Lab: Shrinkage Estimation

1. Predict second-half-of-season first putt percentage from first-half-of-szn first putt percentage

index all putters in our dataset by $i = 1, \ldots, K$

training dataset
$$\begin{cases} X_i = \text{first-half-of-szn first putt success percentage} \\ N_i = \text{num first putts in first-half-of-szn} \end{cases}$$

held-out test dataset
$$\left\{ X_i' = \text{second-half-of-szn first putt success percentage} \right.$$

do NOT use this dataset except in evaluating the various predictions at the very end. do not cheat by using X_i' to make your predictions, no regression!

Your task is to predict $\{X_i'\}_{i=1}^K$ from just $\{(X_i, N_i)\}_{i=1}^K$ without looking at the X_i'.

You will evaluate a prediction $\{\hat{P}_i\}_{i=1}^K$ by its loss $L(\hat{P}_i, X_i') = \sum_{i=1}^K (\hat{P}_i - X_i')^2$, aggregate squared error.

First, consider the MLE $\hat{p}_i^{(MLE)} = X_i$

and the overall mean $\hat{p}_i^{(mean)} = \frac{1}{K}\sum_{i=1}^{K} X_i$.

Which is a better predictor? Thoughts?

As we learned in the empirical Bayes and Shrinkage lectures, a better predictor will shrink the first half putt percentage to the overall mean. Your task is to fit an Efron Morris estimator and an Empirical Bayes estimator to do so. We'll need to do some more steps (below) before forming these estimators. You'll fit 2 versions each of the EM and EB estimators.

* You'll end up fitting 6 estimators $\{\hat{p}_i\}_{i=1}^{K}$: MLE, overall mean, EM v1 & v2, EB v1 & v2. Evaluate the loss $L(\hat{p}, X')$ for each. Also, make a scatterplot of true X' (y axis) vs. estimator $\hat{p}$ (x axis) for each of the 6 estimators (color) on one plot; sort the points in order. Add line $y=x$.

* Recall from the Empirical Bayes lecture our original model $X_i \sim \mathcal{N}(p_i, \sigma_i^2)$ where $\sigma_i^2 = \frac{p_i(1-p_i)}{N_i}$ and p_i is the latent (unobserved) "true" putter quality of player i. Our estimators work when the variance is known, which is not the case here, so we need to do some more work. From here, we'll make 2 versions of the EM and EB estimators.

Efron Morris Estimator, version 1

To simplify, let $\sigma_i^2 = \frac{c}{N_i}$, treating c as a known constant instead of using the unknown $p_i(1-p_i)$. Specifically, try $c = \bar{x}(1-\bar{x})$.

Then transform $\tilde{X}_i \leftarrow \frac{X_i}{\sigma_i} = \frac{X_i}{\sqrt{c/N_i}}$ so that $\tilde{X}_i \sim \mathcal{N}(\theta_i, 1)$ where $\theta_i = \frac{p_i}{\sqrt{c/N_i}}$, which matches the assumptions for the EM estimator.

From here, fit $\hat{p}_i^{(EM)} = \sqrt{\frac{c}{N_i}}\,\hat{\theta}_i^{(EM)}$ where

$$\hat{\theta}_i^{(EM)} = \bar{\tilde{x}} + \left(1 - \frac{k-1}{S^2}\right)(\tilde{x}_i - \bar{\tilde{x}}) \quad \text{and} \quad S^2 = \sum_{i=1}^{k} (\tilde{x}_i - \bar{\tilde{x}})^2.$$

Efron Morris Estimator, version 2

Instead of using the previous simplifying assumption, we'll use a variance stabilizing transform to make the variance become known.

Recalling the model $X \sim N(p, \sigma^2)$ where $\sigma^2 = \frac{p(1-p)}{N}$, ignoring the subscript i, we want to find a transformation T so that T(X) has constant variance.

we write a 1^{st} order Taylor approximation at X centered at p,

$$T(X) \approx T(p) + T'(p) \cdot (X-p)$$

$$\rightarrow \operatorname{var}(T(X)) = \operatorname{var}(X) \cdot (T'(p))^2 = (T'(p))^2 \cdot \underbrace{V(p)}_{= \frac{p(1-p)}{N}} \overset{\text{want}}{=} C$$

$$\rightarrow \text{Solve } T'(p) = \sqrt{\frac{C}{V(p)}}$$

$$\rightarrow T(p) = \int_0^p \frac{\sqrt{C}}{\sqrt{V(t)}}\, dt = \sqrt{NC} \int_0^p \frac{dt}{\sqrt{t(1-t)}}$$

$$= \sqrt{NC} \int_0^{\arcsin\sqrt{p}} \frac{2\sin\theta\cos\theta \, d\theta}{\sqrt{\sin^2\theta\,(1-\sin^2\theta)}}$$

letting $t = \sin^2\theta$

$$= 2\sqrt{Nc} \cdot \int_0^{\arcsin\sqrt{p}} d\theta = \underbrace{2\sqrt{Nc}}_{\text{Constant}} \arcsin\sqrt{p}$$

