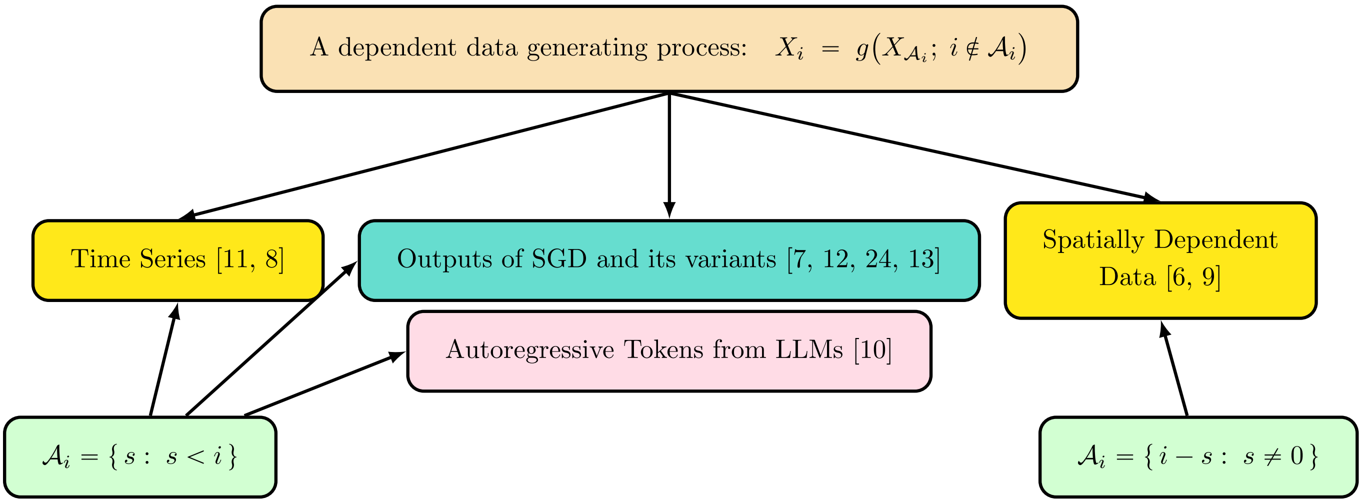
Research Statement

Soham Bonnerjee
Department of Statistics, University of Chicago
sohambonnerjee@uchicago.edu

My core research interests lie at the intersection of time-series and machine learning, with an emphasis on tackling complex non-stationarity arising in various modern applications. I develop valid statistical tools to perform statistical inference for dependent datasets, which have been increasingly commonplace as outputs of iterative optimization algorithms such as Stochastic Gradient Descent (SGD), autoregressive next token generation by large language models (LLM), as well as more classical temporal and spatial datasets from disciplines such as epidemiology, geography, climate science and archaeology. Interpreting the dependent datasets through a distinctly time-series lens while leveraging their unique properties, my theoretical excursions often go beyond central limit theory to obtain *sharp Gaussian approximations* with finite-sample guarantees, facilitating valid bootstrap procedures, construction of confidence bands, and hypothesis tests in non-i.i.d. settings. So far, a majority of my research can be categorized into three intertwined directions.



- **Gaussian approximation and its applications for time series.** I develop strong invariance principles (also called *Komlos-Major-Tusnady* approximations) for multivariate stationary and non-stationary processes [11], augmented with explicit construction of said approximations, along with practically usable valid bootstrap strategies. These approximations have numerous applications such as in change-point literature [8, 40], simultaneous inference [8, 56], and conditional independence testing[53]. In particular, we employ such approximations to solve the novel problem of testing synchronization of change-points [8].

- **Inference for spatially dependent data.** I develop Gaussian-approximation based valid inference strategy for spatially dependent datasets. Most of the literature in this area employs Bayesian, Gaussian process-based methodology to capture dependence. Instead, we employ a notion of spatial dependence that generalizes naturally, but does not require the notion of temporal ordering. Deriving valid Gaussian approximation under such general conditions helps in valid inference for *spatial random effect* models. In an ongoing work, I have also come up with a notion of *spatial change-point*, and have introduced fast, accurate algorithms to localize such change-points.

