36-771 Fall 2018

Martingales 2 : Sequential Analysis

Discrete mixtures and inverted stitching

Lecturer : Aaditya Ramdas

1 Numerical bounds using discrete mixtures.

In applied use, there is often no need for an explicit closed-form expression so long as the bound can be easily computed numerically. Our discrete mixture method gives a straightforward and efficient technique for numerical computation of curved boundaries whenever Assumption 1 is satisfied. It permits arbitrary mixture densities and thus can produce boundaries growing at the asymptotically-optimal $\mathcal{O}(V_t \log\log V_t)$ rate.

Recall that the shape of the stitching bound was determined by the user-specified function h. For the discrete mixture bound, one instead specifies a distribution F. We then discretize F using a series of support points λ_k, geometrically spaced according to successive powers of some $\eta > 1$, and an associated set of weights w_k:

$$\lambda_k := \frac{\lambda_{\max}}{\eta^{k+1/2}} \quad \text{and} \quad w_k := \frac{\lambda_{\max}(\eta-1)f(\lambda_k\sqrt{\eta})}{\eta^{k+1}} \quad \text{for} \quad k = 1, 2, \ldots. \tag{1}$$

With the above definitions in place, we have a discrete mixture bound as follows.

Theorem 1 (Discrete mixture bound) *Fix* $\psi : [0, \lambda_{\max}) \to \mathbb{R}$ *and* $\alpha \in (0,1)$. *Employing any continuous distribution* F *with density* f *that is nonincreasing and positive on a nonempty interval* $(0, \lambda_{\max}]$, *if we define*

$$\widetilde{\mathcal{M}}_\alpha(v) := \inf\{s \in \mathbb{R} : \sum_{k=0}^{\infty} w_k \exp\{\lambda_k s - \psi(\lambda_k)v\} \geq \frac{\mathbb{E}L_0}{\alpha}\}, \tag{2}$$

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

We suppress the dependence of $\widetilde{\mathcal{M}}_\alpha$ on F, $\mathbb{E}L_0$, $\lambda_{\max}$ and η for notational simplicity. Though the above theorem is a straightforward consequence of the method of mixtures, our choice of discretization makes it effective, broadly applicable, and easy to compute numerically.

To see heuristically why the exponentially-spaced grid $\lambda_k = \mathcal{O}(\eta^{-k})$ makes sense, observe that the integrand $\exp\{\lambda s - \lambda^2 v/2\}$ is a scaled normal density in λ with mean s/v and standard deviation $1/\sqrt{v}$. In the regime relevant to our curved boundaries, s is of order $\sqrt{v}$, ignoring logarithmic factors. Hence the integrand at time v has both center and spread of order $1/\sqrt{v}$, so as $v \to \infty$, the relevant scale of the integrand shrinks. With the grid

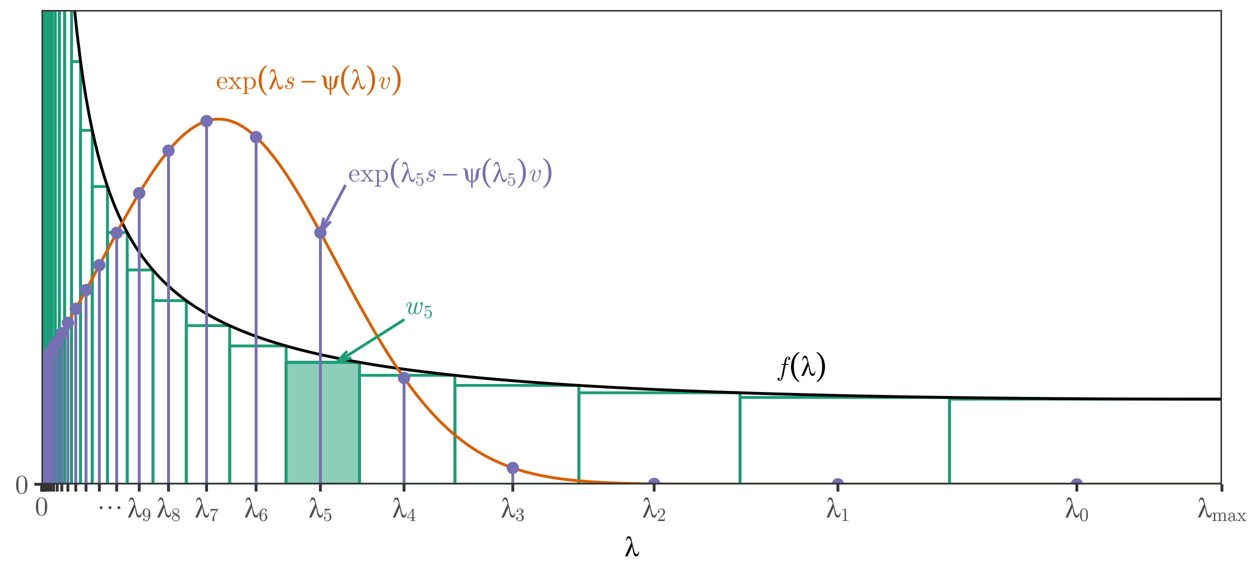

Figure 1: Illustration of the discrete mixture method. Mixture density $f(\lambda)$ is discretized on a grid $(\lambda_k)_{k=0}^{\infty}$ which gets finer as $\lambda \downarrow 0$. Resulting discrete mixture weights are represented by areas within green bars. Integrand $\exp\{\lambda s - \psi(\lambda)v\}$ is evaluated at grid points λ_k, illustrated by purple points. Multiplying one integrand evaluation $\exp\{\lambda_k s - \psi(\lambda_k)v\}$ by the corresponding weight w_k gives one term of the sum (2).

