1 Setup.

Independently
\[ y_i \sim N[\mu_i, V_i = \sigma^2/n_i], \quad i = 1, \ldots, k. \]

For weighted squared error loss, the usual estimator \( \mathbf{y} \), e.g. the sample mean vector, has constant and minimax risk:

\[ E \sum_{i=1}^{k} W_i \ast (y_i - \mu_i)^2 = \sum_{i=1}^{k} W_i \ast V_i. \]
2 Equal variances, James-Stein shrinkages.

Stein’s 1962-3 decision theory course proved for \( k \geq 3 \) and equal variances that

\[
\hat{\mu}_{JS} \equiv (1 - (k - 2) * V/ \|y\|^2)
\]

has smaller MSE than \( y \) for every mean vector \( \mu \). Amazement. How would anyone think of that?

"Shrinkage" of \( y_i \) could be toward any \( \mu_{0,i} \) (and to estimates of such)

\[
\hat{\mu}_{JS,i} = (1 - \hat{B}_{JS}) * y_i + \hat{B}_{JS} * \mu_{0,i}
\]

\[
\hat{B}_{JS} \equiv (k - 2)/S, \quad S \equiv \sum_{i=1}^{k} (y_i - \mu_{0,i})^2/V
\]

\[
\hat{R} = V \left( k - (k - 2)\hat{B}_{JS} \right)
\]
3 Stein’s main shrinkage papers; 1st time topics.

Also 1960 (Regression, Stanford), 1985 (confidence set approximations).

1. 1954-6 Inadmissibility – Berk Symposium

   Inadmiss, Admiss \( k \leq 2 \), Non-Normal, component risk

2. 1961 James-Stein – Berk Symposium

   JS rule, risk calcs, \( \hat{R} = V \left( k - (k - 2)\hat{B}_{JS} \right) \) is minimax
   
   \( V \) estimated, bounded loss, Pitman’s estimator
   
   Willard James

3. 1962 Confidence Sets – JRSS

   Confidence ellipsoids, Stein’s harmonic prior (SHP), admissibility.

4. 1966 BIB Designs – Neyman Festschrift

   Applications, subspace shrinkage risks, Bayesian heuristic

5. 1973,81 SURE Stein’s Unbiased Risk Est. – Ann Math Stat

   Unequal variances, risk estimates,
   
   \( \hat{R}_{SHP} = SURE \) (SHP risk) < Bayes risk < \( k \ast V \), i.e. \( \hat{R}_{SHP} < R^*_{SHP} < k \ast V \)
4 Two-level model-II

Level-II model:
\[ \mu_i \sim N(\mu_{0,i}, A) \]

Given \( y \) and \( A \) (unknown) and so shrinkage constants
\[ B_i \equiv V_i/(V_i + A) \]

\[ \mu_i \sim N( (1 - B_i) * y_i + B_i * \mu_{0,i}, V_i * (1 - B_i) ) \]

With equal variances \( V_i \) and \( \mu_{0,i} = 0 \), this simplifies to
\[ y_i \sim N[ \mu_i, V_i + A ] \], \( i = 1, \ldots, k. \)

and gives a model-II likelihood for \( B \leq 1 \) that is truncated Chi-square.
\[ L(B) = B^{k/2} \exp(-BS/2). \]

Then, given the data and \( A \)
\[ \mu_i \sim N[ (1 - B) * y_i, V * (1 - B) ], \; i = 1, \ldots, k. \]

Some priors on \( A > 0 \):
1. \( A \sim Unif(-V, \infty) \) Gives James-Stein estimator
2. \( A \sim Unif(0, \infty) \) Stein’s harmonic prior
3. \( B \sim Beta(k0 - 2, 1) \) Conjugate to Likelihood, Strawderman
4. \( A \sim A(c/2) * dA/A \) Scale invariant \( c = k + k_0 - u > 0 \)
5. \( A \sim dA/A \) Scale invariant
5 1975 Graph, Formal priors on $\mu$ and $A$

Figure 1: Proper and formal priors that produce minimax shrinkage estimators.

1. Minimax = shaded: $0 < u < k - 2$
2. Scale invariant: $A^{c/2}dA/A$, $k_0 = 0$
3. 45 degree line: Conjugate; Strawderman
4. Stein, SHP: Intersection, admissible boundary
6 Post-Stein, More Models

Level II and Level III models:

1. Unequal variances, as $V_i$ always differ in applications
2. minimaxity, component-wise. Then weights $W_i$ don’t matter.
3. admissibility, component-wise,
4. interval coverages exceed 95%, every component, all $A \geq 0$
5. Level-III priors on $A$. SHP seems central
6. Extend to other models: GLMs, NEFs. SHP is transportable.
7. MCMC construction vs. decision theoretic verification.
8. Shrinkage factors needed: diagnostics