Recent advances on inference for the area under the ROC curve with small samples

Giuliana Cortese, Laura Ventura

Department of Statistical Sciences, University of Padova

ISCB 30
Prague, August 23-27th, 2009

Outline

- Introduction:

  Diagnostic test
  Area under the ROC curve (AUC)

- Two examples:
  (1) Ultrasound measurements of abdominal aorta Aneurysm diameter
  (2) ALCL lymphoma

- Small sample size: Point estimation and confidence interval for AUC

Introduction

- Two populations: healthy and diseased individuals
- Diagnostic characteristic (biomarker) on the two populations: Continuous measurement on healthy subjects (X) and on diseased subjects (Y)
- Parametric assumptions on the distribution family for X and Y (e.g. X, Y normally distributed).

ROC curve and the area AUC

- A diagnostic test based on a continuous response, with cut-off point T, classifies patients as non-diseased ($T-$) or diseased ($T+$)
- The accuracy of the test T is measured by
  - Sensitivity $= \mathrm{P}(T + |\text{diseased}) = \mathrm{P}(Y > T)$,
  - Specificity $= \mathrm{P}(T - |\text{non-diseased}) = \mathrm{P}(X < T)$.
- Varying T, the global performance of the diagnostic test (accuracy in discriminating):
  - ROC curve (plot of sensitivity versus 1-specificity)
  - area under the ROC curve (AUC), $P(X < Y)$

Data example (Sprouse et al., 2003) : Determining seriousness of aneurysms from ultrasound diameter measurements

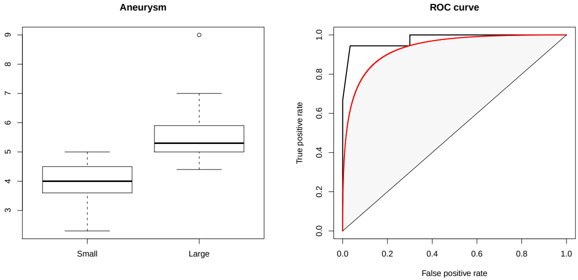
Abdominal aortic aneurysm = localized blood-filled dilation of the abdominal aorta.
Accurate diameter measurements are essential for screening and planning surgical intervention. As the size increases, there is an increased risk of rupture.

- **Aim**: performance of a diagnostic test based on ultrasound (US) measurements of the diameter in distinguishing between small and large aneurysms.
- Patients has been classified in two groups: small (S) or large (L) aneurysm (by using a gold standard measure).
- Two samples of US measurements from S and L, with small size, $n = 10$ and $m = 10$.

Detecting Aneurysm ...

Parametric assumption (Kotz et al., 2003)

- US on a small aneurysm patient $\rightarrow X \sim N(\mu_x, \sigma^2)$
- US on a large aneurysm patient $\rightarrow Y \sim N(\mu_y, \sigma^2)$



Inference for $\Psi = P(X < Y)$?

- point estimate $\hat{\Psi}$, confidence interval $[\hat{\Psi}_l, \hat{\Psi}_u]$.
- accuracy with small sample size

Inference for $\Psi = \text{P}(X < Y)$

$\theta = (\Psi, \lambda)$, Ψ = parameter of interest, λ = nuisance parameter

Classical likelihood based inference procedures

$\hat{\theta} = (\hat{\Psi}, \hat{\lambda})$ MLE of θ $\hat{\lambda}_\Psi$ MLE of λ for a given Ψ

The profile log-likelihood function: $\ell_p(\Psi) = \ell(\Psi, \hat{\lambda}_\Psi)$

- Wald statistic: $Z(\Psi) = (\hat{\Psi} - \Psi)\hat{j}_p^{1/2}$, with j_p observed profile Fisher information
- Signed log-likelihood ratio: $r_p(\Psi) = \text{sgn}(\hat{\Psi} - \Psi)(2(\ell_p(\hat{\Psi}) - \ell_p(\Psi)))^{1/2}$

(null distribution $N(0,1)$ with order of accuracy $O(n^{-1/2})$)

(Severini, 2000)

Inference for $\Psi = \mathrm{P}(X < Y)$

Likelihood-based estimation can be inaccurate when the sample size is small (Brazzale et al., 2007; Obuchoski and Lieber, 1998) .
High-order likelihood-based inference (Severini, 1999) based on the quantity

$$r^*(\Psi) = r_p + \frac{1}{r_p} \log \frac{c \cdot u}{r_p} \qquad \text{(modified log-likelihood ratio)}$$

provides more precise inference.

