Gaussian Process Models for Mortality Improvement Factors

Longevity 12

Jimmy Risk

Dept of Statistics & Applied Probability UC Santa Barbara

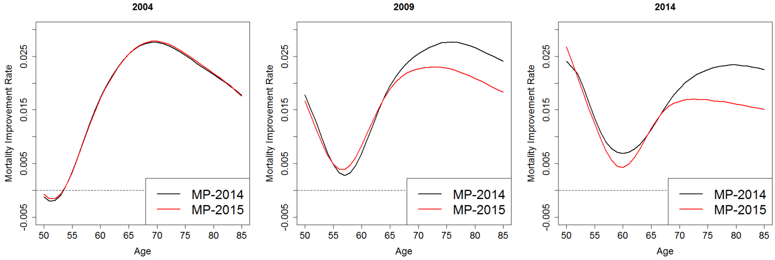
September 29 2016

Joint with M. Ludkovski (UCSB) and H. Zail (Elucidor)

US Mortality Improvement

MP-2014 / MP-2015

- Published by SOA
- US CDC Data
- MP-2014 uses years 1950-2009
- Plans to update scales at least triennially; two years of additional CDC data shows drastic change in later years
  - MP-2015 emerges



Goal:

Model US Mortality data using Gaussian Process (GP) regression

- Bayesian
- Provides posterior Gaussian distribution for input of any age and year
- Offers easy analysis of both mortality and mortality improvement simultaneously
- Gaussian distribution implies one-year mortality improvement factors remains Gaussian
- Differentiable: can provide instantaneous mortality improvement (still Gaussian)
- Spatial approach inherently handles missing and edge data

Typical Regression Assumption

Hypothesis:

$$y = f(x) + \varepsilon$$

- Observe $\boldsymbol{y} = y^{1:N}$ for input locations $\boldsymbol{x} = x^{1:N}$
- Want to understand the function f
  - e.g. $f(x) = \beta_0 + \beta_1 x$ (simple linear regression)
- ε is noise: can't observe $f(x)$ directly
- Assume $\varepsilon \sim \mathcal{N}\left(0, \sigma^2(x)\right)$ (often $\sigma(x) \equiv \sigma \in \mathbb{R}^+$)
- Our assumption: f is a Gaussian Process (modeling log-mortality)

Gaussian Process

- Defined as a set of random variables $\{f(x) | x \in \mathbb{R}^d\}$
- Any finite subset has a multivariate Gaussian distribution with covariance $C(\cdot, \cdot)$:

$$f(x_1), \ldots, f(x_n) \sim \mathcal{N}\left((m(x_1), \ldots, m(x_n)), C(\boldsymbol{x}, \boldsymbol{x}^T)\right).$$

- Fix mean function m and covariance kernel C; this provides a prior distribution

Posterior

- Observe pairs $(\boldsymbol{y}, \boldsymbol{x}) = ((y, x)^{1:N})$
  - (e.g. $y =$ historic log–mortality and $x =$ (age, year))
- Gaussian assumptions imply that marginally for any input x

$$f(x)|(\boldsymbol{y}, \boldsymbol{x}) \sim \mathcal{N}\left(m_*(x), s_*^2(x)\right)$$

Posterior

- Observe pairs $(\boldsymbol{y}, \boldsymbol{x}) = ((y, x)^{1:N})$
  - (e.g. y = historic log–mortality and x = (age, year))
- Gaussian assumptions imply that marginally for any input x

$$f(x)|(\boldsymbol{y}, \boldsymbol{x}) \sim \mathcal{N}\left(m_*(x), s_*^2(x)\right)$$

- m_* and s_*^2 are the posterior mean and variance functions

$$\begin{cases} m_*(x) \doteq \boldsymbol{c}(x)^T(\boldsymbol{C} + \Sigma)^{-1}\boldsymbol{y}; \\ s_*^2(x) \doteq C(x, x) - \boldsymbol{c}(x)^T(\boldsymbol{C} + \Sigma)^{-1}\boldsymbol{c}(x), \end{cases} \tag{1}$$

where

$$\begin{cases} \boldsymbol{c}(x) \doteq \left(C(x, x^i)\right)_{1 \le i \le N} & \text{(covariances between } x \text{ and inputs } \boldsymbol{x}) \\ \boldsymbol{C} \doteq \left(C(x^i, x^j)\right)_{1 \le i,j \le N} & \text{(covariances between inputs } \boldsymbol{x}) \\ \Sigma \doteq \operatorname{diag}\left(\sigma^2(x^i)\right) & \text{(diagonal matrix of noise variance)} \end{cases}$$