- **Inference for dependent datasets in modern machine learning.** I develop distributional theory such as *stable* convergence, *Berry-Esseen* bounds or time-uniform couplings for the iterates of various stochastic approximation algorithms (SA) such as *Stochastic Gradient Descent* (SGD), *Q-learning* and *Federated Learning*. Concurrently, on the statistical theory of large language models (LLM), I have leveraged comparatively obscure change-point tools to perform provably consistent watermark detection in mixed-source (containing both human and LLM generated) texts. The dependence prevalent in these datasets (such as iterates of SA algorithms, or words of a text) cannot be easily classified into any class of "weak" or "strong" dependence vis-à-vis classical time series literature, and usually requires a case-by-case treatment [12, 13, 10, 7]. This line of work can be characterized as a happy melting pot where tools of classical statistics, often much less-used and/or forgotten, can prove to be useful to solve pertinent, modern problems in machine learning.

In parallel to my core interests as described above, I have worked on **(i)** *privacy-aware* training of Transformers using novel variants of *Differentially-private SGD* (DP-SGD) [14] and **(ii)** stable convergence of SGD for non-convex objectives [7]. In the following, I describe my main completed/ongoing research projects in detail, modularized by their respective themes.

Inference for time series: theory and application

Sharp Gaussian approximations for non-stationary time series with explicit construction.

In this project, joint with Sayar Karmakar and Wei Biao Wu [11] and published in *Annals of Statistics*, we tackle the problem of *strong invariance principle* in non-stationary time series. Concretely, suppose $X_1, \ldots, X_n$ be a mean-zero non-stationary time series. Previous results indicated that, with a mild condition on its dependence there exists a Gaussian process G_t that can uniformly approximate partial sums $S_t = \sum_{s=1}^{t} X_s$ while maintaining an optimal rate. However, no concrete structure about such processes were known, preventing their use in statistical inference. In this article, we present an explicit construction of a Gaussian coupling for non-stationary time-series by showing that there exists a Brownian motion $\mathbb{B}$ such that

$$\max_{1 \leqslant i \leqslant n} |S_i - \mathbb{B}(\mathbb{E}[S_i^2])| = o_{\mathbb{P}}(n^{1/p}),$$

where we assume that $\sup_i \mathbb{E}[|X_i|^p] < \infty$. Moreover, going beyond Brownian motion-based approximation, we provide the *first ever* construction of a *covariance-matching* Gaussian coupling; there exists a Gaussian process $\{Y_t\}_{t \geqslant 1}$ with $\mathrm{Cov}(Y_s, Y_t) = \mathrm{Cov}(X_s, X_t)$ such that

$$\max_{1 \leqslant i \leqslant n} |S_i - \sum_{j=1}^{i} Y_j| = o_{\mathbb{P}}(n^{1/p}).$$

Additionally, we propose a uniformly consistent estimation strategy of the said covariance structure, thereby facilitating Gaussian bootstrap-based inference. This paper has applications in change-point analysis, constructing simultaneous confidence band and wavelet analysis. In particular, our methods back up recent questions against long-held assumptions regarding a popular archaeological dataset [3, 18, 58, 20, 49], which is highly non-stationary in nature.

Testing for synchronization of change-points in multiple time-series.

As a non-trivial use-case of the strong invariance principle for practically relevant problems, we tackle the problem of testing for synchronization of multiple change-points in this joint work with Sayar Karmakar, Maggie Cheng and Wei Biao Wu [8]. Consider a multivariate time-series

$$\mathbf{X}_i = \boldsymbol{\mu}_i + \mathbf{e}_i = (\mu_{i,1}, \ldots, \mu_{i,d})^T + \mathbf{e}_i,\ i = 1, \ldots, n,\ \mathbf{X}_i = (X_{ij})_{j=1}^{d},$$

where $\mathbf{e}_i \in \mathbb{R}^d$ is a mean-zero stationary time series. Assume that for each $j \in [d]$, $\boldsymbol{X}_{\cdot j}$ has a potential change-point, namely,

$$\mu_{ij} = \begin{cases} \mu_j^L, & \text{if } i/n \leqslant \tau_j, \\ \mu_j^R, & \text{if } i/n > \tau_j \end{cases},\ 1 \leqslant i \leqslant n,$$

where $\tau_j \in (0,1)$ is the (re-scaled) change-point. Cross-sectional dependence in $\mathbf{X}_i$'s might lead to shared or clustered change-points in different covariates; however, usual literature on change-point analysis seems to assume a shared change-point for the most part. In this work, we describe a valid, Gaussian-bootstrap based algorithm to test such hypotheses of shared change-points $H_0 : \tau_1 = \tau_2 = \ldots = \tau_d$. We propose (i) a valid test statistic to perform this composite-vs.-composite test, (ii) establish the asymptotic validity as well as uniform asymptotic power of a bootstrap-procedure to estimate the null distribution of this test statistic. Along the way, we also introduce a change-point-agnostic algorithm to estimate the long run covariance structure of $\mathbf{X}_i$'s.