$\Rightarrow$ $T(p) = \arcsin\sqrt{p}$ is a variance stabilizing transformation,

so $\tilde{X}_i = \arcsin\sqrt{X_i} = \arcsin\sqrt{\frac{H_i}{N_i}}$ works.

In practice, we use a slightly different transformation, which yields a known variance as follows:

Let $\tilde{\tilde{X}}_i = \arcsin\sqrt{\frac{H_i + 1/2}{N_i + 1/4}}$

$\Rightarrow \tilde{\tilde{X}}_i \sim \mathcal{N}(\tilde{\tilde{\theta}}_i, \nu_i^2)$ where $\begin{cases} \tilde{\tilde{\theta}}_i = \arcsin\sqrt{p_i} & \text{unknown} \\ \nu_i^2 = \frac{1}{4N_i} & \text{known} \end{cases}$

Then transform $\tilde{X}_i \leftarrow \frac{\tilde{\tilde{X}}_i}{\nu_i}$ so that

$\tilde{X}_i \sim \mathcal{N}(\theta_i, 1)$ where $\theta_i = \frac{\tilde{\tilde{\theta}}_i}{\nu_i}$,

which matches the assumptions for the EM estimator.

From here, fit $\hat{p}_i^{(EM)}$ via $\hat{\tilde{\tilde{\theta}}}_i^{(EM)}$ via $\hat{\theta}_i^{(EM)}$.

Empirical Bayes, version 1

Suppose $X_i \sim \mathcal{N}\left(p_i, \frac{c}{N_i}\right)$ where $c = \bar{x}(1-\bar{x})$.

A Bayesian adds the prior $p_i \sim \mathcal{N}(\mu, \tau^2)$.

The Bayesian estimate of p_i is the posterior mean

$$\hat{p}_i^{(\text{Bayes})} = \mathbb{E}[p_i \mid X_i] = \mu + \left(\frac{\tau^2}{\tau^2 + \frac{c}{N_i}}\right)(X_i - \mu).$$

μ and τ^2 are unknown, so we need to estimate them to form the Empirical Bayes estimator $\hat{p}_i^{(EB)} = \hat{\mu} + \left(\frac{\hat{\tau}^2}{\hat{\tau}^2 + \frac{c}{N_i}}\right)(X_i - \hat{\mu})$.

We could find the MLE $\hat{\mu}_{MLE}$, $\hat{\tau}^2_{MLE}$ using the iterative convergence algorithm discussed in lecture, but that feels a bit overkill. Instead, consider the marginal distribution of X_i:

$$\left.\begin{aligned} X_i &\sim N\left(p_i, \frac{c}{N_i}\right) \\ p_i &\sim N(\mu, \tau^2) \end{aligned}\right\} \Longrightarrow X_i \sim N\left(\mu, \tau^2 + \frac{c}{N_i}\right)$$

Thus $\hat{\mu} = \bar{X}$ is an unbiased estimator.

Also, $\quad \mathrm{var}(X_i) = \tau^2 + \dfrac{c}{N_i}$,

so $\quad \tau^2 = \mathrm{var}(X_i) - \dfrac{c}{N_i}$

so a reasonable estimator of τ^2 is

$\hat{\tau}^2 = S_X^2 - c \cdot \mathrm{mean}\left(\left\{\frac{1}{N_i}\right\}\right)$ where $S_X^2 = \dfrac{1}{K-1}\sum_{i=1}^{K}(X_i - \bar{X})^2$

is the sample variance of $\{X_i\}_{i=1}^{K}$.

Empirical Bayes, version 2

Let $\tilde{\tilde{X}}_i = \arcsin\sqrt{\dfrac{H_i + 1/2}{N_i + 1/4}}$

$\rightarrow \tilde{\tilde{X}}_i \sim N(\tilde{\tilde{\theta}}_i, \nu_i^2)$ where $\begin{cases} \tilde{\tilde{\theta}}_i = \arcsin\sqrt{p_i} & \text{unknown} \\ \nu_i^2 = \dfrac{1}{4N_i} & \text{known} \end{cases}$

PRIOR $\tilde{\tilde{\theta}}_i \sim N(\mu, \tau^2)$.

Go from here.