

threshold level for the energy signal was set to 80 keV. The threshold level for the pulse shape spectrum for dividing the front layer and the back layer of the GSO scintillators was set at the valley of the two peaks of the pulse shape spectrum calculated from the ratio of the partial integration value to full integration value.

2.3. Performance of the blood sampling detector

The energy and pulse shape spectra of the blood sampling detector was measured for the 511 keV gamma photons and F-18 positrons. We evaluated the sensitivity, count rate characteristics and background rejection capability. We also measured the input function of a rat using the blood sampling detector.

2.3.1. Energy and pulse shape spectra measurements for 511 keV gamma photons and F-18 positrons. The energy and the pulse shape spectra for the 511 keV gamma photons were measured by irradiation with 511 keV gamma photons from the F-18 solution from the upper side of the detector. With this configuration, since the positrons from F-18 were absorbed outside of the shield, only 511 keV gamma photons could be measured by the detector.

The energy and pulse shape spectra for F-18 positrons were measured using F-18 solution filled in a PE-10 tube without irradiation of the 511 keV gamma photons from outside the detector.

2.3.2. Sensitivity. Sensitivity is one of the most important characteristics of the blood sampling detector for small animals because the amount of blood and the injected radioactivity are limited. The sensitivity of the blood sampling detector was measured for the F-18 and C-11 solutions. Approximately 37 kBq (1 μ Ci) F-18 or C-11 solution was contained in the PE-10 tube and set in the center of the detector. The positron solution was contained in the tube all the way across the field of view of the detector (20 mm long). The position of the positron solution in the tube was set at the center of the detector by fixing the tube at the tungsten shield. Measurements were made for the count rates of the front layer and back layer, and the subtracted count rate was also calculated for positrons. Sensitivity was derived by calculating the ratio of the count rate to the radioactivity of the F-18 or the C-11 in the tube.

2.3.3. Count rate characteristics. The count rate characteristic is important because it relates to the linearity of the system. Approximately 1 MBq ($\sim 30 \mu$ Ci) of the F-18 solution was contained in a PE-10 tube and set in the center of the detector. The count rates for the front layer (positrons + background gamma photons), back layer (background gamma photons) and front layer minus back layer (positrons) were measured following the decay of F-18.

2.3.4. Background gamma photons rejection capability. The background gamma photon rejection capability is important to ensure that the dual-layer phoswich concept really works. Approximately 15 kBq ($\sim 0.4 \mu$ Ci) of the F-18 solution was contained in a PE-10 tube and set in the center of the detector. 4 MBq ($\sim 100 \mu$ Ci) of the external 511 keV gamma source (F-18) contained in a glass container was positioned on the left-hand side of the detector surface and removed, then moved to the right-hand side and removed, and finally moved to the upper side and removed. This process was repeated twice. The schematic drawing of the position of the external 511 keV gamma source is shown in figure 4. The count rates for the front layer (positrons + background gamma photons), the back layer (background gamma photons) and the front layer minus back layer (positrons) were measured.

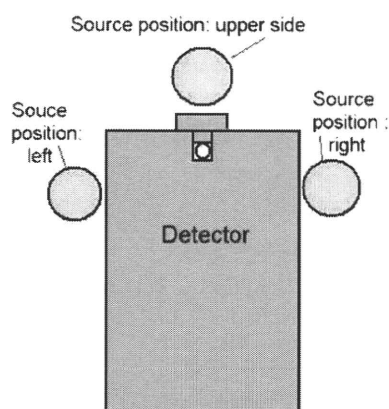


Figure 4. Schematic drawing of the position of the external 511 keV gamma source measured for background rejection capability.

2.3.5. Continuous blood sampling study of a rat using the detector. A continuous blood sampling study of a rat was performed to measure the arterial time activity curve (TAC) and determine whether the blood sampling detector is really useful for small animal measurements. Approximately 30 MBq (0.8 mCi) of F-18-FDG was intravenously administered to normal male Wistar rats (body weight: ~ 600 g) from the tail vein under anesthesia with ~ 1 min injection duration. These studies were performed under the guidelines of the Laboratory Investigation Committee of the Osaka University Graduate School of Medicine. Blood sampling was performed from the femoral artery using a PE-10 tube. The length of the tube from the rat to the detector was ~ 10 cm. The blood in the tube passed through the detector and exited from the end of tube. The exited blood was dropped on a piece of paper and collected in a plastic vial for weighing and counting by a well counter (Packard, Cobra Gamma counter, USA). We did not use any pump for blood sampling, but spontaneous output from the artery of the rat was measured.

The measurement by the blood sampling detector was performed for 1 s and repeated for 1750 s. The tube was flushed three times during blood sampling from ~ 700 to 850 s, ~ 1100 to 1200 s and ~ 1300 to 1600 s, to avoid clotting of the blood. The total volume of the sampled blood from the rat was ~ 1.2 ml. The average flow rate of the blood in the tube was ~ 1 cm s^{-1} or ~ 0.6 μl s^{-1} . The measured count rate was corrected for the decay of F-18.

The blood sampling for the well counter was performed at ~ 10 s intervals in the first 12 points, ~ 30 s intervals in the middle 5 points, and ~ 10 min intervals at the last 3 points. Each measured count rate was corrected for the weight of the blood and decay of F-18.

3. Results

3.1. Performance of the blood sampling detector

3.1.1. Energy and pulse shape spectra measurements for 511 keV gamma photons and F-18 positrons. Figure 5 shows the energy spectrum of the detector for 511 keV gamma photons from F-18. This spectrum was derived from the full integration signal (320 ns) without pulse shape discrimination. Although the GSO scintillators have a dual-layer configuration, the spectrum showed a single peak. The energy resolution was 18.5% FWHM.

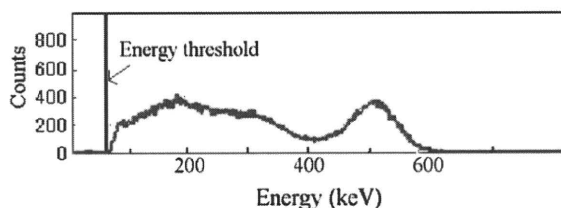


Figure 5. Energy spectrum of detector for 511 keV gamma photons. The vertical line at the low channel of distribution is the energy threshold level.

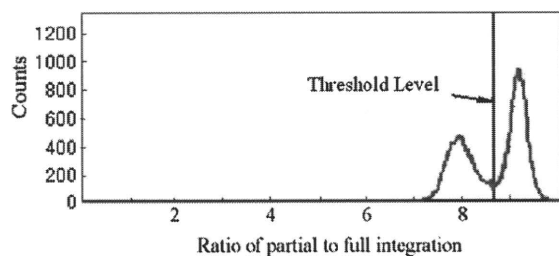


Figure 6. Pulse shape spectrum of the detector for 511 keV gamma photons. The vertical line between peaks of distribution is the threshold level for dividing the count of the front layer (right peak, GSO with 1.5 mol%, 35 ns) and back layer (left peak, GSO with 0.5 mol%, 60 ns).

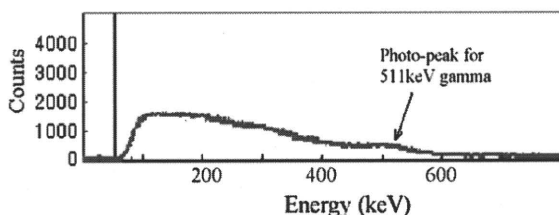


Figure 7. Energy spectrum of the detector for F-18 obtained in a tube. The arrow indicates the photo-peak of 511 keV gamma photons.

In figure 6, the pulse shape spectrum of the detector for the 511 keV gamma photons is shown. This spectrum does not contain any positrons. The right peak corresponds to the front layer GSO (1.5 mol% Ce) and the left peak is the back layer GSO (0.5 mol% Ce). The peak-to-valley (P/V) ratio of these two peaks was ~ 6 . The threshold level for these two peaks was divided at the valley of the peaks of the pulse shape distribution. Threshold level setting to divide the two types of GSO was performed with this spectrum measured for the gamma irradiation from the upper side of the detector. The threshold level was set so as to keep the count rate difference of these two areas less than 1%.

Figure 7 shows the energy spectrum for F-18 contained in the tube. This spectrum was derived from the full integration signal (320 ns) without pulse shape discrimination. The spectrum has a broad energy distribution of beta (positrons) with a small peak for the 511 keV gamma photons from the tube (arrow in figure 7).

In figure 8, the pulse shape spectrum of the detector for F-18 in the tube is shown. The right peak corresponds to the decay time of the front layer GSO (1.5 mol%) and the left peak is that of the back layer GSO (0.5 mol%). Because the front layer detects positrons and gammas

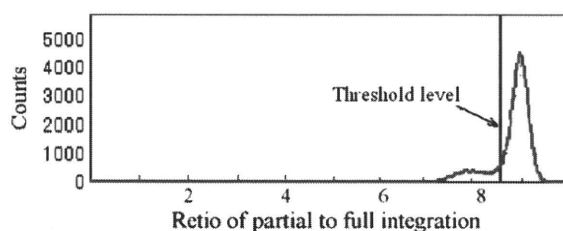


Figure 8. Pulse shape spectrum of the detector for F-18 in a tube. The vertical line between peaks of distribution is the threshold level for dividing the count of the front layer.

Table 1. Sensitivity of the blood sampling detector for F-18 and C-11.

