# Gaussian Process Models in Actuarial Science

Mike Ludkovski & Jimmy Risk

Dept of Statistics & Applied Probability UC Santa Barbara

PSTAT Seminar
May 4 2016

# Outline

- Gaussian Process Regression
- Example: Modeling Mortality Surface
- Example: Pricing Deferred Annuities

# Statistical Learning

- Input-Output "black box": $\boxed{Y(x) = f(x) + \varepsilon(x)}$
- Learn the latent $f$ through employing a stochastic sampler $Y$
  1. The inputs $X$ are the sampled locations/initial simulation states
  2. The response $Y$ are observations
  3. $Y$ has intrinsic noise $\varepsilon$ due to measurement error, or randomness that is part of the simulation
  4. $\mathbb{E}[\varepsilon] = 0$: so $f(x)$ is the expected response
- Goal: Recover $f$ based on the data ($x^{1:N}$ and $y^{1:N}$)
- Solution: choose an approximation architecture $\hat{f} \in \mathcal{H}$
- Choose loss function $L$
- (Choose a simulation design: $x^{1:N}$)
- Set $\hat{f} = \arg\min_{f \in \mathcal{H}} L(f|(x, y)^{1:N})$

# Response Surface Modeling

$$Y(x) := f(x) + \varepsilon(x).$$

- Conditional Expectation $f = \mathbb{E}[Y|x]$: lots of applications in finance: American options, sequential stochastic control, XVA, et cetera.
- Nested Simulation
- Backward stochastic differential equations
- Capital Requirements/Insurance

Also appears in other fields:

- Metamodeling/Statistical Emulation/Surrogates
- OR: Simulation Optimization

# What is an Emulator?

- Classical regression – data is given and try to fit the “best curve”
- In metamodeling generating data (through efficient simulations) is part of the solution
- Also, typically look for a non-parametric model (dense $\mathcal{H}$)
- Used extensively in machine learning; simulation optimization, computational statistics
- See eg Kleijnen (2015), Williams and Rasmussen (2006), Powell and Ryzhov (2012)
- Connects to CS, OR, stats communities (language barriers!)
- Can be applied in non-stochastic contexts: DACE (design and analysis of computer experiments)

# Modeling $f$

Must impose some structure on $f$ ($X$ is a "nice" process, so $f$ is "smooth")

- Project onto basis functions: $f(x) = \sum_i a_i H_i(x)$
- Smoothing spline (piecewise cubic)
- Piecewise linear
- Piecewise constant $f(x) = \sum_i a_i 1_{\{x \in R_i\}}$
- Fully nonparametric (kernel): $f(x) = \sum_i K(x, x^i) y^i$
- Gaussian process

# Gaussian Process Model

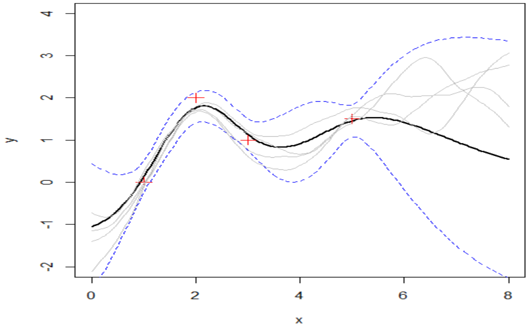
- The latent $f$ lives in the function space $\mathcal{H}_K$ – Gaussian RKHS
- Means $f(\cdot)$ is a realization of a Gaussian random field with a covariance structure defined by $K$, $\mathcal{H} = span(K(\cdot, x) : x \in \mathbf{X})$
- $K(x, x') := \mathbb{E}[f(x)f(x')]$ controls the spatial decay of correlation, i.e. smoothness of $f$
- e.g Gaussian kernel $K(x, x') = \tau^2 \exp(-\frac{\|x-x'\|^2}{2\theta^2})$ – elements of $\mathcal{H}_K$ are $C^\infty$, with lengthscale $\theta$ and fluctuation scale $\tau$.
- Penalized $L^2$ projection: $\hat{f} = \arg\min_{f \in \mathcal{H}} \sum_{i=1}^N (f(x^i) - y^i)^2 + \|f\|_K^2/2$;
- Representer theorem implies that $\hat{f}(x) = \sum_{i=1}^N y^i w_i K(x, x^i)$ – linear model in the infinite basis expansion defined by $K$

