# Anytime-Valid Tests of Conditional Independence under Model-X

Alexander Henzi (ETH),

Joint work with Peter Grünwald and Tyron Lardy,
CWI, Amsterdam, The Netherlands

ETH-UCPH-TUM Workshop on Graphical Models

October 11, 2022

# Anytime-valid inference

It would be nice to have statistical tests which...

- ...allow to stop an experiment early as soon as there is enough evidence against the null hypothesis,

# Anytime-valid inference

It would be nice to have statistical tests which...

- ...allow to stop an experiment early as soon as there is enough evidence against the null hypothesis,

  $\rightarrow$ optional stopping

# Anytime-valid inference

It would be nice to have statistical tests which...

- ...allow to stop an experiment early as soon as there is enough evidence against the null hypothesis,
  $\rightarrow$ optional stopping
- ...allow to continue accumulating evidence if the results so far are inconclusive,
  $\rightarrow$ optional continuation

# Anytime-valid inference

It would be nice to have statistical tests which...

- ...allow to stop an experiment early as soon as there is enough evidence against the null hypothesis,

  $\rightarrow$ optional stopping
- ...allow to continue accumulating evidence if the results so far are inconclusive,

  $\rightarrow$ optional continuation
- ...are valid even if the sample size is dependent on the data in some way.

# Anytime-valid inference

It would be nice to have statistical tests which...

- ...allow to stop an experiment early as soon as there is enough evidence against the null hypothesis,
  → optional stopping
- ...allow to continue accumulating evidence if the results so far are inconclusive,
  → optional continuation
- ...are valid even if the sample size is dependent on the data in some way.
  → anytime-valid inference

Peter Grünwald, Rianne de Heide, and Wouter Koolen (2022+). "Safe Testing". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. To appear. Preprint: arXiv:1906.07801

Aaditya Ramdas et al. (2020). *Admissible Anytime-Valid Sequential Inference Must Rely on Nonnegative Martingales*. arXiv: 2009.03167

Vladimir Vovk and Ruodu Wang (2021). "E-values: Calibration, Combination and Applications". In: *The Annals of Statistics* 49, pp. 1736 –1754

Glenn Shafer (2021). "Testing by Betting: A Strategy for Statistical and Scientific Communication". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184, pp. 407–431

Aaditya Ramdas et al. (2022). *Game-theoretic statistics and safe anytime-valid inference*. arXiv: 2210.01948

Let $(D_n)_{n \in \mathbb{N}}$ be a data stream and $D^n = (D_1, \ldots, D_n)$.

$\mathcal{H}_0$ is a set of distributions for $(D_n)_{n \in \mathbb{N}}$, the null hypothesis.

- In this talk: $D_i = (X_i, Y_i, Z_i)$ i.i.d. and $\mathcal{H}_0 = \{P \colon X \overset{P}{\perp\!\!\!\perp} Y \mid Z\}$

Let $(D_n)_{n\in\mathbb{N}}$ be a data stream and $D^n = (D_1, \dots, D_n)$.

$\mathcal{H}_0$ is a set of distributions for $(D_n)_{n\in\mathbb{N}}$, the null hypothesis.

- In this talk: $D_i = (X_i, Y_i, Z_i)$ i.i.d. and $\mathcal{H}_0 = \{P\colon X \overset{P}{\perp\!\!\!\perp} Y \mid Z\}$

A process $S_n = S_n(D^n)$, $n \in \mathbb{N}$, is called e-process if

$$S_n \geq 0,\ n \in \mathbb{N}, \quad \mathbb{E}_P[S_\tau] \leq 1 \text{ for all stopping times } \tau,\ P \in \mathcal{H}_0.$$

Let $(D_n)_{n \in \mathbb{N}}$ be a data stream and $D^n = (D_1, \dots, D_n)$.

$\mathcal{H}_0$ is a set of distributions for $(D_n)_{n \in \mathbb{N}}$, the null hypothesis.

- In this talk: $D_i = (X_i, Y_i, Z_i)$ i.i.d. and $\mathcal{H}_0 = \{P \colon X \overset{P}{\perp\!\!\!\perp} Y \mid Z\}$

A process $S_n = S_n(D^n)$, $n \in \mathbb{N}$, is called e-process if

$$S_n \geq 0, \; n \in \mathbb{N}, \quad \mathbb{E}_P[S_\tau] \leq 1 \text{ for all stopping times } \tau, \; P \in \mathcal{H}_0.$$

Implications:

- $P(\exists n \colon S_n \geq 1/\alpha) \leq \alpha$ for all $P \in \mathcal{H}_0$, $\alpha \in (0, 1)$,
- $p_n = 1/(\max_{i=1,\dots,n} S_i)$, is an anytime-valid p-value.