	Sensitivity (%)
C-11	24.0
F-18	8.6

and the back layer detects only gammas, the right peak is higher than the left peak. The counts of the left peak (background gamma photons) are subtracted from the right peak (positrons + background gamma photons) to obtain the background subtracted counts (positrons). The count rate of right peak and that of the left peak are divided at the threshold level determined from the distribution of the 511 keV gamma photons shown in figure 8. Each count rate in the areas is calculated automatically by the data acquisition electronics.

3.1.2. Sensitivity. Table 1 shows the sensitivity of the detector for F-18 and C-11 in the PE-10 tube. The sensitivity for F-18 was smaller because the positron energy is smaller than that of C-11.

3.1.3. Count rate characteristics. Figure 9 shows the count rate characteristics of the detector. The horizontal as well as vertical axes are logarithmic scales. The count rate should be the linear function of activity. By comparing the count rate for front minus back (positron) with the estimated ideal decay curve of F-18, count rates are linear up to ~ 50 k cps within an error of 3%.

3.1.4. Background gamma photon rejection capability. Figure 10 demonstrates the background gamma photon rejection capability. The six peaks in the figure observed in the front and back layers were made when the background 511 keV gamma source was positioned near the detector. Valleys between these peaks occurred when the background source was removed from the detector surface (no background activity). The variation of the measured count rate of the front layer minus back layer (count rate for positrons) with and without background activity was less than 3%.

3.1.5. Continuous blood sampling study of a rat. Figure 11 shows the count rate curves of the F-18-FDG rat study measured by the continuous blood sampling detector. Due to the gamma photons from the rat and the syringe, the count rate curves of the front and back layers have significant offset counts. However, by subtracting the latter from the former, these offset counts were corrected. The increase in the count rate of the front layer and back layer between

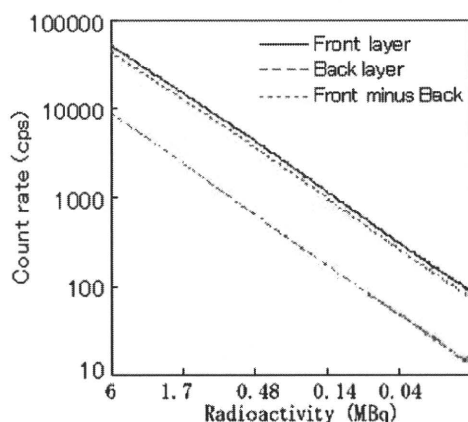


Figure 9. Count rate characteristic of the detector measured using F-18 contained in the PE-10 tube.

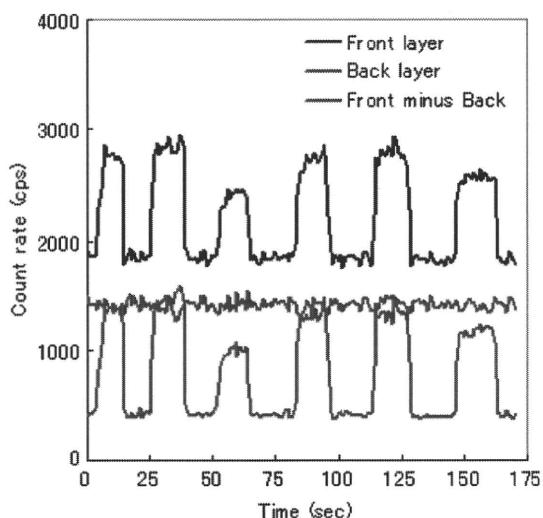


Figure 10. Background gamma rejection capability of the positron detector system. First and fourth peaks for the front layer and back layer occurred when the background source was positioned at the left-hand side of the detector, second and fifth peaks occurred when the source was at the right-hand side, third and sixth peaks occurred when the source was at the upper side. Valleys between peaks for the front and back layers occurred when the background source was removed from the surface.

500 s and 700 s is because the syringe used for the FDG injection, in which some FDG remained, was accidentally removed from the injector support and rolled near the detector and the background activity was increased. Even with these accidental background activity changes, the subtracted count rate did not change, showing the accuracy of the background subtraction method of the detector.

The smoothed count rate curves are shown in figure 12. The curve was smoothed by calculating the running average for 20 neighboring points. The smoothed count rate with the counts measured by the well counter is shown in figure 13, where the blood sampling detector count rate was converted to cps g^{-1} measured by the well counter by multiplying the

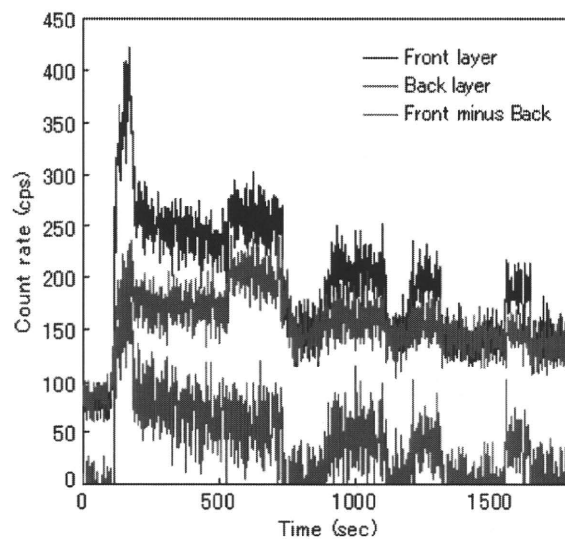


Figure 11. Count rate curves of rat artery blood measured by the blood sampling detector.

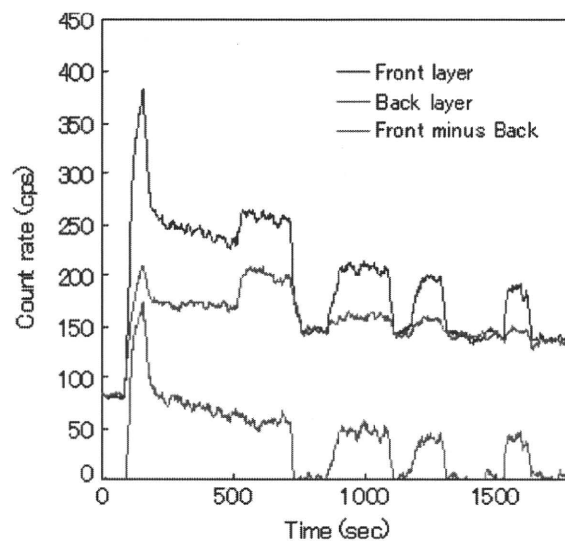


Figure 12. Smoothed count rate curves of rat artery blood measured by the blood sampling detector.

conversion factor between the developed detector and the well counter. The conversion factor was measured using a F-18 solution filled in a PE-10 tube after this animal study. The time difference between continuous blood sampling detector and well counter was adjusted from the rise point to be the same.

4. Discussion

We successfully developed a high-sensitivity and compact blood sampling detector. Excellent background gamma rejection capability of the detector is found in the results shown in

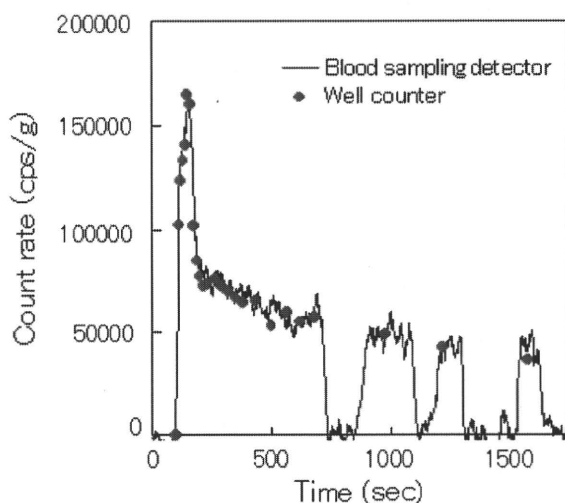


Figure 13. Count rate curves of rat artery blood measured by the blood sampling detector (blue lines) and well counter (red circles).

figures 11–13. The count rate of the background gamma photons in the animal study showed that it was impossible to precisely measure artery TAC without background gamma photon rejection. From figures 11 and 12, the background gamma photon count rate from outside of the detector was larger than the positron count rate from the blood sampling tube. Correction of the background count rate cannot be performed by such simple methods as subtracting a single value from the count rate of the front layer because the background count rate significantly changed with time.

One advantage of the developed blood sampling detector is its high sensitivity, which is high because the energy threshold level can be set low. The sensitivity of our previously developed phoswich blood sampling system was 1.7% for F-18 (Yamamoto *et al* 2001) and 7% for the coincidence blood sampling system (Kudomi *et al* 2003). With this low energy threshold, some lower energy positrons (beta particles) can be detected. To increase the sensitivity, smaller diameter and thinner tubes will be needed because the major sensitivity loss comes from positron absorption in the tube.

Another advantage of the detector is its small size, which can be placed close to the small animals to minimize the amount of the blood taken out. The small size can be realized with the thin gamma shields as well as the simple and compact configuration of the detector with thin scintillators coupled with a small PMT. The thin gamma shields are adequate for this detector because it can accurately correct for the external gamma photons. Without correction for the background gamma photons, precise measurement of the TAC of the artery blood, as shown in figures 11–13, would not be possible.

