Balancing Total and Individual Risk in the Simultaneous Estimation of Poisson Means

Emil Aas Stoltenberg

August 30, 2015

The problem

Have p independent Poisson observations,

$$Y_i \sim \mathcal{P}(\theta_i),\ 1 \leq i \leq p.$$

Want to estimate $\theta_1, \ldots, \theta_p$. Often we will also be interested in estimating $\gamma = \sum_{i=1}^{p} \theta_i$.

Shrinkage estimators

Usual loss functions when estimating $\theta_1, \ldots, \theta_p$ are,

$$L_m(\delta, \theta) = \sum_{i=1}^{p} \frac{1}{\theta^m} (\delta_i - \theta_i)^2,$$

with $m = 0, 1$.

Peng (1975) derived

$$\delta_i^P(Y) = Y_i - \frac{\max\{0, N_0(Y) - 2\}}{\sum_{j=1}^{p} h_j^2(Y_j)} h_i(Y_i)$$

where $N_0(Y) = \#\{i : Y_i > 0\}$ and $h_i(Y_i) = \sum_{k=1}^{Y_i} k^{-1}$ if $Y_i > 0$, zero otherwise. δ^P dominates the MLE under L_0 when $p \geq 3$.

Clevenson and Zidek (1975) showed that

$$\delta^{CZ}(Y) = \left(1 - \frac{p-1}{p-1+Z}\right) Y,$$

where $Z = \sum_{i=1}^{p} Y_i$ dominates the MLE under L_1 when $p \geq 2$.

Shrinkage estimators

Usual loss functions when estimating $\theta_1, \ldots, \theta_p$ are,

$$L_m(\delta, \theta) = \sum_{i=1}^{p} \frac{1}{\theta^m}(\delta_i - \theta_i)^2,$$

with $m = 0, 1$.
Peng (1975) derived

$$\delta_i^P(Y) = Y_i - \frac{\max\{0, N_0(Y) - 2\}}{\sum_{j=1}^{p} h_j^2(Y_j)} h_i(Y_i)$$

where $N_0(Y) = \#\{i \,:\, Y_i > 0\}$ and $h_i(Y_i) = \sum_{k=1}^{Y_i} k^{-1}$ if $Y_i > 0$, zero otherwise. δ^P dominates the MLE under L_0 when $p \geq 3$.
Clevenson and Zidek (1975) showed that

$$\delta^{CZ}(Y) = \left(1 - \frac{p-1}{p-1+Z}\right) Y,$$

where $Z = \sum_{i=1}^{p} Y_i$ dominates the MLE under L_1 when $p \geq 2$.

Possible deficiencies

Both δ^P and δ^{CZ} shrink the MLE ($= Y$) towards zero. As a consequence, they might give

- Implausible estimates of individual θ_i's.
- A bad estimate of $\gamma = \sum_{i=1}^{p} \theta_i$.

Both of these deficiencies can be adressed in the same way, that is by "limiting translation" (Efron and Morris, 1972) away from the MLE.

Consider problem of estimating γ under the loss function $(\delta - \gamma)^2/\gamma$. Here the MLE Z is the unique minimax solution (hence admissible).

$$L_c(\delta, \theta) = \sum_{i=1}^{p} \frac{1}{\theta_i}(\delta_i - \theta_i)^2 + \frac{c}{\gamma}\left(\sum_{i=1}^{p} \delta_i - \gamma\right)^2,$$

where $c \geq 0$ is a user-defined constant.

Possible deficiencies

Both δ^P and δ^{CZ} shrink the MLE ($= Y$) towards zero. As a consequence, they might give

- Implausible estimates of individual θ_i's.
- A bad estimate of $\gamma = \sum_{i=1}^p \theta_i$.

Both of these deficiencies can be adressed in the same way, that is by "limiting translation" (Efron and Morris, 1972) away from the MLE. Consider problem of estimating γ under the loss function $(\delta - \gamma)^2/\gamma$. Here the MLE Z is the unique minimax solution (hence admissible).

$$L_c(\delta, \theta) = \sum_{i=1}^{p} \frac{1}{\theta_i}(\delta_i - \theta_i)^2 + \frac{c}{\gamma}\left(\sum_{i=1}^{p} \delta_i - \gamma\right)^2,$$

where $c \geq 0$ is a user-defined constant.

Improved estimators

$$R(Y, \theta) = E_\theta L_c(Y, \theta) = p + c.$$

Seek an estimator with risk $< p + c$.

Using Stein's idea (e.g. Stein (1981)) we derive difference inequalities, solve these and obtain a class of estimators given by

$$\delta^c(Y) = \left(1 - \frac{\psi(Z)}{p - 1 + (1 + c)Z}\right) Y,$$

where $0 < \psi(z) \leq 2(p - 1)$ is non-decreasing for all $z \geq 0$. The optimal ψ in terms of minimizing risk is $\psi = p - 1$.

Improved estimators

$$R(Y, \theta) = E_\theta L_c(Y, \theta) = p + c.$$

Seek an estimator with risk $< p + c$.
Using Stein's idea (e.g. Stein (1981)) we derive difference inequalities, solve these and obtain a class of estimators given by

$$\delta^c(Y) = \left(1 - \frac{\psi(Z)}{p - 1 + (1 + c)Z}\right) Y,$$

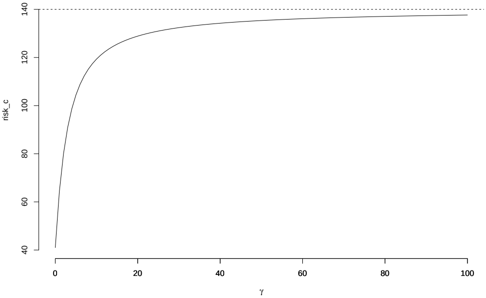
where $0 < \psi(z) \leq 2(p-1)$ is non-decreasing for all $z \geq 0$. The optimal ψ in terms of minimizing risk is $\psi = p - 1$.