# Bayesian Interpretation

- Think of $f \sim GP(0, K)$ as a random element in $\mathcal{H}_K$
- Gaussian prior $f(x) \sim N(0, \tau^2)$
- Below: $\theta = 2, \tau = 1.5$ (solid = mean; dashed blue = 95% CI)

# Bayesian Interpretation

- Condition on observations $\mathcal{G} \equiv (x, y)^{1:N}$ $Y(x) = f(x) + \varepsilon(x)$ where $\varepsilon(x) \sim N(0, \sigma^2(x))$ (below $\sigma \equiv 0.2$)
- The posterior is a measure on $\mathcal{H}_K$ (i.e function-valued)
- Posterior is still a Gaussian process



# GP Equations

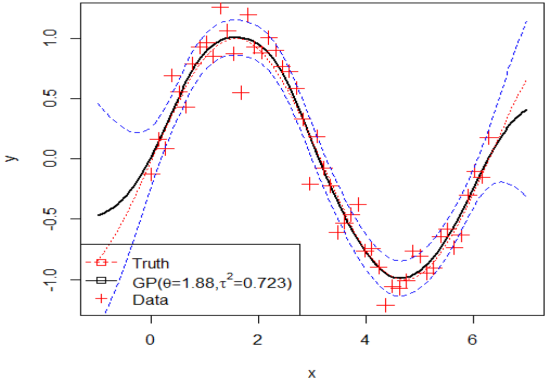
- Have analytic formulas for the posterior distribution of $f(x)|\mathcal{G} \sim N(m(x), v^2(x))$

$$\text{mean} \quad m(x) = \vec{k}(x)^T(\mathbf{K}+\Sigma)^{-1}\vec{y}$$
$$\text{cov} \quad v(x,x') = K(x,x') - \vec{k}(x)^T(\mathbf{K}+\Sigma)^{-1}\vec{k}(x')$$

- $K_{ij} = K(x^i, x^j)$, $\Sigma = diag(\sigma^2(x^i))$, $k_i = K(x, x^i)$
- Visually has a "football" shape– $v^2(x)$ has local minima at $x^i$'s.
- The mean $m(x)$ is a linear combination of kernel eigenfunctions centered at design sites
- Outside the domain $\mathbf{X}'$, revert to prior $m(x) \to 0, v^2(x) \to \tau^2$

# More Data

- Global consistency – converge to the truth as $N \to \infty$
- Optimized Matern-5/2 kernel
  $K(x, x'; \tau, \theta) = \tau^2 (1 + (\sqrt{5} + 5/3)\|x - x'\|_\theta^2) \cdot e^{-\sqrt{5}\|x - x'\|_\theta}$



# GP Model

- Given the kernel, the posterior is in closed-form
- Automatically determines spatial smoothness of $f$: lengthscale $\theta$ controls correlation decay
- Coherently deals with in-sample smoothing and out-of-sample forecasting (+uncertainty quantification)
- Can incorporate a non-zero mean function, or estimate a “trend” like in least-squares

# Fitting a GP

- Need to know the kernel hyperparameters – $\tau, \theta$'s, et cetera.
- Solution I: Use MLE (nonlinear optimization problem).
- Solution II: Specify priors and use a fully Bayesian method (requires MCMC)
- Need the sampling noise $\sigma^2(x)$ – use batching/replications to estimate
- GP is expensive compared to e.g LM; complexity is $O(N^3)$ for a design of size $N$
- We used `DiceKriging` package in `R` – off-the-shelf use