The use of GSO scintillators with different Ce concentrations is ideal for a phoswich positron detector, with background gamma photon detection capability because the detection efficiencies of the front and back layers for the gamma photons are the same, and precise correction is possible by the subtraction method.

For the quantitative measurement in human studies, the developed detector will also be useful. The small size, high sensitivity and background gamma photon correction capability will minimize the tube diameter and shorten the distance between subject and detector to

reduce the delay and dispersion (Kanno *et al* 1987, Iida *et al* 1987) as well as the total amount of blood.

The time activity curve shown in figure 11 is noisy because the count rate is around ~ 100 cps in this study. Also, the noise was enhanced by the subtraction processes. To reduce the noise, the 1 s sample data were smoothed over 20 s. Although the smoothing process reduced some temporal information of the curve, it would not be a serious problem in this study because the temporal change of the FDG study is not so fast. For the other studies using shorter half-life isotopes such as C-11, N-13 and O-15 with a bolus injection, the smoothing process will reduce the temporal information of the input function. However, because C-11, N-13 and O-15 have higher positron energy, sensitivity becomes higher for the developed blood sampling system. In addition, since the injection dose of these shorter half-life isotopes is usually higher for PET imaging, the count rate for blood sampling also becomes higher. Thus, it is expected that the smoothing of the data will be smaller in these studies.

For the animal study presented in this paper, a rat of a relatively larger size (600 g) was used. The limitation of the amount of blood from the small animals is roughly 1% of the body weight (McGuill and Rowan 1989, Morton 1993). In the study of a typical rat weighing 200 g, it will not be a problem because the limitation of the amount of blood sampling will be ~ 2 ml, larger than the amount of blood presented in this study (1.2 ml). However, it is not clear at this time whether the blood sampling from typical mice (20 g) to measure the input function is possible or not because its amount of blood is limited to be ~ 0.2 ml.

5. Conclusion

Our developed positron detector is compact with high sensitivity and is suitable for a blood sampling system for small animal PET studies. With these results, we conclude that the developed positron detector is promising for the quantitative measurement of small animals.

Acknowledgments

The authors thank Dr H Ishibashi of Hitachi Chemical for developing thin GSOs with different Ce concentrations. This work was partly supported by the National Institute of Biomedical Innovation, Osaka, Japan.

References

- Convert L, Morin-Brassard G, Cadorette J, Archambault M, Bentourkia M and Lecomte R 2007 A new tool for molecular imaging: the microvolumetric beta blood counter *J. Nucl. Med.* **48** 1197–206
- Eriksson L, Holte S, Bohm C, Kesselberg M and Hovander B 1988 Automated blood sampling systems for positron emission tomography *IEEE Trans. Nucl. Sci.* **35** 703–7
- Ferl G Z, Zhang X, Wu H M and Huang S C 2007 Estimation of the ^{18}F -FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data *J. Nucl. Med.* **48** 2037–45
- Hutchins G D, Hichiwa R D and Koeppe R A 1986 A continuous flow input function detector for blood flow studies in positron emission tomography *IEEE Trans. Nucl. Sci.* **33** 546–9
- Iida H *et al* 1987 Error analysis of a quantitative cerebral blood flow measurement using an H_2^{15}O autoradiography and positron emission tomography, with respect to the dispersion of the input function *J. Cereb. Blood Flow Metab.* **6** 535–42
- Ishibashi H, Shimizu K, Susa K and Kubota S 1989 Cerium doped GSO scintillators and its application to position sensitive detectors *IEEE Trans. Nucl. Sci.* **36** 170–2
- Kanno I *et al* 1987 A system for cerebral blood flow measurement using an H_2^{15}O autoradiographic method and positron emission tomography *J. Cereb. Blood Flow Metab.* **7** 143–53

- Kim J, Herrero P, Sharp T, Laforest R, Rowland D J, Tai Y C, Lewis J S and Welch M J 2006 Minimally invasive method of determining blood input function from PET images in rodents *J. Nucl. Med.* **47** 330–6
- Knoll G F 1997 *Radiation Detection and Measurement* (New York: Wiley)
- Kudomi N *et al* 2003 Development of a GSO detector assembly for a continuous blood sampling system *IEEE Trans. Nucl. Sci.* **50** 70–3
- McGuill M W and Rowan A N 1989 Biological effects of blood loss: implications for sampling volumes and techniques *ILAR News* **31** 5–18
- Morton D B *et al* 1993 Removal of blood from laboratory mammals and birds: first report of the BVA/FRAME/RSPCA/UFAW joint working group on refinement *Lab. Anim.* **27** 1–22
- Pain F, Lanière P, Mastriplolito R, Gervais P, Hantraye P and Besret L 2004 Arterial input function measurement without blood sampling using a beta-microprobe in rats *J. Nucl. Med.* **45** 1577–82
- Su Y, Welch M J and Shoghi K I 2007 The application of maximum likelihood factor analysis (MLFA) with uniqueness constraints on dynamic cardiac microPET data *Phys. Med. Biol.* **52** 2313–34
- Votaw J and Shulman S D 1998 Performance evaluation of the pico-count flow-through detector for use in cerebral blood flow PET studies *J. Nucl. Med.* **39** 509–15
- Yamamoto S 2008a A dual layer DOI GSO block detector for a small animal PET *Nucl. Instrum. Methods A* **598** 480–4
- Yamamoto S 2008b Optimization of the integration time of pulse shape analysis for dual layer detector with different amount of Ce *Nucl. Instr. Methods A* **587** 319–23
- Yamamoto S and Ishibashi H 1998 A GSO depth of interaction detector for PET *IEEE Trans. Nucl. Sci.* **45** 1078–82
- Yamamoto S, Tarutani K, Suga M, Minato K, Watabe H and Iida H 2001 Development of a phoswich detector for a continuous blood sampling system *IEEE Trans. Nucl. Sci.* **48** 1408–11

ベキ正規分布に基づく ROC 曲線の構成

下川敏雄*・後藤昌司**

要 旨

医学分野では、適切に疾病の有無を予測するための臨床検査値の探索が重要な要件の一つである。このとき、有用な統計的方法の一つが受信者動作特性 (ROC: Receiver Operating Characteristic) 曲線である。ここでは、解釈の簡便さ、あるいはその後の統計的推測の観点から、正規分布に基づく ROC 曲線が広く用いられている。ただし、このような検査値が正規分布に従う現象は稀である。このとき、検査値の正規性を満たすために、検査値にベキ変換を施し、そのうえで正規分布に基づく方法が適用されている。しかしながら、この変換に基づく方法では、ROC 曲線の推定から曲線の解釈までの一貫性を保持できない。本論文では、包括的な接近法として、検査値の潜在基礎分布にベキ正規分布を想定したベキ正規 ROC 曲線を提案した。そのうえで、疾病の有無を識別する最適カットオフ値を選定した。さらに、ベキ正規 ROC 曲線の性能を、事例および若干の数値検証により評価した。その結果、ベキ正規 ROC 曲線は、2群が異なる形状を示す場合にも適合結果が良好であった。

1. 序に代えて

医学分野において、臨床検査に基づいて疾病の有無を診断することは、重要な要件の一つである。とくに、診断医学では、疾病の有無を適切

に予測するための良好な臨床検査値の探索が研究主題となっている。ここでは、種々の臨床検査値の評価だけでなく、最適なカットオフ値の選定も議論されている。このとき、有用な統計的方法の一つが受信者動作特性 (ROC) 曲線である。

ROC 曲線は、統計的決定論 (Wald, 1950) に基づいて開発され、シグナル検出理論 (Peterson *et al.*, 1954) によって有用性が認められたグラフィカル手法である。その後、ROC 曲線は、品質管理あるいは計量心理学といった分野で広く活用され、Lusted (1960, 1968, 1971) によって診断医学に応用された (Molodianovitch *et al.*, 2006)。現在、ROC 曲線は、診断医学だけでなく、放射線医学 (Metz, 1986, 1989)、精神医学 (Hsiao *et al.*, 1989) など、医学分野で広範に適用されている。

ROC 曲線は、1-特定度 (specificity)、すなわち、ある疾患にかかっていない被験者を疾患にかかっていると予測する確率に対して、感度 (sensitivity)、すなわち、疾患患者を正しく同定する確率を座標軸上にプロットする。このとき、定量値に対する ROC 曲線の推定には、主に3種の方法、すなわち、(i) 経験分布関数に基づく方法、(ii) 正規分布に基づく方法 (パラメトリック接近法)、(iii) 核密度推定に基づく方法 (ノンパラメトリック接近法) が広く知られている。(i) は、疾患群および健常者群の経験累

* 山梨大学 大学院医学工学総合研究部 〒400-8511 甲府市武田4-3-11 (Tel. 055-220-8395),
E-mail: shimokawa@yamanashi.ac.jp

