Statistical Learning for Portfolio Tail Risk Measurement

Mike Ludkovski

joint with **Jimmy Risk** (Cal Poly Pomona)

Data Science in Insurance Conference, Cass Business School, July 16, 2018

UC Santa Barbara

STATISTICS & APPLIED PROBABILITY
CENTER FOR FINANCIAL MATHEMATICS & ACTUARIAL RESEARCH

Portfolio Risk Measurement

- Risk Assessment mandated by Solvency II: 99.5% VaR (TVaR in banking) at 1-year horizon
- Practically computed by building N scenarios for market conditions at T ($\mathbb{P}$-measure)
- Then need to evaluate portfolio losses for each scenario ($\mathbb{Q}$-measure) and compute the α-quantile
- No simple way to compute portfolio value. Typical approach: Monte Carlo approximation
- Leads to nested simulations: Generate N_{in} inner simulations at scenario n to compute cashflows $y^{n,i}$; average to estimate portfolio value $\bar{y}^n$

Outline: Finding Needle in a Haystack

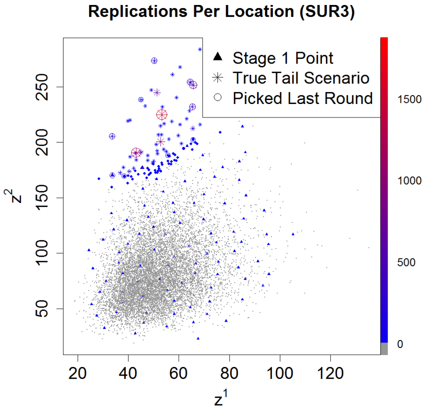
- Intense computation burden: $(N = 100,000) \times (N_{in} \gg 100)$ simulations per scenario $\rightarrow$ tens of millions of simulations
- Ultimately only $0.5\% = 500$ scenarios are relevant for (T)VaR
- **GOAL OF THE TALK**: how to adaptively allocate simulation budget to avoid "wasted" simulations
- Statistics: Spatial emulation and GPs
- Machine Learning: Acquisition functions for sequential design
- Actuarial Science: Case Studies
- Seek two orders-of-magnitude gains

Setting

- Given N scenarios (discretized scenario space)
- Usually comes from economic scenario generators (physical measure)
- VaR: find **level** of the αNth worst loss: order statistic
- TVaR: find the **average** of the αN worst losses
- Have a total limited simulation budget of $\mathcal{N} \simeq \mathcal{O}(N)$
- Will sequentially add inner simulations: rounds $k = 1, 2, \ldots$
- Overall design is $\mathcal{D}_k := (r_k^n)$: r_k^n is the number of inner simulations allocated by round k, $\sum_{n=1}^{N} r_k^n = N_k$ ($N_K = \mathcal{N}$)
- Focus on non-asymptotic performance with low budget $\mathcal{N}$

Take-Away

- Focus effort on the tail scenarios
- Use a statistical surrogate to borrow information from inner simulations of other scenarios
- Skip entirely (most of) the scenarios that are far from the tail
- Improve estimates for scenarios that matter



Spatial Modeling

- Scenarios correspond to realizations of underlying stochastic factors (Z_t)
- Associate scenario to $Z_T = z$
- Portfolio: cashflows with net present value $Y = F(Z_s : s \geq T)$. Portfolio value is $f(z) := \mathbb{E}[Y|Z_T = z]$
- If two scenarios are close, then the portfolio losses $f(z), f(z')$ should be also close
- Build a spatial statistical model for f over the domain of Z —learn the correlation structure of $f(z^{1:N})$
- We use Gaussian Process emulation: quantifies the posterior uncertainty for **allocation of future simulations** + efficient sequential updating

Literature Review

We develop a machine learning framework tailored to portfolio risk measurement

- Large number of scenarios (many emulators are for very expensive simulators)
- Complicated simulation noise (heteroskedastic, non-Gaussian, etc)
- Learning objective is implicit (contrast to thresholding $f(z)$ against a known L)

