

結果の要約

表2に事例検討で得られた結果を要約する。本論文でとり上げた3種類の事例は、すべての観測値が歪んだ分布にしたがっていた。そのため、NROC 曲線の結果は、PTROC 曲線およびPNROC 曲線に比べて極端に悪かった。2群の

分布形状が類似していた事例2では、PTROC 曲線はPNROC 曲線に比べて僅かに良好な結果を示したが、分布形状が異なる、その他の事例ではPNROC 曲線がPTROC 曲線よりも良好であった。

表 2: 事例検討の要約

事例	ROC 曲線の結果	最適カットオフ値の結果
卵巣癌データ	NROC 曲線は EROC 曲線よりも下部に布置しており、適合結果が最も悪かった。PNROC 曲線と PTROC 曲線は、EROC 曲線付近にプロットされた。ただし、PNROC 曲線は PTROC 曲線に比して上部に布置した。これらの曲線の適切性を AIC により評価した結果、PNROC 曲線が最良の適合であることが示唆された。	NROC 曲線は、極端に高い最適カットオフ値を示した。PNROC 曲線は PTROC 曲線に比して、最適カットオフ値が高くなる傾向にあった。最適カットオフ値に対して PNROC 曲線の TPR および FPR は、PTROC 曲線に比べて低くなる傾向にあった。
頭部外傷症データ	検査値に対数変換を施した NROC 曲線は、無変換の場合に比べて EROC 曲線に対する良好な適合を示した。また、PTROC 曲線および PNROC 曲線はほぼ重なっていた。ただし、PTROC 曲線のほうが PNROC 曲線に比べて AIC が僅かに低かった。すなわち、2群の分布形状が類似している場合には、PTROC 曲線で十分なようである。	NROC 曲線は、極端に高い最適カットオフ値を示した。PNROC 曲線と PTROC 曲線の最適カットオフ値には、大きな違いが認められなかった。このとき、PNROC 曲線と PTROC 曲線の最適カットオフ値の違いは、TPR よりも、むしろ FPR で顕著であった。ただし、コスト比が0.85以降では、その差異が殆ど認められなかった。
膵臓癌のバイオマーカーデータ	PNROC 曲線の曲線下面積 (AUC) により、CA19-9のほうが CA125に比べて良好なバイオマーカーであることが示唆された。このとき、PNROC 曲線よりも PTROC 曲線のほうが2個のバイオマーカーの差異を顕著に表した。すべての群に対する、ベキ変換パラメータの推定値が0.0付近の値を示すことから、左に歪んだ形状が示唆された。そのため、NROC 曲線の適合は非常に悪かった。また、いずれのバイオマーカーにおいても、2群の分布形状が異なることから、PNROC 曲線のほうが PTROC 曲線よりも良好な適合を示した。その傾向変化は、CA19-9よりも、CA125のなかで顕著であった。	

4.2 数値検証

文献例では、既存のパラメトリック接近法と対比しながらベキ正規 ROC 曲線 (PNROC 曲線) の有用性を示した。ここでは、次の目標のもとで検証を行う：(1) データが正規分布に従っているときに、3種類の ROC 曲線に違いはあるか。(2) 分布形状が同一のデータにおける PNROC 曲線と PTROC 曲線の違いはどの程度か。(3) 分布形状が異なるデータに対する PNROC 曲線の利点は何か。

ROC 曲線をはじめとする統計的グラフィクスでは、そのままの形式での客観的評価は困難である。そのため、ROC 曲線の適切性を数値検証により評価するには、何らかの要約測度が必要になる。本数値検証では、ROC 曲線の要約に頻繁に利用されている ROC 曲線下面積 (AUC) を用いる。

4.2.1 同一形状の分布に従う臨床検査値に対する評価

事例では、分布形状の異なるデータに対して 3種類の ROC 曲線 (NROC 曲線, PTROC 曲線, PNROC 曲線) を適用した。ただし、これらのデータは、NROC 曲線および PTROC 曲線の仮定を満たしていないことから、PNROC 曲線の十分な評価にはなっていない。そのため、数値検証では、NROC 曲線および PTROC 曲線における仮定が満たされた状況のもとで、PNROC 曲線の性能を吟味する。

数値検証には、生成された10000件のデータに 3種類の ROC 曲線 (NROC 曲線, PTROC 曲線, PNROC 曲線) を個別に適用し、そのうえで、真の AUC (A_0) とそれぞれの ROC 曲線に基づく推定 AUC (\hat{A}) の MSE, すなわち

$$MSE_{AUC} = \frac{1}{10000} \sum_{s=1}^{10000} \left\{ \hat{A}_s - A_0 \right\}^2 \quad (4.1)$$

を用いて評価する。評価基準に AUC を適用した理由は、ROC 曲線による臨床検査値の評価に ROC 曲線の AUC が用いられることが多いためである。たとえば、Molodianovitch *et al.* (2006) は、臨床検査値の比較に AUC を用いた場合での、諸種の ROC 曲線の評価を行っている。さらに、Obuchowski & McClish (1997) は NROC 曲線に基づく標本サイズの選定方法を提案している。本論文では、この標本サイズの選定方法に基づいて数値検証を行う。

さらに、このときの平均の差、標準偏差、標本サイズ、方法といった変動因子が各 ROC 曲線の MSE_{AUC} に及ぼす影響を分散分析の結果の p 値と寄与率により定量的に評価する。このとき、分散分析では、2 因子交互作用までを解釈する。

正規分布に従う場合

デザインおよびデータの生成：ここでは、正規乱数 $x \sim N(\mu, \sigma^2)$, $y \sim N(\mu + \Delta\mu, \sigma^2)$ によってデータを生成し、各 ROC 曲線に個別にあてはめる。このとき、ROC 曲線に影響を及ぼす要因として、平均の差 $\Delta\mu$ 、標準偏差 σ 、標本サイズ N を選定する。実際に、平均の差 $\Delta\mu$ を $\Delta\mu = 0.5, 1.0, 1.5$ の 3 水準とする。また、標準偏差は $\sigma = 0.8, 1.0, 1.2$ の 3 水準とする。これは、真の AUC が、0.70 から 0.85 付近になることを想定したためである。また、Obuchowski & McClish (1997) による AUC に基づく方法を用い、2 群が正規分布に従うと仮定したもとの真の AUC_{A_0} に対して標本サイズを選定する。いま、真の AUC を $A_0 = 0.75$ とする (因に、 $x \sim N(0, 1)$, $y \sim N(1, 1)$ のとき、 $A = 0.73$ である)。このとき、 A_0 に対する 95% 信頼区間を ± 0.05 で構成するためには、標本サイズ 80 が必要である。このことから、ここでの標本サイズ $N (= n_x = n_y)$ を $N = 50, 80, 100, 120$ とする。

