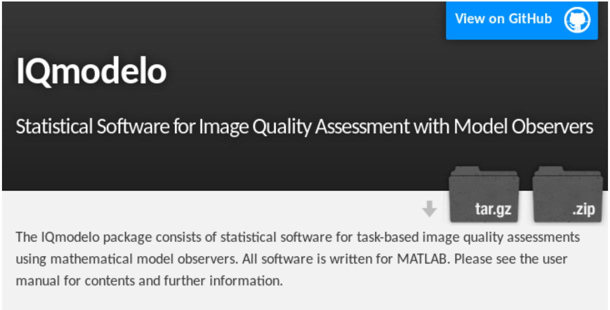
IQmodelo: Statistical software for task-based image quality assessment with model observers

Adam Wunderlich

Center for Devices and Radiological Health
U.S. FDA, Silver Spring, MD

IQmodelo

- Public domain statistical analysis software for task-based image quality assessments with model observers
- Compatible with MATLAB® release R2007b and newer, some functions require Statistics Toolbox™
- https://github.com/DIDSR/IQmodelo
  frontpage: https://didsr.github.io/IQmodelo/



Overview of Contents

- Parametric methods for linear observers
  - useful for ROC assessment of known-location detection/discrimination tasks
  - "exact" confidence intervals
  - channelized Hotelling observers
  - channelized Hotelling observers with known difference of class means
  - known-template linear observers
  - known-template linear observers with known difference of class means
- General methods
  - relaxed assumptions, approximate confidence intervals
  - nonparametric ROC, LROC, and EROC analysis for fixed readers
  - nonparametric analysis of binomial proportions for fixed readers
  - ANOVA-based multi-reader multi-case analysis for random readers
- Demos
- User manual – complete list of functions and references

Parametric Methods: channelized Hotelling observers

- each image multiplied with channel matrix to get channel output vector
- general case:
  - Wunderlich et al., IEEE Trans. Med. Imag., 2015
  - assumptions: channel outputs for signal absent and signal-present images follow multivariate normal distributions with common covariance matrix
    – typically satisfied for known-location signal detection/discrimination
- known difference of class means:
  - Wunderlich and Noo, IEEE Trans. Nucl. Sci., 2013
  - additional assumption: difference of means for signal-present and signal-absent channel outputs is known
    –reconstruction of high-dose data may yield accurate approximation

Parametric Methods: known-template linear observers

- ratings produced by taking inner product of template and image
- general case:
  - Wunderlich and Noo, Med. Phys., 2011
  - assumptions: ratings for signal absent and signal-present images are normally-distributed with common variance
    – typically satisfied for known-location signal detection/discrimination
- known difference of class means:
  - Wunderlich and Noo, IEEE Trans. Nucl. Sci., 2012
  - additional assumption: difference of means for signal-present and signal-absent ratings is known
    –reconstruction of high-dose data may yield accurate approximation

Nonparametric Methods: fixed observers

- ROC
  - Mann-Whitney (trapezoidal) AUC point estimates
  - variance/covariance analysis: DeLong et al., Biometrics, 1988
- LROC
  - generalization of Mann-Whitney and DeLong methods
  - Wunderlich and Noo, IEEE Trans. Med. Imag., 2012
- EROC
  - generalizes LROC to combined detection and estimation tasks
  - Wunderlich and Goossens, SPIE J. Med. Imag., 2014
- Binomial proportions
  - examples: probability of correct decision in MAFC task, probability of correct localization in LROC task
  - Noo et al., Proc. SPIE Med. Imag. Conf., 2013

General ANOVA-based MRMC analysis with random readers

- Obuchowski-Rockette method with Hillis' degrees of freedom (ORH method)

  References:

  Obuchowski and Rockette, Commun. Stat. Simulat. 1995

  Hillis, Stat. Med. 2007, 2014

- applicable to any figure of merit
- requires fixed-reader covariance matrix that can be estimated with functions described previously for nonparametric ROC, LROC, EROC, and MAFC analysis

Confidence Interval Coverage Parameters

The analysis functions return $1 - \alpha$ confidence intervals, where

$\alpha = \alpha_1 + \alpha_2$

α_1 : controls lower bound

α_2 : controls upper bound

Examples:

- 2-sided 95% interval: $\alpha_1 = \alpha_2 = 0.025$
- 1-sided 95% interval: $\alpha_1 = 0.05$, $\alpha_2 = 0$

List of Demos

- known-location signal detection, ROC analysis
  - `demo1a.m` : one imaging scenario, 2-sided CIs
  - `demo1b.m` : same as demo1a, but with 1-sided CIs
  - `demo2a.m` : two imaging scenarios, 2-sided CIs
  - `demo2b.m` : same as demo2a, but with 1-sided CIs
- unknown-location signal detection, LROC analysis
  - `demo3.m` : one imaging scenario
  - `demo4.m` : two imaging scenarios

