War-Winning Capabilities … On Time, On Cost

A Large Sample Confidence Interval for Quantile Differences Between Two Independent Radar Tests

ITEA, May 2014. Las Vegas, NV

Dr. Arnon Hurwitz - Edwards AFB, CA 93524

Mr. Todd Remund – Edwards AFB, CA 93524

Contact: arnon.hurwitz@edwards.af.mil

412th Test Wing

Confidence Intervals for Binary Responses
ITEA – May 2014, Las Vegas, NV

Statistical Methods Group, Edwards AFB

Overview

- Blip-scan radar output returns are: detect/no-detect, or {0, 1}
- Probability of detection **π** (== **p**) increases as Range decreases
- A common metric is R50 – the range at which $100\pi = 50$ (i.e. 50%)
- A common question is: given two flights, what is a confidence interval (C.I.) for the *difference of the two R50's ?*
- Such differences are non-linear functions of parameters of the estimation procedure: its own distribution is hard to derive
- A solution to find a C.I. for a difference is to use a Bootstrap procedure = a non-parametric simulation approach
- Bootstrapping works, but it has to be custom-generated for each different problem at hand. It's sometimes preferable to have a parametric method. We develop such a method here based on the Maximum Likelihood Covariance of the estimation parameters.

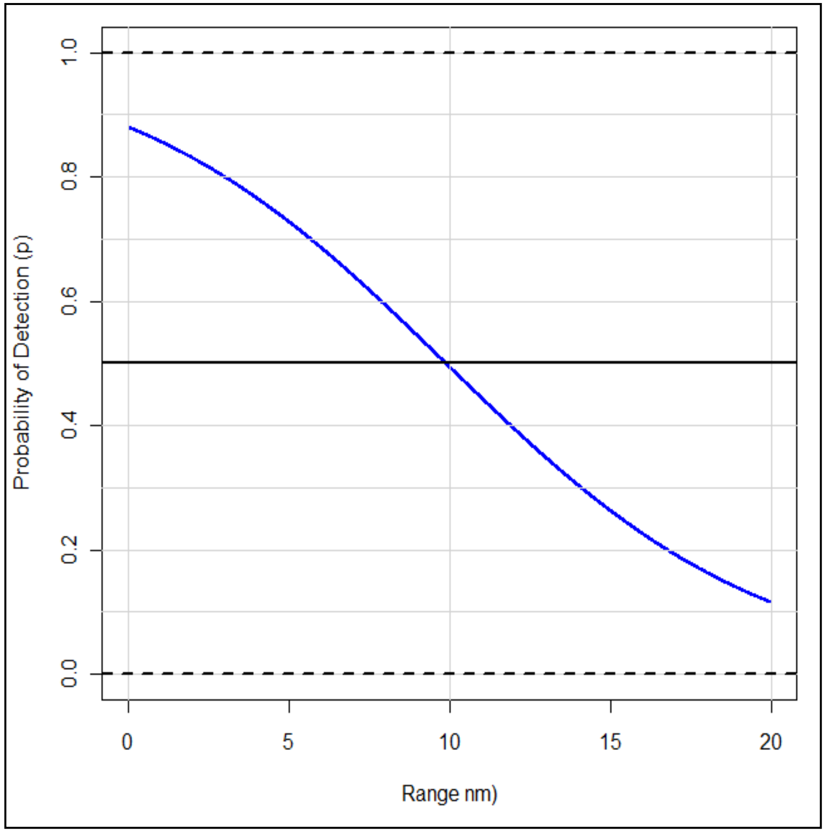
Logistic curve fit to Binary data

- One has the relation:
  - Output = function(Range)
- But output is binary {0, 1}, and we'd rather wish to find something like:
  - π = function(Range)
- Transform the problem:

  $y(R) = \log [\pi/(1 - \pi)] = \alpha + \beta R$

- Now have $-\infty < y < \infty$, and a linear relation of a kind. Note:

  $\pi = \exp(y) / [1 + \exp(y)]$

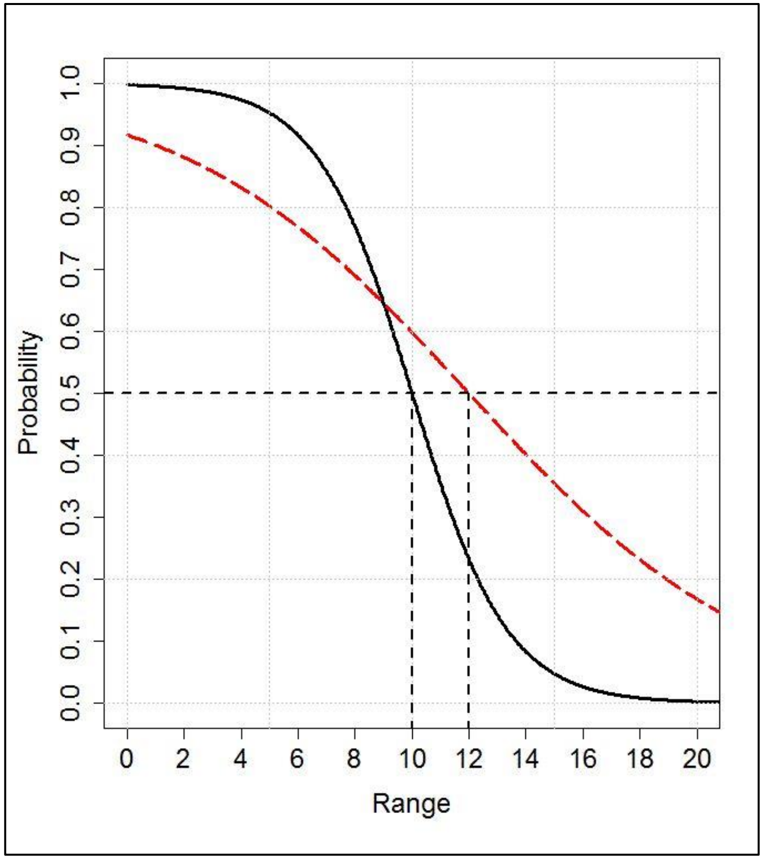


A Logistic curve

Probability π is on the vertical axis

Comparing two logistic curves

- $\log[\pi/(1-\pi)] == \text{logit}(\pi)$ is called the **logit**. Note: log = ln
- The opposite graph shows two such curves (i.e. two flights)
- At $\pi = 0.5$ the black curve shows R50 at about 10nm, the red curve at about 12nm
- The difference is about 2: We want a 95% confidence interval around this difference



Two logistic curves

Here, Prob(detect) π decreases with increasing Range (R)

Estimation

- **Let one curve be estimated with non-linear regression techniques (*generalized linear modeling*) to give the equation**

$$logit(\pi) = \alpha_0 + \alpha_1 R$$

and let the other curve be estimated as

$$logit(\pi) = \beta_0 + \beta_1 R$$

- **At π = 0.5, logit(π) = log(0.5/0.5) = log(1) = 0.**

So for the first curve R0 = $-\widehat{\alpha}_0/\widehat{\alpha}_1$**, and R1=** $-\widehat{\beta}_0/\widehat{\beta}_1$ **for the second.**
Their estimated difference is therefore R0 – R1 = $-(\hat{\alpha}_0/\hat{\alpha}_1 - \hat{\beta}_0/\hat{\beta}_1)$

- Generalized linear modeling uses **maximum likelihood estimation (MLE)** techniques to estimate the coefficients of the models, and also gives us the **Covariance Matrix of the α and β parameters**

- **Call this covariance matrix V. It is a 4x4 symmetric matrix.**

Confidence Interval

- **It can be shown (by MLE large-sample theory) that**

$$(\widehat{R1} - \widehat{R0}) \,\dot{\sim}\, \textbf{Normal}(\mathbf{R1 - R0, hVh'})$$