結果・解釈： 4元分類分散分析の結果を表3に示す。この結果から、標本サイズ(N)の寄与率(74.1%)が最も高かった。次いで、標準偏差(σ)の寄与率(13.2%)が相対的に高かった。手法(M)のp値が0.967であることから、ROC曲線の差は殆ど認められなかった。N、平均値の差($\Delta\mu$)および σ の影響が高いことから、手法毎での影響を評価するために、NとMの組み合わせ水準、および $\Delta\mu$ とMの組み合わせ水準におけるウィンドウ・プロット(交互作用プロット)を図7に示す。ここで、ウィンドウ幅は、個々の水準における95%信頼区間を表している。Nが増加するにつれて、 MSE_{AUC} が緩やかな減少傾向にあった(図7(a))。 $\Delta\mu$ では、 MSE_{AUC} が直線的な減少傾向にあった(図7(b))。一方で、 σ では、 MSE_{AUC} が直線的な増加傾向を示した(図7(c))。ただし、いずれのプロットでも、Mによる差異が認められなかった。

標本サイズ、平均値の差および標準偏差の標本サイズと手法の組み合わせ水準、および平均値の差と手法。いずれのプロットでも、手法による差異が認められなかった。

表 3: 正規分布に従う場合に対する分散分析の結果

因子	p 値	寄与率
$\Delta\mu$	< 0.001	10.0
σ	< 0.001	13.2
N	< 0.001	74.1
M	0.967	0.0
$\Delta\mu \times \sigma$	< 0.001	1.1
$\Delta\mu \times N$	< 0.001	0.4
$\sigma \times N$	< 0.001	1.0
$\Delta\mu \times M$	1.000	0.0
$\sigma \times M$	0.999	0.0
$N \times M$	1.000	0.0

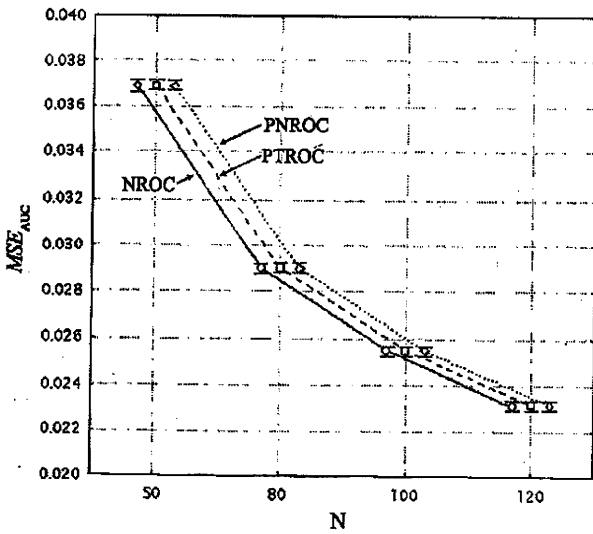
対数正規分布(同一形状の歪んだ分布)に従う場合

デザインおよびデータの生成：ここでは、正規乱数 $x^* \sim N(\mu, \sigma^2)$, $y^* \sim N(\mu + \Delta\mu, \sigma^2)$ によってデータを生成し、それらのデータの指数変換値 $x = \exp(x^*)$, $y = \exp(y^*)$ を各ROC曲線に個別にあてはめる。すなわち、対数正規分布 $LN(\mu, \sigma^2)$, $LN(\mu + \Delta\mu, \sigma^2)$ に対してROC曲線を構成する。このとき、ROC曲線に影響を及ぼす要因には、上述の場合と同様に、平均の差 $\Delta\mu$ 、標準偏差 σ 、標本サイズ $N(=n_x = n_y)$ を用いる。このとき、それぞれの水準も上述の数値検証と同様に設定する。

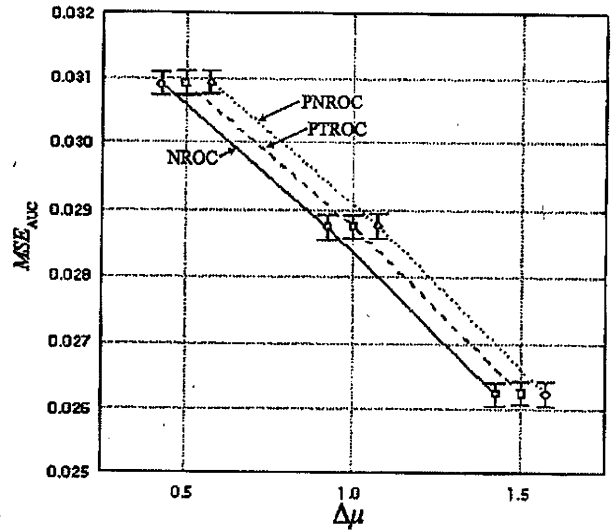
結果・解釈： 4元分類分散分析の結果を表4に示す。この結果から、手法の寄与率(82.1%)が最も高かった。次いで、手法(M)×平均値の差の2因子交互作用の寄与率(10.4%)が相対的に高かった。先の数値検証と同様に、標本サイズ(N)と手法(M)の組み合わせ水準、標準偏差(σ)とMの組み合わせ水準、および平均値の差($\Delta\mu$)とMの組み合わせ水準におけるウィンドウ・プロット(交互作用プロット)を図8に示す。いずれのプロットにおいてもNROC曲線の MSE_{AUC} が他の手法に比して極端に高かつ

