Journal: R-Journal

Title: ROCnReg: An R Package for Receiver Operating Characteristic Curve Inference with and without Covariate Information

Authors: María Xosé Rodríguez-Álvarez and Vanda Inácio

Dear reviewer,

We thank you for your careful review of our paper. Your valuable comments and suggestions have helped us to improve our work, and to prepare a new version of both the paper and the R package (already updated on CRAN). We hope we have successfully addressed all your concerns. When we respond to your queries, we first repeat them in bold font, followed by our response.

Minor comments and curiosities

(1) **Which are the possible consequences, if any, of defining the covariate-adjusted ROC curve according to the distribution of the covariates only in the diseased group?**

Author's response: The point raised by the reviewer is a very interesting one. In our view, there are no consequences but one should be aware of how to interpret the AROC curve. By definition, the covariate-adjusted ROC curve is a weighted vertical average of covariate-specific ROC curves, i.e., the average is done along the TPFs; this is the reason why the weights correspond to the distribution of the covariates in the diseased group. Thus, an operating point in the AROC curve is to be interpreted as the average TPF that is obtained when the thresholds used to define a positive test result are covariate-specific and chosen to ensure the same FPF across all covariate values. Nevertheless, as also pointed out by Janes and Pepe (2009), a (weighted) horizontal average may be more appropriate in certain clinical settings. More specifically, here the AROC curve would be defined as follows

$$\begin{aligned}\text{AROC}_{\text{TPF}}(p) &= \Pr\{Y_{\bar{D}} > F_D^{-1}(1-p \mid \mathbf{X}_{\bar{D}})\} \\ &= \Pr\{1 - F_D(Y_{\bar{D}} \mid \mathbf{X}_{\bar{D}}) \leq p\},\end{aligned}$$

where, by a slight abuse of notation, p here denotes the TPF. Equivalently, the AROC_{TPF} can be defined as follows

$$\text{AROC}_{\text{TPF}}(p) = 1 - \int \text{ROC}_{\text{TNF}}(p \mid \mathbf{x}) \mathrm{d}H_{\bar{D}}(\mathbf{x}),$$

where ROC_{TNF} denotes the covariate-specific TNF-ROC curve (see equation (11) in the manuscript). Note that in this case the weights correspond to the distribution of the covariates in the nondiseased group, and an operating point in the AROC_{TPF} is to be interpreted as the average FPF that is obtained when the thresholds used

to define a positive test result are covariate-specific and chosen to ensure the same TPF across all covariate values.

Although we do not deal specifically with this situation in the package, the user can also fit this curve by simply indicating in argument `tag.h` the value codifying **diseased** individuals. Some care is needed when using the default `plot` function as it assumes that in the x-axis we have the FPF, while now we have the TPF. Below an example of how to fit the curve and obtain the correct plot (Figure 1). As an aside, we note that when the TPF is plotted in the x-axis, the (A)ROC curve will lie below the diagonal. All in all, both definitions of the AROC curve might be of interest, but we have focused in the package on the standard definition since it is the one that has been covered in the literature (to the best of our knowledge the AROC_{TPF} has never been reported nor studied except in the discussion section of the paper by Janes and Pepe (2009)). The study of the most appropriate settings for the use of the standard AROC or the AROC_{TPF} is an interesting topic, but we believe it is beyond the scope of the current paper.

```
R> AROC_bnp_tpf <- AROC.bnp(
+ formula.h = bmi ~ gender + f(age, by = gender, K = c(4,4))
+ group = "cvd_idf", tag.h = 1, data = endosyn, standardise = TRUE,
+ p = seq(0, 1, l = 101), ci.level = 0.95,
+ parallel = "snow", ncpus = 2)

R> plot(AROC_bnp_tpf$p, AROC_bnp_tpf$ROC[,1], xlab = "TPF", ylab = "FPF",
+ xlim = c(0,1), ylim = c(0,1),
+ main = "AROC TPF Bayesian nonparametric",
+ type = "l", cex.lab = 1.3, cex.axis = 1.3)
R> lines(AROC_bnp_tpf$p, AROC_bnp_tpf$ROC[,2], lty = 2, type = "l")
R> lines(AROC_bnp_tpf$p, AROC_bnp_tpf$ROC[,3], lty = 2, type = "l")
R> abline(0,1, col = "grey")
R> legend.text <- paste0("AAUC (95% CI): ",
+  paste(round(AROC_bnp_tpf$AUC[1], 3), " (",
+  round(AROC_bnp_tpf$AUC[2], 3),"",", ",
+  round(AROC_bnp_tpf$AUC[3], 3),")", sep = ""))
R> legend(0, 1, legend.text, bty = "n", cex = 1.3)
```

