Anytime Valid Inference for Multinomial Count Data

Michael Lindon, Alan Malek

michael.s.lindon@gmail.com, alan.malek@gmail.com

Motivation

Many experiments compare count outcomes among treatment groups. Examples include the number of successful signups in conversion rate experiments or the number of errors produced by software versions in canary tests. Observations typically arrive in a sequence and practitioners wish to continuously monitor their experiments, sequentially testing hypotheses while maintaining Type I error probabilities under optional stopping and continuation. These goals are frequently complicated in practice by non-stationary time dynamics. We provide practical solutions through sequential tests of multinomial hypotheses, hypotheses about many inhomogeneous Bernoulli processes and hypotheses about many time-inhomogeneous Poisson counting processes. For estimation, we further provide confidence sequences for multinomial probability vectors, all contrasts among probabilities of inhomogeneous Bernoulli processes and all contrasts among intensities of time-inhomogeneous Poisson counting processes. Together, these provide an "anytime-valid" inference framework for a wide variety of experiments dealing with count outcomes, which we illustrate with several industry applications [1].

Construction

A Sequential u-level Test

- Consider a sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$ of independent $\text{Multinomial}(1, \boldsymbol{\theta})$ random variables and a null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$.
- Test-martingale:

$$O_n(\boldsymbol{\theta}_0) = \frac{\text{Beta}(\boldsymbol{\alpha}_0 + \boldsymbol{S}_n)}{\text{Beta}(\boldsymbol{\alpha}_0)} \frac{1}{\boldsymbol{\theta}_0^{\boldsymbol{S}_n}},$$

where $S_i^n = \sum_{j=1}^n x_{j,i}$, $\boldsymbol{S}_n = (S_1^n, \ldots, S_d^n) \in \mathbb{R}^d$, $\boldsymbol{\theta}_0^{\boldsymbol{S}_n} = \prod_i \boldsymbol{\theta}_{0,i}^{S_i^n}$ and $\text{Beta}(\boldsymbol{v}) := (\prod_i \Gamma(v_i))/\Gamma(\sum_i v_i)$.

- Interpreted as a Bayes factor or mixture sequential probability ratio test statistic with prior/mixture distribution Dirichlet($\boldsymbol{\alpha}_0$)
- u-level sequential test follows from Ville's inequality for nonnegative supermartingales:

$$\mathbb{P}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\exists n \in \mathbb{N} : O_n(\boldsymbol{\theta}_0) \geq 1/u) \leq u$$

- Sequential p-value: Let $p_0 = 1$ and $p_n = \min(p_{n-1}, 1/O_n(\boldsymbol{\theta}_0))$, then

$$\mathbb{P}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\exists n \in \mathbb{N} : p_n \leq u) \leq u,$$

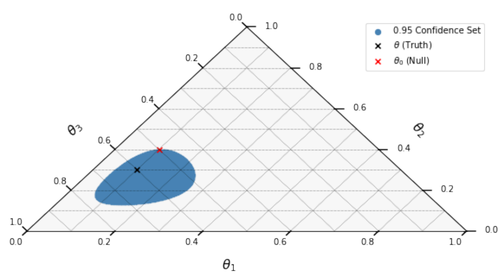
A $1-u$ Confidence Sequence for $\boldsymbol{\theta}$

- Confidence sequence

$$C_n(u) = \{\boldsymbol{\theta} \in \triangle^d : O_n(\boldsymbol{\theta}) < 1/u\}$$

provides a time-uniform $1-u$ coverage guarantee

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in C_n(u) \text{ for all } n \in \mathbb{N}) \geq 1-u,$$



- Confidence sequences on elements of $\boldsymbol{\theta}$ obtained by projecting $C_n(u)$ onto the coordinate axes.
- Let

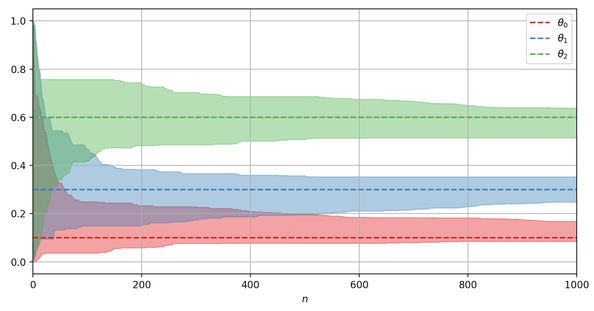
$$j_{n,i}^+(u) = \sup\{\theta_i : \boldsymbol{\theta} \in C_n(u)\}$$
$$j_{n,i}^-(u) = \inf\{\theta_i : \boldsymbol{\theta} \in C_n(u)\}$$

then

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\forall i : \theta_i \in \bigcap_{n=1}^{\infty} [j_{n,i}^-(u), j_{n,i}^+(u)]\right) \geq 1-u.$$

