Lab: Empirical Bayes

1. Rolling Player Quality Estimates for NBA Players

Empirical Bayes is one of the best ways to estimate player quality, especially pre-game rolling player quality which just uses data from previous games.

Today we'll consider an example: estimate an NBA player's "true" scoring talent or skill.

We observe box score data of the form

$$
\begin{cases}
X_{ij} = \text{pts scored by player } i \text{ in game } j \\
N_{ij} = \text{num possessions of player } i \text{ in game } j \\
N_i = \text{num games of player } i, \quad j = 1, \ldots, N_i \\
N = \text{num players}, \quad i = 1, \ldots, N \\
\text{data } \mathcal{D} = \{(X_{ij}, N_{ij}) : 1 \le i \le N,\ 1 \le j \le N_i\}.
\end{cases}
$$

It's not necessarily enough to model the points scored X_{ij} as a draw from player i's "true" scoring quality μ_i because his quality may change over time, for instance due to injury or aging or simply getting better/worse by practice or whatever reason (non-stationarity).

It would be great to capture player i's quality μ_{ij} in game j as j progresses across his career.

To do so, we form a dynamic Bayesian model, which in some form was seen in Mark Glickman's paper from the 1990s,

$$\begin{cases} X_{ij} \mid \mu_{ij} \sim N(N_{ij}\mu_{ij},\ N_{ij}\sigma^2) \\ \mu_{ij} \mid \mu_{i(j-1)} \sim N(\mu_{i(j-1)},\ \tau^2) & \text{if } j>1 \\ \mu_{i1} \mid \mu \sim N(\mu,\ \nu^2) \end{cases}$$

Think about what the parameters represent before looking at the answer (below).

μ_{ij} = latent (unobserved) "true" scoring talent of player i in game j

μ = overall mean player talent

σ^2 = variance in points scored on a possession given a player's true scoring talent

τ^2 = game-to-game variance in a player's scoring quality

ν^2 = variance in player talent before seeing any data (prior variance)

Use empirical Bayes to estimate μ_{ij}, player i's latent scoring talent during game j.

The Bayesian estimate of μ_{ij} is the posterior mean,

$$\hat{\mu}_{ij} = \mathbb{E}\left[\mu_{ij} \,\middle|\, \{X_{ik}\}_{k<j}, \{N_{ik}\}_{j<k}, \mu, \sigma^2, \tau^2, \nu^2\right].$$

Computing this expectation is hard.

To begin, estimate μ_{i1}.

The relevant part of the model is

$$\begin{cases} X_{i1} \mid \mu_{i1} \sim \mathcal{N}(N_{i1}\mu_{i1}, N_{i1}\sigma^2) \\ \mu_{i1} \mid \mu \sim \mathcal{N}(\mu, \nu^2). \end{cases}$$

Write a formula for the posterior mean of this normal-normal model $\hat{\mu}_{i1} = \mathbb{E}(\mu_{i1} \mid X_{i1}, N_{i1}, \mu, \sigma^2, \tau^2, \nu^2)$;

we derived the posterior mean of a normal-normal model in class today, you can use that formula but the parameter names may be different.

Next, estimate μ_{ij} given $\mu_{i(j-1)}$ for $j>1$.

The relevant part of the model is
$$\begin{cases} X_{ij} \mid \mu_{ij} \sim \mathcal{N}(N_{ij}\mu_{ij},\ N_{ij}\sigma^2) \\ \mu_{ij} \mid \mu_{i(j-1)} \sim \mathcal{N}(\mu_{i(j-1)}, \tau^2) \end{cases}$$

Again, write a formula for the posterior mean
$$\hat{\mu}_{ij} = \mathbb{E}[\mu_{ij} \mid \mu_{i(j-1)}, \tau^2, \sigma^2, X_{ij}, N_{ij}].$$

The formulas you obtained for $\{\mu_{ij}\}_{j=1}^{N_i}$ are functions of observed data $\{X_{ij}\}$ and $\{N_{ij}\}$ and unobserved hyperparameters $\mu, \sigma^2, \tau^2, \nu^2$.

The Empirical Bayes approach is to estimate the parameters and then plug in the estimates $\hat{\mu}, \hat{\sigma}^2, \hat{\tau}^2, \hat{\nu}^2$ into the formula for $\hat{\mu}_{ij}$.

So, we must estimate these hyperparameters from the data. It is too difficult to compute the MLE of these params because it is too difficult to even write down the full likelihood of the model. Instead, we'll employ some tricks.

