36-771 Fall 2018

Martingales 1 : Concentration inequalities

Conjugate mixtures

Lecturer : Aaditya Ramdas

1 Conjugate mixtures

Let f be a probability density on $\mathbb{R}$. For appropriate choices of f and ψ, the integral $\int \exp\{\lambda S_t - \psi(\lambda)V_t\} f(\lambda) d\lambda$ will be analytically tractable. Since, under Assumption 1, this mixture process is upper bounded by a mixture supermartingale $\int L_t(\lambda) f(\lambda) d\lambda$, such mixtures yield closed-form or efficiently computable curved boundaries. This approach is known as the method of mixtures, one of the most widely-studied techniques for constructing uniform bounds (Ville 1939, Wald 1945, Darling & Robbins 1968, Robbins 1970, Robbins & Siegmund 1969, 1970, Lai 1976). With the exception of the sub-Gaussian case, most prior work on the method of mixtures has focused on parametric settings. We instead derive a variety of nonparametric uniform boundaries using this approach.

Unlike the stitching bound of two lectures ago, which involves a small amount of looseness in the analytical approximations, mixture boundaries are unimprovable in a sense we make precise later. We present both one-sided and two-sided boundaries. Each conjugate mixture boundary includes a tuning parameter ρ which controls the sample size for which the boundary is optimzed. Such tuning is critical in practice, as we explain later.

In the sub-Gaussian case, the following boundary is well-known (Robbins 1970, example 2).

Proposition 1 (Two-sided normal mixture) *Suppose (S_t) and (V_t) satisfy Assumption 1 with $\psi = \psi_N$ and $\lambda_{\max} = \infty$, and suppose the same holds for $(-S_t)$. Fix $\alpha \in (0, 1)$ and $\rho > 0$, and define*

$$f(v) := \sqrt{(v + \rho) \log \left(\frac{(\mathbb{E}L_0)^2 (v + \rho)}{\alpha^2 \rho} \right)}. \tag{2}$$

Then $\mathcal{P}(\forall t \geq 1 : |S_t| < \mathrm{NM}_2(V_t)) \geq 1 - \alpha$.

When only a one-sided sub-Gaussian assumption holds, the normal mixture can still be used to obtain a sub-Gaussian uniform boundary.

When tails are heavier than Gaussian, the normal mixture boundary is not applicable, However, the follow sub-exponential mixture boundary based a gamma mixing density is universally applicable, as described in the previous lecture. Below we make use of the regularized lower incomplete gamma function $\gamma(a, x) := (\int_0^x u^{a-1} e^{-u} \dot{\text{u}})/\Gamma(a)$, available in standard statistical software packages.

Theorem 1 (Gamma-exponential mixture) *Fix $c > 0, \rho > 0$ and define*

$$\mathrm{GE}_\alpha(v) := \inf\{s \geq 0 : m(s,v) \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{3}$$

$$\textit{where} \quad m(s,v) := \frac{\left(\frac{\rho}{c^2}\right)^{\frac{\rho}{c^2}}}{\Gamma\left(\frac{\rho}{c^2}\right)\gamma\left(\frac{\rho}{c^2}, \frac{\rho}{c^2}\right)} \frac{\Gamma\left(\frac{v+\rho}{c^2}\right)\gamma\left(\frac{v+\rho}{c^2}, \frac{cs+v+\rho}{c^2}\right)}{\left(\frac{cs+v+\rho}{c^2}\right)^{\frac{v+\rho}{c^2}}} \exp\{\frac{cs+v}{c^2}\}. \tag{4}$$

Then GE_α is a sub-exponential uniform boundary with crossing probability α for scale c.

When a sub-exponential condition applies to $(-S_t)$ as well, we may apply these boundaries to both tails and take a union bound, obtaining a two-sided confidence sequence.

2 More examples

The basic idea behind the method of mixtures is as follows. If (S_t), (V_t), and $\psi(\lambda)$ satisfy Assumption 1, and for any probability distribution F on $\mathbb{R}_{\geq 0}$, we have, for all t,

$$\int \exp\{\lambda S_t - \psi(\lambda)V_t\} dF(\lambda) \leq \int L_t(\lambda) dF(\lambda), \tag{5}$$

and the right-hand side is a nonnegative supermartingale with initial expectation $\mathbb{E}L_0$. So defining

$$\mathcal{M}_\alpha(v) := \inf\{s \in \mathbb{R} : \int \exp\{\lambda s - \psi(\lambda)v\} dF(\lambda) \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{6}$$

and invoking Ville's maximal inequality for nonnegative supermartingales, we have the following basic result:

Lemma 2 *$\mathcal{M}_\alpha$ is a sub-ψ uniform boundary with crossing probability α.*

We suppress the dependence of $\mathcal{M}_\alpha$ on ψ, F and $\mathbb{E}L_0$ for notational simplicity, as we did with $\mathcal{S}_\alpha$. With F a point mass at λ we recover the linear uniform bounds.