$\lambda_k = \mathcal{O}(\eta^{-k})$ we have $\lambda_k - \lambda_{k+1} = \mathcal{O}(\lambda_k)$, ensuring that the resolution of the grid around the peak of the integrand matches the scale of the integrand as $v \to \infty$.

The choice of $\lambda_{\max}$ depends on the minimum value of V_t relevant to inference: making $\lambda_{\max}$ larger will make the resulting bound tighter over smaller values of V_t at the cost of a looser bound for all larger values of V_t. In practice, for $\psi = \psi_G$, setting $\lambda_{\max} = [c + \sqrt{v_{\min}/2\log\alpha^{-1}}]^{-1}$ will ensure the bound is tight for $V_t \geq v_{\min}$. Furthermore, in practice the sum can be truncated after $k_{\max} = \lceil \log_\eta(\lambda_{\max}[c + \sqrt{5v/\log\alpha^{-1}}]) \rceil$ terms.

To illustrate the accuracy of the discrete mixture, we compare it to the one-sided normal mixture bound. By using the same half-normal mixing density from last class, and setting $\eta = 1.05$, $\lambda_{\max} = 100$, we may evaluate a corresponding discrete mixture bound $\widetilde{\mathcal{M}}_\alpha$. With $\rho = 14.3$, $\alpha = 0.05$ and $\mathbb{E}L_0 = 1$, numerical calculations show that

$$\sup_{1 \leq t \leq 10^6} \frac{\widetilde{\mathcal{M}}_\alpha(t)}{\mathrm{NM}_\alpha(t)} \leq 1.004, \tag{3}$$

suggesting that the discrete mixture theorem gives an excellent conservative approximation to the corresponding continuous mixture boundary to over a large practical range. Of course, when a closed form is available as in the normal mixture, one should use it in practice. But an exact closed form integral is rarely available as it is here, and substantial looseness often accompanies closed-form approximations which provably maintain crossing probability guarantees. In such cases, unless a closed form is required, the discrete mixture method is preferable.

2 Inverted stitching for arbitrary boundaries.

In the discrete mixture method, we choose a mixture distribution F and the machinery yields a boundary $\widetilde{\mathcal{M}}_\alpha$. Likewise, in the stitching construction from a few lectures ago, we choose an error decay function h and the machinery yields a boundary $\mathcal{S}_\alpha$. In this section we invert the procedure: we choose a boundary function $g(v)$ and numerically compute an upper bound on its S_t-upcrossing probability using a stitching-like construction. For simplicity we restrict to the sub-Gaussian case; we are currently working on extending this idea beyond sub-Gaussianity.

Theorem 2 *For any nonnegative, strictly concave function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and $v_{\max} > 1$, the function*

$$u(v) := \begin{cases} g(1 \vee v), & v \leq v_{\max}, \\ \infty, & \textit{otherwise} \end{cases} \tag{4}$$

is a sub-Gaussian uniform boundary with crossing probability

$$(\mathbb{E}L_0) \inf_{\eta > 1} \sum_{k=0}^{\lceil \log_\eta v_{\max} \rceil} \exp\{-\frac{2(g(\eta^{k+1}) - g(\eta^k))(\eta g(\eta^k) - g(\eta^{k+1}))}{\eta^k(\eta - 1)^2}\}. \tag{5}$$

The proof follows a straightforward idea. We break time into epochs $\eta^k \leq V_t < \eta^{k+1}$. Within each epoch we consider the linear boundary passing through the points $(\eta^k, g(\eta^k))$ and $(\eta^{k+1}, g(\eta^{k+1}))$. This line lies below $g(V_t)$ throughout the epoch, and its crossing probability is determined by its slope and intercept as in the mother theorem. Taking a union bound over epochs yields the result.

A similar idea was considered by Darling & Robbins (1968), using a mixture integral approximation instead of an epoch-based construction to derive closed-form bounds. Inverted stitching requires numerical summation but yields tighter bounds with fewer assumptions. As an example, the above theorem with $\eta = 2.99$ shows that

$$\mathcal{P}\left(\exists t : 1 \leq V_t \leq 10^{20} \text{ and } S_t \geq 1.7\sqrt{V_t(\log\log(eV_t) + 3.46)}\right) \leq 0.025. \tag{6}$$

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.