** 特定非営利活動法人 医学統計研究会 〒560-0085 豊中市上新田2丁目22-10-A411 (Tel. 06-6835-8790),
E-mail: gotoo@bra.or.jp

積分布関数に基づく ROC 曲線である。そのため、得られた ROC 曲線は階段関数で与えられるため、標本サイズが少ない場合には良好な曲線下面積 (AUC: Area Under Curve) の推定値が得られない (Pepe, 2003)。因に、AUC とは、検査値による疾患/非疾患の識別性能の評価基準に用いられる。これに対して、(ii) および (iii) では ROC 曲線は滑らかな曲線によって与えられる。(ii) は疾患群および健常者群の潜在基礎分布に正規分布を仮定する方法である。この方法では、ROC 曲線が 2 群の平均および分散の相対測度を表すパラメータにより要約できることから、解釈が容易である。ただし、実質科学では、検査値が正規分布に従う場合は稀であり、そのような場合には、誤った解釈を導くことがある (Goddard & Hinberg, 1990; Wieand *et al.*, 1989)。(iii) は、核密度によって健常者群および疾患群の確率密度関数を推定し、そのもとで ROC 曲線を推定する方法である (Zou *et al.*, 1997)。この方法では、あらかじめ潜在基礎分布を想定しなくてもよいが、この方法では核密度およびバンド幅を選択しなければならないだけでなく、得られた結果の解釈に難しさを伴う。そのため、本論文では良好な適合性能だけでなく解釈の平易さにも留意するために、正規分布を基調としたパラメトリックな ROC 曲線に議論を限定する。

Molodianovitch *et al.* (2006) は、正規分布に従わない歪んだ分布に対する対策のひとつとして、2 群の母集団が同一形状の分布をもつと仮定したもとで検査値にベキ変換 (Box & Cox, 1964) を施し、変換値のデータに対して正規分布に基づく方法を適用している。ただし、ベキ変換に基づく方法では、ROC 曲線に対する潜在基礎分布の推定から曲線の解釈までの一貫性を保持できない。また、2 群 (健常者群および疾患群) に対して、同一の分布形状を仮定しな

ければならない。そのため、本論文では、それぞれの群の潜在基礎分布にデータ適応型分布、とくにベキ正規分布 (Goto *et al.*, 1979, 1983; Goto & Inoue, 1980) を想定する。ここに、データ適応型分布とは、複数の理論分布を包括できる分布である。これにより、疾患群および健常者群が異なる分布形状を示す場合にも ROC 曲線を描くことができる。さらに、ベキ正規分布が正規分布を包括することから、既存の正規分布に基づく ROC 曲線の適切性の評価にベキ正規分布に基づく ROC 曲線を適用することができる。

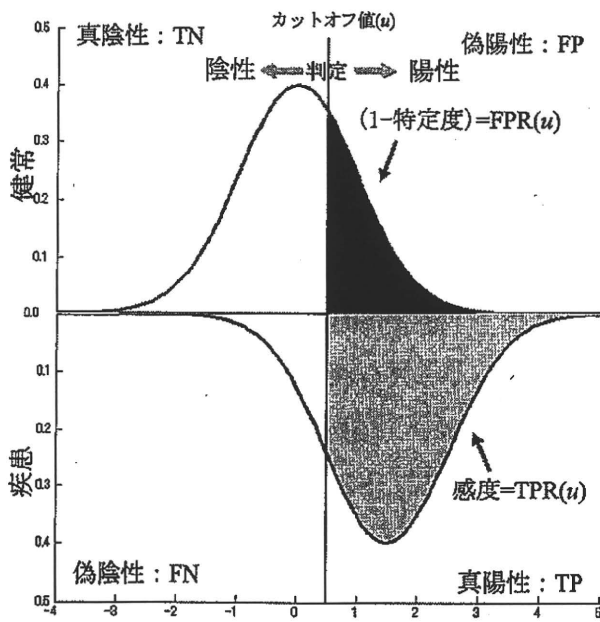
2 章では、既存の ROC 曲線の構成の方法について略説する。3 章では、ベキ正規分布に基づいて ROC 曲線を構成し、最適カットオフ値を選定する。4 章では、事例および数値検証によりベキ正規 ROC 曲線の特長を具体化し、その性能を評価する。5 章では得られた結果および知見を整理し、結びに代える。

2. 既存の ROC 曲線の構成

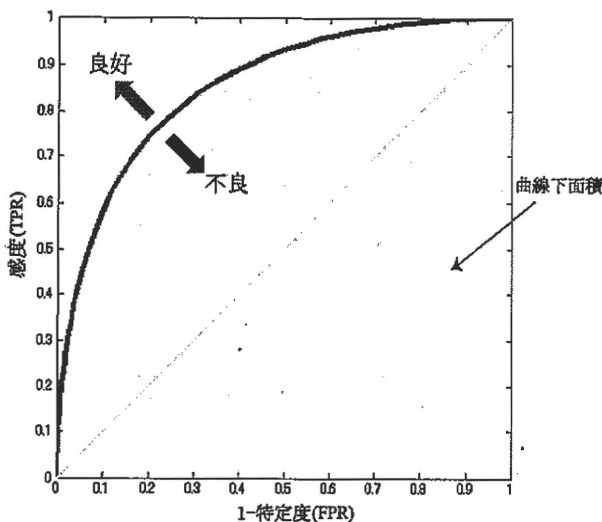
本章では、先ず ROC 曲線の定義について述べる。次に、諸種の ROC 曲線の推定方法のなかから、とくに経験累積分布に基づく接近法、正規分布に基づくパラメトリック接近法、ベキ変換に基づく方法について要約する。

2.1 ROC 曲線の定義

いま、健常者群および患者群に対する検査値 x が、それぞれ、累積分布 $F_x(x)$, $F_y(x)$ をもつ未知の分布に従うとする (図 1)。任意の検査では、カットオフ値 u によって、 $x > u$ ならば陽性、 $x \leq u$ ならば陰性と判定される。このとき、疾患にかかっていることを正しく陽性と判断することを真陽性 (TP: True Positive); 疾患にかかっているのに陰性と判断することを偽陰性 (FN: False Negative); 疾患にかかっていな



(a) 感度および (1-特定度) に関する例示



(b) ROC 曲線の例示

図 1: ROC 曲線の概要

いことを正しく陰性と判断することを真陰性 (TN: True Negative), 疾患にかかっていないにも関わらず陽性と判断することを偽陽性 (FP; False Positive) という。また, u における TP の確率 $TPR(u) = \Pr(x > u | D = 1)$ は感度と呼ばれ, FP の確率 $FPR(u) = \Pr(x > u | D = 0)$ は 1-特定度と呼ばれる。ここに, D は 1 のときに疾患, 0 のときに非疾患を表す。これらは $FPR(u) = 1 - F_x(u)$, $TPR(u) = 1 - F_y(u)$, で

定義される (ここに, $-\infty < u < \infty$ である)。

ROC 曲線は任意の検査値のすべての可能なカットオフ値における感度対 (1-特定度) のプロット $(FPR(u), TPR(u))$ である。

2.2 経験累積分布に基づく ROC 曲線

経験累積分布関数に基づく ROC 曲線 (以下, EROC 曲線と略す) は, 簡単に求めることができるため, 最も広範に利用されている。健康者群において n_x 個の検査値 $\{x_{i_x}\}_{i_x=1}^{n_x}$, 患者群において n_y 個の検査値 $\{y_{i_y}\}_{i_y=1}^{n_y}$ が与えられるとき, $F_x(x)$, $F_y(x)$ の経験累積分布関数に基づく $TPR(u)$, および $FPR(u)$ の推定値は

$$\widehat{FPR}_{\text{emp}}(u_j) = \frac{1}{n_x} \sum_{i_x=1}^{n_x} I(x_{i_x} > u_j),$$

$$\widehat{TPR}_{\text{emp}}(u_j) = \frac{1}{n_y} \sum_{i_y=1}^{n_y} I(y_{i_y} > u_j),$$

である ($j = 1, 2, \dots, n_x + n_y$)。ここに, $I(\cdot)$ は括弧内が真であれば 1, 偽であれば 0 をとる。このとき, ROC 曲線は座標軸上にデータ点 $(\widehat{FPR}_{\text{emp}}(u_j), \widehat{TPR}_{\text{emp}}(u_j))$ をプロットすることで与えられる。

2.3 正規分布に基づく ROC 曲線

EROC 曲線では, その結果が階段関数で与えられるため, 標本サイズが少ない場合に, 適切な AUC および最適カットオフ値が推定できない。また, パラメトリック ROC 曲線のようにパラメータ推定値による要約ができない (Zhou et al., 2002)。パラメトリック接近法では, $F_x(x)$, $F_y(x)$ の潜在基礎分布を予め想定することで滑らかな ROC 曲線を構成することができる。Egan(1975) は, 潜在基礎分布に正規分布を用いた ROC 曲線 (以下, NROC 曲線と略す) を提案している。その利点は, 2 群の位置および尺度の相対測度を表すパラメータで ROC 曲線が要

約できる点にある。

いま、健常者群および患者群の検査値が、それぞれ正規分布 $N(\mu_x, \sigma_x)$ および $N(\mu_y, \sigma_y)$ に従うとする。このとき、任意のカットオフ値 u における $TPR(u)$ 、および $FPR(u)$ は

$$\begin{aligned} FPR_{ND}(u) &= 1 - \Phi\left(\frac{u - \mu_x}{\sigma_x}\right) \\ &= 1 - \Phi(z), \\ TPR_{ND}(u) &= 1 - \Phi\left(\frac{u - \mu_y}{\sigma_y}\right) \\ &= 1 - \Phi(\beta_1 z + \beta_0) \quad (2.1) \end{aligned}$$

