한국통계학회논문집
2010, 17권, 2호, 275-292

대응표본에서 AUC차이에 대한 신뢰구간 추정에 관한 고찰

김희영[1,a]

[a]고려대학교 경제연구소

요약

동일 환자에게 적용된 2가지 진단검사의 정확성을 비교하기 위한 방법들 중에서 두개의 ROC곡선 아래 면적(AUC; Area Under Curve)의 차이는 주요한 잣대 중 하나이다. 본 연구에서는 AUC의 차이를 추정하는 방법으로 비모수적방법, 최대가능도법, 일반화추축량에 의한 방법, 붓스트랩방법의 4가지를 포함확률(coverage probability), 기대길이(expected length) 측면에서 모의실험을 통하여 비교하였다.

주요용어: ROC, AUC, 비모수적방법, 최대우도법, 일반화추축량, 붓스트랩.

1. 연구배경

의학분야에서 정확한 진단검사법을 찾는 것은 중요한 관심사항 중 하나이다. 두 종류의 진단검사법이 있을 경우 이를 비교하기 위해 서로 독립적인 두 그룹의 환자들에 대해 각각 적용하고 평가하기보다는 동일한 환자들에 대해서 서로 다른 두 종류의 진단검사법을 적용하게 되고 그 결과로부터 민감도, 특이도, 정확도 등을 비교하게 된다. 일반적으로 동일 환자에 대해서 두 가지 진단검사가 이루어진 경우, 두 진단검사를 평가하기 위해 ROC(receiver operating characteristic)곡선 아래의 면적인 AUC을 비교하게 된다. 그런데, 현재 사용되고 있는 SPSS 17.0, SAS 9.1과 같은 통계 소프트웨어에서는 대응표본에서의 AUC차이에 대한 검증방법이 제공되고 있지 않다. 널리 사용되고 있는 SAS, SPSS와 같은 통계 패키지에는 포함되어 있지 않으나 그동안 이에 관하여 많은 학자들이 연구를 진행해왔으며, 그 결과 대응표본에서 AUC의 차이의 신뢰구간 추정방법으로 비모수적방법(nonparametric method), 최대우도법(maximum likelihood method), 붓스트랩(bootstrap)을 이용한 방법 등이 제안되었다.

최근 Li 등 (2008)은 일반화추축량(generalized pivotal quantity; GPQ) 개념을 이용해서 정확한 신뢰구간 계산방법을 제안하였다. 그리고, Li 등 (2008)은 DeLong 등 (1988)에 의해 제안된 비모적 방법을 이용한 신뢰구간, McClish (1989)의 정규분포 가정 하에서 최대우도법에 근거한 근사적 신뢰구간, GPQ방법을 이용한 정확신뢰구간을 포함확률(coverage probability)과 신뢰구간 길이의 평균(expected length, 기대길이)를 표본크기, AUC, 상관계수 값들을 달리하여 모의실험한 결과를 보여주었다. 그러나, Li 등 (2008)의 연구에서는 실제 임상시험에서 많이 다루게 되는 데이터 구조와는 다소 거리가 있다. 보통 두 번 반복 측정하게 되는 진단검사 결과 간의 상관관계는 높게 나오는 것이 일반적인데 이들은 상관계수의 값으로 0.01, 0.5, 0.99의 3가지 값만을 고려하였다. 상관계수의 값이 0.01은 상관관계가 매우 극히 낮은 경우로 현실성과 다소 거리가 있고, 상관계수의 값이 0.99는 대단히 극단적인 경우이다.

한편 Bandos 등 (2007)은 AUC의 비모수적 추정량에 대하여 정확한 붓스트랩 분산을 구하기 위한 연구를 진행하였다. Bandos 등 (2007)은 동일 개체에 대해 적용한 평가 척도 간 상관을 고정시켜 놓고 근사적 신뢰구간에 대한 길이와 포함확률을 구하였다. 모의실험에서는 표본의 크기는 고려되지 않았

이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-411-J03302).

[1] (136-701) 서울 성북구 안암동 고려대학교 경제연구소, 연구교수. E-mail: starkim@korea.ac.kr

표 1: 자료구조

환자군	X_1	X_2	정상군	Y_1	Y_2
	x_{11}	x_{12}		y_{11}	y_{12}
	x_{21}	x_{22}		y_{21}	y_{22}
	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	x_{n1}	x_{n2}		y_{m1}	y_{m2}

으며 실제 데이터에서는 평가척도 간 상관계수가 매우 높게 계산되는 경우가 많은데 반해 모의실험에서 적용한 상관계수는 중간이하정도만을 고려하였다.

따라서 본 연구에서는 DeLong 등 (1988)의 비모수적방법, McClish (1989)의 최대우도법, Li 등 (2008)의 GPQ방법, Bandos 등 (2007)가 제시한 붓스트랩방법의 4가지를 보다 정교하게 상관계수의 크기가 0.5에서 0.9까지 0.1씩 증가시키며, 또한 표본 크기를 다르게 하며 비교해보고 평가하고자 한다.

본 논문의 구성은 2절에서는 AUC에 관한 간략한 소개와 더불어 4가지 방법, 즉 비모수적방법, 최대우도법, GPQ방법, 붓스트랩방법을 적용하여 두 가지 진단 검사 결과의 AUC차이에 근사 신뢰구간을 계산하는 방법을 요약한다. 3절에서는 모의실험을 통하여 4가지 방법을 비교하고, 4절에서는 맺음말로 마무리한다.

2. AUC차이에 대한 4가지 추정방법

ROC곡선 플롯에서 세로축은 민감도(또는 진양성률(true positive rate))을 표시하고 가로축에는 1 − 특이도(또는 위양성률(false positive rate))을 표시한다. X, Y를 각각 환자 그룹과 정상자 그룹에 대한 관측을 나타내는 확률변수라고 하자. 그러면, ROC곡선 아래 면적은 $A = P(X \geq Y)$으로 랜덤하게 선택된 환자의 진단결과가 랜덤하게 선택된 정상자의 진단결과보다 크게 나올 확률이 된다. 따라서 AUC차이는 두 진단방법 간 진단의 정확성을 비교하는 측도로 널리 사용되고 있다. 표 1은 환자군 n명과 정상군 m명을 대상으로 서로 다른 두 가지 진단검사가 이루어진 데이터 구조를 나타낸다. 여기서 동일인에게 2가지 진단검사를 실시한 경우를 고려하므로 변수 X_1과 X_2는 서로 상관되어 있으며, 마찬가지로 Y_1과 Y_2는 서로 상관되어 있는 것으로 가정한다. 그리고, X_j와 Y_k는 서로 상관되어 있지 않는 것으로 가정한다.

대응표본에서 두 진단 방법의 ROC곡선에서의 AUC 간 차이(= $A_1 - A_2$)의 연구방법은 아주 다양하게 많으나 그 중 대표적으로 많이 연구되고 있는 4가지 방법을 중심으로 본 연구에서는 이들 방법들을 살펴보고 모의실험을 통해 그 결과들을 비교할 것이다. 먼저 4가지 추정방법을 살펴보면 다음과 같다.

2.1. 비모수적방법

표 1의 자료구조에서와 같이 $X_{ij}\,(i = 1, \ldots, n)$는 n명의 질환자에 대한 j진단검사 결과이고, $Y_{kj}\,(k = 1, \ldots, m)$는 m명의 정상자에 대한 j진단검사의 결과라고 하자. 맨-휘트니(Mann-Whitney) 검정통계량에 근거한 ROC곡선 아래 면적의 비모수적추정량(nonparametric estimate)은 다음과 같다.

$$\hat{A}_j = \frac{1}{nm}\sum_{i=1}^{n}\sum_{k=1}^{m}\left(I\left(X_{ij} > Y_{kj}\right) + \frac{1}{2}I\left(X_{ij} = Y_{kj}\right)\right), \tag{2.1}$$

여기서 I는 표시함수(indicator function)이다. DeLong 등 (1988)에 의해 계산된 $\hat{A}_1 - \hat{A}_2$에 대한 근사적으로 추정된 분산추정량은 다음과 같다.

$$\widehat{\mathrm{var}}\left(\hat{A}_1 - \hat{A}_2\right) = \frac{1}{n}\left(s_{10}^{1,1} + s_{10}^{2,2} - 2s_{10}^{1,2}\right) + \frac{1}{m}\left(s_{01}^{1,1} + s_{01}^{2,2} - 2s_{01}^{1,2}\right), \tag{2.2}$$

여기서

$$s_{10}^{j,j'} = \frac{1}{n-1}\sum_{i=1}^{n}\left(V_{10}^{j}(X_i) - \hat{A}_j\right)\left(V_{10}^{j'}(X_i) - \hat{A}_{j'}\right), \quad j, j' = 1, 2,$$

$$s_{01}^{j,j'} = \frac{1}{m-1}\sum_{k=1}^{m}\left(V_{01}^{j}(Y_k) - \hat{A}_j\right)\left(V_{01}^{j'}(Y_k) - \hat{A}_{j'}\right), \quad j, j' = 1, 2,$$

$$V_{10}^{j}(X_i) = \frac{1}{m}\sum_{k=1}^{m}\left(I\left(X_{ij} > Y_{kj}\right) + \frac{1}{2}I\left(X_{ij} = Y_{kj}\right)\right), \quad j = 1, 2,$$

$$V_{01}^{j}(Y_k) = \frac{1}{n}\sum_{i=1}^{n}\left(I\left(X_{ij} > Y_{kj}\right) + \frac{1}{2}I\left(X_{ij} = Y_{kj}\right)\right), \quad j = 1, 2$$

이다. 따라서, 식 (2.1)과 (2.2) 결과를 기초로 비모수적방법에 의한 $\Delta = A_1 - A_2$의 $100(1-\alpha)\%$ 신뢰구간은 $(\hat{A}_1 - \hat{A}_2 - z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{A}_1 - \hat{A}_2)}, \hat{A}_1 - \hat{A}_2 + z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{A}_1 - \hat{A}_2)})$이다. 여기서 $z_{\alpha/2}$는 표준정규분포의 $(1-\alpha/2)100\%$ 백분위수이다.