In the sub-Gaussian case, we can take the mixture distribution F to be half-normal over the positive reals. The integral in (6) can be evaluated explicitly, yielding the mixture boundary

$$\mathrm{NM}_\alpha(v) = \inf\{s \in \mathbb{R} : \sqrt{\frac{4\rho}{V_t + \rho}} \exp\{\frac{s^2}{2(V_t+\rho)}\} \Phi\left(\frac{s}{\sqrt{V_t+\rho}}\right) \geq \frac{\mathbb{E}L_0}{\alpha}\}. \tag{7}$$

This is easily evaluated to high precision by numerical root finding. Alternatively, we have the following tight analytical upper bound:

$$\mathrm{NM}(v) \leq \widetilde{\mathrm{NM}}_\alpha(v) := \sqrt{2(v+\rho)\log\left(\frac{\mathbb{E}L_0}{2\alpha}\sqrt{\frac{v+\rho}{\rho}} + 1\right)}. \tag{8}$$

Proposition 9 (One-sided normal mixture) *For any $\alpha \in (0,1)$ and $\rho > 0$, the boundaries NM_α and $\widetilde{\mathrm{NM}}_\alpha$ are sub-Gaussian uniform boundaries with crossing probability α.*

Many of our definitions and results have focused on one-sided uniform bounds, which yield one-sided (upper or lower) confidence sequences. Such one-sided bounds can always be combined via a union bound to form a two-sided confidence sequence, and for typical values of α used in statistical practice, such a union bound is hardly wasteful, as the intersection of the two error events will have very small probability. In the method of mixtures, however, it is sometimes convenient to derive a two-sided bound directly using a mixture distribution F with support on both positive and negative values of λ.

In the sub-Bernoulli case, we first rewrite the exponential process $\exp\{\lambda S_t - \psi_B(\lambda) V_t\}$ in terms of the transformed parameter $p = (1 + e^{-(g+h)\lambda})^{-1}$. This is motivated by the transform from the canonical parameter to the mean parameter of a Bernoulli family, but keep in mind that we make no parametric assumption here, these are merely analytical manipulations. Then a truncated Beta distribution on $p \in [g/(g+h), 1]$ yields the one-sided Beta-Binomial uniform boundary. Below, $B_x(a,b) = \int_0^x p^{a-1}(1-p)^{b-1} dp$ denotes the incomplete Beta function, whose implementation is readily available in statistical software packages.

Proposition 10 (One-sided Beta-Binomial mixture) *Fix any $g, h > 0$, $\alpha \in (0,1)$, and $\rho > gh$, let $m = \rho/gh - 1$ and define*

$$f_{g,h}(v) \;:=\; \inf\{s \geq 0 : m_{g,h}(s,v) \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{11}$$

$$\textit{where} \quad m_{g,h}(s,v) \;:=\; \frac{(g+h)^v}{g^{\frac{gv+s}{g+h}} h^{\frac{hv-s}{g+h}}} \frac{B_{h/(g+h)}\left(\frac{h(m+v)-s}{g+h}, \frac{g(m+v)+s}{g+h}\right)}{B_1\{\frac{mg}{g+h}, \frac{mh}{g+h}\}}. \tag{12}$$

Then $f_{g,h}$ is a sub-Bernoulli uniform boundary with crossing probability α and range g, h.

Sub-Bernoulli conditions typically follow from the assumption that centered observations are $[-g, h]$-bounded. In such a case, the following two-sided bound may be preferable. Simpler versions of this boundary have long been studied i.i.d. Bernoulli sampling (Ville 1939, Robbins 1970, Lai 1976, Shafer et al. 2011).