This work has received a *Major Revision* request from *Biometrika*. It has also received a *Hannan Graduate Student Travel Award* from the *Institute of Mathematical Statistics*(IMS).

Inference for spatially dependent data

Inference for spatial random effect model

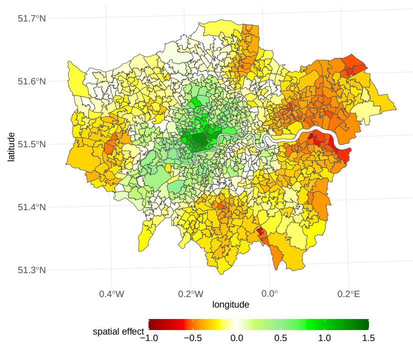


Figure 1: Distribution of spatial effect on London housing price dataset

In spatial data analysis, the spatial dependency is often left unaddressed, or tackled by putting a Gaussian prior with some covariance structure eg. Matérn or squared-exponential kernels. Such assumptions are often un-testable, necessitating a more general treatment. In a joint work with Soudeep Deb and Wei Biao Wu [6], we consider inference on spatial random effect model $Y_{ij} = X_{ij}^\top \beta + U_i + \varepsilon_{ij}$, where (Y_{ij}, X_{ij}) are the observed response-covariate pairs and ε_{ij} are the i.i.d. random errors. Here, the dependency structure of the spatial effects $(U_i)_{i \in \mathbb{Z}^d}$ is simply characterized by $U_i = g(e_{i-s} : s \in \mathbb{Z}^d)$, where $e_i, i \in \mathbb{Z}^d$ are i.i.d. innovations. This characterization is quite general and does not require any form of partial ordering; rather, it arises naturally out of writing out the joint distribution of $(U_k)_{k \in \mathbb{Z}^d}$ in terms of compositions of conditional quantile functions of i.i.d. uniform random variables. Then, under mild assumptions on the covariates X_{ij}, we establish a central limit theory for the least square estimate $\hat{\beta}$ of β, and then proceed to provide a consistent estimate of the corresponding asymptotic variance. Our estimates apart from being theoretically valid, can also be computed efficiently using Fast Fourier Transforms. These results motivate valid inferential procedures to test for presence of spatial effects as well as presence of spatial correlations, which we implement to investigate *London housing price dataset.* Our results indicate a considerable spatial effect, as can also be seen in Figure 1. The preprint will be online soon, and we plan to submit it to *Biometrika*.

Localization of spatial change-points

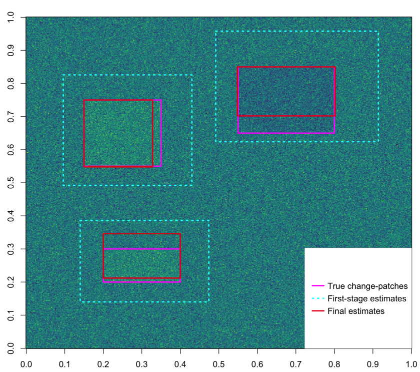
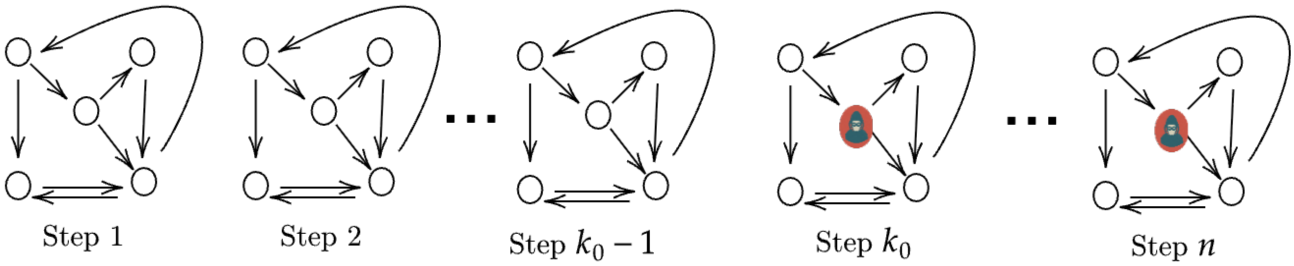