(2) **Between equation (10) and (11) the following is written: "Thus, the covariate-specific ROC curve is an important tool that helps to understand and determine the optimal and suboptimal populations where to apply the tests on". What is the meaning of 'optimal and suboptimal populations' here?**

Author's response: Note that when using the covariate-specific ROC curve we may obtain a different ROC curve for each covariate value $\mathbf{x}$ in the range of $\mathbf{X}$. In other words, the covariate-specific ROC curve depicts the discriminatory capacity

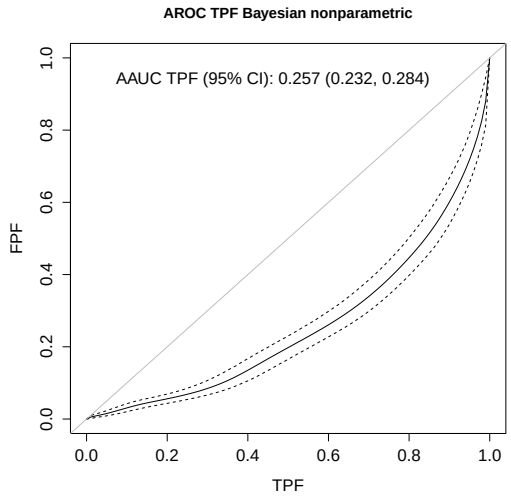

FIGURE 1. Age and gender-adjusted TPF-ROC curve: posterior mean and 95% pointwise credible band.

of the test in populations which are homogeneous with respect to $\mathbf{x}$. As such, it may be used as a tool to help to identify the covariate values (and thus the populations defined by such covariate values) for which the diagnostic test has a 'good' discriminatory capacity ('optimal' populations) or those for which the diagnostic test has a 'poor' discriminatory capacity ('suboptimal' populations). In the revised version of the paper, we have slightly modified the phrasing of this sentence, to make its meaning clearer:

"In this case, a number of possibly different ROC curves (and therefore accuracies) may be obtained for different values of $\mathbf{x}$*. Thus, the covariate-specific ROC curve is an important tool that helps to understand and determine the optimal and suboptimal populations where to apply the tests on (i.e., it allows determining the populations, defined by or homogeneous with respect to* $\mathbf{x}$*, where the diagnostic test has a 'good' or 'poor' discriminatory capacity)"*

(3) **Is there any reason to define equation (22) differently from equations (5) and (14) regarding their corresponding ROC curve?**

Author's response: Note that the pAUC over a range of TPFs $(v_1, 1)$ is specified in terms of the 270° rotation of the standard (pooled, covariate-specific or covariate-adjusted) ROC curve. For the pooled ROC curve and the covariate-specific ROC curve, the rotated curve can be expressed in terms of the (conditional) cumulative distribution function (CDF) of the test outcomes in the healthy and diseased populations (see equation (6) for the pooled ROC curve and equation (11) for the covariate-specific ROC curve). These results are used in the package to calculate

the quantity $\text{pAUC}_{\text{TPF}}(v_1)$. However, for the covariate-adjusted ROC curve it is not possible to define neither the standard AROC curve nor its 270° rotation in terms of the (conditional) CDFs of the test outcomes. As a consequence, in this case the $\text{pAUC}_{\text{TPF}}(v_1)$ has to be calculated on the basis of the AROC curve. In the new version of the paper we have included a figure (Figure 1), which graphically exemplifies how partial AUCs over a range of TPFs can be defined and computed. We hope this figure (in particular Figure 1(b)) helps in understanding Equation (22) of the main manuscript.