2.2. 최대우도법

이제 한 걸음 더 나아가 $\mathbf{X}_1, \ldots, \mathbf{X}_n$과 $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$은 각각 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$와 $N_2(\boldsymbol{\nu}, \boldsymbol{\Psi})$로부터의 랜덤표본이라고 가정하자.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2(\boldsymbol{\nu}, \boldsymbol{\Psi}), \tag{2.3}$$

여기서,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \quad \boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} \psi_1^2 & \psi_{12} \\ \psi_{12} & \psi_2^2 \end{pmatrix} \tag{2.4}$$

이다. McClish (1989)는 이변량정규분포에서의 ROC곡선에서의 AUC차이의 최대가능도 추정량과 근사적 분산추정량을 구하였다. 먼저 $X \sim N(\mu_j, \sigma_j^2)$, $Y \sim N(\nu_j, \psi_j^2)$, $j = 1, 2$일때 j번째 진단검사에 대하여 임계값(threshod) c에서의 진양성율(true positive rate; TP(c))은 $\mathrm{TP}_j(c) = P(X_j > c) = \Phi((\mu_j - c)/\sigma_j)$이고, 위양성율(false positive rate, FP(c))은 $\mathrm{FP}_j(c) = P(Y > c) = \Phi((\nu_j - c)/\psi_j)$이다. 그리고, ROC곡선은 모든 가능한 임계값 c에 대하여 점 (FP(c), TP(c))들을 연결하여 작성한 것이므로, ROC곡선 아래 면적은

$$\begin{aligned} A_j &= \int_{-\infty}^{\infty} \mathrm{TP}(c)\mathrm{FP}'(c)dc \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\mu_j - c}{\sigma_j}\right)\phi\left(\frac{\nu_j - c}{\psi_j}\right) \times \left(-\frac{1}{\psi_j}\right)dc \end{aligned} \tag{2.5}$$

가 된다. 식 (2.5)에서 변수변환을 하면 A_j는 다음과 같다.

$$A_j = \int_{-\infty}^{\infty} \Phi(a_j + b_j v)\phi(v)dv, \tag{2.6}$$

여기서, $a_j = (\mu_j - \nu_j)/\sigma_j$, $b = \sigma_j/\psi_j$이다. 따라서, 최대가능도 추정량 $\hat{A}_j$는 식 (2.6)에 대응하는 표본평균들 $\bar{X}_j = \sum_{i=1}^{n} X_{ij}/n$, $\bar{Y}_j = \sum_{k=1}^{m} Y_{kj}/m$과 표본표준편차 $S_{X_j} = \sqrt{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2/(n-1)}$, $S_{Y_j} =$

$\sqrt{\sum_{k=1}^{m}(Y_{kj}-\bar{Y}_j)^2/(m-1)}$을 대입하여 얻어진다. 또한, $\mathrm{Var}(\hat{A}_j)$를 구하기 위하여 $\hat{A}_j$는 확률변수 $\hat{\delta}_j := \bar{X}_j - \bar{Y}_j$, $S^2_{X_j}$, $S^2_{Y_j}$의 연속함수인 점을 인식하면, $\mathrm{Var}(\hat{A}_j)$는 델타방법(delta method)를 이용하여

$$\begin{aligned}\mathrm{Var}\left(\hat{A}_j\right) &= \mathrm{Var}\left(A_j\left(\hat{\delta}_j, S^2_{X_j}, S^2_{Y_j}\right)\right)\\ &= \left(\frac{\partial A_j}{\partial \delta_j}\right)^2 \mathrm{Var}\left(\hat{\delta}_j\right) + \left(\frac{\partial A_j}{\partial \sigma_j^2}\right)^2 \mathrm{Var}\left(S^2_{X_j}\right) + \left(\frac{\partial A_j}{\partial \psi_j^2}\right)^2 \mathrm{Var}\left(S^2_{Y_j}\right), \quad j=1,2 \end{aligned} \tag{2.7}$$

으로 계산된다. 또한, $\mathrm{Var}(\hat{A}_1 - \hat{A}_2) = \mathrm{Var}(\hat{A}_1) + \mathrm{Var}(\hat{A}_2) - 2\mathrm{Cov}(\hat{A}_1, \hat{A}_2)$이고, $\mathrm{Cov}(\hat{A}_1, \hat{A}_2)$을 구하기 위하여 마찬가지로 델타방법(delta method)를 적용하면

$$\mathrm{Cov}\left(\hat{A}_1, \hat{A}_2\right) = \left(\frac{\partial A_1}{\partial \delta_1}\right)\left(\frac{\partial A_2}{\partial \delta_2}\right)\mathrm{Cov}\left(\hat{\delta}_1, \hat{\delta}_2\right) + \left(\frac{\partial A_1}{\partial \sigma_1^2}\right)\left(\frac{\partial A_2}{\partial \sigma_2^2}\right)\mathrm{Cov}\left(S^2_{X_1}, S^2_{X_2}\right) + \left(\frac{\partial A_1}{\partial \psi_1^2}\right)\left(\frac{\partial A_2}{\partial \psi_2^2}\right)\mathrm{Cov}\left(S^2_{Y_1}, S^2_{Y_2}\right) \tag{2.8}$$

으로 계산된다. 식 (2.7)과 (2.8)에서 각 항의 정확한 표현은 McClish (1989)을 참조하기 바란다. 따라서, 식 (2.7)과 (2.8) 결과를 이용하여 최대가능도 추정량에 의한 $\Delta = A_1 - A_2$의 $100(1-\alpha)\%$ 신뢰구간은 $(\hat{A}_1 - \hat{A}_2 - z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{A}_1 - \hat{A}_2)},\ \hat{A}_1 - \hat{A}_2 + z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{A}_1 - \hat{A}_2)})$이다. 여기서 $z_{\alpha/2}$는 표준정규분포의 $(1-\alpha/2)100\%$ 백분위수이다.

2.3. 일반화추축량에 의한 방법

Li 등 (2008)은 일반화추축량(generalized pivotal quantity; GPQ)의 개념에 기초한 $\Delta = A_1 - A_2$에 대한 정확한 구간추정을 발전시켰다. 먼저, GPQ의 정의를 간략히 살펴보도록 하자. W는 미지의 모수벡터 $\zeta = (\theta, \eta)$에 의존하는 확률변수라고 가정하자. 여기서 θ는 관심모수이고 η는 장애모수(nuisance parameter)벡터이다. 그리고, W는 W로부터 뽑힌 랜덤표본이고 w는 W의 관측값이라고 하자. 또한 $R = R(W; w, \zeta)$는 W, w, ζ의 함수라고 하자. 이때 만약 다음과 같은 두가지 조건이 충족되면 R은 GPQ라고 한다.

(C1) R의 분포는 미지모수들에 종속되어 있지 않다.

(C2) R의 관측값 즉, $r = R(w; w, \zeta)$는 장애모수 벡터 η와 무관하다.

특히 $r = \theta$일 경우 GPQ는 FGPQ(fiducial generalized pivotal quantity)라고 불리는데 Hanning 등 (2006)에 의하면 FGPQ에 기초한 GCIs(generalized confidence intervals)는 근사적으로 정확한 전통적 포함율(frequentist coverage probability)을 갖는 것으로 증명되었다. 따라서, θ에 대한 $100(1-\alpha)\%$ GCI는 $(R_{\alpha/2}, R_{1-\alpha/2})$이다. 여기서 $R_{\alpha/2}$와 $R_{1-\alpha/2}$는 R 분포에서의 $100(\alpha/2)\%$ 분위수와 $100(1-\alpha/2)\%$ 분위수이다. R의 분위수는 몬테칼로 모의실험을 이용하여 추정될 수 있다.

본 절에서는 GPQ 개념을 이용해서 정확한 신뢰구간을 계산하는 절차를 간단히 요약하기로 한다. $\mathbf{X}_1, \ldots, \mathbf{X}_n$과 $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$은 각각 식 (2.3)에서와 같이 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$와 $N_2(\boldsymbol{\nu}, \boldsymbol{\Psi})$로부터의 랜덤표본이라고 하자. 이때 $\mathbf{X}_1, \ldots, \mathbf{X}_n$과 $\mathbf{Y}_1, \cdots, \mathbf{Y}_m$의 표본공분산행렬은 다음과 같다.

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)^t = \frac{1}{n-1}\begin{pmatrix} SSX_1 & SSX_{12} \\ SSX_{12} & SSX_2 \end{pmatrix},$$

$$\frac{1}{m-1}\sum_{k=1}^{m}\left(\mathbf{Y}_k - \bar{\mathbf{Y}}\right)\left(\mathbf{Y}_k - \bar{\mathbf{Y}}\right)^t = \frac{1}{m-1}\begin{pmatrix} SSY_1 & SSY_{12} \\ SSY_{12} & SSY_2 \end{pmatrix}.$$

모수 $\Delta = A_1 - A_2$에 대한 $100(1-\alpha)\%$ GCI는 다음과 같은 몬테칼로 알고리즘으로 구할 수 있다.

단계1: 각각 $n-1, n-2, m-1, m-2$의 자유도를 갖는 카이제곱분포로부터 서로 독립적으로 $U_1, U_{2|1}$, $V_1, V_{2|1}$를 임의생성한다. 그리고 표준정규분포로부터 Z_B와 Z_G를 생성한다.

단계2: 단계1에서 생성된 $U_1, U_{2|1}, V_1, V_{2|1}, Z_B$와 Z_G를 이용하여 아래 식과 같이 정의되는 $(R_{\psi_1^2}, R_{\psi_{2|1}^2}, R_\gamma)$, $(R_{\sigma_1^2}, R_{\sigma_{2|1}^2}, R_\beta)$, $(R_{\Sigma/n} + R_{\Psi/m})^{1/2}$을 계산한다.

$$\left(R_{\sigma_1^2}, R_{\sigma_{2|1}^2}, R_\beta\right) = \left(\frac{ssx_1}{U_1}, \frac{ssx_{2|1}}{U_{2|1}}, b - Z_B\sqrt{\frac{1}{U_{2|1}}\frac{ssx_{2|1}}{ssx_1}}\right),$$

$$\left(R_{\psi_1^2}, R_{\psi_{2|1}^2}, R_\gamma\right) = \left(\frac{ssy_1}{V_1}, \frac{ssy_{2|1}}{V_{2|1}}, g - Z_G\sqrt{\frac{1}{V_{2|1}}\frac{ssy_{2|1}}{ssy_1}}\right),$$

여기서, $b = ssx_{12}/ssx_1$, $g = ssy_{12}/ssy_1$, $ssx_{2|1} = ssx_2 - ssx_{12}^2/ssx_1$, $ssy_{2|1} = ssy_2 - ssy_{12}^2/ssy_1$이다. 또한,

$$R_\Sigma = \begin{pmatrix} R_\beta^2 R_{\sigma_1^2} + R_{\sigma_{2|1}^2} & R_\beta R_{\sigma_1^2} \\ R_\beta R_{\sigma_1^2} & R_\beta^2 R_{\sigma_1^2} + R_{\sigma_{2|1}^2} \end{pmatrix}, \qquad R_\Psi = \begin{pmatrix} R_\gamma^2 R_{\psi_1^2} + R_{\psi_{2|1}^2} & R_\gamma R_{\psi_1^2} \\ R_\gamma R_{\psi_1^2} & R_\gamma^2 R_{\psi_1^2} + R_{\psi_{2|1}^2} \end{pmatrix}.$$

단계3: 표준이변량 정규분포 $N_2(\mathbf{0}, \mathbf{I})$로부터 $Z_{\mathbf{D}}$를 생성한다.

단계4: 아래와 같이 정의되는 R_δ와 R_{A_1}, R_{A_2}을 계산한다.

$$R_\delta = \mathbf{d} - \left(\frac{R_\Sigma}{n} + \frac{R_\Psi}{m}\right)^{\frac{1}{2}} Z_{\mathbf{D}},$$

$$R_{A_j} = \Phi\left(\frac{R_{\mu_j - \nu_j}}{\sqrt{R_{\sigma_j^2} + R_{\psi_j^2}}}\right), \quad j = 1, 2.$$

단계5: 단계4에서 계산된 R_{A_1}, R_{A_2}를 이용하여 $R_\Delta = R_{A_1} - R_{A_2}$을 구한다.

