.00	986	958	972
101.6	-0.18	36.67	1000	963	968
101.6	-0.16	41.25	1000	964	974
101.6	-0.14	47.14	1000	970	980
101.6	-0.12	55.00	1000	972	981
101.6	-0.10	66.00	1000	978	983
101.8	-0.20	34.00	991	970	955
101.8	-0.18	37.78	999	969	970
101.8	-0.16	42.50	1000	969	974
101.8	-0.14	48.57	1000	967	970
101.8	-0.12	56.67	1000	965	989
101.8	-0.10	68.00	1000	979	988
102.0	-0.20	35.00	998	978	970
102.0	-0.18	38.89	999	977	969
102.0	-0.16	43.75	1000	981	980
102.0	-0.14	50.00	1000	977	975
102.0	-0.12	58.33	1000	978	988
102.0	-0.10	70.00	1000	980	991

<표 2> True value를 포함한 수의 평균과 표준편차

추정량	MLE	bootstrap	편의 수정 bootstrap
평균	974.1	952.1	975.1
표준편차	74.9	61.3	9.4

의약품 사용기간은 의약품 특성상 유의수준을 정확하게 유지하는 것 보다 조금은 보수적으로 접근하는 것이 타당하다고 생각한다. 따라서 오차항의 분포가 정규성을 가정하기 어려운 경우에 어떤 상황 하에서도 일관된 결과치를 제공하는 통계적 방법이 바람직하다. 기존에 사용되고 있는 MLE 보다는 작은 표본으로 수행되는 특성을 지녀 항상 표본이 치우쳐있을 가능성이 많다면 향후 의약품 사용기간 추론에는 편의수정 bootstrap 추정량을 고려하는 것이 바람직 해 보인다.

5. 결론

모수적 방법은 잔차가 정규 분포를 따르지 않거나 이상치가 존재할 경우 모수 추정에 있어 심각한 영향을 미친다. 이러한 경우에 사용할 수 있는 강건한(robust)한 통계적 방법이 비모수적 방법이며 bootstrap과 편의수정 bootstrap 방법은 비모수적 방법 중 최근 많이 활용되고 있는 방법이다.

본 연구에서는 모의실험을 통하여 현재 의약품 사용기간을 결정하는 통계적 방법인 MLE 추정량과 비모수적 방법인 bootstrap 추정량, 편의수정 bootstrap 추정량을 모의실험을 통해 비교하였다. 의약품의 특성 상 보수적인 접근이 안전성 측면에서 바람직하고, 실험 결과의 일관성이 신뢰성 확보에 바람직하다는 측면에서 편의수정 bootstrap 추정량을 현재 사용되고 있는 MLE를 보완할 수 있는 방법으로 함께 고려한다면 보다 품질 좋은 의약품을 제공하는데 도움이 되리라 생각된다.

참고문헌

- [1] 김태규, 박상규 (2008). 의약품 안정성 연구에서의 사용기간에 관한 통계적 추론. 품질경영학회지 제 36권 제 1호, 1-6.
- [2] 의약품등의 안정성시험 기준 (2007). 식품의약품안전청고시 제 2007-14호
- [3] ICH Guidance for Industry (2003). Q1E Stability Data Evaluation.
- [4] Davidson, A.C. and Hinkley, D.V. (1997). Bootstrap methods and their application. Cambridge University Press, New York.
- [5] Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. Indagationes Mathematicae 12, 85-91.

2010년 1월 26일 접수, 2010년 3월 4일 수정, 2010년 3월 5일 채택