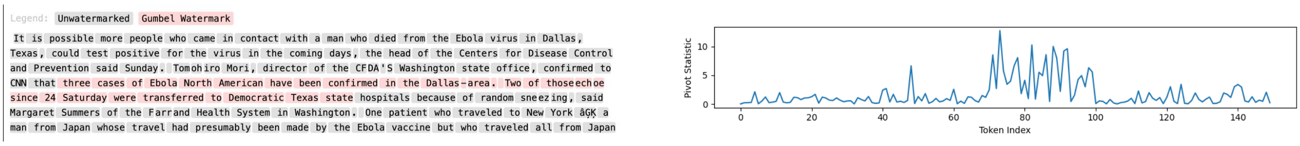
Figure 2: Spatial change-patches localization via two-stage algorithm

Despite a huge literature on change-points in time-series, research on change-points on spatially dependent datasets is almost non-existent. In an *ongoing* work with Sayar Karmakar and George Michailidis [9], we deviate from the scan statistics literature and their Gaussian/i.i.d. assumptions and generalize the notion of epidemic change-points to define spatial *change-patches*. Subsequently, we propose a theoretically valid, two-stage algorithm to localize all rectangular change-patches in a spatially dependent dataset. The first stage produces the efficiency by obtaining a coarse estimate of the patches, before a further fine-tuning produces final estimate (Figure 2). The emphasis on speed is important, since traditional algorithms [23, 17, 51, 52, 30] are prohibitively slow on large spatial datasets. Apart from rigorous theoretical results, our simulation studies have established the efficacy of our method.

Inference for modern machine learning algorithms



(a)



(b)

Figure 3: (a): In a distributed learning system, client(s) may turn malicious at some step; identification of this step as well as the client, are change-point problems [12]. (b) In a text possibly contaminated by LLM-generated watermarked outputs, detecting the watermarked segments is an epidemic change-point problem.[10]

Sharp asymptotic inference for Decentralized Federated Learning (DFL)

In decentralized federated learning systems, multiple clients collaborate periodically through a connection graph $\mathbf{C}$ to perform an optimization problem through *local SGD* algorithm. Concretely, suppose there are K clients collaborating to solve the optimization problem $\theta^\star_K = \arg\min_\theta \sum_{k=1}^K w_k F_k(\theta) \in \mathbb{R}^d$. However, a typical stochastic gradient descent cannot be implemented because (i) combining local gradients at each step is computationally expensive, since K is typically large, and (ii) Even if gradients can be combined, clients may choose to share them only with certain peers, using a connection graph $\mathbf{C}$. Then, one follows the iterative algorithm `local SGD`:

$$\mathbf{\Theta}_t = (\mathbf{\Theta}_{t-1} - \eta_t \mathbf{G}_t)\, C_t, \quad C_t = \begin{cases} \mathbf{C}, & t \in E_\tau, \\ I_K, & \text{otherwise.} \end{cases},$$

where $\mathbf{\Theta}_t = (\theta_t^1, \ldots, \theta_t^K) \in \mathbb{R}^{d \times K}$ denotes the local parameter updates of each client at the t-th step, $\mathbf{G}_t$ denotes the corresponding local gradient updates and η_t are the step-sizes. The final estimate is $Y_n := K^{-1}\mathbf{\Theta}_n \mathbf{1}$. In a joint article [12] with Sayar Karmakar and Wei Biao Wu, we address

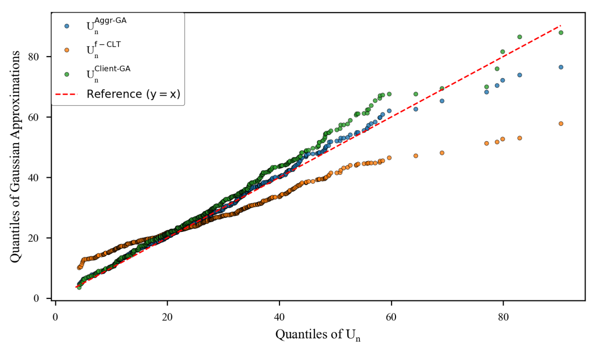


Figure 4: $U_n^{\texttt{Aggr-GA}}$ (blue) and $U_n^{\texttt{Client-GA}}$ (green) represent proposed Gaussian approximations; $U_n^{\texttt{f-CLT}}$ (orange) is the Brownian motion-based approximation. The quantiles of these approximations are plotted against theoretical quantiles based on `local SGD` iterates.