で与えられる。ここに、 $\beta_1 = \sigma_x/\sigma_y$ 、 $\beta_0 = (\mu_x - \mu_y)/\sigma_y$ であり、 $\Phi(\cdot)$ は標準正規分布の累積分布関数である。検査値 $\{x_{i_x}\}_{i_x=1}^{n_x}$ 、および $\{y_{i_y}\}_{i_y=1}^{n_y}$ が与えられたとき、 β_0, β_1 の推定値 $\hat{\beta}_0, \hat{\beta}_1$ は何らかの方法(たとえば、 μ および σ の不偏推定量あるいは最尤推定量)を用いる。NROC 曲線は、カットオフ値 $u_j, j = 1, 2, \dots$ を用いて $\widehat{FPR}_{ND}(u_j), \widehat{TPR}_{ND}(u_j)$ を算出することで、データ点 $(\widehat{FPR}_{ND}(u_j), \widehat{TPR}_{ND}(u_j)), j = 1, 2, \dots$ によって構成される。

実地において、検査値が正規分布に従うことは稀であり、このような場合には、NROC 曲線は誤った解釈を導く恐れがある。Pepe(1997, 1998, 2000) は、式(2.1)を $\Phi(\cdot)$ に関して逆変換することで(Q-Qプロットの範疇で)、一般化線形モデルの枠組みによってROC曲線を捉えている。

2.4 ベキ変換に基づくROC曲線

2.3節では、検査値が正規分布に従うと仮定したが、実際には、この仮定が満たされることは少ない。そのため、Molodianovitch *et al.* (2006) は、正規性の充足を意図して、検査値にベキ変換(Box & Cox, 1964)を施し、そのもとでNROC曲線を適用している。ベキ変換は、対数変換、平

方根変換といった、データ変換を包括する。したがって、NROC 曲線に対するベキ変換の応用は、検査値の潜在基礎分布の分布形状に対して、対数変換(あるいは無変換後の)NROC 曲線よりも柔軟に対処できる。

いま、健常者群において n_x 個の検査値 $\{x_{i_x}\}_{i_x=1}^{n_x}$ 、患者群において n_y 個の検査値 $\{y_{i_y}\}_{i_y=1}^{n_y}$ が与えられている。さらに、 $x_{i_x}^{(\lambda)}, y_{i_y}^{(\lambda)}$ をパラメータ λ によるベキ変換値

$$\begin{aligned} x_{i_x}^{(\lambda)} &= \begin{cases} \frac{x_{i_x}^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \log x_{i_x} & , \lambda = 0 \end{cases} \\ y_{i_y}^{(\lambda)} &= \begin{cases} \frac{y_{i_y}^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \log y_{i_y} & , \lambda = 0 \end{cases} \end{aligned}$$

を与えるとき、ベキ変換に基づく方法(PTROC 曲線)は、 $x_{i_x}^{(\lambda)}, y_{i_y}^{(\lambda)}$ に対してNROC 曲線を構成することで得られる。このとき、ベキ変換パラメータ λ は不明であることから、その最尤推定値 $\hat{\lambda}$ が用いられる。

3. ベキ正規分布に基づくROC曲線

NROC 曲線では、患者群および健常者群の検査値が正規分布に従うことを仮定している。その利点は、正規分布に基づく統計的推測の概念を適用できる点にある。ただし、実際の検査値の多くがこのような仮定を満たすことは少ない。そのため検査値の正規性の仮定の充足を狙って、2.4節で紹介したようにベキ変換に基づくROC曲線が提案されている。ただし、この方法では、健常者群および疾患群の潜在基礎分布が同一形状に従うことを仮定しなければならない。実際に、疾患群と非疾患群に異なるベキ変換を施した場合には、新たな被験者の検査値が得られた場合に、どちらの変換を施せばよいか判らない。これらの対処法として、それぞれの群のベキ変

換前の分布の枠組みに基づいて ROC 曲線を構成することが考えられる。本論文では、それぞれの群の潜在基礎分布にベキ正規分布 (Goto *et al.*, 1979, 1983 : Goto & Inoue, 1980) を想定したベキ正規 ROC 曲線 (以下, PNROC 曲線) を提案する。これにより, 潜在基礎分布が異なる形状をもつ場合にも適用できるだけでなく, ベキ正規分布が正規分布を包括するため, NROC 曲線の適切性の評価にも利用できる。

3.1 ベキ正規分布

いま, 正値をとる確率変数 X がベキ正規分布 $PND(\lambda, \mu, \sigma)$ に従うとき, ベキ変換パラメータ λ によるベキ変換後の確率変数を $X^{(\lambda)}$ とする。このとき, X の確率密度関数は

$$f_{PND}(x) = \frac{x^{\lambda-1}}{A(\lambda, \mu, \sigma)\sqrt{2\pi\sigma}} \exp\left\{-\frac{(x^{(\lambda)} - \mu)^2}{2\sigma^2}\right\} \quad (3.1)$$

で定義される (Goto *et al.*, 1979, 1983 : Goto & Inoue, 1980)。ここに, μ と σ^2 はベキ変換後の検査値 $x^{(\lambda)}$ が近似的に正規分布に従うときの平均と分散を意味する。また, $A(\lambda, \mu, \sigma)$ は

$$A(\lambda, \mu, \sigma) = \begin{cases} \Phi(s), & \lambda \neq 0, \\ 1, & \lambda = 0 \end{cases}$$

で与えられる確率比例定数項であり, $1 - A(\lambda, \mu, \sigma)$ がベキ変換後に想定される片側打ち切り正規分布での打ち切り確率の大きさを表す。ここに, $s = (\lambda\mu + 1)/(\lambda\sigma)$ であり, $\Phi(\cdot)$ は標準正規分布の累積分布関数である (Goto *et al.*, 1979, 1983 : Goto & Inoue, 1980)。

ベキ変換後の検査値の実際の分布は, 検査値が正であるという条件から, 厳密には正規分布に従わない。そのため, ベキ正規分布では確率比例定数項によって打ち切り確率を調整している。このとき, $A(\lambda, \mu, \sigma)$ を考慮した場合

の, ベキ正規分布のパラメータの推定を行うことは, 難しいために (Goto & Inoue, 1980), 通常では, $A(\lambda, \mu, \sigma) \approx 1$ と想定したもとの計算する。さらに, 変換後の検査値が近似的な正規分布に従えば, 既存の正規分布に基づく諸種の統計的方法を活用できることから有用である (後藤他, 1991 : 下川・後藤, 2004)。そのため, 本論文では, 式 (3.1) の $A(\lambda, \mu, \sigma) \approx 1$ を仮定したもとの ROC 曲線を構成する。因に, $A(\lambda, \mu, \sigma) \approx 1$ におけるベキ正規分布でのパラメータの推定量の一致性などについては, 後藤他 (1991), Zarembka (1974), 濱崎・後藤 (1996) に与えられている。

3.2 ベキ正規 ROC 曲線の構成

いま, ベキ正規母集団 $PND(\lambda_x, \mu_x, \sigma_x)$ から抽出された n_x 個の健常者群の検査値 $\{x_{i_x}\}_{i_x=1}^{n_x}$, および, ベキ正規母集団 $PND(\lambda_y, \mu_y, \sigma_y)$ から抽出された n_y 個の疾患群の検査値 $\{y_{i_y}\}_{i_y=1}^{n_y}$ を考える。 $\lambda_x, \lambda_y, \mu_x, \mu_y, \sigma_x, \sigma_y$ が既知のとき, $A(\lambda, \mu, \sigma) \approx 1$ を仮定したもとの, カットオフ値 u に対する感度 $TPR_{PND}(u)$ および $1 -$ 特定度 $FPR_{PND}(u)$ は

$$FPR_{PND}(u) = 1 - \Phi(z)$$

$$TPR_{PND}(u) =$$

$$\begin{cases} 1 - \Phi\left(\frac{\{\zeta(z)\}^{\lambda_y/\lambda_x} - \lambda_y\mu_y - 1}{\lambda_y\sigma_y}\right), & \lambda_x \neq 0, \lambda_y \neq 0 \\ 1 - \Phi\left(\frac{\exp\{\xi(z)\}^{\lambda_y} - \lambda_y\mu_y - 1}{\lambda_y\sigma_y}\right), & \lambda_x = 0, \lambda_y \neq 0 \\ 1 - \Phi((\log \zeta(z) - \lambda_x\mu_x)/\lambda_x\sigma_y), & \lambda_x \neq 0, \lambda_y = 0 \\ 1 - \Phi(\rho - (\mu_x - \mu_y)/\sigma_y), & \lambda_x = 0, \lambda_y = 0 (\lambda_x = \lambda_y \text{ も含む}) \end{cases} \quad (3.2)$$

で与えられる。ここに, $z = (u^{(\lambda_x)} - \mu_x)/\sigma_x$, $\rho = \sigma_x/\sigma_y$, $\xi(z) = \sigma_x z + \mu_x$, $\zeta(z) = \lambda_x \xi(z) + 1$ である。通常, $\lambda_x, \lambda_y, \mu_x, \mu_y, \sigma_x, \sigma_y$ は未知である

ため、これらのパラメータの最尤推定値 $\hat{\lambda}_x, \hat{\lambda}_y, \hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y$ でおきかえることで、1-特定度 $FPR_{PND}(u)$ および感度 $TPR_{PND}(u)$ の推定値 $\widehat{FPR}_{PND}(u), \widehat{TPR}_{PND}(u)$ を得る。したがって、PNROC 曲線は、カットオフ値 $u_j (j = 1, 2, \dots)$ に対して、 $(\widehat{FPR}_{PND}(u_j), \widehat{TPR}_{PND}(u_j))$ を座標軸上にプロットすることで与えられる。