δ^c dominates the MLE

Theorem: Under the loss function L_c

$$R(\delta^c, \theta) < R(\mathrm{MLE}, \theta),$$

for all $\theta \in \Theta$.



δ^c under L_1

Corollary: Under weighted squared error loss L_1,

$$R(\delta^c, \theta) < R(\mathrm{MLE}, \theta),$$

for all $\theta \in \Theta$.

Intuitively, this makes sense because for all $0 < c < \infty$,

$$\delta_i^{CZ} < \delta_i^c < Y_i, \ 1 \leq i \leq p.$$

In the sense, the new estimator is a compromise between δ^{CZ} and the MLE. What value of c gives the desired compromise?

δ^c under L_1

Corollary: Under weighted squared error loss L_1,

$$R(\delta^c, \theta) < R(\text{MLE}, \theta),$$

for all $\theta \in \Theta$.
Intuitively, this makes sense because for all $0 < c < \infty$,

$$\delta_i^{CZ} < \delta_i^c < Y_i, \;\; 1 \leq i \leq p.$$

In the sense, the new estimator is a compromise between δ^{CZ} and the MLE. What value of c gives the desired compromise?

One way to choose c

Since Z is unique minimax in estimating γ under $(\delta-\gamma)^2/\gamma$, and

$$\delta^c \to Y \text{ when } c \to \infty,$$

a big c is preferable when estimating γ with δ^c. In the composite problem of estimating $\theta_1, \ldots, \theta_p$ under

$$L_1(\delta,\theta) = \sum_{i=1}^{p} \frac{1}{\theta_i}(\delta_i - \theta_i)^2,$$

$c = 0$ is the optimal choice.

Pick the smallest $c \geq 0$ that ensures that

$$E_\gamma \frac{1}{\gamma}\left(\sum_i \delta_i^c - \gamma\right)^2 - 1 \leq K,$$

for some value K and a prior guess of γ.

One way to choose c

Since Z is unique minimax in estimating γ under $(\delta-\gamma)^2/\gamma$, and

$$\delta^c \to Y \text{ when } c \to \infty,$$

a big c is preferable when estimating γ with δ^c. In the composite problem of estimating $\theta_1, \dots, \theta_p$ under

$$L_1(\delta, \theta) = \sum_{i=1}^{p} \frac{1}{\theta_i}(\delta_i - \theta_i)^2,$$

$c = 0$ is the optimal choice.
Pick the smallest $c \geq 0$ that ensures that

$$E_\gamma \frac{1}{\gamma}\left(\sum\nolimits_i \delta_i^c - \gamma\right)^2 - 1 \leq K,$$

for some value K and a prior guess of γ.

An example

θ_i	$1/\theta_i(\delta_i^{CZ} - \theta_i)^2$	$1/\theta_i(\delta_i^{c} - \theta_i)^2$
1.48	33.57	14.25
0.68	39.55	14.36
1.84	30.45	13.44
1.24	34.99	13.70
1.46	33.96	14.71
1.38	34.21	13.96
1.64	31.72	13.27
1.46	33.59	14.25
6.98	-11.54	7.15
6.91	-11.25	6.70
L_1	24.93	12.58

Bayesian analysis

Recall the class of estimators

$$\delta^c(Y) = \left(1 - \frac{\psi(Z)}{p - 1 + (1+c)Z}\right) Y,$$

where the estimator with $\psi = p - 1$ is the optimal in terms of minimizing risk. Gain insight into the nature of these estimators by deriving them by Bayesian methods in three different ways.

- Pure: θ_i iid Gamma$(1, b)$, $b \sim \pi(b) \propto b^{\alpha-1}(b+1)^{-(\alpha+\beta)}$.
- Empirical: Estimate b from data, find unbiased estimator.
- Generalized: $\theta_i = \alpha_i \lambda$ with $(\alpha_1, \ldots, \alpha_p) \sim$ Dirichlet$(a_1, \ldots, a_p)$, and λ flat on $[0, \infty)$.

Admissibility

Pure Bayes setup again, θ_i iid Gamma$(1, b)$. Let,

$$b \sim \pi(b) \propto b^{m-2}(b+1)^{-m}.$$

then $\pi(b)$ is proper for $m > 1$, so

$$\delta^c(Y) = \left(1 - \frac{m+p-1}{m+p-1+(1+c)Z}\right) Y,$$

is admissible.
Would like to prove that the optimal estimator in terms of minimizing risk ($m = 0$) is admissible.

Bibliography

Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous Estimation of the Means of Independent Poisson Laws. *Journal of the American Statistical Association*, 70:698–705.

Efron, B. and Morris, C. (1972). Limiting the Risk of Bayes and Empirical Bayes Estimators–Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*, 67:130–139.

Ghosh, M., Hwang, J. T., and Tsui, K.-W. (1983). Construction of Improved Estimators in Multiparameter Estimation for Discrete Exponential Families. *The Annals of Statistics*, 11:351–367.

Peng, J. C.-M. (1975). Simultaneous Estimation of the Parameters of Independent Poisson Disitributions. Technical report no. 48, Stanford University, Departement of Statistics.

Stein, C. (1981). Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9:1135–1151.