- Simulation/OR Literature:
  - Gordy & Juneja (MS 2010), Broadie et al (MS 2011): efficient outer/inner allocation without spatial structure and with continuous scenario space
  - Broadie et al (OR 2012): linear regression plus two stage design
  - Liu and Staum (WSC 2011): three-stage adaptive allocation
- Statistical Emulation
  - Picheny et al (2012, 2015, 2017): surrogate models + active learning for level sets (continuous search space, no replicates)
  - Bauer et al (Astin 2012): LSMC regression for capital requirements (non-adaptive)
  - Binois et al (JCGS 2018): specialized GP surrogate to handle stochastic simulators

Gaussian Process Emulator

- Non-parametric regression, similar to splines or kernel regression
- Multivariate Gaussian structure to describe the shape of $f(\cdot)$: covariance matrix $\mathbf{C}_{i,j} = C(z^i, z^j)$. We used the isotropic Matern-5/2 family.
- MVN posterior $f(z)|\mathcal{D}_k \sim \mathcal{N}(m_k(z), s_k^2(z))$: mean $m_k(z^n)$ is proxy for $\hat{f}(z^n)$); $s_k^2(z^n)$ quantifies credibility
- $\mathbf{R} = \mathrm{Diag}(r_k^1, \dots, r_k^n)$, $\mathbf{\Delta} = \mathrm{Diag}(\tau^2(z^1), \dots, \tau^2(z^n))$

$$m_k(z) \doteq \mathbf{c}(z)^T(\mathbf{C} + \mathbf{R}^{-1}\mathbf{\Delta}_k)^{-1}\overline{\mathbf{y}}_k;$$
$$s_k^2(z) \doteq C(z,z) - \mathbf{c}(z)^T(\mathbf{C} + \mathbf{R}^{-1}\mathbf{\Delta}_k)^{-1}\mathbf{c}(z),$$

- Only need to work with the unique scenarios $z^n \in \mathcal{D}$
- State-dependent simulation variance: Treat the noise terms $\mathbf{\Delta}$'s as a latent spatial process: $\mathbf{\Delta} = \mathbf{C}_g(\mathbf{C}_g + g\mathbf{R}^{-1})^{-1}\mathbf{\Lambda}$ – `hetGP`: package

Budget allocation

I. Initialize $\hat{f}_0$ by generating simulations over a subset of pilot scenarios.

LOOP : predict $\hat{f}_k$ on $\mathcal{Z}$ to determine which scenarios are close to $\mathcal{R}$.

II. Compute acquisition function aka weights $\mathcal{H}(z^n), n = 1, \ldots N$

III. Allocate more inner simulations to scenarios with high weights: Generate cashflows $(Y_t^i(z^n))_{t=T}^{\infty}$ and new $y_k^{n,i}$

IV. Batch of Δr new simulations per round (computational speed-up)

V. Update emulator to $\hat{f}_{k+1}$ based on the new MC output.

ND LOOP

Further Details

Assigning Scenario Weights

- Heuristics $\mathcal{H}(z)$ about information gain from running more simulations at z – greedy but still takes into account Exploration/Exploitation trade-off
- Stepwise Uncertainty Reduction: $\mathcal{H}_{k+1}|_{r^n_{k+1}=r^n_k+1} - \mathcal{H}_k$
- Active learning/simulation optimization/sequential design/knowledge gradient/....
- Take advantage of nesting to improve both accuracy and speed

Estimating Portfolio Risk

- Plug-in estimator based on the ranked posterior means $m_k^{(n)}$
- Risk measure is $R = \sum_{n=1}^N w^n f(z^n)$ with VaR: $w^n = 1_{\{f(z^n)=f^{(\alpha N)}\}}$; TVaR: $w^n = \frac{1}{\alpha N} \cdot 1_{\{f(z^n)<f^{(\alpha N)}\}}$.
- Smooth using Harrell-Davis L-estimator, $\hat{R}_k^{HD,\mathrm{VaR}} \doteq \sum_n \tilde{w}^{(n)} m_k^{(n)}$