ベキ正規分布が正規分布を包括しているため、PNROC 曲線は、NROC 曲線を特別な場合として表すことができる。すなわち、仮説 $H_0: \lambda_x = \lambda_y = 1.0$ が成りたつとき、NROC 曲線の適切性がいえる。これらの選定には、赤池の情報量規準 (AIC: Akaike's Information Criteria) を用いることができる (下川・後藤, 2004)。

3.3 最適カットオフ値の選定

ROC 曲線は、臨床検査値の評価だけでなく、任意の臨床検査値に基づいて疾病の有無を識別するための最適カットオフ値の選定にも用いられる。

Patton & Woolfenden(1989) は、カットオフ値の選択によるコストを

$$C(u) = C_0 + TPR(u)p(C_{TP} - C_{FN}) + FPR(u)(1-p)(C_{FP} - C_{TN}) + pC_{FN} + (1-p)C_{TN} \quad (3.3)$$

と定義し、これを最小にするカットオフ値 u^\dagger を推定することで、最適カットオフ値での感度 $TPR(u^\dagger)$ および1-特定度 $FPR(u^\dagger)$ を算出している。ここに、 p は有症率であり、 C は何らかのコスト (費用など)、そして C_0 は臨床検査値でのコストを表す。最適カットオフ値 u^\dagger の推定は、式 (3.3) をカットオフ値 u に関して微分し、これを0とおくことで与えられる。

Somaza & Mossman(1991) は、NROC 曲線における最適カットオフ値 u^\dagger を

$$u^\dagger = \beta_0 \beta_1 \sqrt{\beta_0^2 + 2(1 - \beta_1^2) \log \left\{ \frac{1-p}{\beta_1 p} R \right\}} / (1 - \beta_1)$$

のように与えている。ここに $R = (C_{FP} - C_{TN}) / (C_{FN} - C_{TP})$ はコスト比と呼ばれ、偽陽性によるコスト/偽陰性によるコストを表す。

同様に、PNROC 曲線に対する最適カットオフ値 u^\dagger も式 (3.3) より計算することができる。すなわち、PNROC 曲線での感度 $TPR(u)$ および1-特定度 $FPR(u)$ を用いて、式 (3.3) で与えられるコストが最小になるように、最適な値 u^\dagger を選定する。ただし、PNROC 曲線ではNROC 曲線のように陽的に最適カットオフ値を表すことができないため、式 (3.3) をカットオフ値 u に関して微分し、これを0とおいた後で、事前に決定されるパラメータ p, R を右辺におき、次式

$$v(u) = \frac{(1-p)}{p} R \quad (3.4)$$

を満たす u を (数値) 計算によって解くことにより、最適カットオフ値 u^\dagger を得る。ここに、 $v(u) = f(u)/g(u)$ は

$$v(u) = \begin{cases} \rho \{ \varsigma(z) \}^{(\lambda_y - \lambda_x) / \lambda_x} \phi \left(\frac{\{ \varsigma(z) \}^{\lambda_y / \lambda_x} - \lambda_y \mu_y - 1}{\lambda_y \sigma_y} \right) / \phi(z), & \lambda_x \neq 0, \lambda_y \neq 0 \\ \rho \exp \{ \lambda_y \xi(z) \} \phi \left(\frac{\exp \{ \lambda_y \xi(z) \} - \lambda_y \mu_y - 1}{\lambda_y \sigma_y} \right) / \phi(z), & \lambda_x = 0, \lambda_y \neq 0 \\ \frac{\rho}{\varsigma(z)} \phi \left(\frac{\log \varsigma(z) - \lambda_x \mu_y}{\lambda_x \sigma_y} \right) / \phi(z), & \lambda_x \neq 0, \lambda_y = 0 \\ \rho \phi \left(\rho z - \frac{1}{\sigma_y} (\mu_x - \mu_y) \right) / \phi(z), & \lambda_x = 0, \lambda_y = 0 \end{cases}$$

である。ここに、 $\phi(z)$ は、標準正規分布の確率密度関数である。

コスト比の選定は、費用あるいは過誤の重みによって決定される。例えば、医学分野では、偽陰性によるコストのほうが偽陽性によるコストよりも高く見積もらなければならない。ただし、

偽陰性によるコストの大きさは、対象疾患の重篤度あるいは罹患率に依存することから、これらの決定には、実際の適用場面における専門家の経験・知識が必要である (Zhou et al., 2002).

4. 事例および数値検証

ここでは、事例および数値検証を通して PNROC 曲線の性能を既存の方法と対比し、評価する。

4.1 事例検討

実際の臨床検査の評価場面では、(1) 患者群および健常者群の潜在基礎分布を予め規定しなくてもよいこと、(2) 構成が簡便であること、から EROC 曲線が頻繁に利用されている。ただし、標本サイズが比較的少ない場面では、EROC 曲線の階段関数のステップ幅が大きくなるため、AUC および最適カットオフ値の結果に疑義がもたれる。また、EROC 曲線では、(正規分布に基づく) 既存の統計的推測の方法を拡張できない。そのため、本論文では、正規分布を基調とした 3 種類のパラメトリック ROC 曲線 (NROC 曲線, PTROC 曲線, および PNROC 曲線) をとりあげた。

パラメトリック ROC 曲線の適合の良さの比較には、AIC などの情報量基準を用いることができる。ただし、ROC 曲線は統計的グラフィックスの一つであり、視覚的にその有用性を評価することが重要である。Huang & Pepe(2007) は、パラメトリック ROC 曲線の視覚的な評価に EROC 曲線からの乖離度を用いている。本論文では、これに倣い、視覚的な評価を行う。

事例 1 : 卵巣癌データ

卵巣癌の有無を識別するために、卵巣癌患者 30 名および健常者 23 名の遺伝子データが観測されている。Pepe(2003) は成書のなかで諸種

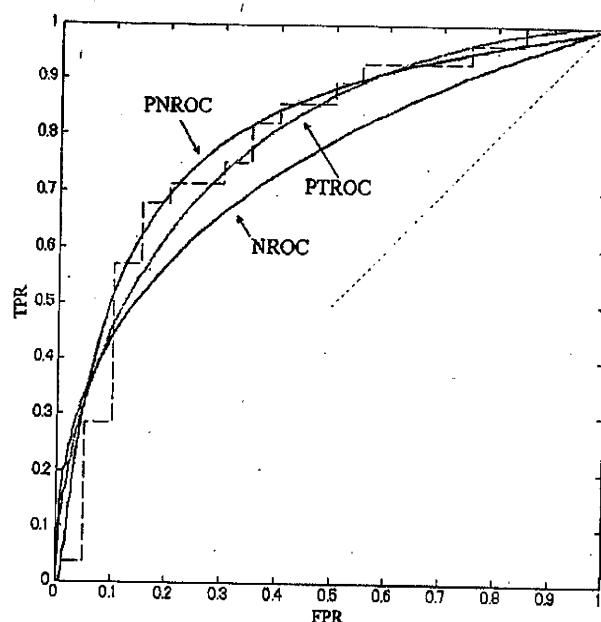


図 2: 卵巣癌データに対する ROC 曲線 (EROC 曲線: 経験 ROC 曲線, PNROC 曲線: ベキ正規 ROC 曲線, NROC 曲線: 正規 ROC 曲線, PTROC 曲線: ベキ変換に基づく ROC 曲線)

の ROC 曲線の説明に本データを用いている。ここでは、NROC 曲線, PTROC 曲線および PNROC 曲線の性能評価に本データを用いる。

図 2 は、卵巣癌データに対する ROC 曲線である。NROC 曲線は、EROC 曲線 (階段関数) に比べて下側に布置しており、NROC 曲線が適切な結果を示さなかった。これに対して、PTROC 曲線および PNROC 曲線は EROC 曲線に適合した形状を示した。

ROC 曲線のモデル (分布) の適切性を評価するために、潜在基礎分布に基づく AIC を計算した。ここで、PTROC 曲線は、ベキ変換に基づいているため、他の方法と AIC を比較することができない。そのため、PTROC 曲線の代替としてベキ正規分布の形状パラメータを $\lambda_x = \lambda_y = \lambda$ としたもとの、PNROC 曲線の AIC を計算した。NROC 曲線での AIC は、568.0 であり、他の 2 手法に比して極端に適合が悪かった。これに対して、 $\lambda_x = \lambda_y = \lambda$ をしたときの PTROC 曲線

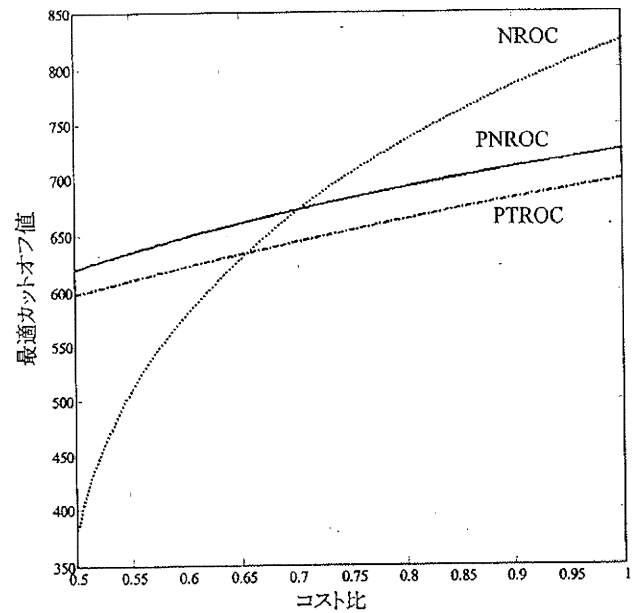
の AIC は 543.2 であり, PNROC 曲線の AIC は, 542.1 であった. すなわち, PNROC 曲線の AIC が最も小さく, 最良の適合を示した. PNROC 曲線の結果, 傾き 45 度の直線 (健常者群と疾患群の分布関数が等しい $F_x(u) = F_y(u)$ ことを表す直線) よりも上部に布置した.

