36-771	Fall 2018
Martingales 1 : Concentration inequalities	
The mother theorem	
Lecturer : Aaditya Ramdas	

1 The mother theorem

To state our main theorem on general exponential line-crossing inequalities, we will make use of the following transforms of ψ:

$$\psi^\star(u) := \sup_{\lambda \in [0, \lambda_{\max})} [\lambda u - \psi(\lambda)] \quad \text{(the Legendre-Fenchel transform)},$$

$$D(u) := \sup \left\{ \lambda \in [0, \lambda_{\max}) : \frac{\psi(\lambda)}{\lambda} \le u \right\} \quad \text{(the "decay" transform), and}$$

$$\mathfrak{s}(u) := \frac{\psi(\psi^{\star\prime}(u))}{\psi^{\star\prime}(u)} \quad \text{(the "slope" transform).}$$

In the definition of $D(u)$, we take the supremum of the empty set to equal zero instead of the usual $-\infty$. This case can arise in general, but not when ψ is CGF-like. Note that $D(u)$ can also be infinite. We call $D(u)$ the "decay" transform because it determines the rate of exponential decay of the upcrossing probability bound in Theorem 1(a) below. We call $\mathfrak{s}(u)$ the "slope" transform because it gives the slope of the linear boundary in Theorem 1(b); this is defined only when ψ is CGF-like.

Our main theorem has four parts, each of which facilitates comparisons with a particular related literature, as we discuss later.

Theorem 1 *If the canonical supermartingale assumption (Assumption 1, previous lecture) holds, then*

(a) For any $a, b \ge 0$, we have

$$\mathcal{P}\{\exists t \in \mathcal{T} : S_t \ge a + bV_t\} \le (\mathbb{E}L_0) \exp(-aD(b)).$$

Additionally, whenever ψ is CGF-like, the following three statements are equivalent to statement (a).

(b) For any $m > 0$ and $x \in [0, m\bar{b})$, we have

$$\mathcal{P}\left(\exists t \in \mathcal{T} : S_t \ge x + \mathfrak{s}\left(\frac{x}{m}\right) \cdot (V_t - m)\right) \le (\mathbb{E}L_0) \exp\left\{-m\psi^\star\left(\frac{x}{m}\right)\right\}.$$

Furthermore, if the slope $\mathfrak{s}(x/m)$ were replaced by any other value, the probability bound on the right-hand side would need to be larger.

(c) *For any* $m \geq 0$ *and* $x \in [0, \bar{b})$, *we have*

$$\mathcal{P}\left(\exists t \in \mathcal{T} : \frac{S_t}{V_t} \geq \mathfrak{s}(x) + \frac{m(x - \mathfrak{s}(x))}{V_t}\right) \leq (\mathbb{E}L_0) \exp\left\{-m\psi^\star(x)\right\}.$$

Furthermore, if $\mathfrak{s}(x)$ *were replaced by any other value, the probability bound on the right-hand side would need to be larger.*

(d) *For any* $m \geq 0$, $x \geq 0$ *and* b *finite in* $\left[0, \bar{b} \wedge \frac{x}{m}\right]$ *(taking* $\frac{0}{0} = \infty$*), we have*

$$\mathcal{P}\left(\exists t \in \mathcal{T} : V_t \geq m \text{ and } S_t \geq x + b(V_t - m)\right) \leq \begin{cases} (\mathbb{E}L_0) \exp\left\{-(x - bm)D(b)\right\}, & m = 0 \text{ or } \mathfrak{s}\left(\frac{x}{m}\right) \geq b \\ (\mathbb{E}L_0) \exp\left\{-m\psi^\star\left(\frac{x}{m}\right)\right\}, & m > 0 \text{ and } \mathfrak{s}\left(\frac{x}{m}\right) \leq \end{cases} \tag{1}$$

We later provide a straightforward proof of the above results that uses only Ville's maximal inequality for nonnegative supermartingales Ville (1939) and elementary convex analysis. We give here several remarks on the theorem, followed by three illustrative examples.