# Mortality Surface

- Raw mortality table has two main inputs: Age and Year
- For each Age/Year have $D_{ij}$ is the # of deaths and $E_{ij}$ the number of exposed-to-risk
- Model the latent mortality state $\mu_{ij}$: $D_{ij}/E_{ij} = e^{\mu_{ij}} + \varepsilon_{ij}$

| Age | Year | Deaths | Exposed | Crude Rate |
|---|---|---|---|---|
| 65 years | 2013 | 25036 | 1616005 | 0.015493 |
| 66 years | 2013 | 27466 | 1683234 | 0.016317 |
| 67 years | 2013 | 22837 | 1231339 | 0.018546 |
| 68 years | 2013 | 23613 | 1208652 | 0.019537 |
| 69 years | 2013 | 25372 | 1173960 | 0.021612 |
| 70 years | 2013 | 27596 | 1198287 | 0.02303 |
| 71 years | 2013 | 26559 | 1036368 | 0.025627 |
| 72 years | 2013 | 25971 | 937562 | 0.027701 |
| 73 years | 2013 | 26303 | 882726 | 0.029797 |
| 74 years | 2013 | 27145 | 829509 | 0.032724 |
| 75 years | 2013 | 27945 | 785563 | 0.035573 |

# Mortality Surface

- Dataset: Age: 50–84, Year: 1999–2011
- In general: mortality is improving (people are living longer)
- How fast (mortality improvement factor)? Will this continue in 2016? In 2020? In 2040?
- What is the uncertainty about e.g life expectancy of a 65 year old today?

# Features of a GP Model

- Historical smoothed mortality curves by calendar year: $m(x^{1:N})$
- Credible interval around such curves: posterior variance $v^2(x^{1:N})$
- Project the curves forward $m(x'), v^2(x')$ for future years/ages
- Generate stochastic future forecasts (sample from $f(x')|\mathcal{G}$ as a future mortality scenario)

# Mortality Modeling

- Work in progress with Howard Zail (Elucidor)
- Believe this is a viable alternative to existing methods and may become a new SOA-sanctioned approach
- Bayesian perspective is a plus:
    - Domain expertise = prior
    - Updating the tables based on new data
    - Working with bespoke pension pools
    - Coherently merging multiple datasets

# Part II: Nested Simulation

# Nested Simulation

- Generate a *multi-level* forest of Monte Carlo simulations
- The outer level are top-level scenarios
- Inner level simulations are used to approximate some quantity of interest given the global scenario
- For example, Value-at-Risk calculations (outer = economic scenarios on $[0, T]$; inner = portfolio value as of $T$)
- Nesting leads to extremely computational demands
- One solution is to build a metamodel for the inner-level simulations

# Valuing Deferred Annuities

- Annuity: pay \$1 as long as the insured is alive
- Survival probability $P(t, u) = \mathbb{E}\left[\exp\left(-\int_0^t \mu(s, u + s)\, ds\right)\right]$
- $a_0 = \sum_t P(t, u)B(0, t)$, where $B(0, t)$ is the bond price
- Model the force of mortality $\mu(t, u)$
- Typically $\mu$ comes from a multi-factor stochastic model (e.g. using auto-regressive time-series or diffusions); plus age-modeling in terms of $a$
- No closed-form expression for $P(t, u)$ – need to be also evaluated

# Valuing Deferred Annuities

- Today: aged 45. Will receive a pension benefit starting at age 65. What is the NPV of the pension?
- This is a 20-year deferred life-annuity
- $P(\tau_u > T \mid \tau_u > t, \mathcal{F}_t) = \mathbb{E}\left[\exp\left(-\int_t^T \mu(s, u+s)\right)\Big| X_t\right] \doteq P(X(t); t, T, u)$ – functional of the trajectory of $X$ between $T$ and $T+s$
- Write $NPV \doteq \mathbb{E}[e^{-rT} \cdot a(X_T, T, u)] = \mathbb{E}[\mathbb{E}[a_T|X_T]]$ – first compute annuity prices for 65-year olds as a function of $X_T$: $f(x) \doteq \mathbb{E}[F(T, X_\cdot)|X_T = x]$
- Then integrate over the distribution of $X_T$: $\mathbb{E}[F(T, X_\cdot)|X_0] = \int_{\mathbb{R}^d} f(x) p_T(x|X_0) dx$
- Metamodeling: construct $\hat{f}(x) \simeq \mathbb{E}[a_T|X_T = x]$