- $c = c(\Psi)$ and $u = u(\Psi)$ non-trivial expressions computed from first and second derivatives of likelihood quantities.
- Null distribution is $N(0,1)$ with order of accuracy $O(n^{-3/2})$.
- A corrected point estimation $\hat{\Psi}^*$ is obtained from the equation $r^*(\Psi) = 0$.

Confidence intervals for the area Ψ based on r^*

- compute $r^*(\Psi)$ for a range of values around the MLE;
- interpolate the point by a smoothing method;
- invert the interpolating function and find the corresponding values Ψ_1^* and Ψ_2^* as solutions to the equations $r^* = z_{\alpha/2}$ and $r^* = z_{1-\alpha/2}$.

(Ψ_1^*, Ψ_2^*) is the confidence interval for Ψ at level $1 - \alpha$.

r^* for the normal case

Assume X, Y ind., $X \sim N(\mu_x, \sigma^2)$ $Y \sim N(\mu_y, \sigma^2)$,
(Φ standard normal d.f.)

Reparameterize from (μ_x, μ_y, σ^2) to $\theta = (\Psi, \lambda_1, \lambda_2)$,

parameter of interest $\Psi = \Phi\left(\frac{\mu_y - \mu_x}{\sqrt{2\sigma^2}}\right)$, (Kotz et al., 2003)

nuisance parameters $\lambda_1 = \frac{\mu_x}{\sqrt{2\sigma^2}}$ $\lambda_2 = \sqrt{2\sigma^2}$.

- $(x_1, \ldots, x_n)$ i.i.d. sample of size n from X,
  $(y_1, \ldots, y_m)$ i.i.d. sample of size m from Y
- MLE for Ψ: $\hat{\Psi} = \Phi(\frac{\hat{\mu}_y - \hat{\mu}_x}{\sqrt{2\hat{\sigma}^2}})$,
  obtained with the invariance property from MLE $(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}^2)$.
- compute $r^*(\Psi)$, with $c(\Psi)$ and $u(\Psi)$ known expressions (Cortese and Ventura, 2009)

Aneurysm: point and interval estimation

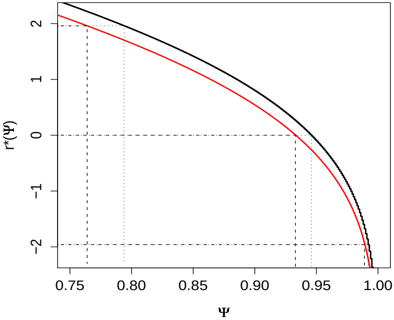
Ultrasound on $\begin{cases} \text{small aneurysm patient} \rightarrow X \sim N(\mu_x, \sigma^2) \\ \text{large aneurysm patient} \rightarrow Y \sim N(\mu_y, \sigma^2) \end{cases}$

Confidence Intervals

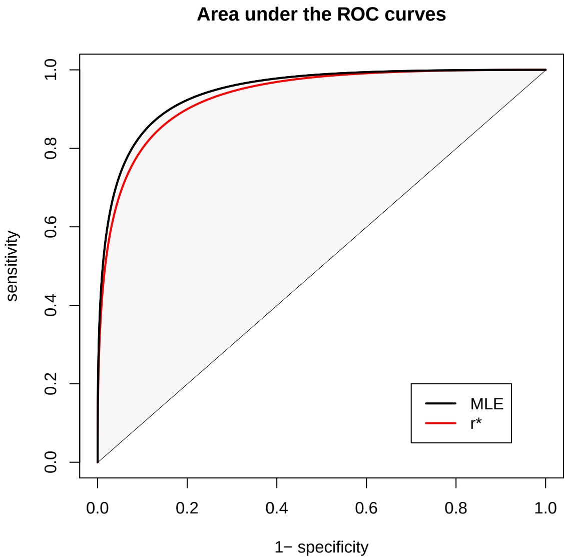
Wald	r_p (signed log-likelihood ratio)	r^*
(0.90, 0.99)	(0.79, 0.99)	(0.76, 0.99)

Point estimates

- $\hat{\Psi} = 0.95$ MLE
- $\hat{\Psi}^* = 0.93$,
  as solution to $r^*(\Psi) = 0$



Aneurysm: ROC curves



Simulation studies: Normal distribution

Two-sided empirical coverage of confidence intervals for Ψ, and non-coverage probabilities on the left and right tail.