단계6: 단계1~단계5을 독립적으로 K번 반복하여 얻어진 K개의 $R_{\Delta,1}, R_{\Delta,2}, \ldots, R_{\Delta,K}$의 $100(\alpha/2)\%$ 분위수와 $100(1-\alpha/2)\%$ 분위수는 각각 $\Delta = A_1 - A_2$의 $100(1-\alpha)\%$ 신뢰구간의 하한과 상한이 된다.

2.4. 붓스트랩방법

정확한 붓스트랩 분산(exact bootstrap variance)은 모든 가능한 붓스트랩표본에 대한 분산을 의미한다. 그러나 실제적으로는 이와 같은 이상적인 붓스트랩 분산은 몬테칼로 추정량으로 임의표본에 대한 분산으로 추정되는 것이 보통이다. 그러나 몬테칼로 붓스트랩 분산 추정량은 컴퓨터 작업량이 많아지는 단점과 재표집에 의한 오차 발생 등의 단점을 지니고 있다. 따라서 만약 정확한 붓스트랩 분산 추정이 가능하다면 이와 같은 문제는 피할 수 있을 것이다. 최근 Bandos 등 (2007)은 정확한 붓스트랩 분산 추정량을 계산하였다. 표 1의 자료구조 하에서 함수 w_{ik}를 아래와 같이 정의한다.

$$w_{ik} = \left(I(X_{i1} > Y_{k1}) + \frac{1}{2}I(X_{i1} = Y_{k1})\right) - \left(I(X_{i2} > Y_{k2}) + \frac{1}{2}I(X_{i2} = Y_{k2})\right),$$

여기서, I는 표시함수이다. 함수 w_{ik}를 이용하여 AUC차이의 정확한 붓스트랩 분산 추정량은

$$\mathrm{Var}_B(A_1^* - A_2^*) = \frac{\sum_{k=1}^{m}(\bar{w}_{\cdot k} - \bar{w}_{\cdot\cdot})^2}{m^2} + \frac{\sum_{i=1}^{n}(\bar{w}_{i\cdot} - \bar{w}_{\cdot\cdot})^2}{n^2} + \frac{\sum_{i=1}^{n}\sum_{k=1}^{m}(w_{ik} - \bar{w}_{i\cdot} - \bar{w}_{\cdot k} + \bar{w}_{\cdot\cdot})^2}{n^2 m^2} \tag{2.9}$$

이다. 식 (2.9)을 이용하여 붓스트랩방법에 의한 $\Delta = A_1 - A_2$의 $100(1-\alpha)\%$ 근사신뢰구간은 $(\hat{A}_1 - \hat{A}_2 - z_{\alpha/2}\sqrt{\mathrm{Var}_B(A_1^* - A_2^*)},\ \hat{A}_1 - \hat{A}_2 + z_{\alpha/2}\sqrt{\mathrm{Var}_B(A_1^* - A_2^*)})$이다. 여기서, $\hat{A}_1, \hat{A}_2$는 식 (2.1)의 비모수추정방법의 추정치와 동일하고, $z_{\alpha/2}$는 표준정규분포의 $(1-\alpha/2)100\%$ 백분위수이다.

3. 모의실험 결과

제 2절에서 간략히 소개한 4가지 방법을 본절에서는 모의실험을 통해 그 성능을 비교해보고자 한다. 표 1의 자료구조에서와 같이 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$은 각각 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $N_2(\boldsymbol{\nu}, \boldsymbol{\Psi})$로부터의 랜덤표본이라고 가정하자.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right), \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \begin{pmatrix} \psi_1^2 & \psi_{12} \\ \psi_{12} & \psi_2^2 \end{pmatrix}\right). \tag{3.1}$$

다음과 같은 3가지 시나리오를 고려하였다.

시나리오1. $(\mu_1, \mu_2) = (2.326, 1.812)$, 따라서, $\Delta = A_1 - A_2 = 0.05 (= \Phi(1.645) - \Phi(1.282) = 0.95 - 0.90)$ 이다.

시나리오2. $(\mu_1, \mu_2) = (2.326, 0.545)$, 따라서, $\Delta = A_1 - A_2 = 0.30 (= \Phi(1.645) - \Phi(0.385) = 0.95 - 0.65)$ 이다.

시나리오3. $(\mu_1, \mu_2) = (0.742, 0.545)$, 따라서, $\Delta = A_1 - A_2 = 0.05 (= \Phi(0.524) - \Phi(0.385) = 0.70 - 0.65)$ 이다.

다른 모수들은 $\sigma_1^2 = \sigma_2^2 = \psi_1^2 = \psi_2^2 = 1$, $\nu_1 = \nu_2 = 0$으로 고정하였다. 환자군의 표본의 크기 n과 정상군의 표본크기 m은 $(n, m) = (10, 10), (10, 20), (10, 50), (10, 100), (20, 20), (20, 50), (20, 100), (30, 30)$ $(50, 50), (100, 100)$으로 하였으며, 환자군에서의 상관계수는 $\sigma_{12} = 0.5, 0.7, 0.9$, 정상군에서의 상관계수는 $\psi_{12} = 0.5, 0.6, 0.7, 0.8, 0.9$을 고려하였다.

위의 3가지 시나리오는 Li 등 (2008)의 연구에서와 동일한 시나리오이다. 그러나, Li 등 (2008)의 연구에서는 상관계수의 값은 0.05, 0.5, 0.99였으며, $\sigma_{12} \le \psi_{12}$로 제한하였으나, 본 연구에서는 동일인의 상관계수의 값을 현실적으로 0.5~0.9를 고려하였다. 또한, 본 연구에서는 정확한 붓스트랩방법을 이용한 추정법을 포함하여 4가지 방법의 비교연구이다.

주어진 모수조합에 대하여 데이터를 생성하여 4가지 방법 각각의 $\Delta = A_1 - A_2$의 95% 신뢰구간을 구하고, Δ를 포함하는지 여부와 신뢰구간의 길이를 계산하였다. 2.3절의 GPQ에 의한 추정법의 단계 6에서 $K = 1000$으로 하였다. 총 1000번을 독립적으로 반복하여 포함확률(coverage probability)과 신뢰구간의 길이의 평균인 기대길이(expected length)를 구하였다. 모의실험은 SAS/IML를 이용하였다.

표 2와 3은 시나리오1에서의 결과를 보여주고 있다. 표 2는 환자군 표본수 $n = 10$, 정상군 표본수 $m = 10, 20, 50, 100$인 경우이며, 표 3은 $n = 20, 30$인 경우이다. 표 2에서 GPQ신뢰구간의 포함확률의 범위는 96.8%~99.4%으로 주어진 명목수준 95% 신뢰수준을 상회하고 있고, NP신뢰구간의 포함확률의 범위는 82.4%~89.4%, ML은 86.2%~93.7%, Bootstrap은 84.7%~93.1%으로 3가지의 신뢰구간은 신뢰수준 95%을 하회하고 있음을 알 수 있다. 특히 NP의 신뢰구간의 범위는 90%에도 미치지 못하고

표 2: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오1

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
10	10	0.5	0.5	0.983	0.890	0.907	0.898	0.332	0.268	0.239	0.270
10	10	0.5	0.6	0.981	0.894	0.916	0.901	0.325	0.265	0.237	0.268
10	10	0.5	0.7	0.987	0.880	0.919	0.894	0.304	0.250	0.218	0.253
10	10	0.5	0.8	0.984	0.869	0.900	0.881	0.285	0.236	0.205	0.240
10	10	0.5	0.9	0.986	0.869	0.877	0.881	0.267	0.226	0.194	0.231
10	10	0.7	0.5	0.978	0.879	0.907	0.884	0.305	0.254	0.221	0.257
10	10	0.7	0.6	0.981	0.879	0.904	0.895	0.291	0.239	0.208	0.243
10	10	0.7	0.7	0.984	0.881	0.900	0.893	0.276	0.229	0.195	0.234
10	10	0.7	0.8	0.981	0.862	0.884	0.878	0.255	0.215	0.182	0.220
10	10	0.7	0.9	0.993	0.861	0.877	0.872	0.235	0.207	0.168	0.213
10	10	0.9	0.5	0.980	0.868	0.888	0.876	0.268	0.224	0.193	0.229
10	10	0.9	0.6	0.988	0.874	0.896	0.879	0.248	0.213	0.179	0.219
10	10	0.9	0.7	0.991	0.847	0.870	0.857	0.231	0.200	0.165	0.206
10	10	0.9	0.8	0.984	0.840	0.867	0.847	0.208	0.189	0.150	0.195
10	10	0.9	0.9	0.994	0.844	0.862	0.855	0.183	0.174	0.134	0.181
10	20	0.5	0.5	0.986	0.894	0.930	0.922	0.274	0.203	0.203	0.224
10	20	0.5	0.6	0.969	0.871	0.905	0.912	0.265	0.198	0.196	0.218
10	20	0.5	0.7	0.987	0.883	0.916	0.917	0.261	0.198	0.192	0.214
10	20	0.5	0.8	0.979	0.885	0.907	0.916	0.253	0.195	0.186	0.208
10	20	0.5	0.9	0.976	0.873	0.887	0.897	0.243	0.188	0.175	0.198
10	20	0.7	0.5	0.976	0.881	0.909	0.924	0.241	0.186	0.183	0.210
10	20	0.7	0.6	0.981	0.880	0.920	0.924	0.233	0.181	0.174	0.203
10	20	0.7	0.7	0.975	0.855	0.884	0.888	0.222	0.173	0.165	0.192
10	20	0.7	0.8	0.979	0.865	0.881	0.900	0.214	0.167	0.157	0.183
10	20	0.7	0.9	0.982	0.837	0.874	0.873	0.210	0.166	0.152	0.179
10	20	0.9	0.5	0.979	0.856	0.903	0.920	0.190	0.154	0.151	0.184
10	20	0.9	0.6	0.986	0.848	0.890	0.904	0.181	0.150	0.142	0.176
10	20	0.9	0.7	0.980	0.834	0.863	0.898	0.172	0.142	0.133	0.167
10	20	0.9	0.8	0.972	0.858	0.885	0.910	0.164	0.143	0.128	0.162
10	20	0.9	0.9	0.982	0.848	0.882	0.892	0.150	0.133	0.115	0.149
10	50	0.5	0.5	0.980	0.865	0.919	0.912	0.246	0.175	0.180	0.189
10	50	0.5	0.6	0.969	0.865	0.909	0.900	0.247	0.175	0.180	0.187
10	50	0.5	0.7	0.979	0.872	0.926	0.906	0.247	0.174	0.177	0.184
10	50	0.5	0.8	0.974	0.855	0.917	0.888	0.241	0.172	0.173	0.179
10	50	0.5	0.9	0.975	0.886	0.923	0.911	0.243	0.177	0.175	0.181
10	50	0.7	0.5	0.977	0.838	0.909	0.915	0.206	0.146	0.152	0.165
10	50	0.7	0.6	0.977	0.837	0.880	0.885	0.205	0.148	0.149	0.163
10	50	0.7	0.7	0.976	0.855	0.903	0.906	0.203	0.146	0.147	0.160
10	50	0.7	0.8	0.971	0.859	0.877	0.888	0.202	0.149	0.145	0.159
10	50	0.7	0.9	0.978	0.857	0.908	0.893	0.199	0.147	0.141	0.153
10	50	0.9	0.5	0.973	0.854	0.910	0.931	0.152	0.115	0.119	0.141
10	50	0.9	0.6	0.968	0.831	0.891	0.909	0.148	0.112	0.115	0.134
10	50	0.9	0.7	0.981	0.837	0.886	0.910	0.145	0.113	0.111	0.132
10	50	0.9	0.8	0.977	0.845	0.892	0.905	0.142	0.112	0.107	0.128
10	50	0.9	0.9	0.982	0.844	0.888	0.890	0.139	0.108	0.102	0.120
10	100	0.5	0.5	0.971	0.886	0.930	0.907	0.244	0.171	0.175	0.176
10	100	0.5	0.6	0.973	0.865	0.917	0.893	0.239	0.166	0.170	0.170
10	100	0.5	0.7	0.969	0.872	0.913	0.891	0.238	0.169	0.170	0.171
10	100	0.5	0.8	0.971	0.865	0.911	0.884	0.240	0.168	0.170	0.168
10	100	0.5	0.9	0.982	0.884	0.937	0.888	0.238	0.165	0.167	0.164
10	100	0.7	0.5	0.982	0.855	0.907	0.898	0.202	0.141	0.144	0.150
10	100	0.7	0.6	0.977	0.840	0.896	0.873	0.199	0.140	0.141	0.148
10	100	0.7	0.7	0.979	0.867	0.916	0.903	0.200	0.139	0.141	0.144
10	100	0.7	0.8	0.977	0.839	0.888	0.872	0.197	0.141	0.138	0.144
10	100	0.7	0.9	0.974	0.852	0.893	0.866	0.198	0.139	0.137	0.140
10	100	0.9	0.5	0.975	0.840	0.891	0.901	0.141	0.103	0.106	0.118
10	100	0.9	0.6	0.989	0.842	0.895	0.909	0.142	0.105	0.105	0.118
10	100	0.9	0.7	0.983	0.828	0.884	0.878	0.136	0.100	0.100	0.111
10	100	0.9	0.8	0.987	0.843	0.895	0.890	0.138	0.105	0.100	0.112
10	100	0.9	0.9	0.982	0.842	0.885	0.876	0.135	0.101	0.096	0.107