表 4: 対数正規分布に従う場合に対する分散分析の結果

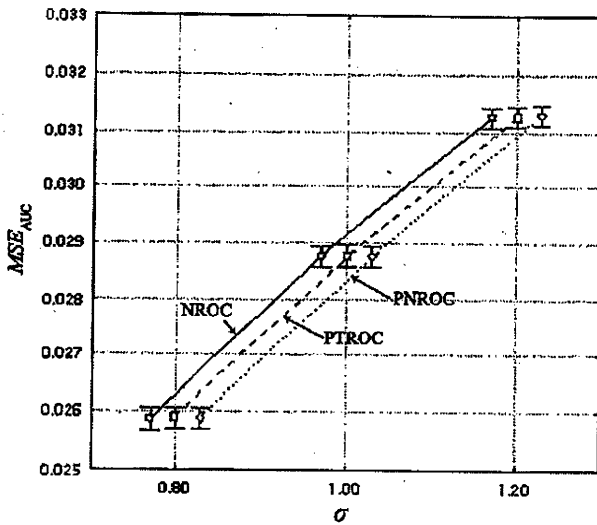
因子	p 値	寄与率
$\Delta\mu$	< 0.001	1.6
σ	< 0.001	1.1
N	< 0.001	1.6
M	< 0.001	82.1
$\Delta\mu \times \sigma$	< 0.001	0.0
$\Delta\mu \times N$	< 0.001	0.2
$\sigma \times N$	0.603	0.0
$\Delta\mu \times M$	< 0.001	10.4
$\sigma \times M$	0.070	0.0
$N \times M$	< 0.001	2.7



(a) M と N の組み合わせ水準



(b) M と μ の組み合わせ水準



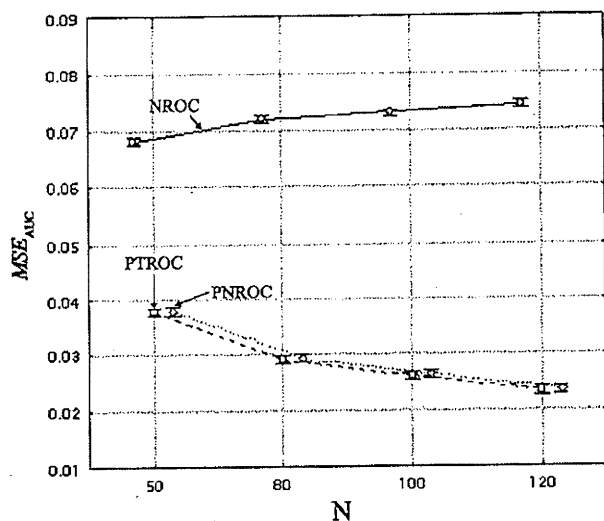
(c) M と σ の組み合わせ水準

図 7: 正規分布に従う場合の数値検証結果に対する交互作用のウィンドウ・プロット (ヒゲは95%信頼区間)

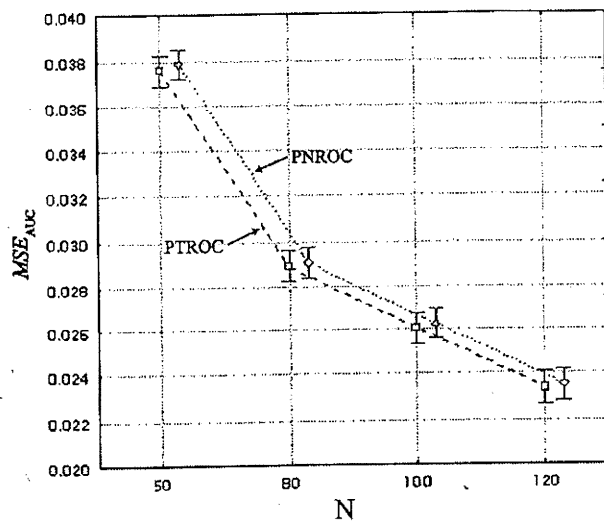
た。さらに、N および $\Delta\mu$ が大きくなるにつれて、他の 2 手法 (PNROC 曲線, PTROC 曲線) の MSE_{AUC} が減少傾向であるにもかかわらず、NROC 曲線は増加した (図 8 (a)(c))。ROC 曲線でも、他の統計的接近法と同様に、NROC 曲線 (正規分布) に基づく多くの方法が提案されている。ただし、臨床検査値が正規性の仮定を満たさない場合には、真値から大きく外れる傾向にあった。そして、N あるいは $\Delta\mu$ が大きくなるほどその傾向変化は大きくなった。 σ にお

いても NROC 曲線の MSE_{AUC} は、PTROC 曲線および PNROC 曲線よりも極端に高かった (図 8 (e))。

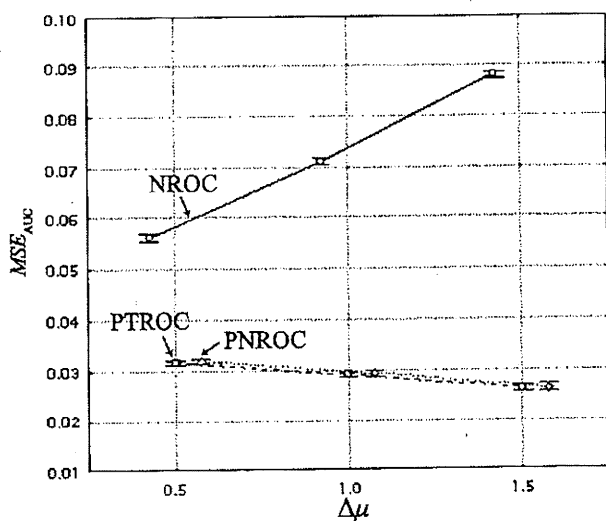
図 8 (a)(c)(e) では NROC 曲線の影響で PNROC 曲線および PTROC 曲線の省察が困難なため、NROC を外してウィンドウ・プロットを描いた。いずれのプロットでも、前述の正規分布に基づく数値検証と同様の傾向変化を示した (図 8 (b)(d)(f))。このとき、いずれの N でも PTROC 曲線が PNROC 曲線より僅かに低