(4) **Are there any recommendations to set the hyperparameters m_0, S_0, a and b in the Bayesian nonparametric approach based on a Dirichlet process mixture of normal distributions (function `priorcontrol.bnp`)?**

In the revised version of the paper we provide some guidance on how to specify these hyperparameters and also the rationale behind the default values (please see page 8 for what we have added on this part). We should note that this review has led us to perform an in-deep sensitivity analysis (not only on simulated datasets but also on several real datasets) of the impact of the default hyperparameters' values specified in the package. Based on these analyses, in the new version of the package, we have modified some of the previously default values.

(5) **The main R functions to estimate the ROC curves allows for standardising the test outcomes by the setting the input parameter `standardise = TRUE`. In this case, are the reported thresholds applicable to the standardised test outcome or to the original one?**

Author's response: For the Bayesian approaches, standardisation is performed to help improving the mixing of the MCMC chains and also to facilitate specification of the hyperparameter values. Nevertheless, regardless standardisation being performed or not, the thresholds values are always returned in the original scale.

(6) **In Figure 1, I understand that the "95% pointwise credible band" for the pooled ROC curve has been computed as the 2.5% and 97.5% quantiles for each FPF in the grid set by the input parameter p. Is this correct?**

Author's response: Yes, it is correct. In all approaches presented in the paper the 95% pointwise credible (or confidence in the case of frequentist approaches) band is computed as the 2.5% and 97.5% quantiles for each FPF in the grid set by the input parameter p. As such, pointwise credible or confidence bands are to be interpreted as credible/confidence intervals for the TPFs or sensitivities. This has been clarified in the paper. Also (see point (8)), in the new version of the package we allow the user to specify the level of the credible (or confidence, in the case of frequentist approaches) intervals.

(7) **Figure 3 illustrates the trace plots of the MCMC draws of the PDFs of the BMI based on model `pROC_dpm` to monitor convergence of the**

MCMC chains. Have those trace plots been intentionally conditioned by "extreme" values of the BMI because those are the more sensitive areas in this regard?

Author's response: No, we have chosen the BMI values randomly, although it is true that we tried that they spanned the range of possible values. We do not expect that convergence, in general, will be worse for extreme values, which is also confirmed by the effective sample sizes (Figure 5 in the manuscript). We should note that in the new version of the manuscript we have changed the BMI values at which the trace plots are shown because, when replicating the results, we were, surprisingly, not able to obtain the same BMI values as in the original submission. To avoid this, and to ensure the same results whenever running the code, we now use `set.seed(, "Mersenne-Twister")`.

(8) **By using the function `compute.threshold.[...]`, different methods may be implemented to estimate ROC-based threshold values and 95% credible interval for them and other related measures (posterior mean for the Youden Index, FPF and TPF). Is it possible to obtain, for instance, all the YI-based threshold values or to change the confidence level for the credible interval?**

Author's response: In the way the function is currently implemented, the whole YI-based threshold values chain is not returned but only some summary statistics (mean and 2.5% and 97.5% quantiles). The reason we have proceeded this way is to avoid returning much information that may be difficult to post-process by the user, as well as to avoid a large size of the returned object. Nevertheless, in the new version of the package submitted to CRAN (1.0-5), we allow the user to specify the level of the credible (or confidence, in the case of frequentist approaches) intervals. This has been done by including a new argument, `ci.level` in all functions of the package. By default, `ci.level = 0.95`. We have modified the paper accordingly.