Wald's (1947) sequential probability ratio test is anytime-valid:

Let $D_i$ have density $f_{0,i}$ under the null hypothesis and $f_{1,i}$ under the alternative. Under independence of $(D_i)_{i \in \mathbb{N}}$,

$$S_n = \prod_{i=1}^{n} \frac{f_{1,i}(D_i)}{f_{0,i}(D_i)} \quad \text{is an } e\text{-process.}$$

Wald's (1947) sequential probability ratio test is anytime-valid:

Let $D_i$ have density $f_{0,i}$ under the null hypothesis and $f_{1,i}$ under the alternative. Under independence of $(D_i)_{i\in\mathbb{N}}$,

$$S_n = \prod_{i=1}^{n} \frac{f_{1,i}(D_i)}{f_{0,i}(D_i)} \quad \text{is an } e\text{-process.}$$

We will need something more general:

If $E_n(D^n) \geq 0$ and $\mathbb{E}_P[E_n(D^n) \mid D^{n-1}] \leq 1$ for all $n$, then

$$S_n = \prod_{i=1}^{n} E_i(D^i) \quad \text{is an } e\text{-process.}$$

# Questions

- How to construct e-processes for testing conditional independence in a general setting?

# Questions

- How to construct e-processes for testing conditional independence in a general setting?
- What are “good” e-processes for this problem?

# Questions

- How to construct e-processes for testing conditional independence in a general setting?
- What are “good” e-processes for this problem?
- Do they actually work?

# Questions

- How to construct e-processes for testing conditional independence in a general setting?
- What are "good" e-processes for this problem?
- Do they actually work?

Everything in the model-X setting (Candès et al., 2018).

# Model-X and conditional randomization test

"a glorified randomized experiment where we know the propensity score"

Assumptions: observations $D_n = (X_n, Y_n, Z_n)$, $n \in \mathbb{N}$, are i.i.d. and the distribution of $X$ given $Z$ is known,

$$X \mid Z = z \ \sim \ Q_z.$$

# Model-X and conditional randomization test

"a glorified randomized experiment where we know the propensity score"

Assumptions: observations $D_n = (X_n, Y_n, Z_n)$, $n \in \mathbb{N}$, are i.i.d. and the distribution of $X$ given $Z$ is known,

$$X \mid Z = z \;\sim\; Q_z.$$

Conditional randomization test (CRT):

- $T(X^n, Y^n, Z^n)$ a function such that $T$ is large if $X$ and $Y$ are conditionally dependent,
- simulate $\tilde{X}_j^n \sim Q^n_{Z^n}$, $j = 1, \ldots, M$, and define

$$p = \frac{1 + \sum_{j=1}^{M} 1\{T(\tilde{X}_j^n, Y^n, Z^n) \geq T(X^n, Y^n, Z^n)\}}{1 + M}.$$

# Anytime-valid tests of conditional independence

For a single $D = (X, Y, Z)$ and any function $h > 0$,

$$E_h(D) = \frac{h(X, Y, Z)}{\int h(x, Y, Z)\, dQ_Z(x)}$$

satisfies $\mathbb{E}_P[E] = 1$ for all $P \in \mathcal{H}_0$.

# Anytime-valid tests of conditional independence

For a single $D = (X, Y, Z)$ and any function $h > 0$,

$$E_h(D) = \frac{h(X, Y, Z)}{\int h(x, Y, Z)\, dQ_Z(x)}$$

satisfies $\mathbb{E}_P[E] = 1$ for all $P \in \mathcal{H}_0$.

Define the e-process

$$S_n = S_n(D^n) = \prod_{i=1}^{n} \frac{h_i(X_i, Y_i, Z_i)}{\int h_i(x, Y_i, Z_i)\, dQ_{Z_i}(x)},$$

where $h_i$ may depend on $D^{i-1}$.

How to choose $h_i$?

How to choose $h_i$?

$S_n$ is "growth rate optimal" (GRO) (Grünwald, de Heide, and Koolen, 2022+) under the alternative hypothesis $\tilde{P}$ if it

$$\text{maximizes} \quad \mathbb{E}_{\tilde{P}}[\log(S_n)].$$

How to choose $h_i$?