		$\psi = 0.8$ $(\mu_x = 5, \mu_y = 6.55, \sigma = 1.3)$		$\psi = 0.95$ $(\mu_x = 5, \mu_y = 8.5, \sigma = 1.5)$	
(n, m)	Method	90%	95%	90%	95%
(5,5)	Z	0.585	0.638	0.516	0.638
		(0.071,0.344)	(0.051,0.311)	(0.010,0.474)	(0.051,0.311)
	r_p	0.845	0.914	0.835	0.914
		(0.045,0.110)	(0.025,0.060)	(0.026,0.139)	(0.025,0.060)
	r^*	0.900	0.952	0.906	0.952
		(0.048,0.053)	(0.025,0.023)	(0.045,0.049)	(0.025,0.023)
(10,10)	Z	0.649	0.717	0.587	0.627
		(0.074,0.277)	(0.041,0.024)	(0.015,0.398)	(0.003,0.370)
	r_p	0.880	0.932	0.869	0.932
		(0.040,0.080)	(0.022,0.046)	(0.032,0.099)	(0.013,0.055)
	r^*	0.899	0.951	0.897	0.951
		(0.049,0.053)	(0.024,0.026)	(0.049,0.054)	(0.024,0.026)
(30,30)	Z	0.692	0.767	0.619	0.699
		(0.100,0.218)	(0.057,0.175)	(0.065,0.315)	(0.027,0.274)
	r_p	0.887	0.943	0.893	0.940
		(0.043,0.070)	(0.024,0.033)	(0.039,0.067)	(0.021,0.039)
	r^*	0.897	0.948	0.899	0.945
		(0.050,0.053)	(0.026,0.026)	(0.053,0.048)	(0.028,0.027)

r^* for the exponential case

Assume X, Y ind., $X \sim Exp(\alpha)$ $Y \sim Exp(\beta)$,

Reparameterize from (α, β) to $\theta = (\Psi, \lambda)$,

parameter of interest $$\Psi = \frac{E_\alpha(X)}{E_\alpha(X) + E_\beta(Y)} = \frac{\alpha}{\alpha + \beta},$$

nuisance parameter $$\lambda = \alpha + \beta.$$

- MLE for Ψ: $\hat{\Psi} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}$ obtained from MLE $(\hat{\alpha}, \hat{\beta})$.
- compute $r^*(\Psi)$, with $c(\Psi) = 1$, and $u(\Psi)$ a known expression (Cortese and Ventura, 2009)

Data example: Anaplastic lymphoma (ALCL)

Hsp70 protein levels are higher in subjects with ALCL lymphoma then in healthy subjects.

- **Aim**: evaluate the accuracy of the diagnostic characteristic, Hsp70 protein, in distinguishing between healthy and diseased subjects
- Assume Hsp70 protein levels are distributed as $X \sim Exp(\alpha)$ (healthy) and $Y \sim Exp(\beta)$ (diseased).
- Two i.i.d. samples from X and Y: "controls" and "cases" groups, with sizes $n = 4$ and $m = 10$.

Confidence Intervals

Wald	r_p (signed log-likelihood ratio)	r^*
(0.72, 0.99)	(0.63, 0.95)	(0.60, 0.92)

Conclusions and outlook

- The Wald statistic tends to overestimate the accuracy of the classifier.
  For small sample size, r^* gives more precise estimates.
  For the area Ψ, r^* tends to shift the confidence interval to the left.
  Thus, r_p slightly overestimates the classifier accuracy.
- Extend to other distribution assumptions
  (Poisson data, truncated distribution)
- pAUC (partial area under the ROC curve),
  for instance when a high specificity is required.

References

Brazzale A.R., Davison A.C., Reid N. (2007): *Applied Asymptotics. Case Studies in Small Sample Statistics*. Cambridge University Press.

Cortese G, Ventura L. (2009): *Likelihood asymptotics for the stress-strength model* $P(X < Y)$. Proceeding of the Italian Statistical Society, XLV Scientific Meeting, Milan, Italy.

Giummolè F., Ventura L. (2002): Practical point estimation from higher-order pivots. *J. Statist. Comput. Simul.*, **72**, 419–430.

Kotz S., Lumelskii Y., Pensky M. (2003): *The Stress-Strength Model and its Generalizations. Theory and Applications*. World Scientific, Singapore.

Obuchoski N.A., Lieber M.L. (1998): Confidence intervals for the receiver operating characteristic area in studies with small samples. *Academic radiology*, *5*, 561–71.

Severini T.A. (1999): An empirical adjustment to the likelihood ratio statistic. *Biometrika*, **86**, 235–247.

Severini T.A. (2000): *Likelihood Methods in Statistics*. Oxford University Press, New York.