Covariance Kernels & Parameter Estimation

- Common choice is squared-exponential (or Gaussian) covariance kernel

$$C(x, x') = \eta^2 \exp\left(-\frac{(x_{ag} - x'_{ag})^2}{2\theta_{ag}^2} - \frac{(x_{yr} - x'_{yr})^2}{2\theta_{yr}^2} \right).$$

- Knowing mortality at x will greatly influence mortality at “neighboring" x's
  - e.g. knowing mortality for a 80 year old in 2015 greatly aids in prediction of a 85 year old's mortality in 2016; knowing a 50 year old's mortality in 2000 has a nearly non-existent effect
- Implies hyperparameter family of $\Theta \doteq (\theta_{ag}, \theta_{yr}, \eta^2, \sigma^2)$
  - Also mean function hyperparameters (if included)
- Estimates are fit using MLE; likelihood can be written out explicitly due to Gaussian assumptions
  - Done using `R` package `DiceKriging`
- Alternatively, can use Bayesian approach with priors on Θ
  - Separate package using STAN language
  - Leads to non-Gaussian posterior

Illustrative Example

Goal: Learn $f(x) = \sin(x)$ over domain $[0, 2\pi]$

- Observe realizations of

$$y = \sin(x) + \varepsilon$$

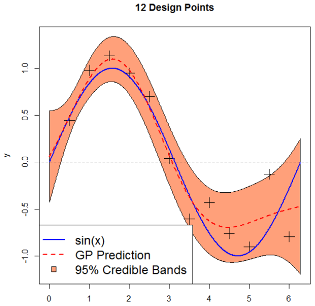
  where $\varepsilon \sim N(0, 0.01x)$
- Try:
  - $x = 0.5, 1, 1.5, \ldots, 5.5, 6;$
  - $x = 0.25, 0.5, 0.75, \ldots, 5.75, 6$

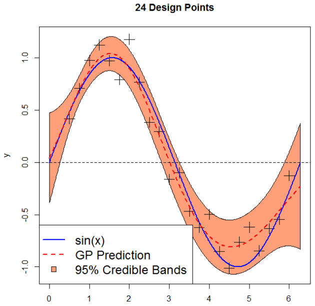
Illustrative Example

$$y = \sin(x) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, 0.01x)$$

1. Uncertainty increases at edges (especially for large x: $\sigma(x) = 0.01x$)
2. Uncertainty decreases as $N = 12 \rightarrow N = 24$
3. Accuracy of fit increases as $N = 12 \rightarrow N = 24$

$x = 0.5, 1, 1.5, \ldots, 5.5, 6$ (12 points)	$x = 0.25, 0.5, 0.75, \ldots, 5.75, 6$ (24 points)





Data

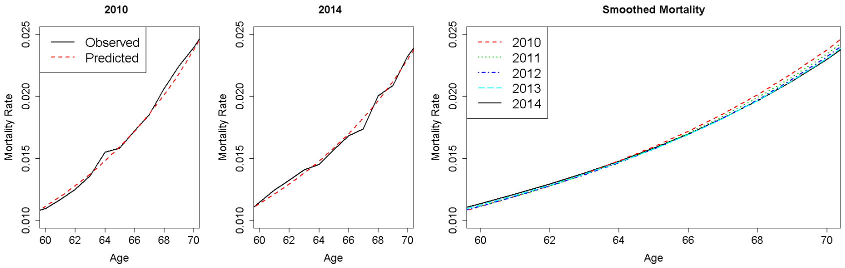
- CDC Data
  - United States
  - Ages 50–84, Years 1999–2014
    - ★ 1360 Data Points ($x = (x_{ag}, x_{yr})$)
    - ★ 84 is maximal age for CDC data
    - ★ 50 chosen as cutoff to minimize mixing lower age behavior
    - ★ 1999 earliest year available on wonder.cdc.gov
    - ★ Could add earlier years, but our analysis suggests they have little effect
    - ★ Most relevant for longevity risk

Model Assumptions

- Observe central mortality rate $e^{-\mu(x_{ag}, x_{yr})} = D(x_{ag,yr})/E(x_{ag,yr})$
- Fit log-mortality rate y to $x = x_{ag}, x_{yr}$ pairs
- Can try $\sigma(x)$ based on Binomial assumption
  - Overdispersion issues ($\mu_{ag,yr}$ is unknown)
  - Minimal change in final model from simply choosing $\sigma := \sigma(x)$
- Use Gaussian covariance kernel
  - Implies f is differentiable
  - Minimal change in final model from other kernel choices