$S_n$ is "growth rate optimal" (GRO) (Grünwald, de Heide, and Koolen, 2022+) under the alternative hypothesis $\tilde{P}$ if it

$$\text{maximizes} \quad \mathbb{E}_{\tilde{P}}[\log(S_n)].$$

Result: If $(X, Y, Z)$ admit a density $f$, then the GRO $S_n$ is obtained for $h_i(x, y, z) = f_{Y|X,Z}(Y_i \mid X_i, Z_i)$, and

$$S_n = \prod_{i=1}^{n} \frac{f_{Y|X,Z}(Y_i \mid X_i, Z_i)}{f_{Y|Z}(Y_i \mid Z_i)}.$$

- By rewriting,

$$S_n = \prod_{i=1}^{n} \frac{f_{X,Y|Z}(X_i, Y_i \mid Z_i)}{f_{X|Z}(X_i \mid Z_i) f_{Y|Z}(Y_i \mid Z_i)}$$

$\mathbb{E}_f[\log(S_n)] = nI_f(X; Y \mid Z)$ (conditional mutual information),

- By rewriting,

$$S_n = \prod_{i=1}^{n} \frac{f_{X,Y|Z}(X_i, Y_i \mid Z_i)}{f_{X|Z}(X_i \mid Z_i) f_{Y|Z}(Y_i \mid Z_i)}$$

$\mathbb{E}_f[\log(S_n)] = nI_f(X; Y \mid Z)$ (conditional mutual information),

- or in a "Wald-form",

$$S_n = \prod_{i=1}^{n} \frac{f_{X|Y,Z}(X_i \mid Y_i, Z_i)}{f_{X|Z}(X_i \mid Z_i)}.$$

- By rewriting,

$$S_n = \prod_{i=1}^{n} \frac{f_{X,Y|Z}(X_i, Y_i \mid Z_i)}{f_{X|Z}(X_i \mid Z_i) f_{Y|Z}(Y_i \mid Z_i)}$$

$\mathbb{E}_f[\log(S_n)] = nI_f(X; Y \mid Z)$ (conditional mutual information),

- or in a "Wald-form",

$$S_n = \prod_{i=1}^{n} \frac{f_{X|Y,Z}(X_i \mid Y_i, Z_i)}{f_{X|Z}(X_i \mid Z_i)}.$$

- The same test statistic, $f_{Y|X,Z}$, yields Neyman-Pearson optimal CRT (Katsevich and Ramdas, 2020).

In practice, $f$ is estimated sequentially.

In practice, $f$ is estimated sequentially.

- Construct estimator $\hat{f}^{i}_{Y|X,Z}$ based on data $D^{i}$, and set

$$S_n = \prod_{i=1}^{n} \frac{\hat{f}^{i-1}_{Y|X,Z}(Y_i \mid X_i, Z_i)}{\int \hat{f}^{i-1}_{Y|X,Z}(Y_i \mid x, Z_i)\, dQ_{Z_i}(x)},$$

- or $\hat{f}^{i}_{X|Y,Z}$, and with the density $q_{X|Z}$ of $Q_z$, set

$$S_n = \prod_{i=1}^{n} \frac{\hat{f}^{i-1}_{X|Y,Z}(X_i \mid Y_i, Z_i)}{q_{X|Z}(X_i \mid Z_i)}.$$

# Robustness

Like the CRT, the proposed tests are valid for any choice of the functions $h_i$ or estimators $\hat{f}^i_{Y|X,Z}$.

# Robustness

Like the CRT, the proposed tests are valid for any choice of the functions $h_i$ or estimators $\hat{f}^i_{Y|X,Z}$.

If the model-X assumption is violated and only approximations $\hat{Q}_z$ of $Q_z$ are available, then

$$P\left(\exists\, n \leq N\colon S_n \geq 1/\alpha \,\middle|\, Y^N, Z^N\right) \leq \alpha + d_{\mathrm{TV}}(Q^N_{Z^N}, \hat{Q}^N_{Z^N}),\;\; N \in \mathbb{N},$$

where $d_{\mathrm{TV}}$ is the total variation distance.

# Robustness

Like the CRT, the proposed tests are valid for any choice of the functions $h_i$ or estimators $\hat{f}^i_{Y|X,Z}$.

If the model-X assumption is violated and only approximations $\hat{Q}_z$ of $Q_z$ are available, then

$$P\left(\exists\, n \leq N \colon S_n \geq 1/\alpha \middle| Y^N, Z^N\right) \leq \alpha + d_{\mathrm{TV}}(Q^N_{Z^N}, \hat{Q}^N_{Z^N}), \;\; N \in \mathbb{N},$$

where $d_{\mathrm{TV}}$ is the total variation distance.

Same bound is known for the CRT (Berrett et al., 2020) without sequential testing (no "$\exists\, n \leq N$").

# Sanity check: logistic regression

Assume that $Y \in \{0, 1\}$ with probabilities

$$p_\theta(y \mid X, Z) = \frac{\exp(y(\beta^\top X + \gamma^\top Z))}{1 + \exp(\beta^\top X + \gamma^\top Z)},$$

with parameter vector $\theta = (\beta, \gamma)$.