Targeting the Quantile Level: ST-GP

- Targeted mean square error (Picheny et al 2012): $\text{tmse}_k(z) \doteq s_k^2(z) W_k(z; L, \varepsilon)$

$$W_k^{\text{VaR}}(z; L, \varepsilon) \doteq \frac{1}{\sqrt{2\pi(s_k^2(z) + \varepsilon^2)}} \exp\left(-\frac{1}{2} \left(\frac{m_k(z) - L}{\sqrt{s_k^2(z) + \varepsilon^2}} \right)^2 \right) = \phi(m_k(z) - L, s_k^2(z) + \varepsilon^2)$$

- High when $m_k \simeq L$ or when posterior variance is large
- ε controls how aggressive is the search. Take $\varepsilon = s(\hat{R}_k^{HD})$ – decreases in k
- Look-ahead variance (want $s_{k+1}^2(z)$):
$V_k(z^n; z^m) \doteq C(z^n, z^n) - \mathbf{c}(z^n)(\mathbf{C} + \mathbf{\Delta}_{k+1}^{cand})^{-1} \mathbf{c}(z^n)^T \Big|_{\mathbf{\Delta}_{k+1}^{cand} = \text{diag}\left(\frac{\hat{\tau}_k^2(z^1)}{r_k^1}, \ldots, \frac{\hat{\tau}_k^2(z^m)}{r_k^m + \Delta r_k}, \ldots, \frac{\hat{\tau}_k^2(z^{N'})}{r_k^{N'}} \right)}$
- Final criterion to minimize is total (integrated) tmse over all of $\mathcal{Z}$ conditional on adding simulations at z^{k+1}: $\widehat{\mathcal{H}}_k^{\text{VaR,timse}}(z) \doteq \frac{1}{N} \sum_{n=1}^{N} V_k(z^n; z) W_k^{\text{VaR}}(z^n; \hat{R}_k^{HD})$
- For TVar: $W_k^{\text{TVaR}}(z; \hat{R}_k^{HD}) \doteq \frac{1}{\sqrt{2\pi(s_k^2(z) + s^2(\hat{R}_k^{HD}))}} \Phi\left(\frac{\hat{R}_k^{HD} - m_k(z)}{\sqrt{s_k^2(z) + s^2(\hat{R}_k^{HD})}} \right)$

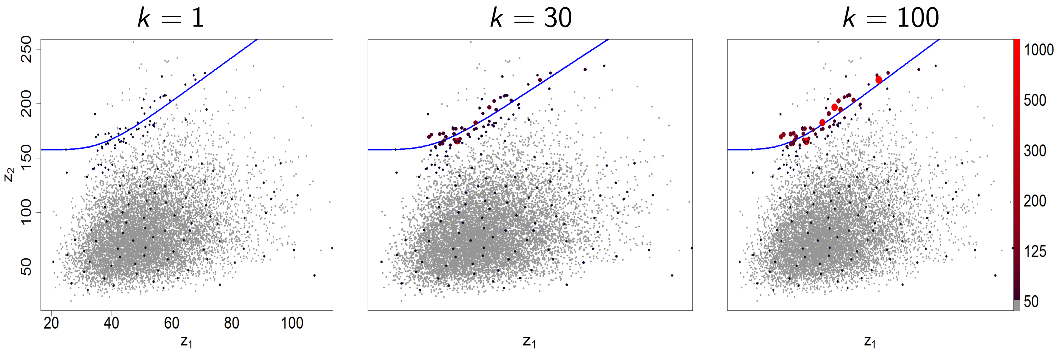
Global Variance Minimization: SV-GP

- Consider the global updating effect of running new inner simulations (Liu and Staum 2010)
- Sample in parallel several different outer scenarios to minimize the posterior estimator variance $s^2(\hat{R}_{k+1})$
- Freeze $\hat{w}^n_{k+1} = \hat{w}^n_k$
- Optimize $\{r'^n_k\}$ such that $\sum_n r'^n_k = \Delta r_k$, $r'^n_k \geq 0$

$$\mathbf{u}_k \boldsymbol{\Delta}^{cand}_{k+1} \mathbf{u}^T_k \qquad \text{where} \quad \mathbf{u}^T_k \doteq (\mathbf{C} + \boldsymbol{\Delta}_k)^{-1} \mathbf{C} \hat{\mathbf{w}}^T_k \to \min!$$

- More exploratory compared to ST-GP
- Also higher overhead and more “diffuse” design