two key theoretical gaps. Firstly, going beyond
central limit theory [25], we derive the *first-ever* Berry-Esseen theorem for Y_n, which also sheds light into optimal choice of the learning rate η_t. Secondly, mirroring the KMT-approximations, we present two, first-of-its-kind *time-uniform* and rate-optimal Gaussian approximations for the local SGD iterates $\mathbf{\Theta}_t$ and Y_t. In particular, our Gaussian approximations are much more accurate (Figure 4 compared to off-the-shelf Brownian-motion approximations provided by functional central limit theory (f-CLT).

We also discuss how these results motivate valid Gaussian bootstrap-based algorithms to identify the onset of adversarial attacks such as Man-In-the-Middle (see Figure 3 (a)). This has been accepted as a *Spotlight* poster (Top 3%) in *NeurIPS 2025*.

Segmentation of watermarked texts

With the advent of Large Language Models, detection of machine-generated texts have become important. A common method uses *Watermarking*, which makes the use of LLM detectable *only* when keys corresponding to each token are present, and un-detectable otherwise. Recent theoretical work in this area has mostly focused on the testing problem of unwatermarked vs. watermarked texts. In a joint research [10] with Sayar Karmakar and Subhrajyoty Roy, we tackle the more difficult problem of identifying such watermarked segments from a mixed-source texts by introducing a novel perspective of *epidemic change-point*. Concretely, given a text $\omega_1 \dots \omega_n$, an administrator will have access to keys $\zeta_1, \dots, \zeta_n$ with the property that if ω_i is human-generated, it is independent of ζ_i; otherwise, it will be correlated with ζ_i conditional on past tokens. This motivates construction of pivot statistics Y_t, whose null distribution is known when ω_t is un-watermarked or human generated; in contrast, if ω_t is watermarked, we expect Y_t to be large. See Figure 3 (b) for an illustration based on a LLM-generated text.

Adapting key ideas from (i) comparatively obscure *epidemic change-point* literature [57, 32, 27, 16] and (ii) recent development in irregular change-point estimation [29], to the novel setting of watermarked texts, we propose `WISER`: a provably consistent and computationally efficient algorithm to identify multiple watermarked segments from a text input. Apart from rigorous theoretical guarantees, our algorithm is shown to out-perform all competitive algorithms-most of which lack theoretical guarantees or are computationally expensive- on benchmark datasets for various watermarking schemes. This work is submitted to *ICLR 2026*.

Sharp asymptotic theory for Q-learning with `LD2Z` learning rate and its generalization

Despite a sustained popularity of *Q-learning* as an online policy learning algorithm, an oft-ignored aspect is choice of step-size in Q-learning algorithm. Motivated by the classical SGD literature [44, 42], the polynomially decaying learning rate $\eta_t = \eta t^{-\alpha}, \alpha \in (0,1)$ is often used in theoretical results leading to inferential procedures. However its theoretical optimality often masks its excruciatingly slow convergence, as also observed by [59]. These criticisms have been echoed by the broad stochastic optimization community, leading to a recent proposal of linearly decaying to zero (`LD2Z`) learning rate $\eta_{t,n} = \eta(1 - t/n)$ [19, 48]. Its empirical success has been well-studied in applications characterized by highly non-smooth or complex optimization landscapes, including state-space models [48], large language models [19, 39, 5], and vision transformers [54]. In contrast, theoretical literature for this learning schedule is quite sparse. To address this, in a joint work [13] with Zhipeng Lou and Wei Biao Wu, we characterize the first-ever theoretical treatment of `LD2Z` learning rate for Q-learning. In particular we establish a step-by-step theory including **(i)** $\mathcal{L}_2$ error bound, **(ii)** central limit theory, and finally, **(iii)** a time-uniform strong invariance principle for Q-learning iterates $\boldsymbol{Q}_t$'s. Of particular interest is the central limit theory; usual literature [34, 37, 33] focus on Polyak-Ruppert (PR) averaged versions of the Q-learning iterates, $\bar{\boldsymbol{Q}}_n := n^{-1}\sum_{t=1}^n \boldsymbol{Q}_t$. Instead, we show that for the `LD2Z` learning rate, $\bar{\boldsymbol{Q}}_n$ is in fact sub-optimal, and the central limit theory can be established only for a *tail Polyak-*

Ruppert averaged version, $\tilde{\boldsymbol{Q}}_n = (n - s_n)^{-1} \sum_{t=s_n}^{n} \boldsymbol{Q}_t$ for some particular $s_n < n$.