- It is useful to think of the parts of 1 as statements about the process (V_t, S_t) or $(V_t, S_t/V_t)$ in $\mathbb{R}^2$. Many of our results are easily understood via this geometric intuition. 1 illustrates the following points.
  - 1(a) takes a given line $a + bV_t$ and bounds its S_t-upcrossing probability.
  - 1(b) takes a point (m, x) in the (V_t, S_t)-plane and, out of the infinitely many lines passing through it, chooses the one which yields the tightest upper bound on the corresponding S_t-upcrossing probability.
  - 1(c) is like part (b), but instead of looking at S_t, we look at S_t/V_t, fix a point (m, x) in the $(V_t, S_t/V_t)$-plane, and choose from among the infinitely many curves $b + a/V_t$ passing through it to minimize the probability bound.
  - The intuition for 1(d) is as follows. If we want to bound the upcrossing probability of the line $(x - bm) + bV_t$ on $\{V_t \geq m\}$, we can clearly obtain a conservative bound from 1(a) with $a = x - bm$. This yields the first case in (1). However, we can also apply 1(b) with the values m, x, obtaining a bound on the upcrossing probability for a line which passes through the point (m, x) in the (V_t, S_t)-plane, and this line yields the minimum possible probability bound among all lines passing through (m, x). If the slope of this line, $\mathfrak{s}(x/m)$, is less than b, then this optimal probability bound is conservative for the upcrossing probability over the original line $x + b(V_t - m)$ on $\{V_t \geq m\}$. This gives the second case in (1), which is guaranteed to be at least as small as the bound in the first case when $\mathfrak{s}(x/m) \leq b$.
- The purpose of excluding ψ being CGF-like from Assumption 1 is to separate the truth of statement (a), which follows solely from the assumption, from its equivalence to (b), (c), and (d), which follows from ψ being CGF-like.

- The factor $\mathbb{E}L_0$ will typically equal one when we have scalar observations, while in matrix cases it generally equals d, the dimension of the matrix observations. As mentioned earlier, in many cases $\lambda_{\max} = \infty$ and $\bar{b} = \infty$, but we allow finite values to handle some cases discussed later.

- 1 yields a uniform extension of many fixed-time or finite-horizon exponential bounds, losing nothing in going from a fixed-time to a uniform bound. We briefly revisit this property later, where we observe that the Dubins-Savage inequality does not possess this property.

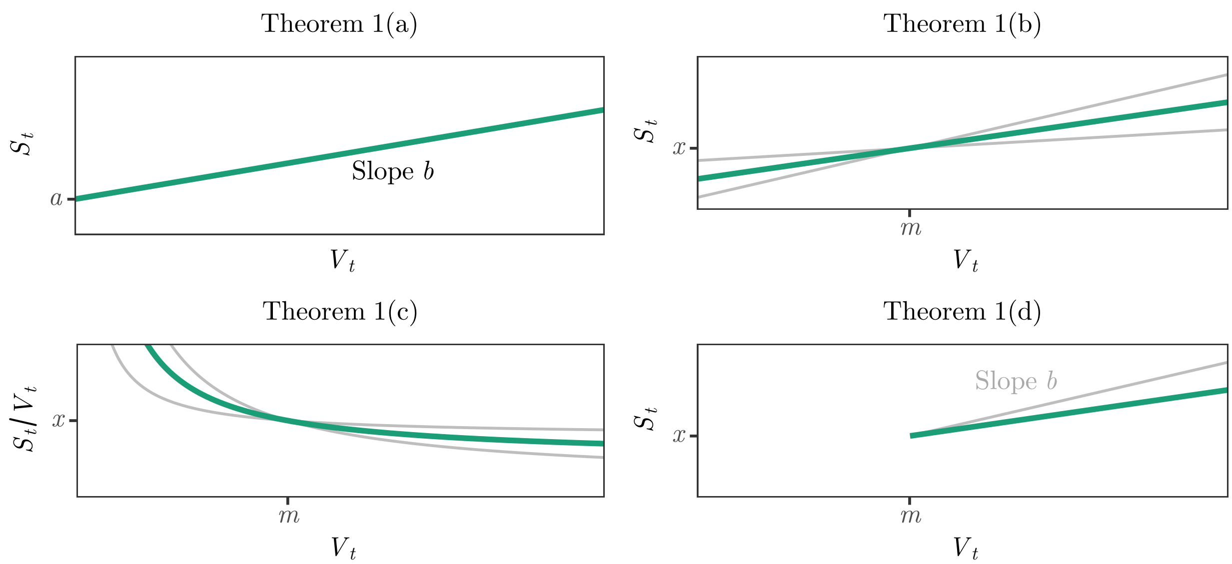

Figure 1: Illustration of the equivalent statements of 1, as described in the text.

References

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