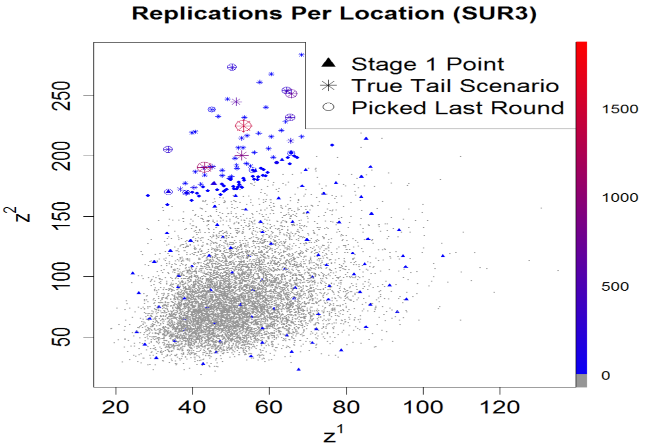
Adaptive Budget Allocation: VaR



Sequential budget allocation by SV-GP at stages $k = 1, 30, 100$ to learn $\text{VaR}_{0.005}$ in the 2-D Black-Scholes case study. The blue line indicates the true quantile contour $f^{(50)}$. Each dot represents an outer scenario z^n (Stage-0 pilot scenarios in **black**); the respective size and color are scaled non-linearly in r_k^n. Some scenarios receive as many as $r_k^n \approx 1200$ scenarios (total budget of $\mathcal{N} = 10^4$).

Adaptive Budget Allocation: TVaR

For TVaR, explore the entire tail, but still non-uniformly due to the spatial structure



Toy Example

- Portfolio consisting of Call options
- Two underlying risks (+correlated):

$$dS_t^1 = S_t^1 \left(\beta - \frac{1}{2}\sigma_1^2 \right) dt + \sigma_1 dW_t^{(1)}, \quad \beta = 0.04$$
$$dS_t^2 = S_t^2 \left(\beta - \frac{1}{2}\sigma_2^2 \right) dt + \sigma_2 (\rho dW_t^{(1)} + \sqrt{1-\rho^2} dW_t^{(2)}$$

Stock	Position	Initial Price	Strike	Maturity	Volatility
S^1	100	50	40	2	25%
S^2	-50	80	85	3	35%

- Explicit formula for portfolio loss $f(z)$ via Black-Scholes

$$\Pi(z^1, z^2) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\beta} 100 \left(S_2^1 - 40 \right)_+ - e^{-2\beta} 50 \left(S_3^2 - 85 \right)_+ \middle| (S_1^1, S_1^2) = (z^1, z^2) \right]$$

- Analytic bias and RMSE

Important Comparisons

- Gain from adaptive allocation – S*-GP vs U1-GP
- Gain from sequential learning – S*-GP vs A3-GP
- Gain from spatial modeling – S*-GP vs BR-SA
- Further considerations: initialization; number of rounds/batch size; variations on acquisition functions; variations on emulators

			Bias				
Approach	Kernel	RMSE	$k=1$	$k=10$	$k=20$	$k=50$	$k=100$
hetGP	Matérn-5/2	57.52	166.065	113.925	97.751	37.158	28.066
hetGP	Gaussian	68.06	91.475	104.427	74.874	52.716	48.249
SK	Matérn-5/2	69.31	914.728	206.267	113.716	69.821	48.670

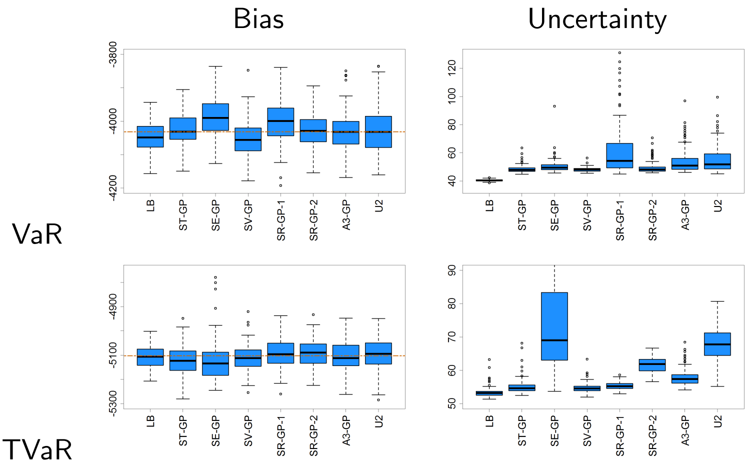
For the 2-D Black Scholes portfolio case study, average RMSE of $\hat{R}_K$ across different GP models/kernel families. We also report average bias bias$(\hat{R}_k)$ across a selection of intermediate stages $k = 1, 10, 20, 50, 100$. All methods use the ST-GP rule and are based on 100 macro-replications.