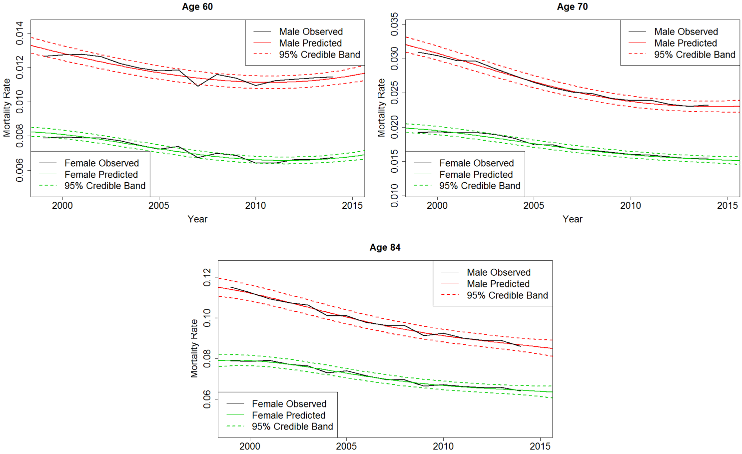
Posterior Predicted Mortality Rates

- Showing $m_*(x)$ for each ages 60–70
- Left panels include historic observations
- Right panel suggests mortality improvement



Mortality Over Time with Credible Bands

- Posterior mean and 95% credibility bands for f_* over calendar year
- Can observe increasing uncertainty at edges
- Observe mortality improvement then decline



Mortality Improvement

- Typical way is to look at the annual backward improvement

$$MI^{obs}_{back}\left(x_{ag}; yr\right) \doteq 1 - \frac{\exp\left(\mu(x_{ag}, yr)\right)}{\exp\left(\mu(x_{ag}, yr-1)\right)}$$

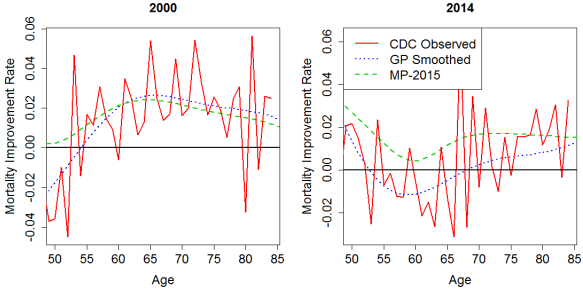
- $f_*(x_{ag}, yr)$ is a random variable, so we have the predicted mean improvement

$$m^{GP}_{back}\left(x_{ag}, yr\right) = \mathbb{E}\left[MI^{GP}_{back}\left(x_{ag}, yr\right)\right] \doteq \mathbb{E}\left[1 - \frac{\exp\left(f_*(x_{ag}, yr)\right)}{\exp\left(f_*(x_{ag}, yr-1)\right)}\right]$$

  - Available in closed form (lognormal distribution)
- Also have $MI^{MP}_{back}\left(x_{ag}; yr\right)$ (published MP-2015 improvement factors)

Comparing Mortality Improvement Methods

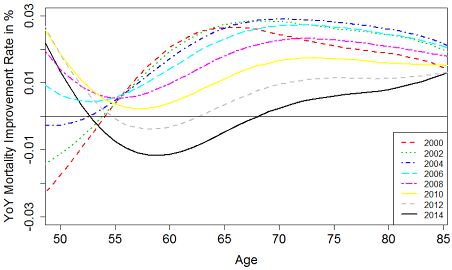
- Raw improvements extremely noisy (unsurprising)
- Smoothed methods both follow data well
- GP implies a stronger decline
  - Additional data suggests mortality deceleration



GP Improvement Over Time

GP Improvements from 2000–2014 (in 2 year increments)

- Shape changes (flips) over time
- Consistent with MP-2015
- Generally decelerating after age 55



Backward Difference & Derivatives

$$1 - \left(\frac{\exp\left(f_*(x_{ag}, yr)\right)}{\exp\left(f_*(x_{ag}, yr - h)\right)} \right)^{1/h} \approx -\frac{f_*(x_{ag}, yr) - f_*(x_{ag}, yr - h)}{h} \qquad (3)$$

- As defined, the annual mortality improvements are backward differences with $h = 1$
- Right side remains a GP by linearity
- Taking limit as $h \to 0$ yields derivative
  - Exists (depending on covariance kernel)
- Closed form expressions for distribution of $\frac{\partial f_*}{\partial x_{yr}}$