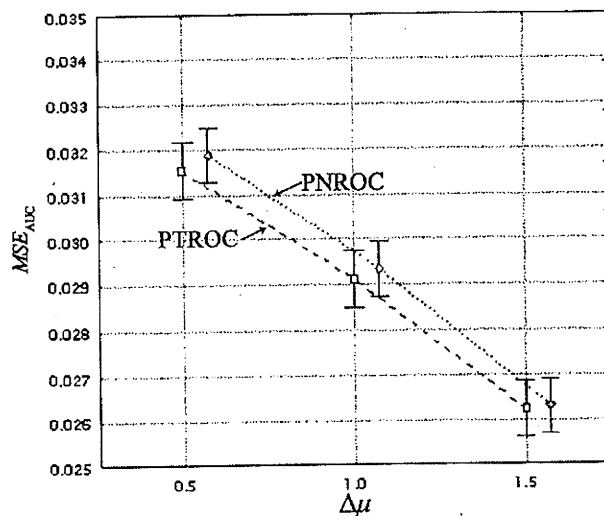
(a) M と N の組み合わせ水準



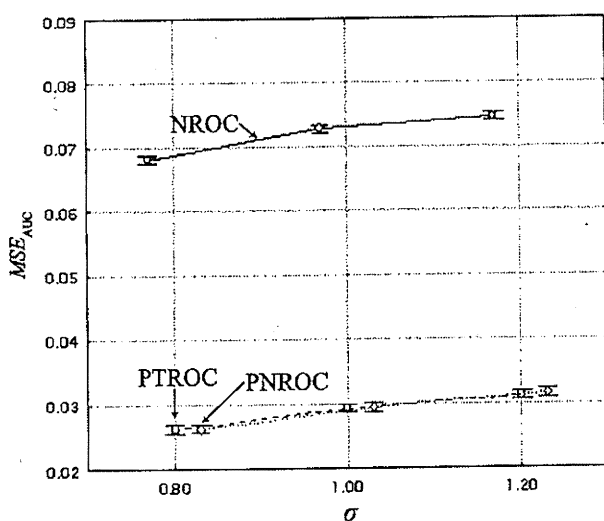
(b) M と N の組み合わせ水準 (NROC 除去)



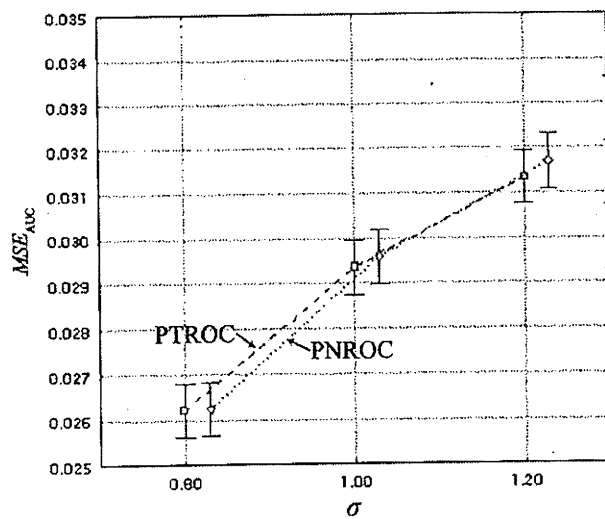
(c) M と μ の組み合わせ水準



(d) M と μ の組み合わせ水準 (NROC 除去)



(e) M と σ の組み合わせ水準



(f) M と σ の組み合わせ水準 (NROC 除去)

図 8: 対数正規分布に従う場合の数値検証結果に対する交互作用のウィンドウ・プロット (ヒゲは95%信頼区間)

い MSE_{AUC} を示したものの、ウィンドウが重なっており、2群の顕著な差は認められなかった(図8(b)). $\Delta\mu$ および σ も同様に、2群の差は認められなかった(図8(d)).

4.2.2 異なる形状の分布に従う臨床検査値に対する評価

4.2.1項では、同一形状の分布に従う臨床検査値に対する評価を行った。ただし、実際の臨床検査値の評価場面ではこのような仮定を満たすことは少ないようである。実際に、事例検討ではいずれのデータもこのような仮定をみたくことはなかった。ここでは、異なる形状の分布に従う臨床検査値に対する評価を行う。このとき、シミュレート・データの潜在基礎分布にはベキ正規分布を想定するが、ベキ変換パラメータ λ の挙動は他のパラメータに多大な影響を与えるため、前節のようなデザインのもとでの数値検証は困難である。そのため、ここでは事例の結果に基づいて数値検証を実施する。このとき、標本サイズ N 、平均の差 μ 、標準偏差 σ による影響は、個々の影響因子を逐次に変化させながら個別に評価される。このときのデータの生成方法を以下に示す：

S1: 臨床検査値が、ベキ正規分布に従うと仮定したもとで、健常者群 $PND(\lambda_x, \mu_x, \sigma_x)$ および疾患群 $PND(\lambda_y, \mu_y, \sigma_y)$ のパラメータを設定する。

S2: N 個の正規乱数 $x^* \sim N(\mu_x, \sigma_x)$, $y^* \sim N(\mu_y, \sigma_y)$ を生成する。

S3: 逆ベキ変換

$$x = \lambda_x(x^* + 1)^{1/\lambda_x}, \quad y = \lambda_y(y^* + 1)^{1/\lambda_y}$$

によりシミュレート・データを生成する。

数値検証では、事例2および事例3の CA19-9 をとりあげる。事例2は、2群の形状があまり

異なっておらず、PTROC 曲線の適切性が示唆されている。これに対して、事例3の CA19-9 は十分な適合結果が得られなかった。ただし、ここではベキ正規分布に従うと仮定したもとでシミュレート・データを生成した。これは、2群の分布形状が大きく異なっている状況をシミュレートするためである。

それぞれの事例のパラメータ推定値に基づき、基準となるパラメータを表5のように設定した(ここに、事例3は CA19-9の結果である)。このときの標本サイズは、文献データの標本サイズに基づき、20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200とした。また、その他の影響因子は、上述の患者群での基準となるパラメータ μ_y, σ_y からの差 $\Delta\mu_y = -1.0, -0.5, 0.0, 0.5, 1.0$, および $\Delta\sigma_y = -0.50, -0.25, 0.00, 0.25, 0.50$ とした。

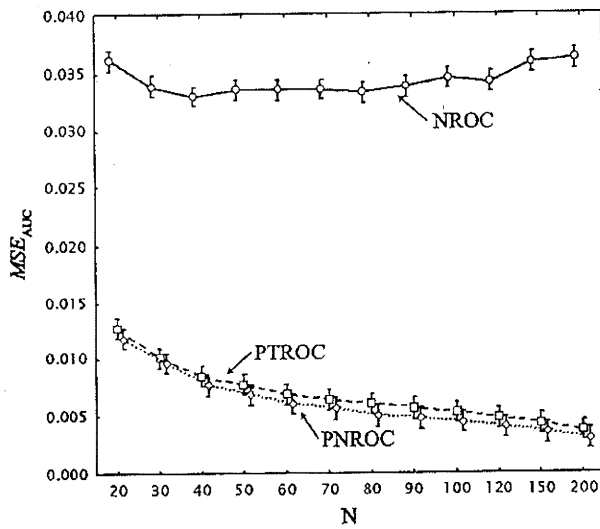
表 5: 事例に基づく数値検証での諸種のパラメータ

	事例2に基づく設定		事例3に基づく設定	
	疾患群	健常者群	疾患群	健常者群
μ	11.50	10.50	1.50	5.50
σ	4.50	3.00	0.3	2.5
λ	0.35	0.20	-0.40	0.00