In the following, I briefly describe my future research interests.

Future research directions

Change-point estimation on Random Objects

Time-series data taking values in metric spaces, which we refer to as *random objects*, are increasingly common in real-world applications, such as graph Laplacians, covariance matrices, probability distributions, and compositional vectors with examples in various domains like brain imaging, social networks, income histograms, microbiome data, and genetics [1, 41, 46, 21, 45, 50, 4, 47]. However, most of the statistical literature have investigated change-point problems in object data either by assuming i.i.d. observations, or assuming an embedding onto Hilbert space or some tangent space. We aim to investigate

(i) the change-point localization problem as well as multiple change-point detection, and,

(ii) change-point-agnostic long-run variance estimation, in a general dependent framework,

by potentially generalizing *functional dependence measure*[55] and establishing guarantees on object-versions of *Wild Binary Segmentation* [22] or *Seeded Binary Segmentation* [31].

Statistical foundations on Watermarking in LLMs

Building on [10], we aim to investigate several pertinent problems in watermarking. For example,

(i) as opposed to the usual offline tests [35, 36], it is relevant to propose a valid watermark detection-scheme that parses a text sequentially, stopping whenever it has determined to have encountered watermarked segments.

(ii) On the other hand, in realistic scenarios it might not be known if a specific watermarking scheme is the only candidate to have been potentially used. In these scenarios, it is useful to employ the concept of *e-values* [43] to combine different watermark-detection tests on the same text input.

(iii) Moreover, in most of the literature, one usually considers the mean of the pivot statistics for a testing procedure; it is conceivable that using a test statistic that captures the change in the entire distribution of the pivot statistic for watermarked tokens (through *reproducing kernel Hilbert space (RKHS)* or *Kernel MMD tests*) might lead to increase in power for many watermarking schemes. In fact, our preliminary simulation experiments show this to be indeed true for inverse watermarking schemes.

Analysis of stochastic optimization algorithms

Continuing my research on stochastic optimization algorithms, I aim to work on the following two immediate problems.

(i) Recently, [15, 38] explored two-stage stochastic algorithms for instrumental variable regression. While assuming fixed instruments, these methods contribute substantial improvements over usual conditional stochastic optimization or Generalized Method of Moments-based approaches. However, empirically it can be observed that these algorithms perform very poorly in presence of weak instruments.

I aim to develop valid, efficient optimization algorithms to address this issue, and perform inference with them.

(ii) Building on my work on decentralized federated learning [12], I am interested in theoretical analysis of *differentially private* `local SGD` algorithms, which are of immense interest in federated learning literature. In particular, we aim to develop valid Gaussian approximations in presence of injected noise to make `local SGD` differentially private.

Chain-of-thought learning and dependence

Chain-of-thought (CoT) supervision, which achieves the final outputs with intermediate reasoning steps, has emerged as a powerful empirical technique driving much of the recent progress in large language models' reasoning abilities. Recently, the statistical literature has focused much on its generalization error [26, 28, 2]. In contrast, research on the effect of length of the reasoning chain has been sparse, and it remains unclear whether longer chains, while being inconvenient, will guarantee increased accuracy. By leveraging tools from time-series literature on the autoregressive chains of CoT-learning, I aim to explore the sweet spot between accuracy and chain length-if it exists.

Some other potential research directions include change-point analysis for *high-dimensional logistic regression*, capturing cliques or clusters in a spatio-temporal setting, and application of social choice theory in preference modeling and *Reinforcement Learning from Human Feedback* (RLHF).

References

[1] A. Alexander-Bloch, J. N. Giedd, and E. Bullmore. Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*, 14(5):322–336, 2013.

[2] A. Altabaa, O. Montasser, and J. Lafferty. Cot information: Improved sample complexity under chain-of-thought supervision. *arXiv preprint arXiv:2505.15927*, 2025.

[3] A. P. Andrews, E. W. Andrews, and F. R. Castellanos. The northern maya collapse and its aftermath. *Ancient Mesoamerica*, 14(1):151–156, 2003.

