Weighting Background-Subtracted Events

Jim Linnemann, MSU

Andrew Smith, U Md

PHYSTAT 2007

June 27, 2007

The Context

Milagro cosmic γ ray experiment

2630 m altitude = 750 g/cm^2 (of 1030) overburden

H_2O Cherenkov pond (+ tank surface array) =

calorimeter after 20.5 Xo, 8.3λ

Task: tell if hadron or γ started the shower

AND: most cosmic rays are hadron-initiated (p, He,…)

No big surprise that $\langle B \rangle \approx 10^3 \langle S \rangle$

Background Subtraction

To see a signal, must subtract background with 10^{-3} precision

We do this: use nearby sky ("sideband")

$$m = n - \hat{B}$$

Consider as a model for large-background LHC signal

Gaussian Significance etc.

$$Z = m/\delta m = m/\sqrt{Var(m)}$$

$$1/Z = \text{fractional error} = \sigma/\mu = \text{Coeff.Variation}$$

$$\mathrm{N_e} = Z^2$$ ***Poisson Events w/o bkg, with same σ/μ***

Ne < *m*, *B*; typical: *m*~1000, Ne~100

Significance Improvement

Let x be a discriminator variable (possibly n-dim)

so pdf's s(x) and b(x) are different

Suppose I selected on $x>x_c$

Define $Q = Z(x>x_c) / Z(\text{no cut})$

A good cut has $Q > 1$

Suppose background is well known:

$$\delta m \approx \sqrt{\langle B \rangle} \quad \text{Then } Q = \varepsilon_s / \sqrt{\varepsilon_b}$$

More stringent than $\varepsilon_s > \varepsilon_b$

I've seen HEP cuts which fail this

Event Weighting

My colleague (Andy Smith) says I should weight
m(x) (***background subtracted*** data)
with
$$\boxed{w(x) = \langle S(x) \rangle \,/\, \langle B(x) \rangle}$$
$$= s(x)/b(x) \quad \text{(within a constant)}$$

event weights defined only to within a constant
constant cancels in wtd averages and Ne

$$\bar{f} = \left(\sum f \bullet w\right) / \left(\sum w\right); \qquad N_e = \left(\sum w\right)^2 / \sum w^2$$

Cheating? Already subtracted B(x)!

But he's right!

Want estimate of M = true photons (Signal mean)

Naïve:

$$\hat{M}_1 = \sum m_i$$

$$Var(\hat{M}_1) = \sum Var(m_i) = \left(\sum V_i\right)$$

Sum: over bins of x for example; or integ. over all x

Better: if know s(x) = shape of x distribution

each bin m_i is an estimate of M

BLUE (Best Linear Unbiased Estimator)

Seek minimum variance estimator of M

Equivalently, χ^2 fit for normalization multiplier

over bins of x

BLUE treatment

Bin contents linear in parameter M:

$$\langle m_i \rangle = M s_i$$

Could have generalized with $s_i \rightarrow c_i\ s_i$

Gauss Markov: best estimator wtd by 1/variance:

$$\hat{M}_i = m_i / s_i; \quad Var(\hat{m}_i) = V_i / s_i^{\,2}; \quad w_i = 1/Var(\hat{m}_i)$$

$$\hat{M} = (\sum \hat{M}_i w_i) / \sum w_i = (\sum m_i s_i / V_i) / \sum s_i^{\,2} / V_i$$

Best = min variance among linear estimators

Using expected variance, not just estimated…

Chi-squared Treatment

Define and minimize a fit to the histogram of x:

$$\chi^2 = \sum \frac{(m_i - Ms_i)^2}{V_i}; \frac{\partial \chi^2}{\partial M} = 0 \text{ for } \hat{M}$$

$$\hat{M} = (\sum m_i s_i / V_i) / (\sum s_i^2 / V_i)$$

Bins could also be x bins over different data sets

BLUE = LLSQ

V_i = Variance of m_i (Careful: use true variance)

s(x) expected normalized signal distribution

$\Sigma s_i = 1$ (= $\int s(x)dx$) ; b(x) same for background

Then expected $m_i = M\, s_i$ and

$$\boxed{\hat{M} =} k \sum m_i \frac{s_i}{V_i} = \boxed{k \sum m_i u_i},$$

$$u_i = s_i / V_i; \quad 1/k = \sum \frac{s_i^{\,2}}{V_i}$$

Notice each m_i has a weight proportional to u_i

Can calculate M estimate just by accumulating weights!

Weight u_i

When $V_i \sim B_i$ (well-determined background)

and $B_i = B\ b_i$

$\boxed{u_i = s_i / b_i}$ in this limit

we have the advertised weight

(within a constant B, which doesn't matter)

When variance of mi and Bi estimated, use better Vi

V_i when B is uncertain

Reasonable: (assume Null Hyp for n in m=n-B; sidebands so $B = N_B/\tau$)

$V(m) \sim (n+N_B)/\tau$ (still close to B)

Better:

Calculate Z_{Bi} as in my PHYSTAT03 talk

Take $V \sim (m/Z_{Bi})^2$ (for m>0)

But: Careful: any variance small due to fluctuations should really use $m_i \rightarrow Ms_i$ (expected m_i) in calculations

(see Louis Lyons book)

Variance Improvement

$$Var(\hat{M}) = k^2 \sum Var(m_i) u_i^2 = k^2 \sum V_i u_i^2$$

$$= k^2 \sum (s_i^2 / V_i) = 1 / \sum (s_i^2 / V_i) = k$$

$$Var(\hat{M}_1) = \sum V_i \quad \text{(larger)}$$

Cf. resistors: importance-weighted $R_\parallel$ vs. R_s

weighted variance ≤ unweighted

The variances are equal if all V_i, s_i equal

With optimum weights, approach Cramer-Rao min variance bound for enough data (Gauss-Markov theorem)

Sensitivity to Assumptions

Since s and b normalized, indep. of absolute normalization assumptions.