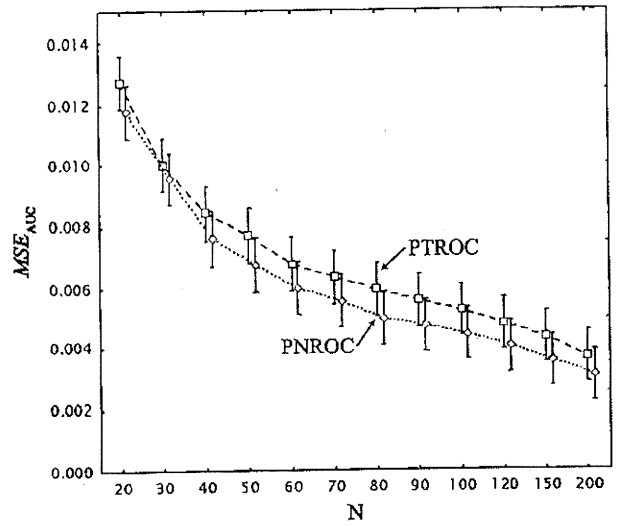
評価の方法には、上述のアルゴリズムによって生成された10000組のデータに3種類の ROC 曲線 (NROC 曲線, PTROC 曲線, PNROC 曲線) を個別に適用したときの、真の AUC_{A_0} とそれぞれの ROC 曲線に基づく推定 $AUC_{\hat{A}}$ の $MSE(4.1)$ を用いる。

事例2に基づく数値検証の結果

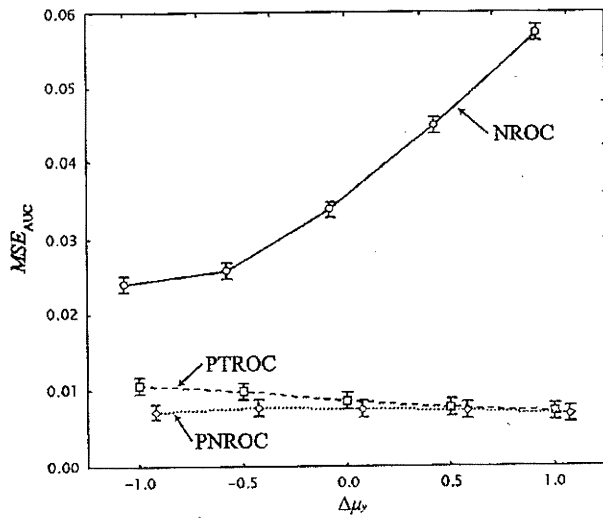
事例2に基づく結果を図9に示す。図9(a)(c)(e)より、NROC 曲線は、他の ROC 曲線に比して極端に悪かった。とくに、 μ_y からの差 ($\Delta\mu_y$) が大きくなるにつれて、NROC 曲線の MSE_{AUC} が急激な増加傾向を示した(図



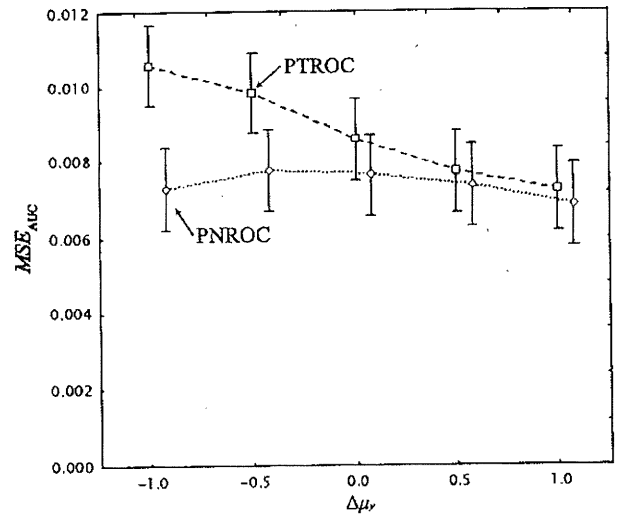
(a) M と N の組み合わせ水準



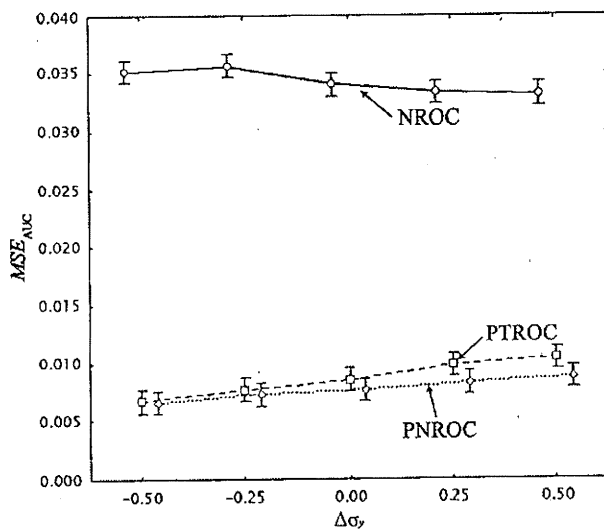
(b) M と N の組み合わせ水準 (NROC 除去)



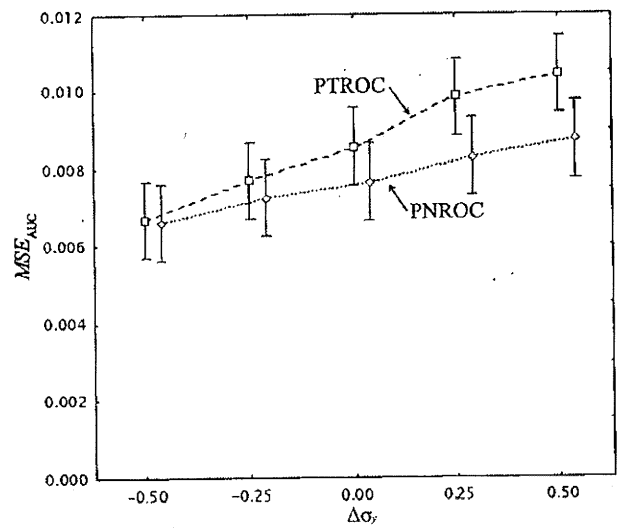
(c) M と $\Delta\mu_y$ の組み合わせ水準



(d) M と $\Delta\mu_y$ の組み合わせ水準 (NROC 除去)



