BROWN UNIVERSITY *Providence, Rhode Island • 02912*

Malcolm Hudson

~~Nice~~ Always good to come to a place where you are not known,

you make so many new friends.

o MY ~~IN~~ RESEARCH ~~A~~ INTERESTS ARE IN THE FIELD OF MULTIPARAMETER ESTIMATION, I'D LIKE TO PRESENT ONE OR TWO IDEAS FROM THAT AREA.

~~EST WE KNOW A LOT ABOUT ESTIMATION OF POISSON MEANS, WHATMORE IS LEFT TO BE SAID?~~

~~MUCH STATISTICAL THEORY IS BASED ON ASYMPTOTICS~~ ~~BOOKS ON LOG-LINEAR MODELS ARE VERY COMPREHENSIVE.~~

~~I AM GOING TO TALK TODAY WITHIN THE~~

~~I A~~

~~MUCH STATISTICAL ESTIMATION THEORY CONCERNS ITSELF WITH OBTAINING EFFICIENT~~

2 o EFFICIENT METHODS FOR ESTIMATING FIXED NUMBER PARAMETERS AS THE AMOUNT OF DATA AVAILABLE INCREASES (ASYMPTOTICS) MLEST

1 o ~~MY~~ ~~INTEREST~~ TALK ~~TODAY CONCERNS~~ ANOTHER CASE

ESTIMATION METHODS FOR ~~LARGE SPARSE DATA SETS~~ DATA

WHERE, IF ASYMPTOTICS ARE APPROPRIATE, $n, p \longrightarrow \infty$. without information building up about individual parame

~~o QUITE DIFFERENT MET~~

EXAMPLES

1. LARGE SPARSE DATA SETS

~~1~~ IN STUDIES COUNTING ~~NUMBERS~~ RATES OF UNFAVOURABLE OUTCOMES AND DETERMINING WHAT RISK FACTORS ARE RELATED WE DIVIDE STUDY SUBJECTS INTO DIFFERENT SUBGROUPS

2. IN IMAGE RECONST ~~TO~~ AS TECHNOLOGY ~~IMPROV~~ DEVELOPES, ~~EQUIPMENT~~ RECORDING EQUIPMENT TYPICALLY COLLECTS MORE PRECISE DATA ~~256~~ 128 DETECTORS NOW WHERE THERE WERE 64 LAST YEAR, BUT THE IMAGE IS REQUIRED IN GREATER RESOLUTION ~~128 x 128~~ INTRODUCING MORE PARAMETERS. 128 x 128 grid replacing 64 x 64.

o QUITE DIFFERENT METHODS ARE REQUIRED, SMOOTHING OR SHRINKAGE.

o ~~FEW MORE~~ WANT NOW TO DEVELOP A MORE SPECIFIC CONTEXT, ~~TO~~ AND ~~EXPRESS~~ FORMULATE MY PROBLEM

EXAMPLE

Numbers of study subjects

Army corps	National servicemen Non-veteran	Veteran
Infantry	5400 (86)	8300 (140)
Engineer	2600 (42)	2800 (44)
Armour/Arty	2700 (44)	2500 (40)
Minor field presence	5900 (93)	2300 (37)
Non-field	9100 (140)	3400 (55)

Response data - number of deaths 10-12 yrs since

Cross classifying factors - Vet. status, Corps, Education...

Estimate - risk of death in each "cell" (p)

Number of deaths

Army corps	Non-veteran	Veteran
Infantry	80	122
Engineer	17	45
Armour/Arty	35	34
Minor field pres.	38	23
Non field	93	36

Elements

x, the cell count, subject to Poisson variation

$\mu = Ex$, the expected count, <u>to be estimated</u>
(varies from one cell to the next)

μ_0, another expected count,

either known constant
or model based estimate } not to be trusted!