因に, このときの PTROC 曲線におけるベキ変換パラメータの最尤推定値は $\hat{\lambda} = 0.911$ であった. これに対して, PNROC 曲線のベキ変換パラメータの推定値は, $\hat{\lambda}_x = -1.750, \hat{\lambda}_y = -0.367$ であった.

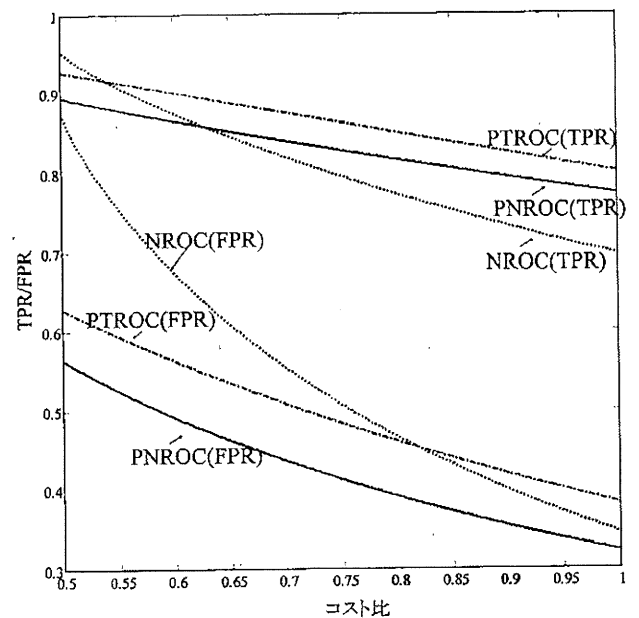
次いで, コスト比に対する最適カットオフ値と FPR および TPR を推定した (図 3 (a)). このとき, 有症率 p は 2 群の比率に基づき $30/(23 + 30) = 0.57$ とした. コスト比として, 0.5 から 1.0 までの範囲をプロットした理由は, 偽陰性によるコスト (すなわち, 疾患を見落とすことによるコスト) のほうが, 偽陽性によるコスト (すなわち, 疾患でないことを見落とすコスト) に比べて被験者に対するリスクが高いためである.

コスト比が 0.65 未満のとき, NROC 曲線の最適カットオフ値の推移は, 他の 2 手法の推移に比して大幅に低い値を示したものの, 0.70 以上では大幅に高い値を示した. 前述したように, NROC 曲線の当てはまりは悪く, これにより, 他の 2 手法とは異なる最適カットオフ値を示した.

これに対して, PTROC 曲線および PNROC 曲線の最適カットオフ値の推移は類似していた. ただし, PNROC 曲線のほうが PTROC 曲線よりも最適カットオフ値が 25 程度高かった. PNROC 曲線において, 偽陰性のコストを重視するとき (コスト比が 0.5 のとき), 最適カットオフ値が 624 であり, 偽陰性と偽陽性のコストが等しいとき (コスト比が 1.0 のとき), 最適カットオフ値は 725 であった. すなわち, PTROC 曲線



(a) コスト比対最適カットオフ値のプロット



(b) コスト比対 FPR, TPR プロット

図 3: 卵巣癌データに対する最適カットオフ値の選定のほうが, PNROC 曲線に比して, 卵巣癌と診断する可能性が高くなる傾向にあった.

コスト比に対する FPR および TPR の変動を図 3 (b) に示す. NROC 曲線での TPR および FPR の変動曲線は, いずれにおいても他の 2 手法に比べて急激な減少傾向にあった. PNROC

曲線と PTROC 曲線の TPR および FPR の推移は、類似した傾向を示したが、PNROC 曲線のほうが、PTROC 曲線に比して、TPR が 0.03 程度低く、そして、FPR が 0.04 程度低かった。すなわち、PNROC 曲線の最適カットオフ値は、PTROC 曲線に比べてコスト比に対する TPR および FPR が低かった。

事例 2：頭部外傷症データ

Zhou *et al.* (2002) は、頭部外傷症の重篤度を識別するために、CK-BB(クレアチン・キナーゼ BB) が有効か否かを検討している。ここでは重篤度が重度および非重度によって与えられている。本解析の目標は、CK-BB が重篤度の判定に有用な指標であるかを検討することである。Zhou *et al.* (2002) は、重症群および非重症群のそれぞれが対数正規分布に従っていることを指摘しており、対数変換後のデータに対して NROC 曲線を適用している。そのため、本解析では、NROC 曲線、PTROC 曲線、PNROC 曲線および対数変換後のデータに対する NROC 曲線 (log(NROC 曲線)) での性能を評価する。

図 4 は、本データに対する ROC 曲線である。NROC 曲線は、EROC 曲線から大きく外れており、Zhou *et al.* (2002) の示唆を裏づける結果 (正規分布に従わない) を示した。log(NROC 曲線) は、対数変換後のデータに NROC 曲線をあてはめた結果である。log(NROC 曲線) は、NROC 曲線に比して EROC 曲線の近くに位置しているが、小さな FPR では EROC 曲線から外れていた。他方、PTROC 曲線、および PNROC 曲線は小さな FPR でも EROC 曲線付近に布置しており、良好な適合を示した。このときの PTROC 曲線のベキ変換パラメータは、 $\hat{\lambda} = 0.178$ であり、PNROC 曲線のベキ変換パラメータは、 $\hat{\lambda}_x = 0.361, \hat{\lambda}_y = 0.195$ ($\hat{\lambda}_x$ は非重症群のベキ変換パラメータの推定値であり、 $\hat{\lambda}_y$

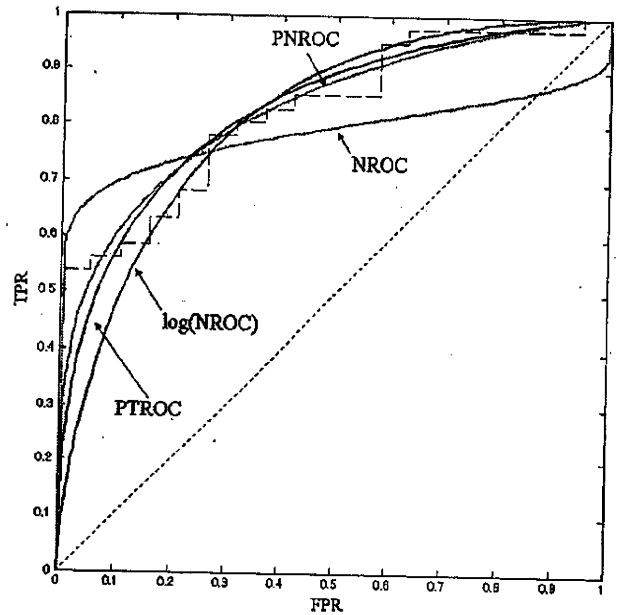
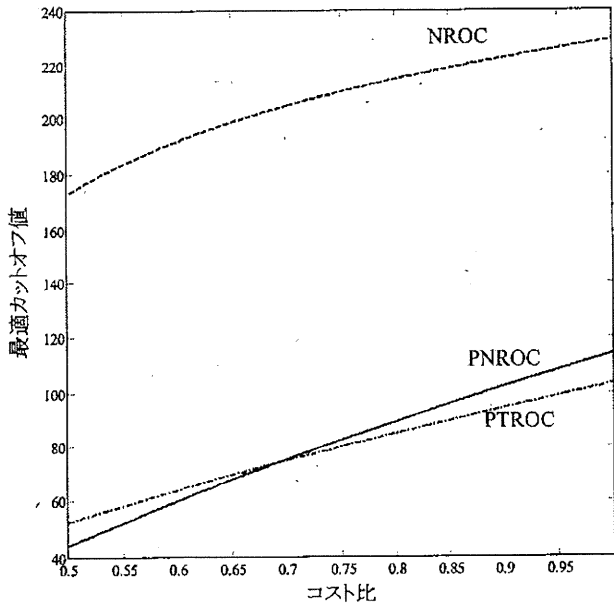


図 4：頭部外傷症データに対する ROC 曲線 (EROC 曲線:経験 ROC 曲線, PNROC 曲線:ベキ正規 ROC 曲線, NROC 曲線:正規 ROC 曲線, log(NROC 曲線):対数変換データに対する正規 ROC 曲線, PTROC 曲線:ベキ変換に基づく ROC 曲線)