Results for 2-D Example

	$\text{VaR}_{0.005}$				$\text{TVaR}_{0.005}$			
	$SD(\hat{R}_K^{HD})$	$\bar{s}$	RMSE	$\lvert\mathcal{D}_K\rvert$	$SD(\hat{R}_K)$	$\bar{s}$	RMSE	$\lvert\mathcal{D}_K\rvert$
LB	44.35	40.42	46.77	1	47.29	53.48	47.42	1
ST-GP	50.57	48.55	**50.59**	121.52	59.12	55.17	61.46	118.27
SE-GP	50.48	50.71	74.03	116.12	93.36	87.83	95.70	111.79
SV-GP	56.50	48.28	60.53	305.03	55.78	54.76	**56.65**	163.08
SR-GP-1	63.27	61.90	69.74	112.43	61.48	55.34	61.86	165.27
SR-GP-2	50.45	49.82	50.52	180.97	61.66	61.56	62.13	193.46
A3-GP	61.07	54.18	60.83	292.83	63.57	59.92	63.18	297.44
U2-GP	68.76	55.91	68.47	194.55	64.92	67.77	64.87	194.64
U1-GP	695.33	560.52	2965.05	10^4	909.17	700.43	3003.07	10^4

For the 2-D Black Scholes case study w/ $N = 10000$: $SD(\hat{R}_K^{HD})$: sample standard deviation (SD) over 100 macro-replications [smaller is better], $\bar{s}$: average GP posterior standard deviation of $\hat{R}_K$ [should be close to SD], RMSE of $\hat{R}_K$ (vis-a-vis exact quantile) [smaller is better], $|\mathcal{D}_K|$: average final design size (100 pilot scenarios) [smaller is faster].

Results for 2-D Example



Var/TVaR estimation for the 2-D Black Scholes case study. Left boxplots display the distribution of the final $\hat{R}_K^{\mathrm{TVaR}}$ estimates; on the right is corresponding GP standard deviation $s(\hat{R}_K^{\mathrm{TVaR}})$. Results are based on 100 macro-replications for each approach.

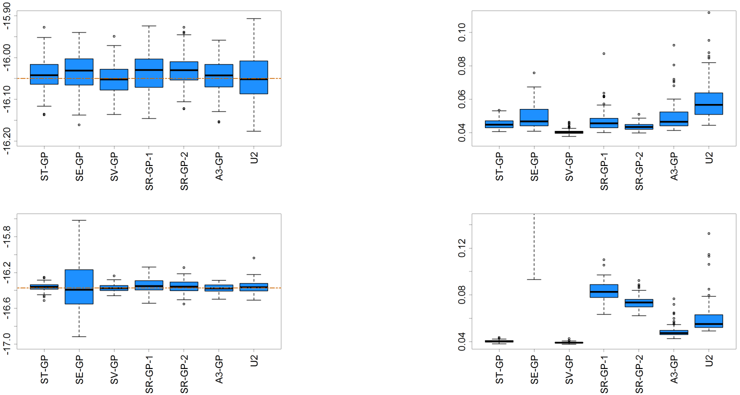
Life Annuity Case Study

Six-dimensional example valuing annuity portfolios: 3 factor longevity model M7 (APC fitted StMoMo), 3 factor interest rates (SIR with SV+stoch $\bar{r}_t$). $N = \mathcal{N} = 10^5$.