The counts in different cells statistically independent

My objectives

Develop a methodology (MAP) for such problems
Mention other applications
Develop theory for MAP estimation free of model assumptions (priors)

Two approaches

(a) Classical statistical theory – unrelated parameters

$$l(x|\mu) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad \hat{\mu} = x \qquad \text{MVUE}.$$

(b) Prior distributions on μ

<u>Ex</u> $\qquad \mu^* = \mu_0^* + \varepsilon$ known

where $\varepsilon \sim N(0, \sigma_0^2)$

STOCHASTIC MODEL or PRIOR

independent errors ε in different cells

MAP estimator will maximize, by choice of μ,

$$p(\mu^* | \text{prior}, x) \propto \underbrace{l(x|\mu^*)}_{\text{LIKELIHOOD}}\ \underbrace{p(\mu^*)}_{\text{PRIOR DENSITY}}$$

<u>Ex</u> (cont.)

$$p(\mu^*|x) \propto e^{-\mu}\mu^{x} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu^*-\mu_0^*)^2\right\}$$

Maximum when

$$\mu = x - \frac{1}{\sigma_0^2}(\mu^* - \mu_0^*) \qquad \text{Implicit}$$

known

EXAMPLE

Contingency table cell estimates

- An insurance company wishes to set fair premiums for different classes of policy holders. Data may be classified by age group, region, ... What is the risk for each subgroup.

In each cell

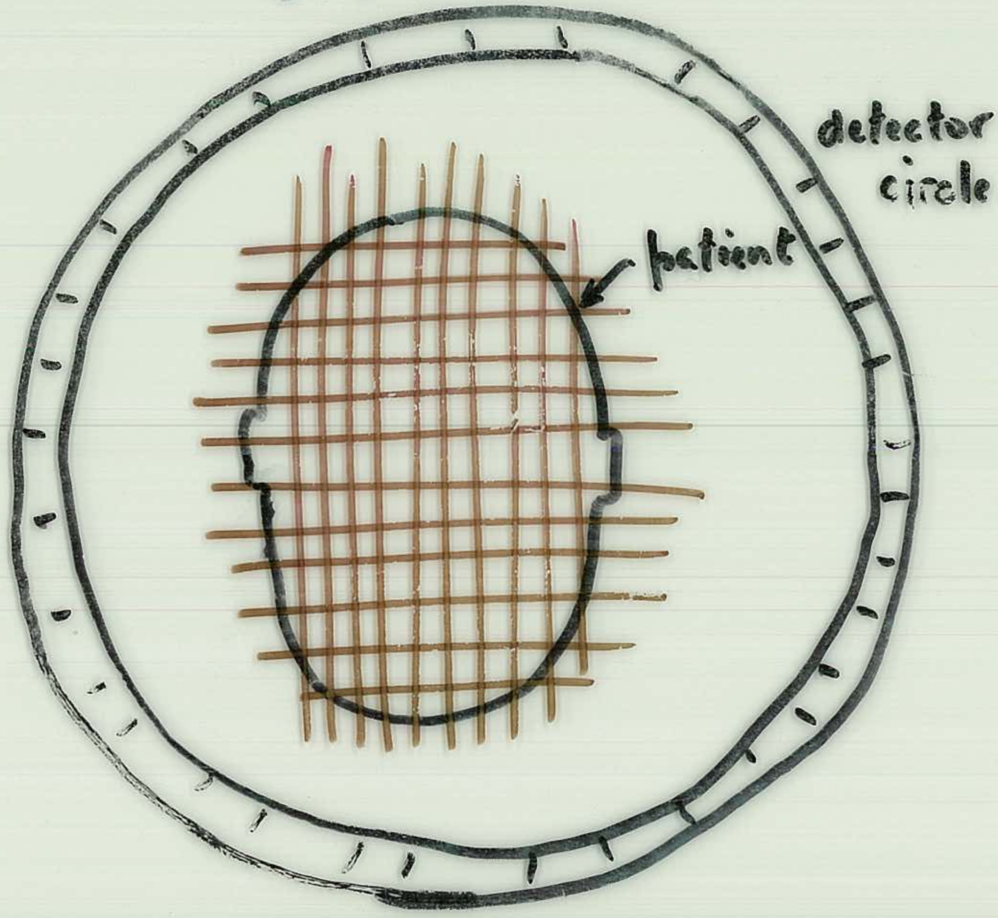
x = number of events (risk)

μ_0 = an expected number of events

(estimated by GLM?)

σ_0^2 measures expected variations in risk between cells.