시나리오1은 $A_1 = 0.95, A_2 = 0.90$인 경우. 여기서, $A_i, i = 1,2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

표 3: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오1

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
20	20	0.5	0.5	0.965	0.919	0.918	0.925	0.200	0.182	0.163	0.184
20	20	0.5	0.6	0.967	0.936	0.916	0.943	0.196	0.182	0.161	0.184
20	20	0.5	0.7	0.952	0.920	0.910	0.924	0.184	0.170	0.150	0.173
20	20	0.5	0.8	0.976	0.931	0.914	0.941	0.173	0.164	0.143	0.166
20	20	0.5	0.9	0.972	0.912	0.918	0.919	0.161	0.153	0.133	0.156
20	20	0.7	0.5	0.970	0.936	0.927	0.945	0.185	0.172	0.152	0.174
20	20	0.7	0.6	0.971	0.945	0.931	0.951	0.177	0.168	0.145	0.171
20	20	0.7	0.7	0.972	0.921	0.913	0.930	0.165	0.157	0.135	0.160
20	20	0.7	0.8	0.974	0.927	0.914	0.931	0.152	0.149	0.126	0.152
20	20	0.7	0.9	0.967	0.914	0.899	0.918	0.139	0.137	0.116	0.141
20	20	0.9	0.5	0.962	0.916	0.908	0.925	0.163	0.154	0.135	0.157
20	20	0.9	0.6	0.977	0.932	0.921	0.936	0.152	0.149	0.127	0.152
20	20	0.9	0.7	0.960	0.918	0.902	0.928	0.140	0.137	0.117	0.141
20	20	0.9	0.8	0.971	0.897	0.894	0.905	0.125	0.127	0.105	0.131
20	20	0.9	0.9	0.973	0.904	0.899	0.920	0.110	0.117	0.095	0.121
20	50	0.5	0.5	0.961	0.918	0.933	0.947	0.165	0.139	0.138	0.152
20	50	0.5	0.6	0.957	0.896	0.910	0.922	0.161	0.138	0.135	0.150
20	50	0.5	0.7	0.963	0.918	0.932	0.947	0.159	0.136	0.132	0.146
20	50	0.5	0.8	0.972	0.909	0.935	0.932	0.152	0.132	0.126	0.139
20	50	0.5	0.9	0.955	0.901	0.917	0.918	0.148	0.132	0.123	0.137
20	50	0.7	0.5	0.960	0.890	0.916	0.935	0.143	0.120	0.120	0.136
20	50	0.7	0.6	0.969	0.885	0.905	0.919	0.139	0.120	0.117	0.134
20	50	0.7	0.7	0.962	0.894	0.909	0.931	0.134	0.116	0.112	0.127
20	50	0.7	0.8	0.970	0.911	0.931	0.936	0.129	0.114	0.108	0.123
20	50	0.7	0.9	0.964	0.881	0.904	0.918	0.124	0.113	0.104	0.119
20	50	0.9	0.5	0.963	0.880	0.916	0.935	0.113	0.096	0.099	0.116
20	50	0.9	0.6	0.965	0.886	0.916	0.936	0.108	0.094	0.094	0.112
20	50	0.9	0.7	0.956	0.877	0.896	0.931	0.102	0.092	0.089	0.107
20	50	0.9	0.8	0.968	0.875	0.899	0.923	0.097	0.089	0.084	0.101
20	50	0.9	0.9	0.954	0.878	0.892	0.915	0.091	0.086	0.079	0.095
20	100	0.5	0.5	0.962	0.920	0.943	0.943	0.152	0.126	0.126	0.135
20	100	0.5	0.6	0.953	0.901	0.931	0.926	0.151	0.127	0.125	0.135
20	100	0.5	0.7	0.960	0.892	0.920	0.918	0.148	0.124	0.122	0.131
20	100	0.5	0.8	0.968	0.904	0.921	0.920	0.147	0.125	0.121	0.129
20	100	0.5	0.9	0.969	0.910	0.930	0.920	0.144	0.123	0.118	0.126
20	100	0.7	0.5	0.962	0.893	0.920	0.929	0.129	0.109	0.108	0.120
20	100	0.7	0.6	0.957	0.869	0.910	0.907	0.126	0.106	0.105	0.116
20	100	0.7	0.7	0.965	0.892	0.920	0.925	0.123	0.106	0.102	0.114
20	100	0.7	0.8	0.965	0.919	0.929	0.940	0.123	0.106	0.102	0.112
20	100	0.7	0.9	0.970	0.906	0.915	0.920	0.120	0.106	0.100	0.110
20	100	0.9	0.5	0.958	0.863	0.898	0.925	0.096	0.080	0.084	0.096
20	100	0.9	0.6	0.970	0.878	0.915	0.931	0.094	0.079	0.081	0.093
20	100	0.9	0.7	0.966	0.884	0.909	0.922	0.091	0.079	0.078	0.091
20	100	0.9	0.8	0.972	0.890	0.901	0.926	0.089	0.079	0.076	0.088
20	100	0.9	0.9	0.966	0.892	0.905	0.920	0.085	0.077	0.072	0.083
30	30	0.5	0.5	0.970	0.951	0.944	0.953	0.159	0.151	0.136	0.152
30	30	0.5	0.6	0.963	0.936	0.932	0.941	0.152	0.144	0.129	0.145
30	30	0.5	0.7	0.958	0.938	0.928	0.941	0.144	0.138	0.123	0.140
30	30	0.5	0.8	0.956	0.929	0.917	0.931	0.134	0.131	0.115	0.132
30	30	0.5	0.9	0.972	0.944	0.932	0.949	0.127	0.124	0.109	0.126
30	30	0.7	0.5	0.970	0.944	0.931	0.946	0.143	0.137	0.122	0.139
30	30	0.7	0.6	0.964	0.927	0.915	0.933	0.135	0.131	0.115	0.133
30	30	0.7	0.7	0.972	0.933	0.913	0.938	0.128	0.127	0.109	0.129
30	30	0.7	0.8	0.962	0.924	0.914	0.929	0.120	0.120	0.103	0.122
30	30	0.7	0.9	0.965	0.926	0.908	0.932	0.108	0.109	0.093	0.111
30	30	0.9	0.5	0.957	0.927	0.918	0.931	0.127	0.125	0.109	0.127
30	30	0.9	0.6	0.967	0.925	0.909	0.931	0.117	0.117	0.101	0.119
30	30	0.9	0.7	0.965	0.933	0.903	0.937	0.108	0.111	0.094	0.113
30	30	0.9	0.8	0.956	0.921	0.899	0.922	0.098	0.103	0.086	0.105
30	30	0.9	0.9	0.977	0.912	0.894	0.921	0.086	0.091	0.077	0.094

시나리오1은 $A_1 = 0.95, A_2 = 0.90$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

있다. 이러한 특징은 표 3에서도 계속 나타나며, 표 2와 차이가 나는 점은 GPQ신뢰구간의 포함확률의 범위는 95.2%~97.7%으로 명목수준 95%를 상회하는 정도가 표 1에 비하여 다소 작고, 또한 NP는 86.3%~95.1%, ML은 89.2%~94.4%, Bootstrap은 90.5%~95.3%으로 명목수준 95%를 하회하는 정도가 표 1에 비하여 다소 작아졌음을 알 수 있다. 그리고, 표 2에서는 NP, ML, Bootstrap의 3가지 방법 중 bootstrap신뢰구간의 포함확률이 명목 신뢰수준인 95%에 보다 근사함을 알 수 있다.

시나리오2는 $A_1 = 0.95, A_2 = 0.65$으로 $\Delta = 0.3$인 경우로 결과는 표 4와 5이다. 표 4는 환자군 표본수 $n = 10$, 정상군 표본수 $m = 10, 20, 50, 100$인 경우이며, 표 5는 $n = 20, 30$인 경우이다. 표 4에서 GPQ신뢰구간의 포함확률의 범위는 95.9%~98.8%, NP는 89.8%~95.0%, ML는 88.8%~93.5%, Bootstrap는 88.8%~94.3%으로 시나리오1의 표 2 결과와 동일하게 GPQ는 명목수준 95%을 상회하는 보수적인 특성을 보이고 그 외 3가지 신뢰구간은 95%를 하회하는 특징을 나타낸다. 표 5에서는 GPQ신뢰구간의 포함확률범위는 94.7%~97.6%, NP는 92.6%~96.2%, ML는 91.4%~95.4%, Bootstrap은 91.7%~95.7%으로 60개의 경우 중 포함확률이 95%미만인 경우는 GPQ는 2/60, NP는 40/60, ML은 57/60, Bootstrap은 50/60이다.