# Sanity check: logistic regression

Assume that $Y \in \{0, 1\}$ with probabilities

$$p_\theta(y \mid X, Z) = \frac{\exp(y(\beta^\top X + \gamma^\top Z))}{1 + \exp(\beta^\top X + \gamma^\top Z)},$$

with parameter vector $\theta = (\beta, \gamma)$.

Estimate $\theta$ sequentially with maximum likelihood method.

- Result: if $(X, Z) \in \mathbb{R}^p \times \mathbb{R}^q$ is subgaussian, then the corresponding test has asymptotic power one if $\beta \neq 0$,

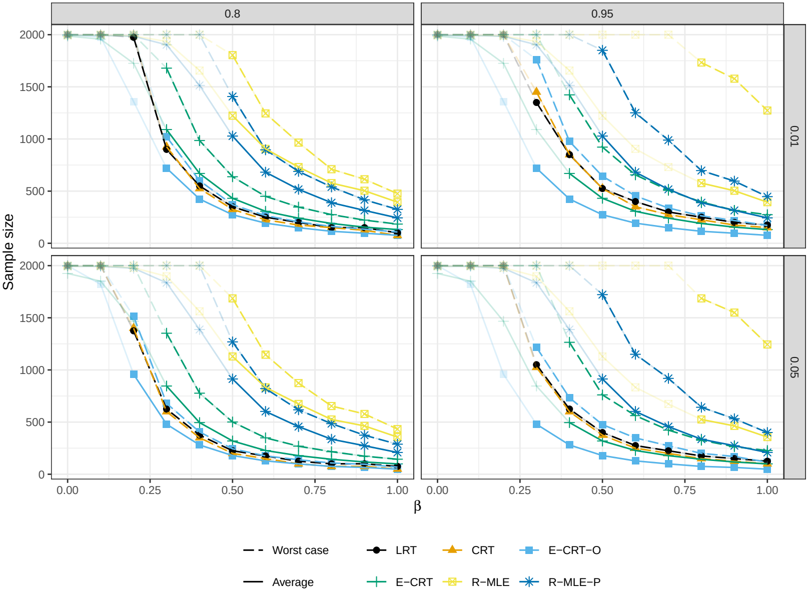
$$S_n = \exp\left(nI(X; Y \mid Z) + r_n\right), \ \ r_n/n \to 0 \text{ a.s.}$$

Simulations:

- $(X, Z) \in \mathbb{R} \times \mathbb{R}^{q-1}$ with multivariate normal distribution
- logistic model for $Y$

Other tests:

- Non-sequential CRT and likelihood ratio test
- Universal inference running maximum likelihood (Wasserman, Ramdas, and Balakrishnan, 2020)


0.8
0.95
0.01
0.05
Sample size
2000
1500
1000
500
0
0.00
0.25
0.50
0.75
1.00
β
Worst case
Average
LRT
CRT
E−CRT−O
E−CRT
R−MLE
R−MLE−P

Preprint on arXiv:

Peter Grünwald, Alexander Henzi, and Tyron Lardy (2022). "Anytime Valid Tests of Conditional Independence Under Model-X". In: *arXiv e-prints*. arXiv: 2209.12637

Berrett, Thomas B. et al. (2020). “The Conditional Permutation Test for Independence While Controlling for Confounders”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82, pp. 175–197.

Candès, Emmanuel et al. (2018). “Panning for Gold: ‘Model-X’ Knockoffs for High Dimensional Controlled Variable Selection”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80, pp. 551–577.

Grünwald, Peter, Rianne de Heide, and Wouter Koolen (2022+). “Safe Testing”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. To appear. Preprint: arXiv:1906.07801.

Grünwald, Peter, Alexander Henzi, and Tyron Lardy (2022). “Anytime Valid Tests of Conditional Independence Under Model-X”. In: *arXiv e-prints*. arXiv: 2209.12637.

Katsevich, Eugene and Aaditya Ramdas (2020). *On the Power of Conditional Independence Testing Under Model-X*. arXiv: 2005.05506.

Ramdas, Aaditya et al. (2020). *Admissible Anytime-Valid Sequential Inference Must Rely on Nonnegative Martingales*. arXiv: 2009.03167.

Ramdas, Aaditya et al. (2022). *Game-theoretic statistics and safe anytime-valid inference*. arXiv: 2210.01948.

Shafer, Glenn (2021). “Testing by Betting: A Strategy for Statistical and Scientific Communication”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184, pp. 407–431.

Vovk, Vladimir and Ruodu Wang (2021). “E-values: Calibration, Combination and Applications”. In: *The Annals of Statistics* 49, pp. 1736 –1754.

Wald, Abraham (1947). *Sequential Analysis*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, pp. xii+212.

Wasserman, Larry, Aaditya Ramdas, and Sivaraman Balakrishnan (2020). “Universal inference”. In: *Proceedings of the National Academy of Sciences* 117, pp. 16880–16890.