(e) M と $\Delta\sigma_y$ の組み合わせ水準



(f) M と $\Delta\sigma_y$ の組み合わせ水準 (NROC 除去)

図 9: 頭部外傷症データに基づく数値検証結果に対する交互作用のウィンドウ・プロット (ヒゲは95%信頼区間)

9(c). 図9(b)は, NROC 曲線を外したときの, 手法(M)と標本サイズ(N)の組み合わせ水準における PTROC 曲線および PNROC 曲線でのウィンドウ・プロット(交互作用プロット)である. PTROC 曲線に比して, PNROC 曲線は僅かに低い MSE_{AUC} を示した. ただし, ウィンドウが重なっており, 良好な結果を示したという十分な結論を導くには至らなかった. 図9(d)は, Mと $\Delta\mu_y$ の組み合わせ水準における PTROC 曲線および PNROC 曲線でのウィンドウ・プロットである. 2群の平均値の差が小さくなるとき($\Delta\mu_y < 0.0$ のとき), PNROC 曲線は PTROC 曲線に比して非常に良好だった. ただし, $\Delta\mu_y$ が増加するにつれて, PTROC 曲線と PNROC 曲線の MSE_{AUC} の差異が減少し, $\Delta\mu_y > 0.0$ では差異が殆ど認められなかった. 図9(f)は, Mと σ_y からの差 $\Delta\sigma_y$ の組み合わせ水準における PTROC 曲線と PNROC 曲線のウィンドウ・プロット(交互作用プロット)である. $\Delta\sigma_y$ が小さいとき, 2個の ROC 曲線には殆ど差異が認められなかったものの, $\Delta\sigma_y$ が大きくなるにつれて, PNROC 曲線が PTROC 曲線に比して良好な結果を示した.

事例3の CA19-9に基づく数値検証の結果

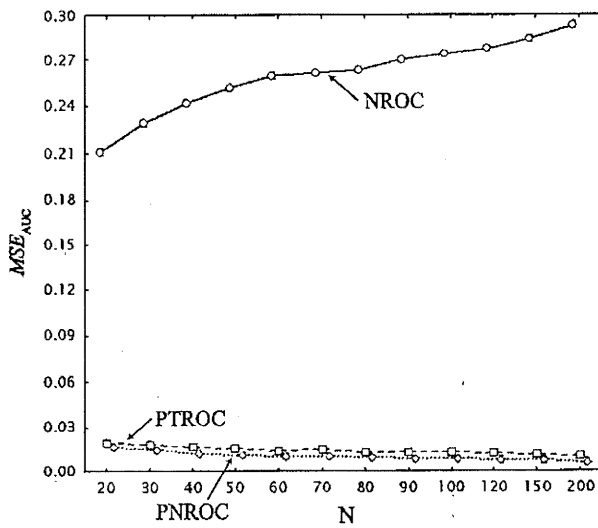
事例3の CA19-9に基づく数値検証の結果を図10に示す. 図10(a)(c)(e)より, NROC 曲線は, これまでの歪んだ分布に対する数値検証結果と同様に, 他の ROC 曲線よりも極端に悪かった. とくに, 標本サイズ(N)および μ_y からの差($\Delta\mu_y$)が大きくなるにつれて, NROC 曲線の MSE_{AUC} が急激な増加傾向を示した(図10(a)(c)). 他方, σ_y からの差 $\Delta\sigma_y$ の増加に対しては, 緩やかな減少傾向を示した. 図10(b)は, NROC 曲線を外したときの, 手法(M)とNの組み合わせ水準でのウィンドウ・プロットである. PTROC 曲線に比して, PNROC 曲

線は良好な結果を示しており, 図9(b)よりもその傾向変化は顕著であった. 図10(d)は, Mと $\Delta\mu_y$ の組み合わせ水準における PTROC 曲線および PNROC 曲線でのウィンドウ・プロットである. μ_y よりも2群の差が大きくなるとき($\Delta\mu_y < 0.0$ のとき), PNROC 曲線は, PTROC 曲線に比して, かなり良好であった. $\Delta\mu_y$ が増加するにつれて, 2個の ROC 曲線の MSE_{AUC} の差が小さくなったものの, 図9(d)に比して, 差の減少傾向は緩やかであった. 図10(f)は, Mと $\Delta\sigma_y$ の組み合わせ水準における PTROC 曲線と PNROC 曲線のウィンドウ・プロット(交互作用プロット)である. いずれの水準でも PNROC 曲線は, PTROC 曲線よりも良好な適合を示しており, $\Delta\sigma_y$ が大きくなるにつれて, PNROC 曲線と PTROC 曲線の MSE_{AUC} の差が顕著であった.

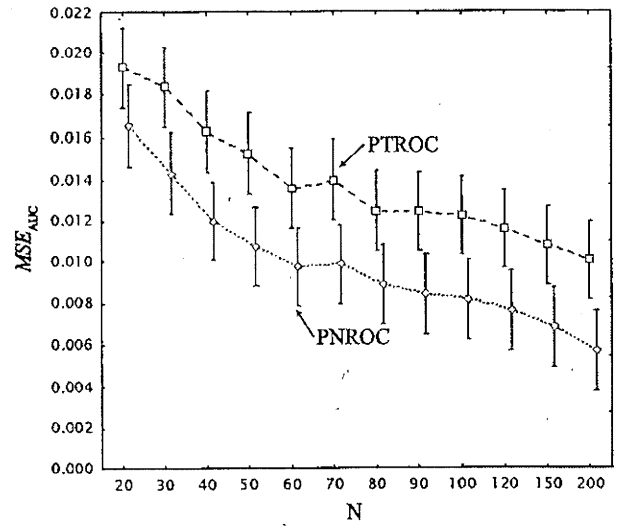
5. 結びに代えて

本論文では, 検査値の潜在基礎分布にベキ正規分布を想定する, ベキ正規 ROC 曲線を提案した. 臨床検査値は, 正規分布に従うことが少なく, 多くの場合に歪んだ分布形状を呈するが, ベキ正規 ROC 曲線は, このような状況においても良好な適合結果を示した. また, ROC 曲線では, 最適カットオフ値を視覚的に表現できない. これに対して, コスト比対最適カットオフ値のプロットは, 最適カットオフ値の挙動を視覚的に解釈できる点で有用であった. ベキ正規 ROC 曲線およびコスト比対最適カットオフ値のプロットの有用性は, 文献事例において提示した.