시나리오3은 $A_1 = 0.70, A_2 = 0.65$으로 $\Delta = 0.05$이며 결과는 표 6과 7이다. 표 6에서 GPQ신뢰구간의 포함확률범위는 96.2%~99.2%으로 시나리오1의 표 2와 시나리오2의 표 4와 동일하게 95%신뢰수준을 항상 상회하지만, NP는 88.4%~96.4%, ML은 90.4%~95.5%, Bootstrap는 91.3%~97.0%으로 3가지 방법은 시나리오1의 표 2와 시나리오2의 표 4와 다르게 95%신뢰수준을 항상 하회하지는 않는다. 60개의 경우 중 포함확률이 95%미만인 경우는 NP는 46/60, ML은 58/60, Bootstrap은 32/60임을 알 수 있다. 또한, 표 7은 GPQ신뢰구간의 포함확률범위는 94.3%~98.0%, NP는 89.4%~97.1%, ML는 91.4%~96.3%, Bootstrap은 92.9%~97.2%으로 60개의 경우 중 포함확률이 95%미만인 경우는 GPQ는 1/60, NP는 44/60, ML은 52/60, Bootstrap은 29/60이다.

기대길이 관점에서 살펴보면, 시나리오1과 시나리오3에서는 GPQ방법의 기대길이가 대체적으로 크고, 시나리오2에서는 NP의 기대길이가 다른 3가지 방법에 비하여 큰 경향이 있다.

그림1~3은 표 2~7의 포함확률의 수치를 그림으로 나타내고 있다. 표본수가 큰 $(n, m) = (50, 50), (100, 100)$ 경우에서 4가지 방법들의 포함확률이 비슷함을 보이고, 특히 그림 2의 시나리오2와 그림 3의 시나리오3에서는 그림 1의 시나리오1에 비하여 포함확률이 방법들 간에 매우 유사함을 확인할 수 있다.

4. 맺음말

본 논문에서는 두 종류의 진단검사가 동일인에게 시행되었을 때 진단검사의 정확성을 재는 기준으로 많이 사용되는 ROC곡선 아래 면적(AUC)의 차이를 비모수적방법, 최대우도법, 일반화추축량을 이용한 방법, 붓스트랩방법의 4가지 방법으로 비교하였다. 동일인의 상관계수의 크기는 현실적으로 0.5~0.9의 범위로 제한하였으며, 환자군과 정상군의 표본의 크기를 다양하게 하면서 4가지 방법의 신뢰구간의 성질을 포함확률과 기대길이측면에서 살펴보았다. 표본의 크기가 적당히 커지면 근사이론에 의하여 4가지 방법들의 성능이 비슷해질 것이므로, 본 연구에서는 환자군의 표본크기는 30이하만을 살펴보았다. 3가지 시나리오를 설정하였으며, 첫 번째 시나리오는 두 진단검사의 AUC가 각각 0.95, 0.90인 경우, 두 번째 시나리오는 두 진단검사의 AUC가 각각 0.95, 0.65인 경우, 세 번째 시나리오는 두 진단검사의 AUC가 각각 0.70, 0.65인 경우이다.

모의실험결과 일반화추축량을 이용한 방법은 명목수준인 95%보다는 포함확률이 상회하는 보수적인 신뢰구간임을 확인할 수 있었고, 나머지 3가지 방법은 포함확률이 명목수준을 하회하는 특성을 나타냄을 알 수 있었다. AUC의 차이가 큰 두 번째 시나리오에서는 GPQ방법을 제외한 3가지 방법이 포

표 4: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오2

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
10	10	0.5	0.5	0.980	0.927	0.917	0.921	0.432	0.473	0.415	0.463
10	10	0.5	0.6	0.977	0.905	0.894	0.904	0.417	0.464	0.404	0.455
10	10	0.5	0.7	0.981	0.913	0.892	0.911	0.410	0.457	0.397	0.449
10	10	0.5	0.8	0.977	0.920	0.902	0.917	0.393	0.445	0.384	0.438
10	10	0.5	0.9	0.981	0.925	0.888	0.917	0.386	0.444	0.379	0.437
10	10	0.7	0.5	0.978	0.924	0.920	0.922	0.410	0.455	0.395	0.447
10	10	0.7	0.6	0.980	0.917	0.907	0.914	0.397	0.449	0.386	0.441
10	10	0.7	0.7	0.975	0.924	0.901	0.920	0.383	0.441	0.375	0.434
10	10	0.7	0.8	0.979	0.922	0.901	0.922	0.375	0.433	0.369	0.428
10	10	0.7	0.9	0.988	0.925	0.933	0.924	0.360	0.431	0.361	0.426
10	10	0.9	0.5	0.972	0.918	0.908	0.917	0.385	0.441	0.377	0.434
10	10	0.9	0.6	0.966	0.921	0.905	0.920	0.373	0.435	0.369	0.429
10	10	0.9	0.7	0.986	0.944	0.921	0.942	0.361	0.432	0.361	0.427
10	10	0.9	0.8	0.982	0.947	0.930	0.942	0.347	0.424	0.350	0.420
10	10	0.9	0.9	0.973	0.935	0.924	0.936	0.334	0.417	0.342	0.415
10	20	0.5	0.5	0.973	0.942	0.925	0.933	0.373	0.412	0.358	0.393
10	20	0.5	0.6	0.977	0.930	0.926	0.924	0.365	0.406	0.351	0.386
10	20	0.5	0.7	0.981	0.948	0.926	0.941	0.363	0.407	0.349	0.384
10	20	0.5	0.8	0.979	0.926	0.904	0.923	0.356	0.402	0.340	0.377
10	20	0.5	0.9	0.967	0.928	0.906	0.918	0.356	0.404	0.339	0.376
10	20	0.7	0.5	0.978	0.922	0.910	0.921	0.343	0.388	0.334	0.372
10	20	0.7	0.6	0.977	0.950	0.932	0.943	0.343	0.397	0.336	0.376
10	20	0.7	0.7	0.971	0.939	0.922	0.931	0.337	0.391	0.328	0.369
10	20	0.7	0.8	0.970	0.925	0.916	0.923	0.331	0.385	0.321	0.362
10	20	0.7	0.9	0.975	0.931	0.913	0.917	0.328	0.385	0.318	0.360
10	20	0.9	0.5	0.961	0.930	0.915	0.920	0.314	0.378	0.318	0.361
10	20	0.9	0.6	0.975	0.937	0.934	0.933	0.310	0.377	0.315	0.358
10	20	0.9	0.7	0.963	0.934	0.919	0.926	0.304	0.374	0.308	0.354
10	20	0.9	0.8	0.963	0.932	0.916	0.926	0.298	0.372	0.302	0.350
10	20	0.9	0.9	0.971	0.947	0.928	0.942	0.292	0.370	0.296	0.347
10	50	0.5	0.5	0.969	0.918	0.913	0.914	0.341	0.356	0.317	0.338
10	50	0.5	0.6	0.976	0.929	0.923	0.918	0.341	0.353	0.316	0.334
10	50	0.5	0.7	0.974	0.919	0.916	0.908	0.340	0.351	0.313	0.331
10	50	0.5	0.8	0.969	0.917	0.905	0.901	0.331	0.345	0.306	0.325
10	50	0.5	0.9	0.975	0.923	0.909	0.909	0.335	0.351	0.308	0.327
10	50	0.7	0.5	0.976	0.937	0.920	0.926	0.315	0.340	0.298	0.322
10	50	0.7	0.6	0.965	0.924	0.919	0.908	0.308	0.332	0.291	0.315
10	50	0.7	0.7	0.975	0.940	0.919	0.925	0.310	0.336	0.291	0.316
10	50	0.7	0.8	0.975	0.926	0.914	0.924	0.309	0.334	0.290	0.313
10	50	0.7	0.9	0.975	0.931	0.918	0.921	0.308	0.334	0.287	0.312
10	50	0.9	0.5	0.959	0.933	0.910	0.916	0.276	0.318	0.273	0.302
10	50	0.9	0.6	0.978	0.939	0.926	0.929	0.273	0.318	0.270	0.301
10	50	0.9	0.7	0.981	0.947	0.930	0.932	0.271	0.316	0.267	0.297
10	50	0.9	0.8	0.972	0.949	0.925	0.936	0.269	0.315	0.264	0.296
10	50	0.9	0.9	0.972	0.939	0.932	0.931	0.270	0.316	0.263	0.294
10	100	0.5	0.5	0.966	0.908	0.903	0.900	0.330	0.329	0.300	0.313
10	100	0.5	0.6	0.980	0.907	0.918	0.904	0.330	0.327	0.300	0.310
10	100	0.5	0.7	0.970	0.918	0.908	0.910	0.331	0.331	0.300	0.313
10	100	0.5	0.8	0.980	0.909	0.903	0.888	0.328	0.326	0.297	0.308
10	100	0.5	0.9	0.966	0.898	0.895	0.891	0.327	0.327	0.295	0.307
10	100	0.7	0.5	0.968	0.907	0.908	0.898	0.300	0.309	0.278	0.294
10	100	0.7	0.6	0.970	0.926	0.908	0.909	0.303	0.313	0.279	0.297
10	100	0.7	0.7	0.971	0.914	0.902	0.902	0.301	0.310	0.276	0.292
10	100	0.7	0.8	0.965	0.922	0.910	0.904	0.298	0.309	0.274	0.291
10	100	0.7	0.9	0.980	0.935	0.922	0.919	0.298	0.307	0.273	0.289
10	100	0.9	0.5	0.960	0.915	0.901	0.901	0.261	0.290	0.253	0.276
10	100	0.9	0.6	0.979	0.933	0.935	0.925	0.263	0.292	0.254	0.276
10	100	0.9	0.7	0.961	0.922	0.911	0.912	0.260	0.289	0.251	0.273
10	100	0.9	0.8	0.968	0.918	0.914	0.912	0.262	0.290	0.251	0.273
10	100	0.9	0.9	0.977	0.940	0.926	0.930	0.259	0.289	0.248	0.271

시나리오2은 $A_1 = 0.95, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1,2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