Demo Structure

Because each demo presents performance estimates obtained from a single random data set, the results

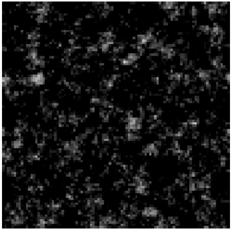
- are subject to statistical variations
- do not justify general conclusions

Demos 1 and 2: Setup

- **Task:** known-location signal detection
- **Observer:**
  - known-template linear observer
    - template: uniform disk roughly same size as signal
  - channelized Hotelling observer (CHO)
    - 3 DOG channels
    - 40 Gabor channels
- **Figure of merit:** area under ROC curve (AUC)

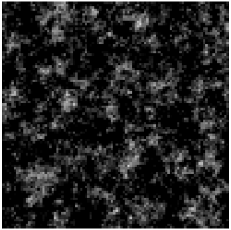
Demos 1 and 2: Images

- image size: 96×96
- colored gaussian noise background following power-law model
  - common noise component across imaging scenarios $\Rightarrow$ correlation
- gaussian signal added at center
- specified set of seeds for random number generator $\Rightarrow$ same results
- 150 signal-absent images, 150 signal-present images



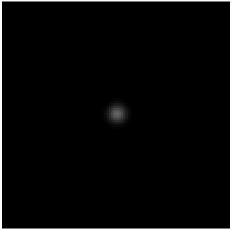
signal absent

grayscale: [0, 50]



signal present

grayscale: [0, 50]



signal

grayscale: [0, 25]

```
demo1a
----------------------- demo 1a ----------------------------------
computing lesion-absent images (m=150) ...
computing lesion-present images (n=150) ...

Displaying two-sided 95% confidence intervals for AUC:

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fixed template linear observer:
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2AFC estimates:

p =

    0.7933


p_CI =

    0.7208    0.8509

nonparametric (Mann-Whitney) estimates:

AUC =

    0.8068


AUC_CI =

    0.7532    0.8511
```

Demo 1a results

observer	method	point estimate	95% CI (2-sided)
disk-template	2AFC	0.793	[0.721, 0.851]
disk-template	Mann-Whitney-DeLong	0.807	[0.753, 0.851]
disk-template	parametric	0.805	[0.754, 0.850]
disk-template	parametric known-means	0.783	[0.765, 0.802]
CHO-DOG	parametric	0.810	[0.757, 0.853]
CHO-DOG	parametric known-means	0.794	[0.775, 0.813]
CHO-Gabor	parametric	0.890	[0.819, 0.907]
CHO-Gabor	parametric known-means	0.880	[0.860, 0.900]

Demo 1b results

observer	method	point estimate	95% CI (1-sided)
disk-template	2AFC	0.793	[0.734, 1]
disk-template	Mann-Whitney-DeLong	0.807	[0.763, 1]
disk-template	parametric	0.805	[0.763, 1]
disk-template	parametric known-means	0.783	[0.768, 1]
CHO-DOG	parametric	0.810	[0.766, 1]
CHO-DOG	parametric known-means	0.794	[0.778, 1]
CHO-Gabor	parametric	0.890	[0.827, 1]
CHO-Gabor	parametric known-means	0.880	[0.863, 1]

Demo 2

- Compare two fixed imaging scenarios with confidence interval for $\mathrm{AUC}_B - \mathrm{AUC}_A$
- Same signal and noise statistics for each scenario
- Common noise component across scenarios $\Rightarrow$ correlated results
- For the parametric methods,
  - known-template, known difference of class means method provides accurate CI for AUC difference (assumes same correlation for signal-absent and signal-present ratings)
  - other parametric methods rely on Bonferroni inequality to obtain (overly) conservative confidence intervals (see appendix of manual)
- The nonparametric methods give accurate confidence intervals for AUC difference, but do not utilize known difference of class means

Common one-sided hypotheses

Non-inferiority test: (δ = non-inferiority margin)

H_0: $\mathrm{AUC}_B - \mathrm{AUC}_A \leq -\delta$
inferiority

H_1: $\mathrm{AUC}_B - \mathrm{AUC}_A > -\delta$
non-inferiority

- used to assess "substantial equivalence"

Superiority test: ($\delta = 0$)

H_0: $\mathrm{AUC}_B - \mathrm{AUC}_A \leq 0$
non-superiority

H_1: $\mathrm{AUC}_B - \mathrm{AUC}_A > 0$
superiority

For either test:

- Can evaluate with one-sided $1-\alpha$ intervals of the form $[\Delta\mathrm{AUC}_L,\ 1]$ by setting $\alpha_1 = \alpha$, $\alpha_2 = 0$
- For a given α, one-sided intervals are more efficient than two-sided intervals

Demo 2a results

observer	method	$\widehat{AUC}_A$	$\widehat{AUC}_B$	95% CI (2-sided)
disk-template	2AFC	0.793	0.827	$[-0.054, 0.121]$
disk-template	Mann-Whitney-DeLong	0.807	0.791	$[-0.080, 0.049]$
disk-template	parametric	0.805	0.784	$[-0.134, 0.091]$
disk-template	parametric known-means	0.783	0.787	$[-0.022, 0.030]$
CHO-DOG	parametric	0.810	0.778	$[-0.147, 0.081]$
CHO-DOG	parametric known-means	0.794	0.789	$[-0.047, 0.038]$
CHO-Gabor	parametric	0.890	0.908	$[-0.072, 0.119]$
CHO-Gabor	parametric known-means	0.880	0.900	$[-0.025, 0.065]$

Demo 2b results

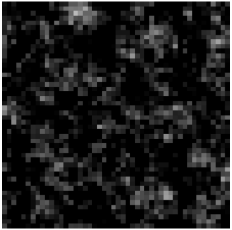
observer	method	$\widehat{AUC}_A$	$\widehat{AUC}_B$	95% CI (1-sided)
disk-template	2AFC	0.793	0.827	$[-0.040, 1]$
disk-template	Mann-Whitney-DeLong	0.807	0.791	$[-0.070, 1]$
disk-template	parametric	0.805	0.784	$[-0.120, 1]$
disk-template	parametric known-means	0.783	0.787	$[-0.018, 1]$
CHO-DOG	parametric	0.810	0.778	$[-0.133, 1]$
CHO-DOG	parametric known-means	0.794	0.789	$[-0.042, 1]$
CHO-Gabor	parametric	0.890	0.908	$[-0.060, 1]$
CHO-Gabor	parametric known-means	0.880	0.900	$[-0.020, 1]$

Demos 3 and 4: Setup

- **Task:** unknown-location signal detection
- **Observer:**
  - known-template scanning-linear observer
    - template: uniform disk roughly same size as signal
    - template scanned over all pixels
    - rating = maximum score
    - mark = location of maximum score
- **Figures of merit:**
  - probability of correct signal localization (P_{CL})
  - area under LROC curve ($\mathcal{A}_L$)

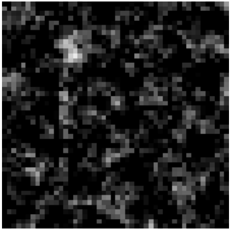
Demos 3 and 4: Images

- image size: 48×48
- colored gaussian noise background following power-law model
  - common noise component across imaging scenarios $\Rightarrow$ correlation
- gaussian signal added at random location (uniformly distributed)
- specified set of seeds for random number generator $\Rightarrow$ same results
- 150 signal-absent images, 150 signal-present images



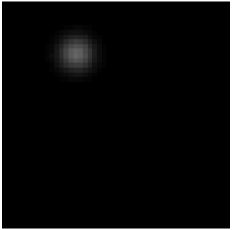
signal absent

grayscale: $[0, 25]$



signal present

grayscale: $[0, 25]$



signal

grayscale: $[0, 25]$

```
----------------------- demo 3 ----------------------------------
computing lesion-absent images (m=150) ...
computing lesion-present images (n=150) ...
scanning linear observer:

Displaying point estimates and two-sided 95% confidence intervals...

Probability of correct localization:

PCL =

    0.5867

PCL_CI =

    0.5060    0.6629

Area under LROC curve:

AL =

    0.4903

AL_CI =

    0.4167    0.5639
```

```
----------------------- demo 4 ----------------------------------
computing lesion-absent images (m=150) ...
computing lesion-present images (n=150) ...
scanning linear observer:

Displaying point estimates and two-sided 95% confidence intervals for performance differences...

Probability of correct localization:

PCL =

    0.5867
    0.6200

PCL_CI =

   -0.0722    0.1389

Area under LROC curve:

AL =

    0.4903
    0.4742

AL_CI =

   -0.1036    0.0714
```

Summary

- https://github.com/DIDSR/IQmodelo
- See manual and demo code for details on function usage
- Always justify assumptions!
- Some approaches for efficiently utilizing sample images:
  - use observer models and estimators that don't require splitting images into separate training and testing sets
  - one-sided confidence intervals for one-sided hypotheses
  - obtain class mean difference and incorporate into performance estimation
  - more complex tasks, e.g., unknown-location signal detection