Estimated $\hat{R}_K$ | GP posterior uncertainty $s(\hat{R}_K)$

$\mathrm{VaR}_{0.005}$

$\mathrm{TVaR}_{0.005}$



Results for 6D

	$\text{VaR}_{0.005}$				$\text{TVaR}_{0.005}$				
	$SD(\hat{R}_K^{HD})$	$\bar{s}$	RMSE	$\lvert\mathcal{D}_K\rvert$	$SD(\hat{R}_K)$	$\bar{s}$	RMSE	$\lvert\mathcal{D}_K\rvert$	Time
ST-GP	0.0394	0.0455	0.0403	151.83	0.0461	0.0404	0.0472	147.10	330
SE-GP	0.0427	0.0493	0.0459	143.27	0.2853	0.2717	0.2850	101.62	295
SV-GP	0.0382	0.0406	0.0380	497.81	0.0408	0.0393	0.0407	254.35	403
SR-GP-1	0.0467	0.0470	0.0485	135.03	0.0430	0.0402	0.0430	184.73	219
SR-GP-2	0.0391	0.0437	0.0434	217.36	0.0447	0.0431	0.0450	224.96	198
A3-GP	0.0434	0.0497	0.0436	298.15	0.0464	0.0490	0.0461	298.55	115
U2-GP	0.0598	0.0598	0.0596	194.03	0.0684	0.0601	0.0689	195.81	112
U1-GP	0.5020	0.4156	0.5853	10^4	0.4940	0.4705	0.6709	10^4	177

Results for the 6-D life annuity case study based on 100 macro-replications. We report sample standard deviation of $\hat{R}_K^{[1:100]}$, average GP posterior standard deviation $\bar{s}$, and RMSE of $\hat{R}_K$, as well as average final design size for each approach.

Conclusions

- New links between machine learning/emulation tools and risk measurement
- Important gains thanks to sequential design + advanced emulator (`hetGP`)
- Tried a variety of acquisition functions, still more work to be done
- Future: make N also adaptive

Spatial model gives a variance reduction of $x2-5$

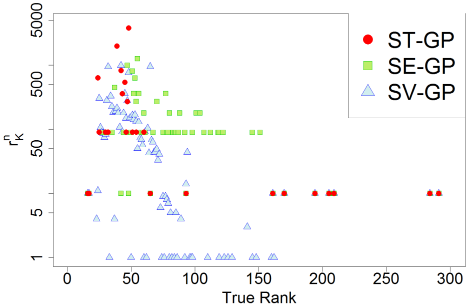
Adaptive allocation gives a speed-up of $x10-40$

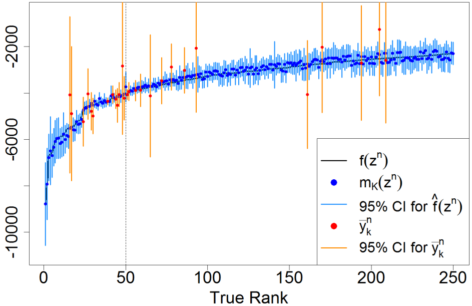
References

- Bect, J., Ginsbourger, D., Li, L., Picheny, V., Vazquez, E., 2012. Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing 22 (3), 773–793.
- Liu, M., Staum, J., 2010. Stochastic kriging for efficient nested simulation of expected shortfall. Journal of Risk 12 (3), 3.
- Binois, M., Gramacy, R. B., Ludkovski, M., 2018. Practical heteroskedastic gaussian process modeling for large simulation experiments. Journal of Computational and Graphical Statistics, to Appear. arXiv preprint arXiv:1611.05902.
- Risk, J., Ludkovski, M., 2016. Statistical emulators for pricing and hedging longevity risk products. Insurance: Mathematics and Economics 68, 45–60.
- Risk, J., Ludkovski, M., 2017. Sequential Design and Spatial Modeling for Portfolio Tail Risk Measurement, Submitted. arxiv.org/abs/1710.05204

Thank You!

Spatial Modeling Gains via Adaptive Replication





Right: Replication counts r_K^n versus true rank of $f^{1:N}$ after the final stage for sequential methods for learning VaR in the 2-D Black-Scholes case study. Left: estimated $\hat{f}(z^n)$ vs true $f(z^n)$.