[4] N. A. Asif, Y. Sarker, R. K. Chakrabortty, M. J. Ryan, M. H. Ahamed, D. K. Saha, F. R. Badal, S. K. Das, M. F. Ali, S. I. Moyeen, et al. Graph neural network: A comprehensive review on non-euclidean space. *Ieee Access*, 9:60588–60606, 2021.

[5] S. Bergsma, N. S. Dey, G. Gosal, G. Gray, D. Soboleva, and J. Hestness. Straight to zero: Why linearly decaying the learning rate to zero works best for llms. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025.

[6] S. Bonnerjee, S. Deb, and W. B. Wu. Inference for spatial random effects model under dependence. *In preparation*, 2025.

[7] S. Bonnerjee, Y. Han, and W. B. Wu. Stable convergence of stochastic gradient descent for non-convex objectives. *Preprint*, 2025.

[8] S. Bonnerjee, S. Karmakar, M. Cheng, and W. B. Wu. Testing synchronization of change-points for multiple time series. *Major Revision from Biometrika*, 2025.

[9] S. Bonnerjee, S. Karmakar, and G. Michailidis. Fast detection of anomalous patches for spatial data. *In preparation*, 2025.

[10] S. Bonnerjee, S. Karmakar, and S. Roy. Wiser: Segmenting watermarked region - an epidemic change-point perspective. *arXiv preprint arXiv:2509.21160*, 2025.

[11] S. Bonnerjee, S. Karmakar, and W. B. Wu. Gaussian approximation for nonstationary time series with optimal rate and explicit construction. *Ann. Statist.*, 52(5):2293–2317, 2024.

[12] S. Bonnerjee, S. Karmakar, and W. B. Wu. Sharp gaussian approximations for decentralized federated learning. *arXiv preprint arXiv:2505.08125; NeurIPS 2025, Spotlight*, 2025.

[13] S. Bonnerjee, Z. Lou, and W. B. Wu. Sharp asymptotic theory for q-learning with `LD2Z` learning rate and its generalization. *arXiv preprint arXiv:2505.08125; NeurIPS 2025, Spotlight*, 2025.

[14] S. Bonnerjee, Z. Wei, A. Asch, S. Nandy, P. Ghosal, et al. How private is your attention? bridging privacy with in-context learning. *arXiv preprint arXiv:2504.16000*, 2025.

[15] X. Chen, A. Roy, Y. Hu, and K. Balasubramanian. Stochastic optimization algorithms for instrumental variable regression with streaming data. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[16] Z. Chen, Z. Li, and M. Zhou. Detecting change-points in epidemic models. *Journal of advanced statistics*, 1(4):181, 2016.

[17] P. Datta and B. Sen. Optimal inference with a multidimensional multiscale statistic. *Electron. J. Stat.*, 15(2):5203–5244, 2021.

[18] A. Demarest. *Ancient Maya: The Rise and Fall of a Rainforest Civilization*. Case Studies in Early Societies. Cambridge University Press, 2004.

[19] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics, 2019.

[20] B. B. Faust. Maya environmental successes and failures in the yucatan peninsula. *Environmental Science & Policy*, 4(4):153–169, 2001.

[21] L. K. Ferreira and G. F. Busatto. Resting-state functional connectivity in normal brain aging. *Neuroscience & Biobehavioral Reviews*, 37(3):384–400, 2013.

[22] P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014.

[23] J. Glaz and Z. Zhang. Multiple window discrete scan statistics. *J. Appl. Stat.*, 31(8):967–980, 2004.

[24] O. Goldreich, Z. Wei, S. Bonnerjee, J. Li, and W. B. Wu. Asymptotic theory of sgd with a general learning-rate. *NeurIPS 2025, Poster*, 2025.

[25] J. Gu and S. X. Chen. Statistical inference for decentralized federated learning. *Ann. Statist.*, 52(6):2931–2955, 2024.

[26] X. Hu, F. Zhang, S. Chen, and Z. Yang. Unveiling the statistical foundations of chain-of-thought prompting methods. *arXiv preprint arXiv:2408.14511*, 2024.

[27] M. Hušková and A. Slabý. Testing for an epidemic change in mean. *Commentationes Mathematicae Universitatis Carolinae*, 36(4):737–747, 1995.

[28] N. Joshi, G. Vardi, A. Block, S. Goel, Z. Li, T. Misiakiewicz, and N. Srebro. A theory of learning with autoregressive chain of thought. *arXiv preprint arXiv:2503.07932*, 2025.