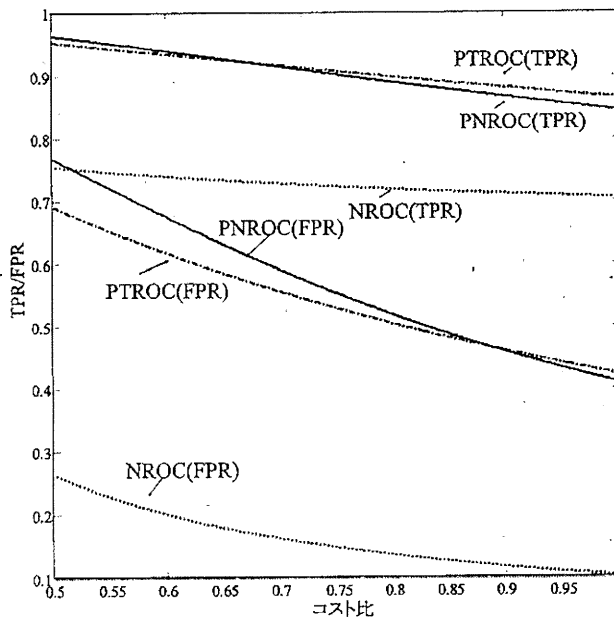
は重症群のベキ変換パラメータの推定値である)であった。すなわち、本データでは、対数変換よりも、むしろ平方根変換 (ベキ変換パラメータ λ が 0.5 のとき、平方根変換を表す) のほうが適切なようである。モデル (分布) の適切性を評価するために、事例 1 と同様の方法で各 ROC 曲線に対して AIC の値を計算した。NROC 曲線での AIC の値は、662.9 であり、 $\lambda_x = \lambda_y = \lambda$ としたときの PTROC 曲線の AIC は 634.01 であり、PNROC 曲線の AIC は、635.40 であった。すなわち、PTROC 曲線が PNROC 曲線に比して僅かに AIC の値が小さかった。PTROC 曲線および PNROC 曲線の結果、傾き 45 度の直線 (重篤患者群と非重篤患者群の分布関数が等しい $F_x(u) = F_y(u)$ ことを表す直線) よりも上部に布置した。

さらに、コスト関数 (3.3) を用いて各 ROC 曲線に基づく最適カットオフ値を推定した (図

5). ここで、有症率は、本データにおける重症群の割合 $41/(19+41) = 0.683$ とした。NROC 曲線は、いずれのコスト比でも NROC 曲線の最適カットオフ値が他の 2 手法に比して極端に高かった。コスト比が 0.70 未満のとき、PNROC 曲線の最適カットオフ値は、PTROC 曲線よりも僅かに低かったが、コスト比が 0.7 以上のと



(a) コスト比 対 最適カットオフ値のプロット



(b) コスト比 対 FPR, TPR プロット

図 5: 頭部外傷症データに対する最適カットオフ値の選定

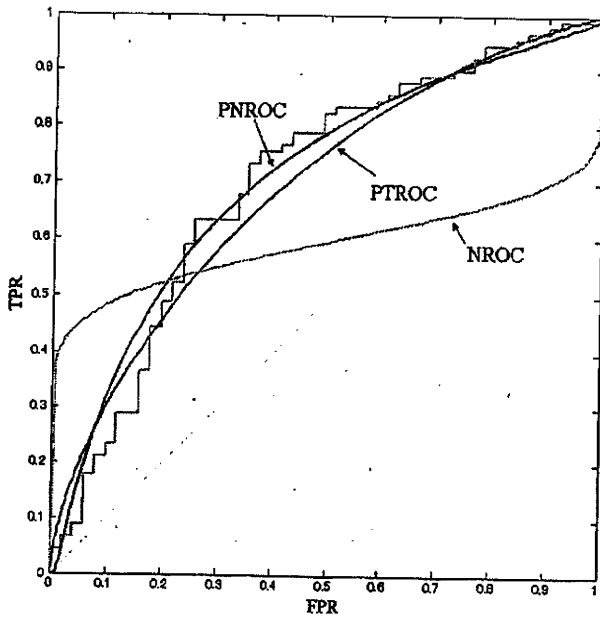
き、その傾向は逆転した。ただし、2 個の ROC 曲線の最適カットオフ値には、顕著な差異が認められなかった。

図 5 (b) はコスト比に対する最適カットオフ値での FPR および TPR である。最適カットオフ値が他の 2 手法に比べて非常に大きかった NROC 曲線の TPR および FPR が極端に低く、FPR においてその傾向が顕著であった。また、PNROC 曲線と PTROC 曲線の TPR には、殆ど差異が認められなかった。コスト比が小さいとき、PNROC 曲線と PTROC 曲線の FPR に若干の差異が認められたものの、コスト比が大きくなるにつれて、その差異は減少した。そして、コスト比が 0.85 以上では、PNROC 曲線と PTROC 曲線の差異は殆ど認められなかった。

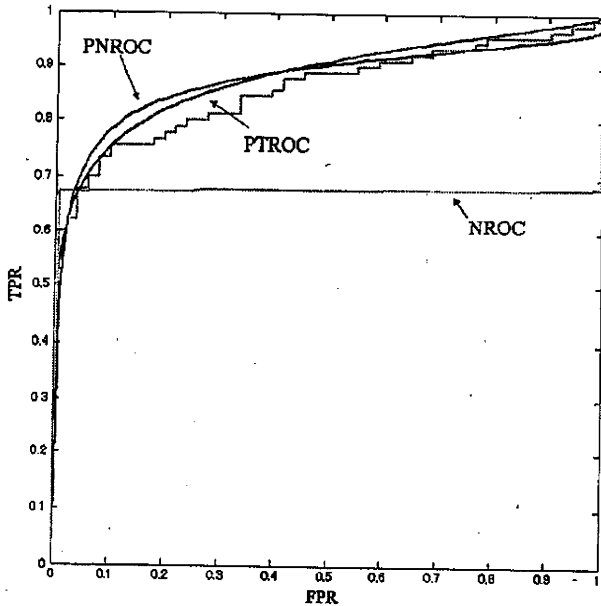
事例 3: 膵臓癌のバイオマーカ・データ

膵臓癌に対する適切な血清バイオマーカを評価するために、90名の膵臓癌患者および51名の健常者が採血され、血清バイオマーカ CA125 および CA19-9が測定されている (Wieand *et al.*, 1989)。CA125は、卵巣癌および子宮癌のバイオマーカとして知られており、他方、CA19-9は、膵臓癌のバイオマーカとして用いられている。本データは、いずれのバイオマーカも 2 群の分布形状が異なっていることが指摘されている (Pepe, 2003; Zhou *et al.*, 2002)。ここでの目標は、2 群の分布形状が異なる場合における PNROC 曲線の適切性を提示することにある。

図 6 (a) は、CA125に対する ROC 曲線であり、図 6 (b) は CA19-9に対する ROC 曲線である。NROC 曲線は、CA125および CA19-9のいずれにおいても EROC 曲線から外れていた。CA125の PTROC 曲線は、EROC 曲線から僅かに外れていた。このときの、ベキ変換パラメータの推定値および AIC の値を表 1 に示す。PNROC 曲線の AIC はいずれのバイオマーカ



(a) CA125



(b) CA19-9

図 6: ROC 曲線に基づく膵臓癌のバイオマーカーデータの比較

でも最良であることから (CA19-9のとき1369.7であり, CA125のとき806.7である), PNROC 曲線の適切性が示唆された。このときの CA125 におけるベキ変換パラメータは, 膵臓癌患者群で $\hat{\lambda}_{x1} = -0.745$ であり, 健常者群で $\hat{\lambda}_{y1} = -0.342$ であった。ただし, CA19-9において, 1-特定度

表 1: 膵臓癌のバイオマーカーデータに対する諸種の推定値。ここに, 括弧は $\lambda_x = \lambda_y = \lambda$ と仮定したもとの PNROC 曲線での AIC である。

		PNROC	PTROC	NROC
CA125	$\hat{\lambda}_{x1}$	-0.745	-0.425	-
	$\hat{\lambda}_{y1}$	-0.342		
	AIC	806.7	(808.5)	1241.9
	AUC	0.713	0.696	0.593
CA19-9	$\hat{\lambda}_{x2}$	-0.388	-0.149	-
	$\hat{\lambda}_{y2}$	0.009		
	AIC	1369.7	(1380.9)	1793.6
	AUC	0.880	0.885	0.677

が0.1から0.4の範囲では, PNROC 曲線および PTROC 曲線が EROC 曲線よりも上部に布置した。すなわち, ベキ正規分布の適合度が低い場合には, 適切な結果が得られないことが示唆された。

さらに, CA19-9におけるベキ変換パラメータは, 膵臓癌患者群の場合に $\hat{\lambda}_{x2} = -0.388$, および健常者群の場合に $\hat{\lambda}_{y2} = 0.009$ であった。したがって, いずれの群も左に大きく歪んだ分布に従っていた。

CA125に対する PNROC 曲線は, CA19-9に比して傾き45度の直線 (健常者群と疾患群の分布関数が等しい $F_x(u) = F_y(u)$ ことを表す直線) の近くに布置していることから, CA19-9のほうが CA125よりも良好なバイオマーカーであることが示唆される。

次いで, AUC を用いて2個のバイオマーカーを比較した。PNROC の AUC は, CA125 のときに0.713であり, CA19-9のときに0.880であることから, AUC の差は顕著であり ($\Delta AUC = 0.167$), CA19-9の適切性が裏づけられる。