GP Derivative

Proposition

For the Gaussian Process f_* with a twice differentiable covariance kernel C, the limiting random variables

$$\frac{\partial f_*}{\partial x_{yr}}(x_{ag}, yr) \doteq \lim_{h \to 0} \frac{f_*(x_{ag}, yr+h) - f_*(x_{ag}, yr)}{h} \tag{4}$$

exist in mean square and form a Gaussian process $\frac{\partial f_*}{\partial x_{yr}} \sim GP(m_{diff}, s_{diff})$. Given the training set $\mathcal{D} = (\mathbf{x}, \mathbf{y})$, the posterior distribution of $\frac{\partial f_*}{\partial x_{yr}}(x_*)$ has mean and variance

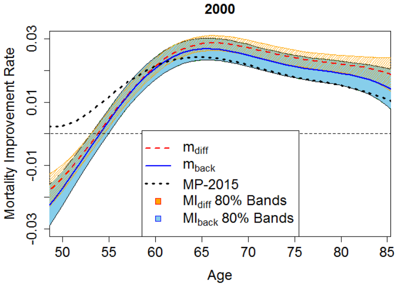
$$\begin{cases} m_{diff}(x_*) = \mathbb{E}\left[\frac{\partial f_*}{\partial x_{yr}}(x_*) \middle| \mathbf{x}, \mathbf{y}\right] = \frac{\partial C}{\partial x'_{yr}}(\mathbf{x}, x_*)(\mathbf{C} + \Sigma)^{-1}\mathbf{y}, \\ s^2_{diff}(x_*) = Var\left(\frac{\partial f_*}{\partial x_{yr}}(x_*) \middle| \mathbf{x}, \mathbf{y}\right) = \frac{\partial^2 C}{\partial x_{yr}\partial x'_{yr}}(x_*, x_*) - \frac{\partial C}{\partial x'_{yr}}(\mathbf{x}, x_*)(\mathbf{C} + \Sigma)^{-1}\frac{\partial C}{\partial x_{yr}}(x_*, \mathbf{x}), \end{cases}$$

where $\frac{\partial C}{\partial x'_{yr}}(\mathbf{x}, x_*) = \left[\frac{\partial C}{\partial x'_{yr}}(x^1, x_*), \ldots, \frac{\partial C}{\partial x'_{yr}}(x^N, x_*)\right]$ and each component is computed as the partial derivative of $C(x, x')$.

See Theorem 2.2.2 in Adler (2010) for more details/proof.

Comparing Other Methods with GP Derivative

- Blue is backwards mortality difference (as before); red is GP derivative; black is MP-2015
- Analysis of other years shows deceleration begins around 2010
  - Implies mortality evolution is convex
    - Justifies accelerating divergence between yearly difference and derivative methods
  - MP-2014 and MP-2015 begin to diverge around 2010
    - Suggests that later years are crucial to mortality forecasts



Conclusions

- GP's provide a variety of benefits to modeling mortality and mortality improvement
  - Bayesian approach (data driven)
  - Posterior distribution for any location
    - Including distribution of mortality improvement (both yearly difference and instantaneous)
    - Credible bands (historic and forecasting)
- Relatively consistent results with MP-2015
  - Four years of additional data pushes GP results in the direction that MP-2015 took compared to MP-2014
  - Differences in results is likely due to data differences than model issues
- GP framework easily handles joint analysis of mortality rates and mortality improvement

Future Work

- Modeling annual mortality improvement directly with GP
- Monotonicity constraint
- Multiple populations
  - Jointly modeling male & female mortality
  - Multivariate GP of multiple countries
- Modeling by cause of death
- More detailed backtesting
  - Analysis on other countries (individually)

Future Work

- Modeling annual mortality improvement directly with GP
- Monotonicity constraint
- Multiple populations
  - Jointly modeling male & female mortality
  - Multivariate GP of multiple countries
- Modeling by cause of death
- More detailed backtesting
  - Analysis on other countries (individually)

THANK YOU!

References

Williams, C. K. and Rasmussen, C. E. 2006.
Gaussian processes for machine learning, the MIT Press.

Adler, Robert J. 2010
The geometry of random fields, Siam

Roustant, O., Ginsbourger, D., Deville, Y., et al. 2012.
Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization.
Journal of Statistical Software, 51(1):1–55.

M. Ludkovski, J. Risk, H. Zail
Gaussian Process Models for Mortality Rates and Improvement Factors
https://arxiv.org/abs/1608.08291
Submitted to North American Actuarial Journal