# Traditional Nested Approach

- Outer: $\mathbb{E}[F(T, X_\cdot)|X_0] \simeq \frac{1}{N_{out}} \sum_{m=1}^{N_{out}} f(x^{(m)})$
- Inner: $f(x^{(m)}) \simeq \frac{1}{N_{in}} \sum_{n=1}^{N_{in}} F(T, x_\cdot^{(m),n}), \qquad m = 1, \dots, N_{out}$
- $x_t^{(m),n}, t \geq T$ are $N_{in}$ independent trajectories of $X$
- Simulation budget is $\mathcal{O}(N_{out} \cdot N_{in})$
- An emulation framework generates a fitted $\hat{f}(\cdot)$ by solving regression equations over a training dataset $\{x^{(n)}, F(T, x^{(n)}(\cdot))\}_{n=1}^{N_{tr}}$ of size $N_{tr}$
- Emulator budget $N_{tr} + N_{out}$, plus regression overhead
- provides a principled statistical framework for optimizing, assessing and improving such two-level simulations

# Contribution

- Practitioners often compute $\hat{f}(x)$ by an approximation such as

$$\mathbb{E}[\exp(-m(t,x))] \approx \exp(-\mathbb{E}[m(t,x)])$$

  which allows closed-form expressions for annuity prices
- No assessment of the approximation error
- New proposal (L-Risk 2016 IME): use GP to build a flexible metamodel for $f$
- Related: computing Value-at-Risk of the annuity

# Challenges

- Challenge 1: stochastic mortality models are often multi-dimensional (e.g. APC = Age-Period-Cohort version of Lee-Carter has 3 factors)
- $m(t,u) = \beta_u^{(1)} + \frac{1}{n_a}\kappa^{(2)}(t) + \frac{1}{n_a}\gamma^{(3)}(t-a)$, with $\kappa, \beta, \gamma$ stochastic processes
- Challenge 2: calibration of stochastic mortality is nontrivial. Cairns et al (2011) advocate re-calibration using the simulated trajectory $X_{0:T}$ (breaking Markov structure)
- In other words $\beta_u = \beta_u(X_{0:T})$;, et cetera.

# GP for Annuities

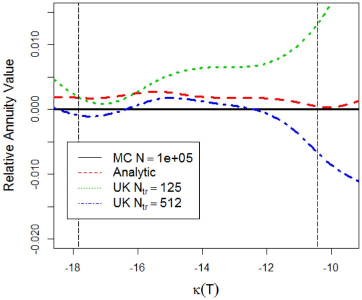
- Generate a design $x^{1:N_{tr,1}}$
- At each $x^{1:N_{tr,1}}$ generate $N_{tr,2}$ trajectories from $X_{T:\bar{T}}$ and corresponding NPV $a(X_T)$
- Average/compute variance to be entered into the GP metamodel:

$$y^{(n)} \doteq \frac{1}{N_{tr,2}} \sum_{j=1}^{N_{tr,2}} F(T, x^{(n),j}(\cdot))$$

$$\hat{\sigma}^2(x^{(n)}) \doteq \frac{1}{N_{tr,2}-1} \sum_{j=1}^{N_{tr,2}} \left\{ y^{(n)} - F(T, x^{(n),j}(\cdot)) \right\}^2$$

- $\hat{\sigma}^2(x^{(n)})/N_{tr,2}$ is proxy for the variance of the batch mean $y^{(n)}$
- Fit a GP for $a(X_T)$ (total of $N_{tr,1} \times N_{tr,2}$)
- Generate $N_{out}$ scenarios $X_{0:T}$ and average: $\hat{a}_0 = \frac{1}{N_{out}} \sum_{n=1}^{N_{out}} \hat{f}(x_T^n)$

# Illustration: 1D $\mu_T$ following a jump-diffusion




- Analytic approximation has a fixed, unknown bias
- Monte Carlo approximations will converge to the truth as $N \to \infty$
- Also asymptotically unbiased

# GPs with Trend

- Generalize to $f(x) = X(x) + \beta_0 + \sum_{j=1}^{p} \beta_j h_j(x)$ where $\beta_j$ are constants to be estimated, and $h_j(\cdot)$ are given basis functions
- Trend component allows to incorporate domain knowledge about the response, while the mean-zero GP component $X$ offers a flexible nonparametric model for the residuals
- Option 1: variance reduction by using the analytic mean $F(\mathbb{E}[X.])$ as trend
- Option 2: estimate $\beta_j' s$ simultaneously with the GP equations