Proposition 13 (Two-sided Beta-Binomial mixture) *Suppose (S_t) and (V_t) satisfy Assumption 1 with $\psi = \psi_B$ for range g, h and $\lambda_{\max} = \infty$, and suppose the same holds for $(-S_t)$ with range h, g. Fix any $\rho > gh$, let $m = \rho/gh - 1$ and define*

$$f_{g,h}(v) \;:=\; \inf\{s \geq 0 : m_{g,h}(s,v) \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{14}$$

$$\textit{where} \quad m_{g,h}(s,v) \;:=\; \frac{(g+h)^v}{g^{\frac{gv+s}{g+h}} h^{\frac{hv-s}{g+h}}} \frac{B_1\left(\frac{g(m+v)+s}{g+h}, \frac{h(m+v)-s}{g+h}\right)}{B_1\left(\frac{mg}{g+h}, \frac{mh}{g+h}\right)}. \tag{15}$$

Then $\mathcal{P}(\forall t \geq 1 : -f_{h,g}(V_t) < S_t < f_{g,h}(V_t)) \geq 1 - \alpha$.

The Beta mixing density is chosen so that the corresponding mixture on λ is approximately mean zero with precision ρ, making the boundary comparable to a normal mixture bound with the same precision. This allows the user to choose ρ following the same logic as for the normal mixture boundary, as described later, and indeed this is true by construction for all of our conjugate mixture boundaries.

The gamma-exponential mixture is the result of evaluating the mixture integral (6) with mixture density

$$\frac{dF}{d\lambda} = \frac{1}{\gamma(\rho/c^2, \rho/c^2)} \frac{(\rho/c)^{\rho/c^2}}{\Gamma(\rho/c^2)} (c^{-1} - \lambda)^{\rho/c^2 - 1} e^{-\rho(c^{-1} - \lambda)/c}. \tag{16}$$

This is a gamma distribution with shape ρ/c^2 and scale ρ/c applied to the transformed parameter $u = c^{-1} - \lambda$, truncated to the support $[0, c^{-1}]$. The distribution has mean zero and variance equal to $1/\rho$, making it comparable to the normal mixture distribution used above. As $\rho \to \infty$, the gamma mixture distribution converges to a normal distribution and concentrates about $\lambda = 0$, the regime in which $\psi_E(\lambda) \sim \psi_N(\lambda)$, which gives some intuition for why the gamma mixture recovers the normal mixture when $\rho \gg c^2$. Like the normal mixture, the gamma mixture is unimprovable and is effective in practice.

A similar mixture boundary holds in the sub-Poisson case:

Proposition 17 (Gamma-Poisson mixture) *Fix $c > 0, \rho > 0$ and define*

$$\mathrm{GP}_\alpha(v) := \inf\{s \geq 0 : m(s, v) \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{18}$$

$$\textit{where} \quad m(s, v) := \frac{\left(\frac{\rho}{c^2}\right)^{\frac{\rho}{c^2}}}{\Gamma\left(\frac{\rho}{c^2}\right)\gamma\left(\frac{\rho}{c^2}, \frac{\rho}{c^2}\right)} \frac{\Gamma\left(\frac{cs+v+\rho}{c^2}\right)\gamma\left(\frac{cs+v+\rho}{c^2}, \frac{v+\rho}{c^2}\right)}{\left(\frac{v+\rho}{c^2}\right)^{\frac{cs+v+\rho}{c^2}}} \exp\{\frac{v}{c^2}\}. \tag{19}$$

Then GP_α is a sub-Poisson uniform boundary with crossing probability α for scale c.

References

Darling, D. A. & Robbins, H. (1968), ‘Some Further Remarks on Inequalities for Sample Sums’, *Proceedings of the National Academy of Sciences* **60**(4), 1175–1182.

Lai, T. L. (1976), ‘On Confidence Sequences’, *The Annals of Statistics* **4**(2), 265–280.

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

Robbins, H. & Siegmund, D. (1969), ‘Probability Distributions Related to the Law of the Iterated Logarithm’, *Proceedings of the National Academy of Sciences* **62**(1), 11–13.

Robbins, H. & Siegmund, D. (1970), 'Boundary Crossing Probabilities for the Wiener Process and Sample Sums', *The Annals of Mathematical Statistics* **41**(5), 1410–1429.

Shafer, G., Shen, A., Vereshchagin, N. & Vovk, V. (2011), 'Test Martingales, Bayes Factors and p-Values', *Statistical Science* **26**(1), 84–101.

Ville, J. (1939), *Étude Critique de la Notion de Collectif.*, Gauthier-Villars, Paris.

Wald, A. (1945), 'Sequential Tests of Statistical Hypotheses', *Annals of Mathematical Statistics* **16**(2), 117–186.