36-771 Fall 2018
Martingales 1 : Concentration inequalities

Proof of the mother theorem

Lecturer : Aaditya Ramdas

1 Proof of the mother theorem

Ville's maximal inequality for nonnegative supermartingales (Ville (1939); Durrett (2017), exercise 4.8.2), often attributed to Doob, is the foundation of all uniform bounds in this paper. It is an infinite-horizon uniform extension of Markov's inequality, asserting that a nonnegative supermartingale (L_t) has probability at most $\mathbb{E}L_0/a$ of ever crossing level a: $\mathcal{P}(\exists t : L_t \geq a) \leq \mathbb{E}L_0/a$ for any $a > 0$. Applying this inequality to Assumption 1 gives, for any $\lambda \in (0, \lambda_{\max})$ and $z \in \mathbb{R}$,

$$\mathcal{P}(\exists t \in \mathcal{T} : \exp\{\lambda S_t - \psi(\lambda)V_t\} \geq e^z) \leq \mathcal{P}(\exists t \in \mathcal{T} : L_t \geq e^z) \leq (\mathbb{E}L_0)e^{-z}. \tag{1}$$

To derive Theorem 1(a) from (1), fix $a, b \geq 0$ and choose $\lambda \in [0, \lambda_{\max})$ such that $\psi(\lambda) \leq b\lambda$, supposing for the moment that some such value of λ exists. Then

$$\begin{aligned}\mathcal{P}\left(\exists t \in \mathcal{T} : S_t \geq a + bV_t\right) &= \mathcal{P}\left(\exists t \in \mathcal{T} : \exp\left\{\lambda S_t - b\lambda V_t\right\} \geq e^{a\lambda}\right) \\ &\leq \mathcal{P}\left(\exists t \in \mathcal{T} : \exp\left\{\lambda S_t - \psi(\lambda) V_t\right\} \geq e^{a\lambda}\right) \\ &\leq (\mathbb{E}L_0)e^{-a\lambda},\end{aligned}$$

applying (1) in the last step. This bound holds for all choices of λ in the set $\{\lambda \in [0, \lambda_{\max}) : \psi(\lambda)/\lambda \leq b\}$, so to minimize the final bound, we take the supremum over this set, recovering the stated bound $(\mathbb{E}L_0)e^{-aD(b)}$ by the definition of $D(b)$. If no value $\lambda \in [0, \lambda_{\max})$ satisfies $\psi(\lambda) \leq b\lambda$, then $D(b) = 0$ by definition, so that the bound holds trivially. This shows that Assumption 1 implies Theorem 1(a).

To complete the proof we will show that the four parts of Theorem 1 are equivalent whenever ψ is CGF-like. We require some simple facts about $\psi(\lambda)/\lambda$:

Lemma 1 *Suppose ψ is CGF-like with domain $[0, \lambda_{\max})$.*

1. *$\psi(\lambda)/\lambda < \psi'(\lambda)$ for all $\lambda \in (0, \lambda_{\max})$.*
2. *$\lambda \mapsto \psi(\lambda)/\lambda$ is continuous and strictly increasing on $\lambda > 0$.*
3. *$\inf_{\lambda \in (0,\lambda_{\max})} \psi(\lambda)/\lambda = 0$*
4. *$\sup_{\lambda \in (0,\lambda_{\max})} \psi(\lambda)/\lambda = \bar{b}$.*

5. $\psi(D(b))/D(b) = b$ *for any* $b \in [0, \bar{b})$. *That is,* $D(b)$ *is the inverse of* $\psi(\lambda)/\lambda$.

Proof: To see (i), write $\psi(\lambda) = \int_0^\lambda \psi'(t)\mathrm{d}t < \lambda\psi'(\lambda)$, where the inequality follows since ψ is strictly convex so that ψ' is strictly increasing. For (ii), the function is continuous because ψ is continuous, and differentiating reveals it to be strictly increasing by part (i). L'Hôpital's rule implies (iii) along with the assumptions $\psi(\lambda) = \psi'(\lambda) = 0$, and implies (iv) along with the assumption $\sup_\lambda \psi(\lambda) = \infty$. Part (v) follows from the definition of $D(\cdot)$ and parts (ii), (iii) and (iv). ■

We also repeatedly use the well-known fact about the Legendre-Fenchel transform that $\psi'^{-1}(u) = \psi^{\star\prime}(u)$, which follows by differentiating the identity $\psi^\star(u) = u\psi'^{-1}(u) - \psi(\psi'^{-1}(u))$.