# Two Population Longevity Hedge Portfolio with stochastic mortality factors

| | $N_{tr} = 1000$ | | $N_{tr} = 8000$ | |
|---|---|---|---|---|
| Type | Bias | $\sqrt{\text{IMSE}}$ | Bias | $\sqrt{\text{IMSE}}$ |
| Analytic A1 | -2.101e-02 | 3.460e-02 | -2.101e-02 | 3.460e-02 |
| Analytic A2 | 3.629e-03 | 3.733e-03 | 3.629e-03 | 3.733e-03 |
| Thin Plate Spline | -1.050e-03 | 1.437e-02 | 4.431e-04 | 3.294e-03 |
| Universal Kriging | -1.156e-03 | 1.872e-02 | 2.556e-03 | 1.454e-02 |
| Simple Kriging | 2.148e-03 | 2.308e-03 | 9.229e-04 | 1.469e-03 |
| Least Squares MC | -1.050e-03 | 1.437e-02 | 5.324e-04 | 3.295e-03 |

Tab: Performance of analytic estimates and surrogate models for hedge portfolio values in the two-population model case study. Numbers reported are based on $N_{out} = 1000$ simulations of $Z(T)$ with a Monte Carlo benchmark. $N_{tr}$ is allocated into $N_{tr,1} = N_{tr}^{2/3}$ training points and $N_{tr,2} = N_{tr}^{1/3}$ Monte Carlo batches per training point. Simple kriging model uses A2 estimator as trend. For comparison purposes, the average value of the hedge portfolio was 0.1995.

# Variable Annuity with Mortality, Stochastic Interest Rate, Stock Price Factors

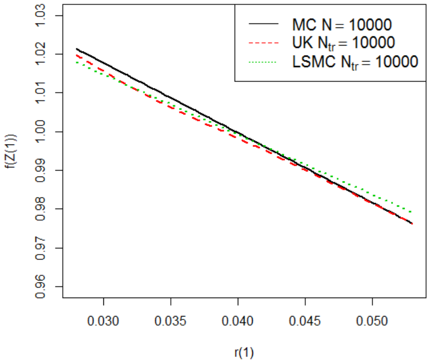



Fig: Marginal dependence plot of $f(Z(1))$ versus $r(1)$, where $f(Z(1))$ and is estimated through a smoothed Monte Carlo benchmark with $N_{out} = 25$, $N_{in} = 10,000$. The two emulator models used $N_{tr,1} = N_{tr,2} = 100$. The 2-dim. experimental design $\mathcal{D}$ for the emulators was empirical.

# Take-Aways

- GP is a flexible, off-the-shelf regression framework
- Many opportunities to apply emulation in insurance
- Lots more opportunities in this direction

# Take-Aways

- GP is a flexible, off-the-shelf regression framework
- Many opportunities to apply emulation in insurance
- Lots more opportunities in this direction

Metamodeling = Regression + Stochastic Grid

THANK YOU!

# References


- JPC Kleijnen *Design and Analysis of Simulation Experiments*, 2nd Edition, Springer, 2015.
- CKI Williams and CE. Rasmussen *Gaussian processes for machine learning*, MIT Press, 2006.
- M. Kohler, *A review on regression-based Monte Carlo methods for pricing American options*, in Recent Developments in Applied Probability and Statistics, Springer, 2010, pp. 37–58.
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E., *Sequential design of computer experiments for the estimation of a probability of failure*, Statistics and Computing, 22(3), 773–793, (2012).
- R. Gramacy and M. Ludkovski *Sequential Design for Optimal Stopping Problems* SIFIN 6(1), 2015, pp. 748–775
- M. Ludkovski *Kriging Metamodels for Bermudan Option Pricing*, preprint, 2016 arXiv:1509.02980
- R. Hu and M. Ludkovski *Sequential Design for Ranking Response Surfaces*, preprint, 2015. arXiv:1509.00310