표 5: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오2

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
20	20	0.5	0.5	0.961	0.933	0.930	0.930	0.297	0.322	0.294	0.320
20	20	0.5	0.6	0.959	0.930	0.914	0.929	0.290	0.317	0.289	0.315
20	20	0.5	0.7	0.976	0.942	0.938	0.941	0.285	0.314	0.284	0.312
20	20	0.5	0.8	0.957	0.938	0.927	0.937	0.277	0.308	0.277	0.306
20	20	0.5	0.9	0.962	0.936	0.930	0.937	0.268	0.302	0.270	0.301
20	20	0.7	0.5	0.960	0.938	0.925	0.938	0.283	0.311	0.282	0.310
20	20	0.7	0.6	0.967	0.930	0.922	0.927	0.276	0.307	0.276	0.305
20	20	0.7	0.7	0.963	0.937	0.927	0.937	0.269	0.302	0.270	0.300
20	20	0.7	0.8	0.961	0.950	0.927	0.951	0.263	0.298	0.265	0.297
20	20	0.7	0.9	0.963	0.944	0.930	0.945	0.253	0.292	0.258	0.291
20	20	0.9	0.5	0.963	0.944	0.924	0.943	0.269	0.301	0.271	0.300
20	20	0.9	0.6	0.963	0.939	0.929	0.939	0.261	0.297	0.264	0.296
20	20	0.9	0.7	0.967	0.948	0.950	0.950	0.254	0.291	0.258	0.291
20	20	0.9	0.8	0.962	0.947	0.930	0.948	0.247	0.287	0.251	0.287
20	20	0.9	0.9	0.962	0.936	0.932	0.936	0.239	0.282	0.245	0.282
20	50	0.5	0.5	0.969	0.954	0.932	0.944	0.251	0.272	0.246	0.262
20	50	0.5	0.6	0.974	0.955	0.928	0.950	0.248	0.273	0.243	0.261
20	50	0.5	0.7	0.964	0.940	0.933	0.936	0.246	0.270	0.241	0.257
20	50	0.5	0.8	0.963	0.938	0.926	0.935	0.243	0.271	0.238	0.256
20	50	0.5	0.9	0.960	0.947	0.926	0.934	0.239	0.269	0.234	0.252
20	50	0.7	0.5	0.964	0.952	0.939	0.946	0.235	0.262	0.233	0.251
20	50	0.7	0.6	0.967	0.951	0.938	0.942	0.232	0.261	0.230	0.249
20	50	0.7	0.7	0.947	0.933	0.914	0.935	0.228	0.261	0.227	0.247
20	50	0.7	0.8	0.963	0.962	0.935	0.950	0.226	0.260	0.224	0.244
20	50	0.7	0.9	0.951	0.950	0.933	0.941	0.223	0.257	0.221	0.240
20	50	0.9	0.5	0.965	0.961	0.944	0.950	0.215	0.249	0.218	0.239
20	50	0.9	0.6	0.951	0.942	0.935	0.935	0.212	0.249	0.215	0.236
20	50	0.9	0.7	0.952	0.943	0.928	0.939	0.208	0.246	0.211	0.233
20	50	0.9	0.8	0.964	0.957	0.944	0.948	0.205	0.245	0.207	0.230
20	50	0.9	0.9	0.958	0.954	0.941	0.945	0.202	0.245	0.204	0.228
20	100	0.5	0.5	0.958	0.946	0.938	0.941	0.235	0.249	0.228	0.239
20	100	0.5	0.6	0.962	0.949	0.938	0.937	0.234	0.250	0.227	0.239
20	100	0.5	0.7	0.965	0.948	0.930	0.935	0.232	0.249	0.225	0.237
20	100	0.5	0.8	0.962	0.942	0.923	0.931	0.230	0.247	0.222	0.234
20	100	0.5	0.9	0.958	0.950	0.929	0.937	0.231	0.250	0.223	0.236
20	100	0.7	0.5	0.962	0.949	0.936	0.940	0.216	0.235	0.212	0.226
20	100	0.7	0.6	0.952	0.938	0.929	0.935	0.215	0.235	0.210	0.225
20	100	0.7	0.7	0.950	0.933	0.918	0.917	0.213	0.234	0.208	0.222
20	100	0.7	0.8	0.966	0.960	0.937	0.945	0.212	0.234	0.207	0.222
20	100	0.7	0.9	0.963	0.943	0.927	0.931	0.209	0.233	0.204	0.220
20	100	0.9	0.5	0.956	0.938	0.937	0.937	0.195	0.220	0.196	0.212
20	100	0.9	0.6	0.955	0.954	0.929	0.944	0.193	0.222	0.194	0.211
20	100	0.9	0.7	0.960	0.953	0.933	0.942	0.193	0.222	0.193	0.210
20	100	0.9	0.8	0.958	0.952	0.927	0.942	0.190	0.220	0.190	0.208
20	100	0.9	0.9	0.958	0.948	0.939	0.936	0.190	0.221	0.190	0.208
30	30	0.5	0.5	0.948	0.926	0.925	0.926	0.243	0.259	0.241	0.258
30	30	0.5	0.6	0.963	0.949	0.930	0.949	0.237	0.256	0.236	0.255
30	30	0.5	0.7	0.960	0.941	0.938	0.941	0.232	0.252	0.232	0.251
30	30	0.5	0.8	0.968	0.947	0.935	0.946	0.228	0.248	0.228	0.248
30	30	0.5	0.9	0.957	0.951	0.937	0.951	0.221	0.244	0.222	0.243
30	30	0.7	0.5	0.957	0.937	0.931	0.937	0.232	0.252	0.232	0.251
30	30	0.7	0.6	0.959	0.937	0.928	0.936	0.226	0.248	0.227	0.247
30	30	0.7	0.7	0.955	0.934	0.929	0.933	0.221	0.244	0.222	0.243
30	30	0.7	0.8	0.971	0.950	0.954	0.950	0.216	0.239	0.217	0.239
30	30	0.7	0.9	0.963	0.944	0.943	0.944	0.210	0.235	0.212	0.235
30	30	0.9	0.5	0.953	0.939	0.931	0.938	0.220	0.243	0.222	0.242
30	30	0.9	0.6	0.965	0.952	0.952	0.952	0.215	0.240	0.217	0.239
30	30	0.9	0.7	0.957	0.940	0.938	0.940	0.209	0.234	0.212	0.234
30	30	0.9	0.8	0.962	0.953	0.942	0.953	0.203	0.231	0.207	0.231
30	30	0.9	0.9	0.963	0.957	0.940	0.957	0.197	0.226	0.202	0.227

시나리오2은 $A_1 = 0.95, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

표 6: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오3

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
10	10	0.5	0.5	0.988	0.953	0.942	0.950	0.534	0.517	0.445	0.509
10	10	0.5	0.6	0.992	0.957	0.937	0.952	0.513	0.491	0.421	0.486
10	10	0.5	0.7	0.987	0.955	0.938	0.956	0.487	0.466	0.395	0.462
10	10	0.5	0.8	0.982	0.954	0.936	0.955	0.462	0.442	0.372	0.440
10	10	0.5	0.9	0.980	0.951	0.927	0.957	0.431	0.415	0.345	0.415
10	10	0.7	0.5	0.986	0.959	0.935	0.957	0.488	0.463	0.396	0.460
10	10	0.7	0.6	0.990	0.958	0.936	0.958	0.473	0.445	0.376	0.443
10	10	0.7	0.7	0.992	0.962	0.955	0.962	0.441	0.415	0.347	0.414
10	10	0.7	0.8	0.985	0.955	0.935	0.960	0.412	0.387	0.317	0.389
10	10	0.7	0.9	0.987	0.956	0.933	0.962	0.376	0.354	0.285	0.358
10	10	0.9	0.5	0.985	0.954	0.939	0.956	0.432	0.411	0.343	0.412
10	10	0.9	0.6	0.990	0.947	0.932	0.954	0.407	0.386	0.318	0.388
10	10	0.9	0.7	0.992	0.959	0.930	0.966	0.375	0.351	0.283	0.356
10	10	0.9	0.8	0.987	0.964	0.938	0.970	0.340	0.316	0.245	0.322
10	10	0.9	0.9	0.992	0.944	0.928	0.955	0.294	0.272	0.204	0.281
10	20	0.5	0.5	0.972	0.925	0.927	0.936	0.456	0.396	0.385	0.426
10	20	0.5	0.6	0.973	0.936	0.929	0.947	0.440	0.385	0.369	0.409
10	20	0.5	0.7	0.981	0.947	0.946	0.956	0.433	0.378	0.356	0.397
10	20	0.5	0.8	0.975	0.926	0.920	0.933	0.423	0.373	0.344	0.385
10	20	0.5	0.9	0.972	0.945	0.932	0.944	0.404	0.361	0.325	0.367
10	20	0.7	0.5	0.980	0.929	0.950	0.958	0.399	0.339	0.331	0.377
10	20	0.7	0.6	0.975	0.924	0.925	0.942	0.385	0.329	0.317	0.363
10	20	0.7	0.7	0.979	0.931	0.933	0.941	0.371	0.320	0.299	0.347
10	20	0.7	0.8	0.983	0.931	0.928	0.953	0.362	0.311	0.284	0.332
10	20	0.7	0.9	0.982	0.955	0.941	0.961	0.348	0.300	0.266	0.312
10	20	0.9	0.5	0.985	0.917	0.947	0.960	0.317	0.263	0.268	0.319
10	20	0.9	0.6	0.980	0.908	0.938	0.957	0.301	0.251	0.248	0.300
10	20	0.9	0.7	0.984	0.917	0.929	0.966	0.285	0.238	0.228	0.281
10	20	0.9	0.8	0.983	0.928	0.931	0.966	0.270	0.223	0.206	0.258
10	20	0.9	0.9	0.991	0.947	0.936	0.968	0.248	0.207	0.179	0.232
10	50	0.5	0.5	0.978	0.905	0.924	0.927	0.418	0.347	0.343	0.365
10	50	0.5	0.6	0.971	0.914	0.930	0.925	0.415	0.340	0.333	0.353
10	50	0.5	0.7	0.962	0.921	0.918	0.922	0.411	0.342	0.329	0.350
10	50	0.5	0.8	0.968	0.917	0.910	0.919	0.402	0.339	0.319	0.342
10	50	0.5	0.9	0.977	0.917	0.919	0.920	0.403	0.334	0.312	0.331
10	50	0.7	0.5	0.975	0.925	0.948	0.948	0.352	0.277	0.280	0.306
10	50	0.7	0.6	0.976	0.918	0.923	0.939	0.351	0.279	0.274	0.301
10	50	0.7	0.7	0.984	0.921	0.927	0.933	0.347	0.277	0.266	0.293
10	50	0.7	0.8	0.976	0.904	0.924	0.915	0.340	0.271	0.256	0.282
10	50	0.7	0.9	0.977	0.914	0.921	0.924	0.338	0.268	0.249	0.271
10	50	0.9	0.5	0.978	0.892	0.944	0.952	0.253	0.184	0.203	0.233
10	50	0.9	0.6	0.979	0.890	0.936	0.949	0.253	0.183	0.194	0.226
10	50	0.9	0.7	0.980	0.893	0.925	0.955	0.245	0.177	0.181	0.212
10	50	0.9	0.8	0.984	0.895	0.924	0.935	0.238	0.172	0.168	0.199
10	50	0.9	0.9	0.987	0.914	0.927	0.951	0.229	0.166	0.154	0.183
10	100	0.5	0.5	0.969	0.908	0.925	0.917	0.405	0.329	0.323	0.334
10	100	0.5	0.6	0.974	0.908	0.915	0.913	0.401	0.333	0.320	0.334
10	100	0.5	0.7	0.977	0.924	0.921	0.924	0.406	0.334	0.316	0.331
10	100	0.5	0.8	0.970	0.920	0.924	0.918	0.399	0.331	0.311	0.325
10	100	0.5	0.9	0.976	0.927	0.911	0.919	0.405	0.333	0.310	0.324
10	100	0.7	0.5	0.979	0.905	0.923	0.929	0.345	0.265	0.262	0.278
10	100	0.7	0.6	0.973	0.908	0.904	0.917	0.341	0.263	0.254	0.272
10	100	0.7	0.7	0.975	0.910	0.927	0.924	0.339	0.267	0.254	0.272
10	100	0.7	0.8	0.979	0.923	0.919	0.925	0.337	0.262	0.248	0.263
10	100	0.7	0.9	0.982	0.924	0.919	0.927	0.340	0.264	0.245	0.260
10	100	0.9	0.5	0.983	0.884	0.931	0.939	0.241	0.165	0.176	0.195
10	100	0.9	0.6	0.983	0.890	0.937	0.950	0.235	0.162	0.168	0.188
10	100	0.9	0.7	0.985	0.888	0.936	0.938	0.232	0.161	0.161	0.180
10	100	0.9	0.8	0.985	0.907	0.930	0.942	0.232	0.160	0.155	0.174
10	100	0.9	0.9	0.990	0.923	0.933	0.940	0.229	0.158	0.148	0.165