- $(a) \implies (b)$: Fix $m > 0$ and $x \in [0, m\bar{b})$. Any line with slope $b \in [0, x/m]$ and intercept $x - mb$ passes through the point (m, x) in the (V_t, S_t) plane, and part (a) yields

$$\begin{aligned}\mathcal{P}\left(\exists t \in \mathcal{T} : S_t \geq x + b(V_t - m)\right) &\leq (\mathbb{E}L_0) \exp\left\{-(x - mb)D(b)\right\} \\ &= (\mathbb{E}L_0) \exp\left\{-m\left(\frac{x}{m} \cdot D(b) - \psi(D(b))\right)\right\}\end{aligned}$$

  using 1(v) in the second step. Now we choose the slope b to minimize the probability bound. The unconstrained optimizer $b_\star$ satisfies $\psi'(D(b_\star)) = x/m$, and a solution is guaranteed to exist by our restriction on x. This solution is given by $D(b_\star) = \psi'^{-1}(x/m) = \psi^{\star\prime}(x/m)$. Hence $b_\star = \mathfrak{s}(x/m)$ using the definition of $\mathfrak{s}(\cdot)$. 1(i) shows $x/m = \psi'(D(b_\star)) > \psi(D(b_\star))/D(b_\star) = b_\star$, verifying that $b_\star$ is in the allowed range for part (a). Identify the Legendre-Fenchel transformation $\psi^\star(x/m) = \sup_b[(x/m)D(b) - \psi(D(b))]$ to complete the proof of part (b). The fact that this bound is exact for Brownian motion shows that, since we have optimized over the slope, the bound could not hold in general for any other slope.

- $(b) \implies (c)$: Fix $m \geq 0$ and $x \in [0, \bar{b})$ and observe that

$$\mathcal{P}\left(\exists t \in \mathcal{T} : \frac{S_t}{V_t} \geq \mathfrak{s}(x) + \frac{m(x - \mathfrak{s}(x))}{V_t}\right) = \mathcal{P}\left(\exists t \in \mathcal{T} : S_t \geq mx + \mathfrak{s}(x) \cdot (V_t - m)\right).$$

  Now applying part (b) with values m and mx yields part (c).

- $(c) \implies (a)$: Fix $a \geq 0$ and $b \in [0, \bar{b})$. Set $x = \psi'(D(b))$ and $m = a/(x - \mathfrak{s}(x))$. Recalling $\psi^{\star\prime} = \psi'^{-1}$ we see that $\mathfrak{s}(x) = \psi(D(b))/D(b) = b$. Now apply part (c) to obtain

$$\begin{aligned}\mathcal{P}\left(\exists t \in \mathcal{T} : S_t \geq a + bV_t\right) &\leq (\mathbb{E}L_0) \exp\left\{-a \cdot \frac{\psi^\star(x)}{x - \mathfrak{s}(x)}\right\} \\ &= (\mathbb{E}L_0) \exp\left\{-a \cdot \frac{\psi^\star(x) \cdot \psi^{\star\prime}(x)}{x\psi^{\star\prime}(x) - \psi(\psi^{\star\prime}(x))}\right\}.\end{aligned}$$

Recognizing the Legendre-Fenchel transform in the denominator of the final exponent, we see that the probability bound equals $(\mathbb{E}L_0)\exp\{-a\psi^{\star\prime}(x)\}$. Again using $\psi^{\star\prime}(x) = \psi'^{-1}(x) = D(b)$ yields (a).

- $(a, b) \implies (d)$: Clearly $\{\exists t \in \mathcal{T} : V_t \geq m, S_t \geq x + b(V_t - m)\} \subseteq \{\exists t \in \mathcal{T} : S_t \geq x + b(V_t - m)\}$, and the probability of the latter event is upper bounded by $(EL_0)\exp\{-(x - bm)D(b)\}$ from part (a); the intercept $x - bm$ is nonnegative by our restriction on b. However, if $m > 0$ and $b \geq \mathfrak{s}(x/m)$, then $\{\exists t \in \mathcal{T} : V_t \geq m, S_t \geq x + b(V_t - m)\} \subseteq \{\exists t \in \mathcal{T} : V_t \geq m, S_t \geq x + \mathfrak{s}(x/m)(V_t - m)\}$; that is, we may replace the slope b by the smaller slope $\mathfrak{s}(x/m)$. The probability of the latter event is upper bounded by $(\mathbb{E}L_0)\exp\{-m\psi^\star(x/m)\}$ by part (b).

- $(d) \implies (a)$: set $m = 0$ and $x = a$ to recover part (a).

It is worth noting here that, unlike the proofs of Freedman (1975), Khan (2009), Tropp (2011), and Fan et al. (2015), we do not explicitly construct a stopping time in our proof. While an optional stopping argument is hidden within the proof of Ville's inequality, the underlying stopping time here is different from that in the aforementioned citations.

References

Durrett, R. (2017), *Probability: Theory and Examples*, 5a edn.

Fan, X., Grama, I. & Liu, Q. (2015), 'Exponential inequalities for martingales with applications', *Electronic Journal of Probability* **20**(1), 1–22.

Freedman, D. A. (1975), 'On Tail Probabilities for Martingales', *The Annals of Probability* **3**(1), 100–118.

Khan, R. A. (2009), 'L_p-Version of the Dubins–Savage Inequality and Some Exponential Inequalities', *Journal of Theoretical Probability* **22**(2), 348.

Tropp, J. A. (2011), 'Freedman's inequality for matrix martingales', *Electronic Communications in Probability* **16**, 262–270.

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