Ex Emission tomography (ideal)



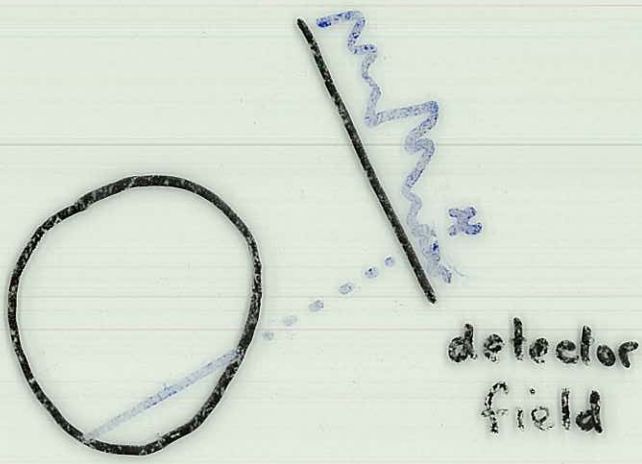
x_j = count of the photon emissions detected which originated in pixel j

$x_1, x_2, \ldots$ are indept. Poisson r.v.'s with means $\mu_1, \mu_2, \ldots$

$\{\mu_j\}$ is the image we wish to reconstruct

Model: independent variations of μ_j around a known "expected" image (e.g. a uniform background).

EXAMPLE : EMISSION TOMOGRAPHY (existing capability)



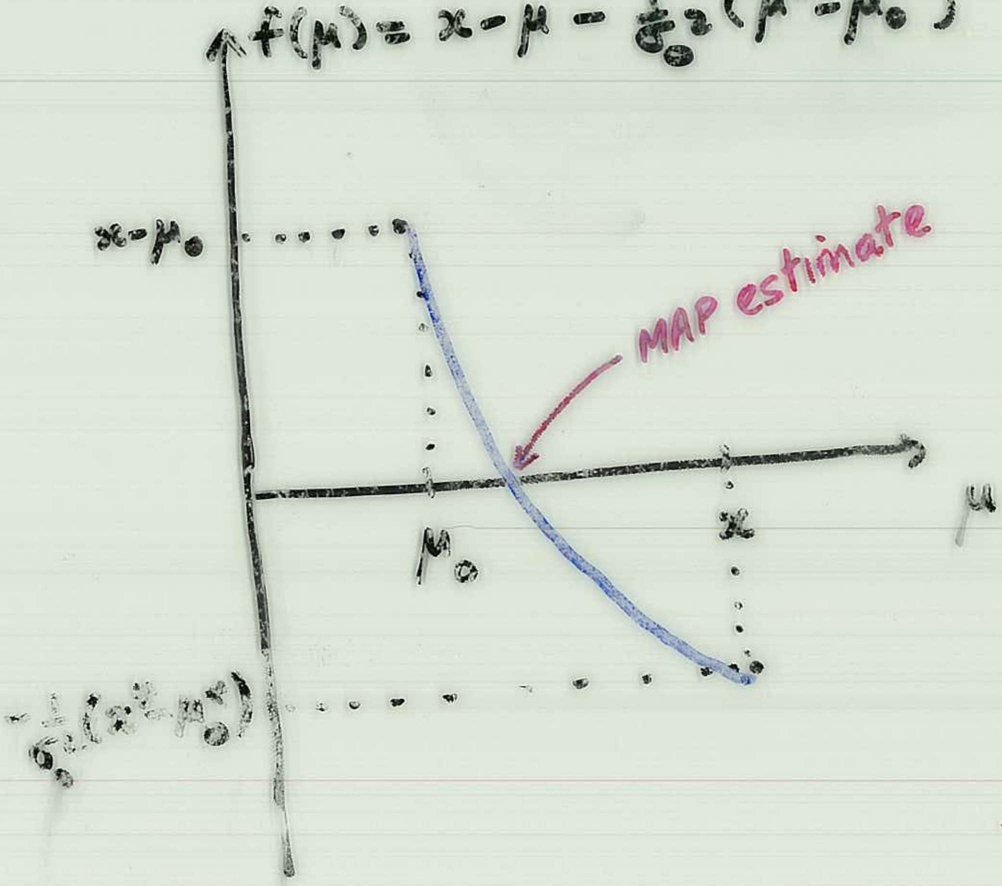
x is the count of the number of emissions directed towards one detector on the detector field.

The ~~x~~ raw projection data includes Poisson variation, causing the (M.L.) image to be too rough.

Smoothing, by MAP, before reconstructing the image (Fourier techniques) may be appropriate. Set μ_0 to be an ~~aver~~ expected count. σ_0^2 is a smoothing parameter.

Properties of MAP estimator

MAP is a compromise between the data and the prior value.



$\mu_0^* = \log \mu_0$

As $\sigma_0^2 \uparrow \infty$, MAP approaches MLE.