시나리오3은 $A_1 = 0.70, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

표 7: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률과 기대길이- 시나리오3

n	m	σ_{12}	ψ_{12}	포함확률				기대길이			
				GPQ	NP	ML	Bootstrap	GPQ	NP	ML	Bootstrap
20	20	0.5	0.5	0.976	0.947	0.943	0.946	0.348	0.351	0.319	0.350
20	20	0.5	0.6	0.961	0.952	0.936	0.951	0.332	0.337	0.304	0.336
20	20	0.5	0.7	0.970	0.947	0.941	0.949	0.316	0.317	0.285	0.317
20	20	0.5	0.8	0.964	0.958	0.933	0.961	0.295	0.300	0.267	0.301
20	20	0.5	0.9	0.967	0.949	0.943	0.949	0.275	0.278	0.247	0.280
20	20	0.7	0.5	0.980	0.971	0.963	0.972	0.321	0.320	0.288	0.320
20	20	0.7	0.6	0.971	0.950	0.947	0.951	0.296	0.299	0.266	0.300
20	20	0.7	0.7	0.975	0.947	0.944	0.948	0.276	0.281	0.248	0.282
20	20	0.7	0.8	0.971	0.968	0.939	0.971	0.259	0.259	0.228	0.262
20	20	0.7	0.9	0.971	0.955	0.934	0.957	0.236	0.240	0.206	0.243
20	20	0.9	0.5	0.961	0.942	0.925	0.944	0.275	0.281	0.250	0.283
20	20	0.9	0.6	0.968	0.961	0.942	0.963	0.256	0.259	0.227	0.261
20	20	0.9	0.7	0.967	0.952	0.942	0.958	0.231	0.233	0.202	0.237
20	20	0.9	0.8	0.968	0.938	0.924	0.944	0.209	0.210	0.178	0.214
20	20	0.9	0.9	0.974	0.955	0.938	0.960	0.176	0.178	0.146	0.183
20	50	0.5	0.5	0.955	0.918	0.937	0.942	0.292	0.263	0.267	0.286
20	50	0.5	0.6	0.967	0.945	0.946	0.956	0.281	0.261	0.259	0.279
20	50	0.5	0.7	0.965	0.936	0.935	0.947	0.276	0.255	0.251	0.269
20	50	0.5	0.8	0.962	0.925	0.935	0.937	0.269	0.253	0.243	0.263
20	50	0.5	0.9	0.960	0.937	0.935	0.945	0.263	0.250	0.235	0.254
20	50	0.7	0.5	0.964	0.928	0.949	0.951	0.248	0.218	0.227	0.248
20	50	0.7	0.6	0.969	0.932	0.957	0.964	0.239	0.214	0.217	0.238
20	50	0.7	0.7	0.967	0.931	0.939	0.952	0.232	0.208	0.207	0.228
20	50	0.7	0.8	0.958	0.907	0.927	0.929	0.221	0.203	0.196	0.217
20	50	0.7	0.9	0.943	0.921	0.914	0.929	0.214	0.201	0.187	0.208
20	50	0.9	0.5	0.953	0.896	0.930	0.944	0.192	0.158	0.178	0.199
20	50	0.9	0.6	0.961	0.903	0.944	0.953	0.180	0.150	0.164	0.186
20	50	0.9	0.7	0.968	0.925	0.955	0.964	0.170	0.144	0.153	0.175
20	50	0.9	0.8	0.971	0.923	0.927	0.964	0.158	0.138	0.137	0.160
20	50	0.9	0.9	0.973	0.926	0.939	0.961	0.144	0.129	0.122	0.143
20	100	0.5	0.5	0.955	0.914	0.927	0.934	0.270	0.244	0.245	0.259
20	100	0.5	0.6	0.972	0.941	0.956	0.957	0.268	0.245	0.243	0.257
20	100	0.5	0.7	0.959	0.921	0.929	0.936	0.262	0.241	0.236	0.249
20	100	0.5	0.8	0.960	0.935	0.938	0.940	0.263	0.240	0.233	0.245
20	100	0.5	0.9	0.956	0.929	0.930	0.930	0.256	0.240	0.228	0.241
20	100	0.7	0.5	0.969	0.921	0.933	0.944	0.225	0.196	0.201	0.217
20	100	0.7	0.6	0.962	0.923	0.942	0.943	0.219	0.194	0.195	0.211
20	100	0.7	0.7	0.967	0.938	0.942	0.952	0.219	0.196	0.193	0.209
20	100	0.7	0.8	0.953	0.928	0.931	0.940	0.211	0.192	0.185	0.200
20	100	0.7	0.9	0.950	0.929	0.922	0.934	0.210	0.190	0.180	0.193
20	100	0.9	0.5	0.962	0.894	0.939	0.955	0.161	0.127	0.145	0.161
20	100	0.9	0.6	0.967	0.896	0.951	0.954	0.155	0.124	0.137	0.153
20	100	0.9	0.7	0.959	0.907	0.938	0.946	0.151	0.122	0.129	0.145
20	100	0.9	0.8	0.966	0.922	0.948	0.955	0.145	0.119	0.121	0.136
20	100	0.9	0.9	0.967	0.929	0.935	0.950	0.137	0.117	0.112	0.126
30	30	0.5	0.5	0.962	0.940	0.938	0.940	0.279	0.284	0.262	0.283
30	30	0.5	0.6	0.959	0.945	0.932	0.945	0.263	0.271	0.249	0.270
30	30	0.5	0.7	0.963	0.960	0.946	0.960	0.250	0.256	0.234	0.256
30	30	0.5	0.8	0.972	0.957	0.946	0.959	0.236	0.240	0.219	0.241
30	30	0.5	0.9	0.967	0.960	0.954	0.962	0.217	0.224	0.203	0.225
30	30	0.7	0.5	0.958	0.947	0.946	0.947	0.249	0.255	0.233	0.255
30	30	0.7	0.6	0.958	0.943	0.943	0.943	0.235	0.241	0.219	0.242
30	30	0.7	0.7	0.956	0.947	0.937	0.948	0.221	0.226	0.204	0.227
30	30	0.7	0.8	0.967	0.956	0.949	0.958	0.201	0.206	0.185	0.208
30	30	0.7	0.9	0.957	0.947	0.939	0.948	0.183	0.188	0.166	0.191
30	30	0.9	0.5	0.961	0.966	0.942	0.967	0.219	0.225	0.204	0.227
30	30	0.9	0.6	0.956	0.941	0.929	0.943	0.202	0.207	0.186	0.209
30	30	0.9	0.7	0.964	0.948	0.939	0.950	0.184	0.190	0.168	0.192
30	30	0.9	0.8	0.969	0.964	0.953	0.965	0.160	0.167	0.145	0.169
30	30	0.9	0.9	0.970	0.963	0.950	0.964	0.136	0.141	0.120	0.144

시나리오3은 $A_1 = 0.70, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두 진단결과간의 상관계수; GPQ는 GPQ에 의한 방법; NP는 비모수적방법; ML은 최대우도법; Bootstrap은 붓스트랩방법.

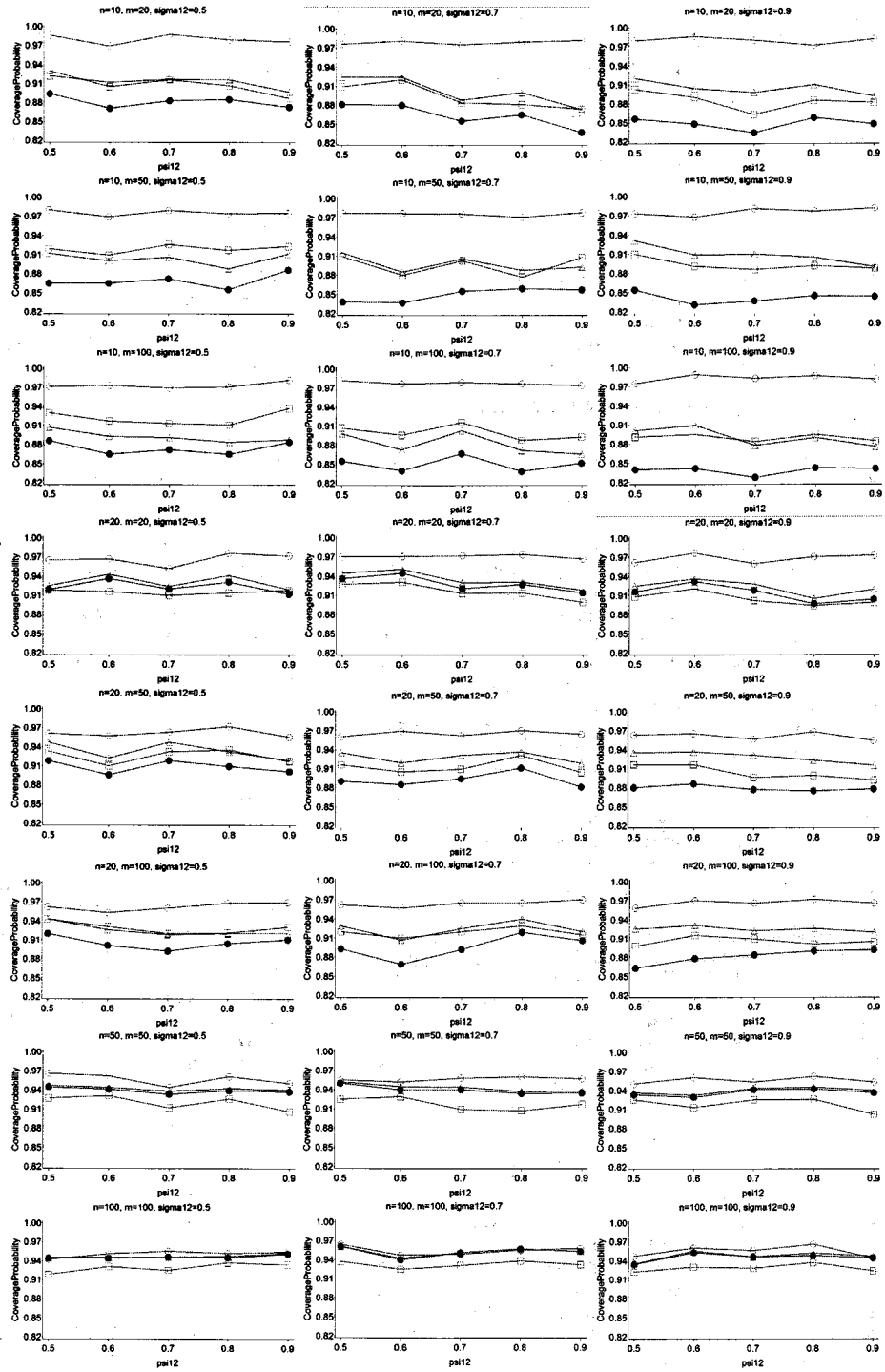

그림 1: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률- 시나리오1은 $A_1 = 0.95, A_2 = 0.90$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두진단결과간의 상관계수; (○) GPQ에 의한 방법, (•)비모수적방법, (□) 최대우도법, (△) 붓스트랩방법.