However, sensitive to shape of s, b.

We know b accurately, fortunately:

b from data, so just use to check MC.

But s from MC: depends on

shower physics, and source energy spectrum

Test fit by χ^2 and pulls of fit of m's to s, M.

A surprising application

Consider a map of counts vs. 2-d position xy: sky map.

Solve for sources by ML: consider all candidate positions, fit to photon excess * point spread function (angular resol)

many candidate pixels, events: ML infeasible

OR: weighting *all events* by

$$w(x) = s(xy)/(b(xy) + \alpha s(xy))$$

s(xy) = point spread function

b(xy) ~ flat; so w(xy) ~ s(xy) ~ 2d Gaussian (ideal)

So $\sum w$, $\sum w^2$ at each sky position (ideogram/kernel est.)

"ugh, you smeared the map" —but it approaches ML!

Modest (10%) gain in Z over "optimal" s/√b bin size

BIG gains when 3d: {xy, z} where s(xy,z) varies with z

much more weight to events with good psf resolution!

General weighted event solution

Roger Barlow, J. Comp. Phys 72 (1987) p202

Write expected average weight in terms of parameter(s) and solve (Barlow):

$$p(x) = \alpha s(x) + (1-\alpha) b(x), \text{ so expect}$$

$$\overline{w}_d \equiv \frac{1}{N} \sum w = \alpha \overline{w}_s + (1-\alpha)\overline{w}_b; \text{ where}$$

$$\overline{w}_s = \int w(x) s(x) dx; \quad \overline{w}_b = \int w(x) b(x) dx$$

$$\text{solve for } \alpha \quad (\text{unbiased for any } w):$$

$$\hat{\alpha} = (\overline{w}_d - \overline{w}_b) / \left(\overline{w}_s - \overline{w}_b\right)$$

Why is weighting good?

Textbooks shows method of moments inefficient

ML typically has min var for parameters *a*

moments: generally above min var bound

A “moment” is just some weighting function whose data average you calculate

Then solve for the parameters *a* by equating to expected moments as f(*a*)

Typically weights not chosen optimally

$w(x) = x^k$ (classical moments)

need not be good for estimating your parameters!

Barlow Optimal Weights

Calculated above ***unbiased*** solution for parameters for general weight function w(x), and its variance

Calculus of variations: find function w(x) giving minimum variance on parameter α (actually, on M)

Finds for large number of events, w(x) solution gives *same* variance as ML (*if* w(x) is close to optimal).

But: with weighting, unlike ML,

you do *NOT* need to iterate through all events!

Shows variance less than cut on ***same*** distribution w(x)

Comment: a fit to the distribution (histogram) of w(x) is also close to optimal

Barlow's Optimal solution:

$w(x) = s(x) / (b(x) + \alpha_o\, s(x))$, $\alpha_o = M/B$

$= r(x) / (1 + \alpha\, r(x)) = 1/(\alpha + 1/\, r(x))$,

where $r(x) = s(x)/b(x)$

$w(x)\ \varepsilon\ [0,1]$; truly optimal if $\alpha_o = \alpha$

Cf. Neyman-Pearson best test variable:

$r(x) = s(x)/b(x)$

And discriminant variable

$d(x) =$ posterior prob(s|x)

$= s / (b + \alpha\, s)$, $\alpha = \pi_s/(1 - \pi_s)$

What if weights are wrong?

Barlow: Near (quadratic) optimum, parameter variance and Z estimates only slightly worse
Note; MUST guess initial value for alpha, in order to estimate α: need α_o near true α

But: wrong s or b => biased estimate of M

you are fitting normalization to wrong shape

Relationship with BLUE

Barlow: knowing B reduces variance of M

Still: using same w(x) is optimal.

Now compare with subtraction:

$w(x) = s(x)/(b(x) + \alpha s(x))$

When $\alpha \ll 1$, we recover our s/b above.

(i.e. for small α, s/b is near optimal)

F. Tkachov Optimal Weight

phyiscs/0001019=Part.Nucl.Lett.111(2002)28
physics/0604127

Elegant general principle for choosing w(x)

Again calculus of variations for minimum variance of parameter estimate

General:

$$w(x,a)_{\text{opt}} = C(a)\frac{\partial Ln[p(x;a)]}{\partial a} + D(a)$$

$$ML : \Sigma \frac{\partial Ln[p(x;a)]}{\partial a} = 0$$

$$\text{Let } p = (as+b)/(1+a)$$

$$w = s/(as+b) - 1/(1+a) \rightarrow s/(as+b)$$

Caution: He is cavalier with normalization of p(x)

Simpler ML/moments solution

Parameterize p = (as+b)/(1+a); a = (α/(1- α))

Then

$$\langle w_d \rangle = \int w(x)\,p(x)dx = \int \frac{(as+b)}{a+1}\,\frac{s\,dx}{(a_o s+b)} \approx 1/(a+1)$$

Compare ML Solution :

$$\sum w(x,a) = \frac{N}{a+1}$$

Summary

An optimal weight function can achieve ML accuracy

Weighting methods are powerful and simple

There is a rational scheme leading to choice of optimal weight functions

Weighting (or fitting to weight distributions) is more accurate than cuts