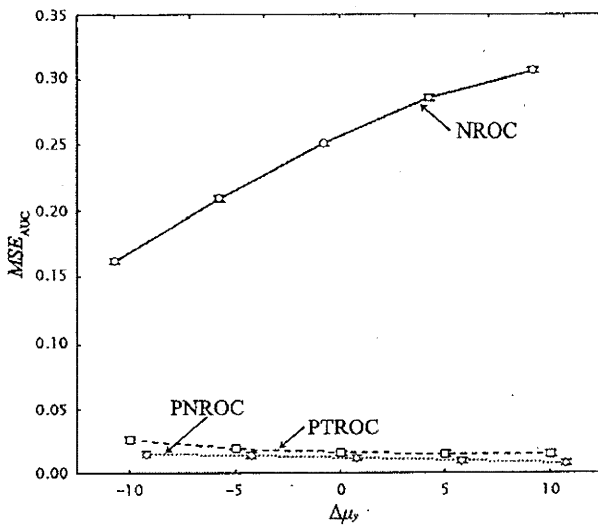
さらに, ベキ正規 ROC 曲線と既存の ROC 曲線の性能を若干の数値検証により評価した. その結果, 次の知見が得られた. (1) 観測値が歪んだ分布に従うとき, 正規分布に基づく ROC 曲線は, 真値から大きく外れており, 平均値の差



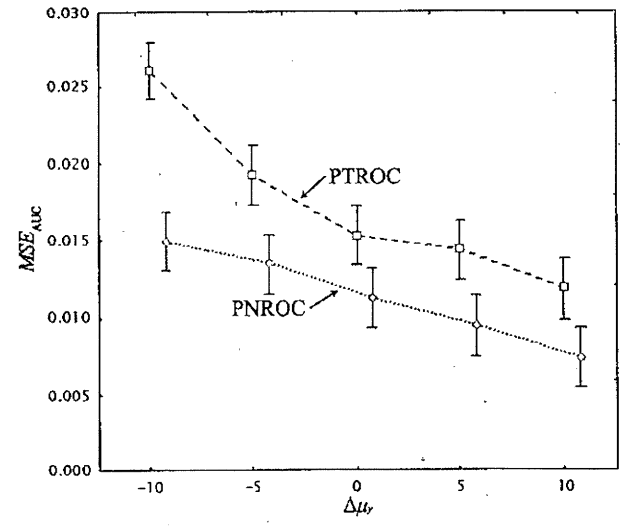
(a) M と N の組み合わせ水準



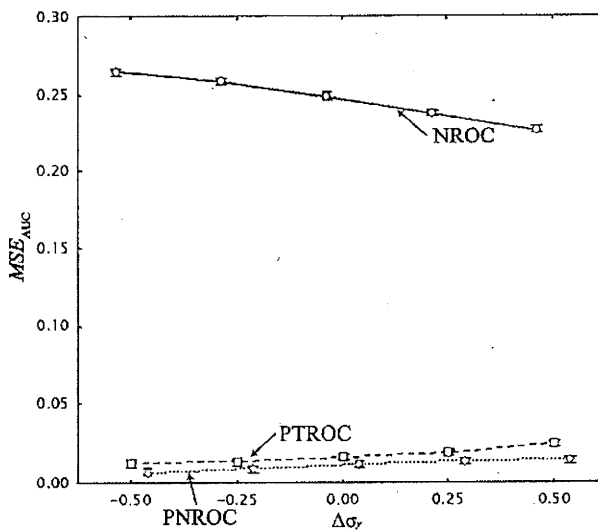
(b) M と N の組み合わせ水準 (NROC 除去)



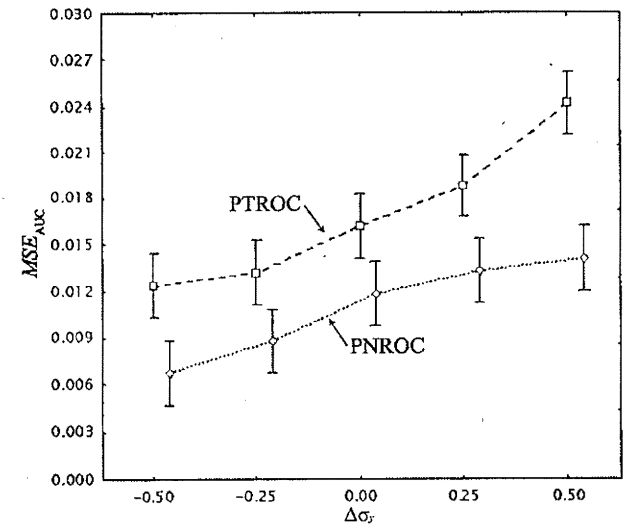
(c) M と $\Delta\mu_y$ の組み合わせ水準



(d) M と $\Delta\mu_y$ の組み合わせ水準 (NROC 除去)



(e) M と $\Delta\sigma_y$ の組み合わせ水準



(f) M と $\Delta\sigma_y$ の組み合わせ水準 (NROC 除去)

図 10: 膵臓癌データに基づく数値検証結果に対する交互作用のウィンドウ・プロット (ヒゲは95%信頼区間)

および標準偏差が増加するにつれてその傾向変化は顕著になる, (2) 事例2あるいは2群が対数正規分布に従う場合の数値検証のように, 2群の分布形状が歪んでおり, かつ2群の分布形状の差が小さいとき(あるいは同一のとき), PTROC 曲線の方が PNROC 曲線よりも僅かに良好な結果を示すものの, 形状の差が広がると, PNROC 曲線のほうが良好な結果を示した。(3) ただし, ROC 曲線の AUC が小さい状況(臨床検査値の標準偏差が大きいか, あるいは相対的な位置が近い)では, 分布形状が類似していても PNROC 曲線のほうが PTROC 曲線に比して良好であった(事例に基づく数値検証)。これは, AUC が小さな場合に, ROC 曲線がデータの挙動に敏感になるため, データに対して柔軟で, かつ適応的な PNROC 曲線が良好であったと考えられる。

謝辞

丁寧な査読と不備な点のご指摘を頂戴した, 本論文の審査員の先生方に深甚の謝意を捧げます。ありがとうございました。また, 本稿の修正には, 大阪大学大学院基礎工学研究科博士後期課程山口祐介君(D1)の手を煩わせた。ここに記して感謝いたします。

参考文献

Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations (with discussions). *J. Roy. Statist. Soc.*, B26, 211-252.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. Academic Press.