As $\sigma_0^2 \downarrow 0$, MAP approaches prior value μ_0.

Limitations of MAP

- Model dependent in choice of prior (intensities a random sample from a log normal distn). How do we know the process generating the intensities?

- Hyper parameters ~~...~~ (μ_0, σ_0^2) are assumed known. How are they to be chosen? What properties will MAP have if they are data determined (e.g. fitted values).

- MAP estimates require the solution, cell by cell, of an implicit equation.

Unbiased estimator of error mean square

- Fundamental identity for Poisson

$$\mu E^{\mu} f(x) = E^{\mu} x f(x-1) \qquad E|f(x)| < \infty$$

i.e. $$E^{\mu}(x-\mu) f(x) = E^{\mu} x (f(x) - f(x-1))$$

- Application 1-dimension

$$E^{\mu}(x-\mu)^2 - E^{\mu}(g(x)-\mu)^2$$

$$= E^{\mu}\{-2(x-\mu)f(x) - f^2(x)\} \qquad g(x) = x + f(x)$$

$$= E^{\mu}\{\underbrace{-2x[f(x)-f(x-1)] - f^2(x)}_{\psi(x)}\}$$

- More than 1 mean

$$X = (x_1, \dots, x_p)\,; \quad x_i \sim P(\mu_i) \text{ indept}\,; \quad \mu = (\mu_1, \dots, \mu_p)$$

$$E^{\mu}\|X-\mu\|^2 - E^{\mu}\|\underset{\sim}{g}(X) - \mu\|^2$$

$$= E^{\mu}\{\underbrace{-2X.\nabla f - \|f(X)\|^2}_{\psi(X)}\} \qquad g_i(X) = x_i + f_i(X)$$

We can assess the performance of <u>any</u> estimator g on a given set of data X, by evaluating $\psi(X)$.

Application to MAP

- consider a simplified approximation to MAP.

$$\hat{\mu} = x - \tau\,(x^* - \mu_0^*)$$

explicit calculation

τ is a constant controlling smoothing

Euler's constant .58...

$$x^* = \log\left(\frac{x+\gamma}{\gamma}\right)$$

- how should we choose τ ?

my choice $\quad \tau = \dfrac{p-2}{\Sigma(x^* - \mu_0^*)^2} = \dfrac{p-2}{S}$ where

the sum extends over all cells, and p represents the number of cells with non-zero counts.

- this estimator has smaller error mean square than MLE:

  for any configuration of means (almost)

  for any expected values μ_0

  for any number of cells, provided 3 or more.

because

$$E^{\mu}\|X-\mu\|^2 - E^{\mu}\|\hat{\mu}-\mu\|^2 \geq E^{\mu}\left\{\frac{(p-2)^2}{S}\right\}$$

Choosing expected values (μ_0) which are data determined

- Choose a simple log-linear model

$$\mu^*_{0i} = \underset{\sim}{a}_i' \underset{\sim}{\beta} \qquad \beta = (\beta_1, \cdots, \beta_q)$$

$\underset{\sim}{a}_i$ known

- Estimate β by ordinary least squares to minimize

$$S = \sum (x^*_i - \underset{\sim}{a}_i' \underset{\sim}{\beta})^2$$

i.e. $\hat{\beta} = (A'A)^{-1} A' \underset{\sim}{X}^*$ for given design A.

- Then the estimator

$$\hat{\mu} = x - \frac{(p-q-2)_+}{S} (x^* - \underset{\sim}{a}_i' \hat{\beta})$$

has error mean square reduction exceeding

$$E^\mu \left\{ \frac{(p-q-2)_+^2}{S} \right\}.$$

Conclusion

- In problems with large numbers of parameters, standard statistical theory is unhelpful. **Stein's contribution**.
- Smoothing procedures must be employed.
- The unbiased estimator of error mean square can provide estimates of smoothing parameters.
- This choice leads to estimation of cell means that are

(a) **efficient** in that they incorporate prior information

(b) **safe** the extent of use of prior information depends on how consistent it is with data.

- "Log-linear estimators" have explicit, small sample error mean square properties.

$$c = \sum_{i=1}^{D} c_{ij} \qquad \text{pixel } j.$$

$$= \sum_{i=1}^{D} 1_{\{j \text{ projects onto } i\}} = NANG.$$

Uniform grey image – bad news for patient – apparent ~~extra~~ variability in projections (random fluctuations).