Either use the following trick or figure out a better way.

TRick:

- treat ν^2 as a tuning parameter
- Supposing we know ν^2, estimate τ^2 and σ^2 from the data as follows:

The marginal distribution of $Z_{ij} := \frac{X_{ij}}{N_{ij}}$ is

$Z_{ij} \sim N(\mu, W_{ij})$ where $W_{ij} := \frac{\sigma^2}{N_{ij}} + \nu^2 + \tau^2(j-1)$.

Notice $\{Z_{i_1 j}\}$ is independent of $\{Z_{i_2 j}\}$ $\forall i_1 \neq i_2$.

$V_j := \mathrm{VaR}(Z_{1j}, \dots, Z_{Nj})$ variance across all players' j^{th} games.

By lemma (below), given ν^2, 2 eqns 2 unknowns, solve for (σ^2, τ^2):

1. $\overline{V_j - \nu^2} = \sigma^2 \overline{N_j} + \tau^2 \overline{(j-1)}$ where $\overline{N_j} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{N_{ij}}$
2. $\overline{V_{j+1} - V_j} = \sigma^2 (\overline{N_{j+1}} - \overline{N_j}) + \tau^2$

Lemma If $Y_j \overset{ind}{\sim} N(\mu, \sigma_j^2)$, then $\mathbb{E}\left[\frac{1}{n-1} \sum_{j=1}^{n} (y_j - \bar{y})^2\right] = \bar{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \sigma_j^2$.

Write a function where the argument is ν^2 that estimates τ^2 and σ^2 using the aforementioned method. Then, write a function that

Computes $\{\hat{\mu}_{ij} : 1 \le i \le N, 1 \le j \le N_i\}$ given ν^2.

First, tune ν^2 by splitting the data into test/train and minimize out-of-sample MSE.

Then, make a $\boxed{\text{spaghetti plot}}$:

plot estimated scoring talent $\hat{\mu}_{ij}$ (y axis) vs. time (x axis) for some players of your choosing (color).

Play around with different values of ν^2.

2. A simpler version of dynamic Empirical Bayes illustrated via Kicker Quality

The Empirical Bayes model from the previous question wasn't even that complex to write down, but it was an absolute pain to solve for the posterior mean and especially the hyperparameters.

Sometimes, such a formal model is overkill and we can estimate a player quality trajectory of similar quality but in a much easier way. We'll illustrate this by estimating kicker quality trajectories for NFL kickers.

We'll define kicker quality by a weighted sum of his field goal probability added over all his previous kicks in his career.

To begin, fit a field goal probability model $\mathbb{P}_{FG}^{(0)}$ as a function of just yard line.

Use logistic regression and a spline.
Ignore the selection bias...
Define the field goal probability added of the j^{th} field goal by

$$FGPA_j = \mathbb{1}\{n^{th} \text{ field goal is made}\} - \mathbb{P}^{(0)}_{FG}\left(\text{yardline of } n^{th} \text{ field goal}\right),$$

Note $\mathbb{1}\{x\} = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{else} \end{cases}$ is the indicator function.

Now we define kicker quality using Raven's kicker Justin Tucker for concreteness. Index all his field goals in order by j. We define his quality prior to field goal j by

$$(*) \quad Kq_j := \alpha \cdot Kq_{j-1} + FGPA_{j-1},$$

$Kq_0 = 0$ and $FGPA_0 = 0$.

Find a formula for Kq_j purely in terms of α and $\{FGPA_\ell\}_{\ell < j}$. You'll notice that α plays a similar role as τ^2 in the Empirical Bayes model of the previous section.

$\alpha < 1$ is an exponential decay weight that upweighs more recent field goals, thus accounting for non-stationarity. For instance, $\alpha = 0.995$ weighs the 138^{th} field goal in the past half as much as the previous field goal, $\alpha^{138} = 0.50$.

Our estimator of kicker quality is equivalent to that of an Empirical Bayes model! We just never wrote out the model. What's the prior?

Write a function that, given α, estimates each player i's kicker quality prior to each kick j, $\{kq_{ij} : 1 \le i \le N, \ 1 \le j \le N_i\}$. Use a for loop with formula (*), not the closed form formula, to make it fast.

Then, tune α by selecting the α that minimizes out-of-sample logloss when you make a field goal probability model $\mathbb{P}_{FG}(kq, ydl)$.

Finally, make a $\boxed{\text{spaghetti plot}}$:

plot kicker quality KQ_{ij} (y axis) vs. time (x axis) for some kickers of your choosing (color).