(9) **With the objective of estimating the covariate-specific ROC curve modelling the effect of continuous covariates in a nonlinear way using B-spline basis expansions, one optional input parameter in function f is the number of internal knots to be used. In the provided example, `K = c(0,0)` has been fixed for computing f in formula.h. What does 0 mean here? Regarding also the number of interior knots, in the text it is indicated that such number is assisted by the WAIC, DIC andor LPML. But this is only the case if the input parameter K is not specified by the user, isn't it?**

Author's response: By specifying `K = 0`, a B-spline basis with no internal knots is constructed. Although not obvious, we note that this is equivalent to a cubic polynomial function. Regarding the use of WAIC, DIC and/or LPML to assist in the selection of the number of knots, we shall note that this is not done by the

`ROCnReg` package in an automatic fashion, but the user does need to fit the model for a varying number of interior knots and use the returned WAIC, DIC and/or LPML to choose the most appropriate number of interior knots, i.e., the number for which the WAIC/DIC is the lowest or the LPML the highest (from all fitted models). We agree with the reviewer that this point was not clear in the previous version of the manuscript, and we have modified this part, which now reads as follows:

"We also note that to assist in the selection of the number of interior knots (in `ROCnReg` the location is always based on the quantiles of the corresponding covariates), the user can make use of the WAIC, DIC, and/or LPML. For instance, for this application, we fitted different models with different number of internal knots and we have chosen the model that provided the lowest WAIC (this was done in both the healthy and diseased populations and we remark that the number of knots does not need to be the same in the two populations) The final model is shown below."

(10) **Does the R function `AROC.bnp` perform the same computations as `cROC.bnp` and finally compute the AROC curve? If this was the case, wouldn't it be better for the `AROC.bnp` function to accept an output from `cROC.bnp` if it has been previously computed?**

Author's response: We note that equations (18) and (19) in the main manuscript suggest two approaches to estimate the AROC curve. The first (based on equations (19)) exploits the representation of the AROC as the cumulative distribution function of $U_D = 1 - F_D(Y_{\bar{D}} \mid \mathbf{X}_D)$. Under this approach, estimation is based on first estimating $F_{\bar{D}}(\cdot \mid \mathbf{x})$ and then the outer probability in (19) (or, equivalently, the cumulative distribution function of U_D) is estimated empirically. An alternative approach (based on equations (18)) is based on first estimating the covariate-specific ROC curves, and then averaging them according to the distribution of the covariates in the diseased population. This approach, although direct, requires the estimation of the (possibly multivariate) distribution function $H_D(\mathbf{x})$, which may not be an easy task. Therefore, in the package, the estimation of the AROC curve for all implemented methods is based on the first approach. As a result, there is no need to first estimate the covariate-specific ROC curve (which may also be time consuming; please see also next point) in order to estimate the AROC. This is the reason why we have created different functions to estimate the covariate-specific ROC curve and the AROC curve.

(11) **I miss some comments about computation times in the manuscript. In the R code provided as a separate file, there is a note of "time consuming" for creating the object called `cROC_bnp` by using the function `cROC.bnp` with `L = 10` and nonlinear effects of the covariate age. I tried to compute it and I stopped it after 15 minutes. There were some other computational intensive lines of code, but not as much as the referred one. It may be interesting to provide some notes in this field.**

Author's response: Following your suggestion, in the revised version of the manuscript we have included some results regarding computing times. In particular, a new section called **Computational aspects** has been included. Please see pages 26/27 for its content.

(12) **Typos: `nthin` instead of `nskip` on page 11; absolute value in equation (17); $pAUC_{TPF}(v_1)$ in equation (22); FFP on page 3; `predictive.ckecks` on page 17**

Author's response: We thank the reviewer for letting us know of these typos. They have been corrected in the revised version of the manuscript. We should, however, note that we need the absolute value when defining the Youden index for the covariate-specific ROC curve. The reason is that, since in this case a 'different' ROC curve is estimated for each covariate value, it may happen that for some covariate values the associated ROC curve may lie below the diagonal (classification is always based on assuming that larger values of the test are more indicative of disease), and thus the YI may be below 0. Therefore, to take that into account (and correct for it) when computing the YI and associated threshold value we use the absolute value.

References

Janes, H. and M. S. Pepe (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika 96*(2), 371–382.