Goddard, M. J. & Hinberg, I. (1990). Receiver operating characteristic (ROC) curves and non-normal data: an empirical study. *Statistics in Medicine*, 9, 325-337.

Goto, M., Matsubara, Y. & Tsuchiya, Y.

(1983). Power-normal distribution and its applications. *Rep. Statist. Appl. Res., JUSE.*, 30(3), 8-28.

Goto, M., Uesaka, H. & Inoue, T. (1979). Some linear models for power transformed data. *Invited paper at The 10th International Biometric Conference*, August, 6-10 (Res. Rep. No.93, Res. Instit. Fund. Infor. Sc., Kyushu University).

Goto, M. and Inoue, T. (1980). Some properties of power-normal distribution. *Bulletin of the Biometric Soc.*, 1, 28-54.

Huang, Y. & Pepe, M.S. (2007). A parametric ROC model based approach for evaluating the predictiveness of continuous markers in case-control study. *UW Biostatistics Working Paper, University of Washington*.

Hsiao, J.K., Bartko, J.J. & Potter, W.Z. (1989). Diagnosing diagnoses: Receiver operating characteristic methods and psychiatry. *Arch Gen Psychiatry*, 46, 664-667.

Lusted, L. B. (1960). Logical analysis in Roentgen diagnosis. *Radiology*, 74, 178-193.

Lusted, L. B. (1968). *Introduction to Medical Decision Making*. Thomas Springfield [野村裕・中村正彦訳(1976). 臨床診断への新しい道: 意思決定の理論と実際. コロナ社].

Lusted, L. B. (1971). Decision-making studies in patient management. *New England Journal of Medicine*, 284, 416-424.

Metz, C. E. (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, 21, 720-733.

- Metz, C. E. (1989). Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*, **24**, 234-245.
- Molodianovitch, K., Faraggi, D. & Reiser, B. (2006). Comparing the area under two correlated ROC curve: parametric and non-parametric approaches. *Biometrical Journal*, **48(3)**, 745-757.
- Obuchowski, N.A. & McClish, D.K. (1997). Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Statistics in Medicine*, **16**, 1529-1542.
- Pepe, M.S. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, **84**, 595-608.
- Pepe, M.S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics*, **54**, 124-135.
- Pepe, M.S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, **56**, 352-359.
- Pepe, M.S. (2003). *The statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Press.
- Patton, D.D. & Woolfenden, J.M. (1989). A utility-based model for comparing the cost-effectiveness of diagnostic studies. *Invest. Radiol.*, **46**, 653-660.
- Peterson, W.W., Birdsall, T.G. & Fox, W.C. (1954). The theory of signal detectability. *IRE Transactions*, PGIT-4, 171-212.
- Somoza, E., & Mossman, D. (1991). "Biological markers" and psychiatric diagnosis: risk-benefit balancing using ROC analysis. *Biological Psychiatry*, **29**, 811-826.
- Wald, A. (1950). *Statistical Decision Functions*, Wiley.
- Wieand, S., Gail, M.H., James, B.R. & James, K.L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, **76**, 585-592.
- Zarembka, P. (1974). Transformation of variables in economics. *Frontiers in Econometrics*(ed. Zarembka, P., Chap.3), 81-104, *Academic Press*.
- Zhou, X.H., Obuchowski, N., & McClish, D.K. (2002). *Statistical Methods in Diagnostic Medicine*, Wiley.
- Zou, K.H., Hall, W.J. & Shapiro, D.E. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, **16**, 21-43
- 後藤昌司・山本成志・井上俊昭 (1991). ベキ正規分布のパラメータの推定: 推定量の漸近挙動について. *計算機統計学*, **4**, 45-60.
- 下川敏雄・後藤昌司 (2004). データ適応型判別解析とその診断. *計算機統計学*, **17(2)**, 87-108.
- 濱崎俊光・後藤昌司 (1996). ベキ変換の変換尺度の不変調整. *計算機統計学*, **9(1)**, 37-53.

(2008年4月10日受付, 2010年8月11日採択)

RECEIVER OPERATING CHARACTERISTIC CURVE BASED ON POWER-NORMAL DISTRIBUTION

Toshio Shimokawa* and Masashi Goto**

*Division of Graduate School of Medicine and Engineering, University of Yamanashi,
4-3-11 Takeda, Kofu City 400-8511 Japan

**Biostatistical Research Association, NPO,
4-3-11 Kamishinden, Toyonaka City 560-0085, Japan

In medical science, it is one of important subjects to explore the diagnostic test based on which we can discriminate the disease and non-disease groups. Especially, for evaluation of appropriateness of the biomarkers, one of the most useful statistical tools is the receiver operating characteristic (ROC) curve. For example, the ROC curve is applied in radiological (Metz, 1986, 1989) and psychiatric medicine (Hsiao *et al.*, 1989). Ordinary the ROC curve is usually based on the normality of distribution of observations for each group. In fact, it is easy to carry out ordinary statistical inference based on the normal distribution. However, observations obtained in real world rarely satisfy this strict assumption. Molodianovitch *et al.* (2006) have proposed the ROC curve based on the power-normal transformation (Box & Cox, 1964) of the observations, in order to satisfy normality of observations. However, in the approaches which depend on the framework of this “transformation”, the ROC curve does not have consistency from estimation of ROC curve to inspection of optimal cutoff value. Then, we provide the power-normal ROC curve assuming the power-normal distribution (Goto *et al.*, 1979, 1983; Goto & Inoue, 1980) as the underlying distribution of observations of each group, where the power-normal distribution is defined as the distribution specified before the power-normal transformation.

Key words: Power transformation, Diagnostic tests, Normal distribution, Medical tests, Clinical investigations

