Outline of Methods for the 3d DLNM

John Nicklas: April 2, 2024

Problem – including a third day dimension makes the state space is very large, dramatically increasing the complexity of the glm (“binomial” – logit link function) fitting procedure. The basis functions of this large state space (patient \* exposure level \* lag \* post-disc-day) are no longer fit in a way that makes sense using the standard glm procedure.

Workaround:

1. Initial estimate of risk not dependent on temperature
   1. Use glm (“binomial” ) to fit the expected risk of individual patients based on individual-level characteristics: health information, surgical procedure risks, mean pollution exposure, ~~day of the week of discharge, season of discharge~~.
   2. Fit additional risk functions for the day # post-discharge and ~~day of the week~~ on this ~~7d \*~~ 30d space. These events are rare, can use log risk
   3. Generate predictions (low + high) for each patient \* post-disc-day of risk
2. Compute “counts / voxel” density histogram of all observed data within the triangular prism space. (#/lag\_dag / post-disc-day / 1°F). Integrating this function along the temperature axis (step size 1°F) returns the total # of patients.

$total numbers: ~4000pts \* 450 history days per patient

Into a space of 45,000 “bins”

* 1. Linear fit (LM) the DLNM 3d-basis functions to these densities, generating a smooth interpolated density – reduces to 300 dof

1. Repeat for “events / voxel” density histogram of all observed data within the triangular prism space. (#/lag\_dag / post-disc-day / 1°F). Integrating this function along all axes (step size 1°F in temperature) returns the total # of observed readmission + morality events.

$total numbers: ~100pts \* 450 history days per patient

Into a space of 45,000 “bins”

* 1. Linear fit (LM) the DLNM 3d-basis functions to these densities, generating a smooth interpolated density - reduces to 300 dof “SVD inverse – wait to be assigned until later”
  2. Compute the interpolated non-events / voxel
  3. Perform Bayesian binomial regression, integrating over the interpolated # of events per each ?larger? voxel, say ~5000 total (like running many global significance tests). This generates a Beta pdf curve automatically by performing a local average smoothing operation at the same scale as the larger voxel size.

1. Repeat for “expected events / voxel” density histogram based on step 1, finding the mean expected number of events by summing each patient’s expectation and dividing by the total count in part 2.
   1. Linear fit (LM) to find smooth interpolated expected density
   2. Repeat for low range of expected events
   3. Repeat for high range of expected events
   4. Fit (at each voxel) a log-normal curve describing the assumed (whole-population) probability of that event happening
2. Find the difference between these two probability distributions (in regular p-space, not log-space). where

X~lognormal (tight variance assumption from knowing only stuff about patient step 1 and fitting which patients landed on that day/lag/exposure in step 4)

Y~beta(Bayesian binomial from integrated events step 2 and counts step 3)

* 1. Difference in ?medians? Not sure if this is the same as difference in expected values of the distributions
  2. Most positive difference: 97.5% CI
  3. Most negative difference: 2.5% CI
  4. Re-transform to log-space

1. Re-calculate additional risk for each patient from their exposure history.
   1. Update the individual-level risks in step 1 to reflect this new source of explanation.
   2. Repeat all steps until convergence? (What if it doesn’t converge?)

Outcome Variable Name: 'Composite\_Readmit\_Mort'

Daily Temp Records are in separate columns named 'Temp\_Post-Disch\_Day\_0' (to 'Day\_29')

That's probably all you need for testing the method, but if you want to add in the other variables as linear predictors, every other variable in the dataframe other than 'static\_pt\_idx' should be ready to go. My tentative plan was to run a separate multivariate logistic regression and only keep those that are borderline significant (p-val < 0.1) as well as the mean pollution exposures during the post-discharge periods (since, regardless of significance, these seem to be correlated with the temperature).

glm(formula = modelspec, family = binomial(), data = final\_df,

weights = final\_df$weight\_numpd, na.action = na.exclude)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.2186380 0.4818333 0.454 0.650000

Disch\_Home1 -0.4003026 0.1184509 -3.379 0.000726 \*\*\*

RF.Last.Hematocrit -0.0501818 0.0095873 -5.234 1.66e-07 \*\*\*

Post.Op.Other.Pericardiocentesis1 1.8985849 1.2691969 1.496 0.134681

Post.Op.Surgical.Site.Infection1 2.1105045 0.4528766 4.660 3.16e-06 \*\*\*

NO2\_Post\_Means -0.0135757 0.0057205 -2.373 0.017636 \*

Total.ICU.Hours 0.0006237 0.0003852 1.619 0.105433

Post.Op.Pulm.Pneumonia\_Modified1 -0.4363133 0.3771513 -1.157 0.247327

BMI 0.0144708 0.0089416 1.618 0.105581

curDOW1 0.4639709 0.2159836 2.148 0.031700 \*

curDOW2 0.0861040 0.2309580 0.373 0.709288

curDOW3 0.2052612 0.2241649 0.916 0.359840

curDOW4 0.2345779 0.2231460 1.051 0.293153

curDOW5 0.6693038 0.2070430 3.233 0.001226 \*\*

curDOW6 0.2852164 0.2213292 1.289 0.197519

curFupDb0 -3.0856437 0.5922543 -5.210 1.89e-07 \*\*\*

curFupDb2 -0.0594475 0.0070715 -8.407 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2588.8 on 113526 degrees of freedom

Residual deviance: 2358.1 on 113510 degrees of freedom

(287 observations deleted due to missingness)

AIC: 1677.1

Number of Fisher Scoring iterations: 7