[29] T. Kley, Y. P. Liu, H. Cao, and W. B. Wu. Change-point analysis with irregular signals. *The Annals of Statistics*, 52(6):2913–2930, 2024.

[30] C. König, A. Munk, and F. Werner. Multidimensional multiscale scanning in exponential families: limit theory and statistical consequences. *Ann. Statist.*, 48(2):655–678, 2020.

[31] S. Kovács, P. Bühlmann, H. Li, and A. Munk. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023.

[32] B. Levin and J. Kline. The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Statistics in Medicine*, 4(4):469–488, 1985.

[33] G. Li, C. Cai, Y. Chen, Y. Wei, and Y. Chi. Is q-learning minimax optimal? a tight sample complexity analysis. *Operations Research*, 72(1):222–236, 2024.

[34] X. Li, J. Liang, and Z. Zhang. Online statistical inference for nonlinear stochastic approximation with markovian data. *arXiv preprint arXiv:2302.07690*, 2023.

[35] X. Li, F. Ruan, H. Wang, Q. Long, and W. J. Su. A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules. *The Annals of Statistics*, 53(1):322–351, 2025.

[36] X. Li, G. Wen, W. He, J. Wu, Q. Long, and W. J. Su. Optimal estimation of watermark proportions in hybrid ai-human texts. *arXiv preprint arXiv:2506.22343*, 2025.

[37] X. Li, W. Yang, J. Liang, Z. Zhang, and M. I. Jordan. A statistical analysis of polyak-ruppert averaged q-learning. In *International Conference on Artificial Intelligence and Statistics*, pages 2207–2261. PMLR, 2023.

[38] H. Liang, Y. Jin, K. Balasubramanian, and L. Lai. Differentially private two-stage gradient descent for instrumental variable regression. *arXiv preprint arXiv:2509.22794*, 2025.

[39] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[40] F. Mies. Strong gaussian approximations with random multipliers. *arXiv preprint arXiv:2412.14346*, 2024.

[41] Y. Nie, L. Wang, and J. Cao. Estimating time-varying directed gene regulation networks. *Biometrics*, 73(4):1231–1242, 2017.

[42] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992.

[43] A. Ramdas and R. Wang. Hypothesis testing with e-values. *arXiv preprint arXiv:2410.23614*, 2024.

[44] D. Ruppert. Efficient estimations from a slowly convergent robbins-monro process. *Technical Report*, 1988.

[45] R. Sala-Llonch, D. Bartrés-Faz, and C. Junqué. Reorganization of brain networks in aging: a review of functional connectivity studies. *Frontiers in psychology*, 6:663, 2015.

[46] J. Scealy and A. Welsh. Colours and cocktails: Compositional data analysis 2013 lancaster lecture. *Australian & New Zealand Journal of Statistics*, 56(2):145–169, 2014.

[47] T. A. Snijders. Models for longitudinal network data. *Models and methods in social network analysis*, 1:215–247, 2005.

[48] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[49] B. L. Turner and J. A. Sabloff. Classic period collapse of the central maya lowlands: Insights about human–environment relationships for sustainability. *Proceedings of the National Academy of Sciences*, 109(35):13908–13914, 2012.

[50] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov. Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, 68(2):261–275, 2009.

[51] G. Walther. Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.*, 38(2):1010–1033, 2010.

[52] G. Walther and A. Perry. Calibrating the scan statistic: finite sample performance versus asymptotics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 84(5):1608–1639, 2022.

[53] M. Wieck-Sosa, M. F. Haddad, and A. Ramdas. Conditional independence testing with a single realization of a multivariate nonstationary nonlinear time series. *arXiv preprint arXiv:2504.21647*, 2025.

[54] S. Wu, G. Zhang, and X. Liu. SwinSOD: Salient object detection using swin-transformer. *Image Vis. Comput.*, 146(105039):105039, June 2024.

[55] W. B. Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154, 2005.

[56] W. B. Wu and Z. Zhao. Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):391–410, 2007.

[57] Q. Yao. Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):179–191, 1993.

[58] N. Yoffee and G. Cowgill. *The Collapse of Ancient States and Civilizations*. Book collections on Project MUSE. University of Arizona Press, 1991.

[59] Y. Zhang and Q. Xie. Constant stepsize q-learning: Distributional convergence, bias and extrapolation. *arXiv preprint arXiv:2401.13884*, 2024.