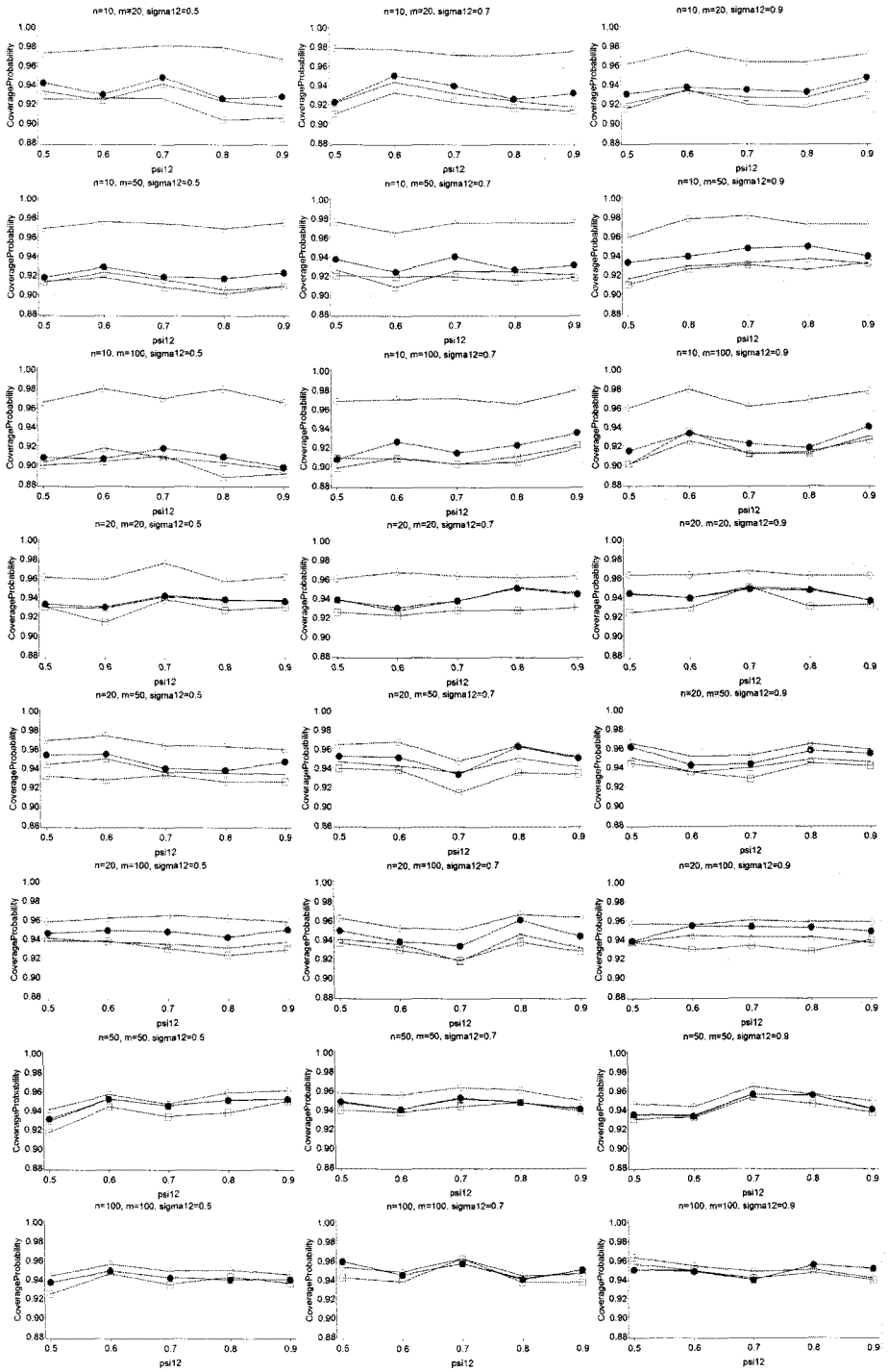

그림 2: $\Delta = A_1 - A_2$의 95% 신뢰구간의 포함확률- 시나리오2은 $A_1 = 0.95, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두진단결과간의 상관계수;(○) GPQ에 의한 방법, (●)비모수적방법, (□) 최대우도법, (△) 붓스트랩방법.

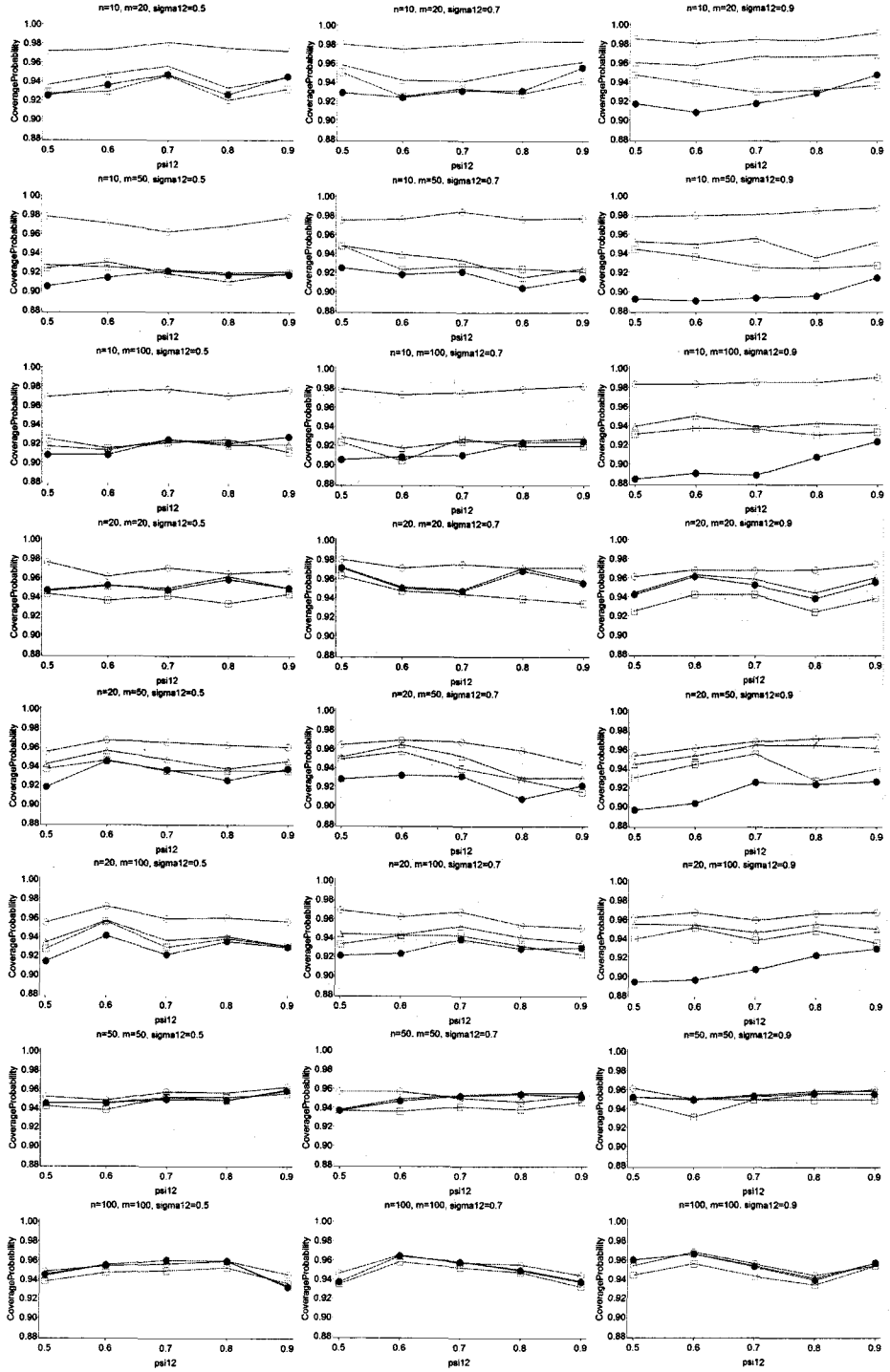

그림 3: $\Delta = A_1 - A_2$의 95%신뢰구간의 포함확률- 시나리오3은 $A_1 = 0.70, A_2 = 0.65$인 경우. 여기서, $A_i, i = 1, 2$은 i번째 진단방법의 ROC곡선 아래 면적; n은 환자군의 표본수; m은 정상군의 표본수; σ_{12}, ψ_{12}은 각각 환자군, 정상군에서 동일개체의 두진단결과간의 상관계수;(○) GPQ에 의한 방법, (•)비모수적방법, (□) 최대우도법, (△) 붓스트랩방법.

함확률의 측면에서 명목수준 95%를 하회하는 정도가 모두 비슷함을 보였다. AUC의 차이가 작은 첫 번째 시나리오와 세 번째 시나리오에서는 GPQ방법을 제외한 3가지 방법 중에서 붓스트랩방법이 명목수준 95%를 하회하는 정도가 제일 작았으며, 그 크기는 우려할 수준은 아니었다.

참고 문헌

Bandos, A. I., Rockette, H. E., and Gur, D. (2007). Exact bootstrap variances of the area under ROC curve, *Communications in Statistics-Theory and Methods,* **36**, 2443–2461.

DeLong, E. R, DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics,* **44**, 837–845.

Hanning, J., Iyer H. K., and Patterson, P. (2006). Fiducial generalized confidence intervals, *Journal of the American Statistical Association*, **101**, 254–269.

Li, C. R., Liao, C. T., and Liu, J. P. (2008). On the exact interval estimation for the difference in paired areas under the ROC curves, *Statistics in Medicine,* **27**, 224–242.

McClish, D. K. (1989). Analyzing a portion of the ROC curve, *Medical Decision Making,* **9**, 190–195.

2009년 12월 접수; 2010년 3월 채택

A Comparison of the Interval Estimations for the Difference in Paired Areas under the ROC Curves

Hee-Young Kim[1,a]

[a]Institute of Economics, Korea University

Abstract

Receiver operating characteristic(ROC) curves can be used to assess the accuracy of tests measured on ordinal or continuous scales. The most commonly used measure for the overall diagnostic accuracy of diagnostic tests is the area under the ROC curve(AUC). When two ROC curves are constructed based on two tests performed on the same individuals, statistical analysis on differences between AUCs must take into account the correlated nature of the data. This article focuses on confidence interval estimation of the difference between paired AUCs. We compare nonparametric, maximum likelihood, bootstrap and generalized pivotal quantity methods, and conduct a monte carlo simulation to investigate the probability coverage and expected length of the four methods.

Keywords: ROC, AUC, nonparametric method, maximum likelihood method, generalized pivotal quantity, bootstrap.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund)(KRF-2007-411-J03302).

[1] Research professor, Institute of Economics, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-701, Korea. E-mail: starkim@korea.ac.kr