Where V is the covariance matrix, and where

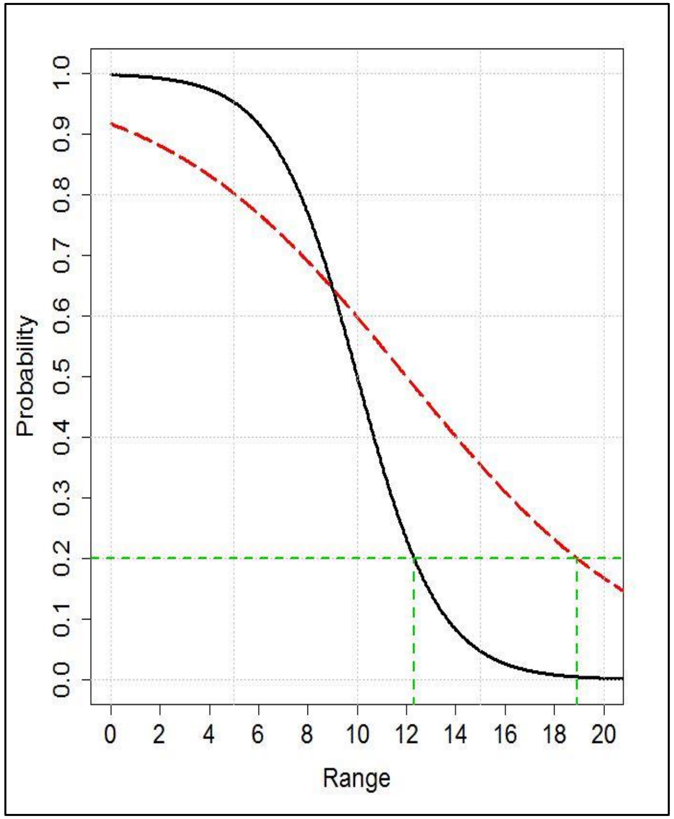
$$\mathbf{h} = \left(\frac{1}{\alpha_1}, \frac{-\alpha_0}{{\alpha_1}^2}, \frac{-1}{\beta_1}, \frac{\beta_0}{{\beta_1}^2}\right)$$

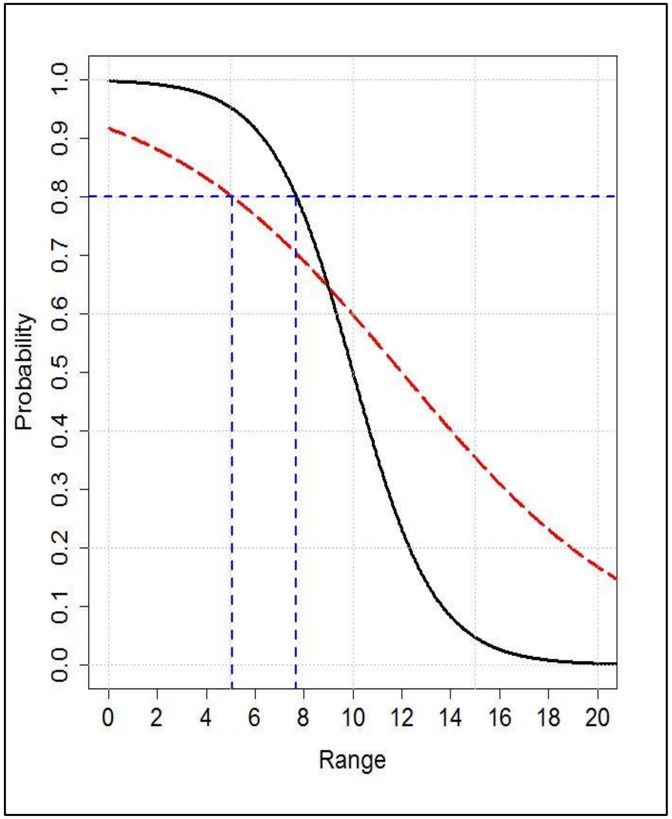
This gives us the (95%) confidence interval that we desire as:

$$(\widehat{R1} - \widehat{R0}) - 1.96 \times \hat{h}\hat{V}\hat{h}' < R1 - R0 < (\widehat{R1} - \widehat{R0}) + 1.96 \times \hat{h}\hat{V}\hat{h}'$$

C.I. for the General Case (cont.)