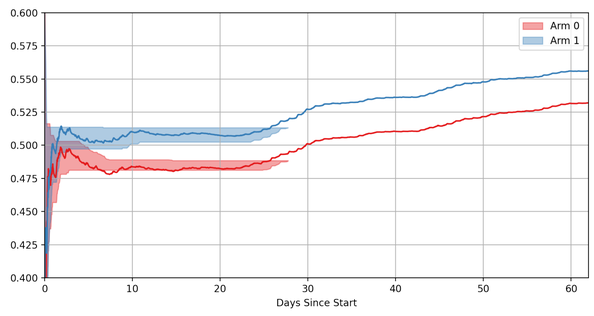
- $j_{n,i}^+(u)$ and $j_{n,i}^-(u)$ can be computed by solving the following convex optimization program

$$\begin{aligned} \max \quad & \theta_i \\ \text{s.t.} \quad & c + \log u \leq \sum_i S_i^n \log \theta_i \\ & \sum_i \theta_i = 1 \end{aligned}$$



Sequential Tests of Inhomogeneous Bernoulli Processes

- Outcomes in A/B Tests often incorrectly modelled as stationary Bernoulli trials
- Under stationarity assumptions, Binomial-Beta confidence sequences become an empty set:



- Behaviour of CS and point estimates above indicate a time-varying success probability
- Instead, model as a *inhomogeneous Bernoulli process*
- Bernoulli probability for arm i at time t

$$p_i(t) = e^{\mu(t)} e^{\delta_i}$$

- Improvement of arm j over arm i at any time is $p_j(t)/p_i(t) = \exp(\delta_j - \delta_i)$
- $\delta_j = \delta_i \Rightarrow p_i(t) = p_j(t)\ \forall t$ under null hypothesis.
- Propensity scores $\boldsymbol{\rho} \in \triangle^d$, (conditional) probability the next Bernoulli success comes from arm i is

$$\theta_i = \sigma_{\boldsymbol{\rho}}(\boldsymbol{\delta})_i := \frac{\rho_i e^{\delta_i}}{\sum_{j=1}^d \rho_j e^{\delta_j}},$$

independent of the time-varying effect $\mu(t)$.

- Sequences of Bernoulli successes a sequence of $\text{Multinomial}(1, \boldsymbol{\theta})$ random variables.
- Under null $\boldsymbol{\theta} = \boldsymbol{\rho}$. Test equality by comparing the *counts of successes*
- Confidence sequence for $\boldsymbol{\delta}$:

$$\mathbb{P}[\boldsymbol{\delta} \in K_n(u) \text{ for all } n \in \mathbb{N}] \geq 1-u,$$

where $K_n(u) = \sigma_{\boldsymbol{\rho}}^{-1}(C_n(u))$

- Confidence Sequence for Contrasts: Let $\mathcal{A}^d = \{\boldsymbol{a} \in \mathbb{R}^d : \sum_i a_i = 0\}$ denote the set of all d-dimensional contrasts. For all $\boldsymbol{a} \in \mathcal{A}^d$ define

$$l_{n,\boldsymbol{a}}^+(u) = \sup\{\sum_i a_i \delta_i : \boldsymbol{\delta} \in K_n(u)\},$$
$$l_{n,\boldsymbol{a}}^-(u) = \inf\{\sum_i a_i \delta_i : \boldsymbol{\delta} \in K_n(u)\},$$

then

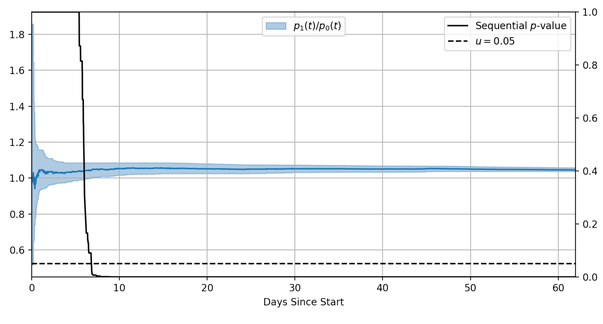
$$\mathbb{P}_{\boldsymbol{\theta}}\left(\forall \boldsymbol{a} \in \mathcal{A}^d : \sum_i a_i \delta_i \in \bigcap_{n=1}^{\infty} [l_{n,a}^-(u), l_{n,a}^+(u)]\right) \geq 1-u.$$

