36-771	Fall 2018
Martingales 2 : Sequential Analysis	

Sequential testing, always valid p-values

Lecturer : Aaditya Ramdas

We have organized our presentation around confidence sequences and their closely related uniform concentration bounds. We have emphasized confidence sequences due to our belief that they offer a useful "user interface" for sequential inference. However, our methods may alternatively be viewed as sequential hypothesis tests or always-valid p-values processes (Johari et al. 2015). Indeed, a slew of related definitions from the literature are equivalent or dual to one another. Here we briefly discuss these connections, building upon the definitions and dualities of Johari et al. (2015). Recall Lemma 1 from lecture 15, which gives equivalent formulations of certain common definitions in sequential testing.

First, let us mention that our definition of confidence sequence based on Darling & Robbins (1967*a*) and Lai (1984), differs from that Johari et al. (2015), who require that $\mathcal{P}(\theta_\tau \in \mathrm{CI}_\tau) \geq 1-\alpha$ for all stopping times τ. They allow $\tau = \infty$ by defining $\mathrm{CI}_\infty := \liminf_{t\to\infty} \mathrm{CI}_t$. By taking $A_t := \{\theta_t \notin \mathrm{CI}_t\}$ in Lemma 1 from lecture 14, we see that the distinction is immaterial, and furthermore that we could equivalently define confidence sequences in terms of arbitrary random times, not necessarily stopping times. This generalizes Proposition 1 of Zhao et al. (2016).

As an alternative to confidence sequences, Johari et al. (2015) define an *always-valid p-value process* for some null hypothesis H_0 as an adapted, $[0,1]$-valued sequence $(p_t)_{t=1}^\infty$ satisfying $\mathcal{P}_0(p_\tau \leq \alpha) \leq \alpha$ for all stopping times τ, where $\mathcal{P}_0$ denotes probability under the null H_0. Taking $A_t := \{p_t \leq \alpha\}$ in Lemma 1 from lecture 14 shows that we may replace this definition with an equivalent one over all random times, not necessarily stopping times, or with the uniform condition $\mathcal{P}_0(\exists t \in \mathcal{N} : p_t \leq \alpha) \leq \alpha$. By analogy to the usual dual construction between fixed-sample p-values and confidence intervals[1], one can see that confidence sequences are dual to always-valid p-values, and both are dual to sequential hypothesis tests, as defined by a stopping time and a binary random variable indicating rejection (Johari et al. 2015, Proposition 5). In particular, for the null $H_0 : \theta = \theta^\star$, if (CI_t) is a $(1-\alpha)$-confidence sequence for θ, it is clear that a test which stops and rejects the null as soon as $\theta^\star \notin \mathrm{CI}_t$ controls type I error: $\mathcal{P}_0(\text{reject } H_0) = \mathcal{P}_0(\exists t \in \mathcal{N} : \theta^\star \notin \mathrm{CI}_t) \leq \alpha$. Typically, then, a confidence sequence based on any of the curved uniform bounds in this paper with radius $u(v) = o(v)$ will yield a *test of power one* (Darling & Robbins 1967*b*, Robbins 1970). In particular, for a confidence sequence with limits $\bar{X}_t \pm u(V_t)$, it is sufficient that $\bar{X}_t$ converges a.s. to θ

[1] Indeed, if (CI_t^α) is a $(1-\alpha)$-level confidence sequence for some constant parameter θ, for each $\alpha \in (0,1)$, then $p_t := \inf\{\alpha \in (0,1) : \theta^\star \notin \mathrm{CI}_t^\alpha\}$ gives an always-valid p-value process for the null hypothesis $H_0 : \theta = \theta^\star$. Conversely, if $(p_t^{\theta^\star})$ is an always-valid p-value process for the null hypothesis $H_0 : \theta = \theta^\star$, for each $\theta^\star$ in some domain Θ, then $\mathrm{CI}_t := \{\theta^\star \in \Theta : p_t^{\theta^\star} > \alpha\}$ gives a $(1-\alpha)$-level confidence sequence for θ.

and $\limsup_{t\to\infty} V_t/t < \infty$ a.s., conditions that will typically hold. These conditions imply that the radius of the confidence sequence, $u(V_t)/t$, approaches zero, while the center $\bar{X}_t$ is eventually bounded away from $\theta^\star$ whenever $\theta \neq \theta^\star$, so that the confidence sequence will eventually exclude $\theta^\star$ with probability one.

In the one-parameter exponential family case, the exponential process $\exp\{\lambda S_t(\mu) - t\psi_\mu(t)\}$ is exactly the likelihood ratio for testing $H_0 : \theta = \theta(\mu)$ against $H_1 : \theta = \theta(\mu) + \lambda$. When using a mixture uniform boundary, a sequential test which rejects as soon as the confidence sequence excludes $\mu^\star$ can be seen as equivalently rejecting as soon as either of the mixture likelihood ratios $\int \exp\{\lambda S_t - \psi_{\mu^\star}(\lambda)t\} dF(\lambda)$ or $\int \exp\{-\lambda S_t - \psi_{\mu^\star}(-\lambda)t\} dF(\lambda)$ exceeds $2/\alpha$. Thus a sequential hypothesis test built upon a mixture-based confidence sequence is equivalent to a mixture sequential probability ratio test (Robbins 1970) in the parametric setting. As we have discussed, stitching bounds can also be viewed as approximations to certain mixture bounds, so that hypothesis tests based on stitching bounds are also approximations to mixture SPRTs. Importantly, the confidence sequences defined in this paper are natural nonparametric generalizations of the mixture SPRT, recovering various mixture SPRTs in the parametric cases.

References

Darling, D. A. & Robbins, H. (1967*a*), 'Confidence Sequences for Mean, Variance, and Median', *Proceedings of the National Academy of Sciences* **58**(1), 66–68.

Darling, D. A. & Robbins, H. (1967*b*), 'Iterated Logarithm Inequalities', *Proceedings of the National Academy of Sciences* **57**(5), 1188–1192.

Johari, R., Pekelis, L. & Walsh, D. J. (2015), 'Always valid inference: Bringing sequential analysis to A/B testing', *arXiv preprint arXiv:1512.04922* .

Lai, T. L. (1984), 'Incorporating scientific, ethical and economic considerations into the design of clinical trials in the pharmaceutical industry: a sequential approach', *Communications in Statistics - Theory and Methods* **13**(19), 2355–2368.

Robbins, H. (1970), 'Statistical Methods Related to the Law of the Iterated Logarithm', *The Annals of Mathematical Statistics* **41**(5), 1397–1409.

Zhao, S., Zhou, E., Sabharwal, A. & Ermon, S. (2016), Adaptive Concentration Inequalities for Sequential Decision Problems, *in* '30th Conference on Neural Information Processing Systems (NIPS 2016)', Barcelona, Spain.