- So far, we've developed a CI for R50; that is, where π = p = 0.5
- Sometimes percentiles other than R50 are required; e.g. p=0.2, p=0.8





C.I. for the General Case

- We can get a CI for any value ‘p’ in (0,1) by replacing the ‘**h**’ in our formula with

$$\mathbf{h} = \left(\frac{1}{\alpha_1}, \frac{(y_c - \alpha_0)}{{\alpha_1}^2}, \frac{-1}{\beta_1}, \frac{-(y_c - \beta_0)}{{\beta_1}^2}\right)$$

where y_c = logit(p) = log[p/(1-p)]

- This enables us to get a CI for the difference between any quantile $\boldsymbol{R}_{100p}$, not just R50
- The assumptions here are: we have a large sample, and the flights are independent; this is a typical scenario for blip-scan radar

Monte-Carlo Simulation Test

1. We specify flights A and B with model: $y = logit(\pi) = \theta_0 + \theta_1 R$, where $\{\theta_0, \theta_1\} = \{6.0, -0.6\}$ for A and $\{\theta_0, \theta_1\} = \{2.4, -0.2\}$ for B.
2. We specify a maximum range length of 30nm, and divided it into 4,000 evenly spaced Range values ($R_i, i = 1, \ldots, 4000$)
3. Each flight A, and B, then gets 4,000 y_i values corresponding to each R_i value via the appropriate model for A or B.
4. Each of these y_i values are turned into a probability (π_i) value via the logistic function $\pi_i = \exp(y_i)/(1 + \exp(y_i))$
5. Loop 10,000 times
    1. A blip-scan result is simulated for each range point by taking the π_i corresponding to that range point and generating a 'detect' or 'no detect' as the outcome of a Bernoulli trail; this is done for A and for B.
    2. Logistic regressions of 'detect' as a function of 'Range' for A and for B
    3. Confidence Interval (CI) for the difference of $\boldsymbol{R}_{100p}$ computed & saved
6. Count of CI's containing true $\boldsymbol{R}_{100p}$ difference to get % coverage

Simulation Results

- We ran simulations for R_{100p} values for p = {0.1, 0.2, 0.5, 0.8, 0.9}
- For each value of ‘p’ (i.e. π) the coverage was within 1% of 95%

π	% Coverage
0.1	94.74
0.2	94.94
0.5	95.08
0.8	95.22
0.9	95.20

Summary

- We looked at an analytic large-sample confidence interval (CI) on the difference between the R50 range values for two independent flights
- We extended the results to a CI for the difference between two R_{100p} Range values, where $0 < p < 1$
- Our 'Analytic' method is based on the covariance matrix generated from the MLE procedure of generalized regression
- We compared CI's of our method to the 'true' CI by generating 10,000 Monte-Carlo estimates
- In all cases examined, the 'Analytic' method produced good results for a 95% CI – probability coverage was within 1% of 95% for a range of p values in [0.1 , 0.9] ; that is, R_{100p} between R_{10} and R_{90}

REFERENCES

- Agresti, A. (2002). Categorical Data Analysis, 2nd Ed. John Wiley & sons.
- Efron, B. (1987). Better Bootstrap Confidence Intervals. J. of the American Statistical Association, Vol. 52, No. 397. Theory and Methods.
- Hosmer, D.W., Lemeshow, S. (1989). Applied Logistic Regression. Wiley.
- Kendall, M.G., Stuart, A.S. (1961). The Advanced Theory of Statistics. Vol.1. 2nd edition. Griffin
- McCullagh P., Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall
- Morgan, B.J.T. (1992). Analysis of Quantal Response Data. Chapman & Hall.
- Pawitan, Y. (2001). In All Likelihood. Clarendon Press, Oxford.
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org
- Schwenke, J.R., Milliken, G.A. (1991). On the Calibration Problem Extended to Nonlinear Models. Biometrics 47: 563-574
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc.: 54: 426-482.