- Upper bound $l_{n,a}^+(u)$ solution to convex optimization

$$\begin{aligned} \max \quad & \sum_i a_i \delta_i \\ \text{s.t.} \quad & c \leq \sum_i S_i^n \left(\delta_i + \log \rho_i - \log \sum_j \rho_j e^{\delta_j}\right) \end{aligned}$$

where $c = \log \text{Beta}(\boldsymbol{\alpha}_0 + \boldsymbol{S}_n) - \log \text{Beta}(\boldsymbol{\alpha}_0) + \log u$.

- Confidence sequence for the contrast $\exp(\delta_1 - \delta_0)$ visualized below.



Sequential Tests of Time-Inhomogeneous Poisson Processes

- Consider d inhomogeneous Poisson point processes with intensity functions

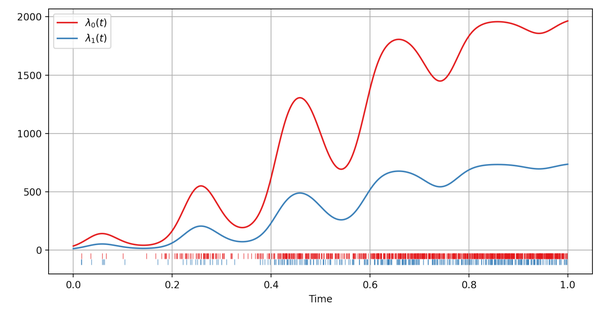
$$\lambda_i(t) = \rho_i e^{\delta_i} \lambda(t)$$

- Each point from process i marked with label i
- Probability next point has label i

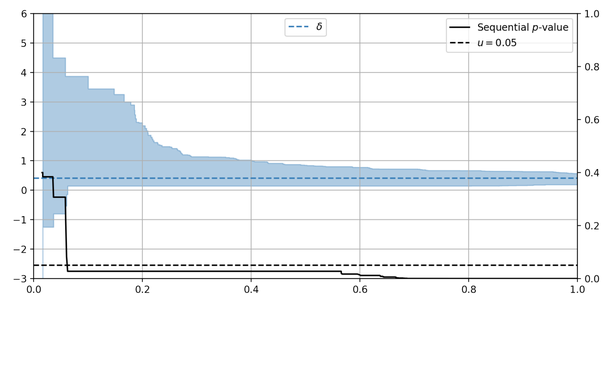
$$\theta_i = \frac{\rho_i e^{\delta_i}}{\sum_{j=1}^d \rho_j e^{\delta_j}}.$$

- Sequence of point labels a sequence of $\text{Multinomial}(1, \boldsymbol{\theta})$ random variables. Independent of $\lambda(t)$
- Test for equality in intensity functions with $H_0 : \boldsymbol{x}_i \sim \text{Multinomial}\left(1, \left(\frac{1}{d}\right) 1_d\right)$.
- Test hypotheses by comparing counting processes
- Example:

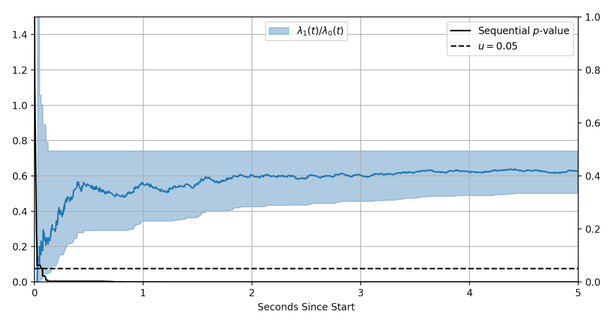
$$\lambda_0(t) = 2000 \text{sigmoid}(\sin(10\pi t) + 8t - 4)$$
$$\lambda_1(t) = \frac{1}{4} e^{\frac{3}{2}} \lambda_0(t)$$



- Confidence Sequence for $\log \delta$ and sequential p-value for $\delta = 0$



- Netflix uses A/B tests to release new software
- Monitors *successful play starts* and *error codes* in new software.
- Decrease or Increase indicates bad software.
- Following figure from a software rollout containing a bug that prevented 60% of streams from starting. Detected and interruped in < 1 second.



References

[1] Michael Lindon and Alan Malek. Anytime valid inference for multinomial count data. In *Advances in Neural Information Processing Systems*, volume 36, 2022.

Neurips 2022, New Orleans