From Bandits to Dishonest Statisticians

Robert J. Vanderbei

2014 April 25

20th Applied Probability Day: In Honor/Memory of Larry Shepp
Columbia University

To view animations: Download and open in Acrobat!

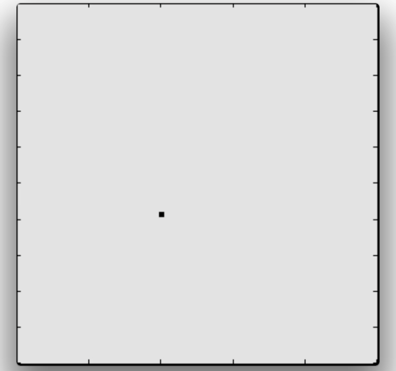
http://www.princeton.edu/∼rvdb

A Two-Armed Bandit Problem

Let $X_i(t_i)$, $i = 1, 2$, denote two Brownian motions on $[0, 1]$ with absorbing end points.

At any given instant in time, we can run one of the two Brownian motions while the other one idles. We can switch between them as often as we like.

The switching process is modeled as an *optional increasing path*: $T(t) = (T_1(t), T_2(t))$.



The T_i's have appropriate measurability properties and, in addition, $T_1(t) + T_2(t) = t$ for all t.

Intuitively, $T_i(t)$ represents the time that process i has "run" over the interval $[0, t]$.

Selecting this random time change amounts to determining a control policy for the process.

Given a payoff function $f : [0, 1]^2 \to \mathbb{R}$, the problem is to select a control policy and a stopping time so as to maximize the expected reward at the stopping time:

$$v(x_1, x_2) = \sup_{T,\tau} \mathbf{E}_{x_1,x_2} f(X_1(T_1(\tau)), X_2(T_2(\tau))).$$

Optimal Solution

The *value function* v is the *smallest bi-excessive majorant* of f:

$$\frac{\partial^2 v}{\partial x_1^2} \leq 0$$

$$\frac{\partial^2 v}{\partial x_2^2} \leq 0$$

$$v \geq f$$

For simplicity, we henceforth assume that f is zero in the interior of the state space and is given by nonnegative concave functions on the four sides of the state space.

With this assumption on f it is easy to see that, given a switching strategy T, the optimal stopping time τ is precisely the first exit time from the interior of the state space.

Larry's Principle of Smooth Fit

Assume that the concave functions on the "north" and "east" sides are identically zero.

Let $\gamma_1(x_1)$ denote the nonnegative concave function along the bottom side.
Let $\gamma_2(x_2)$ denote the nonnegative concave function along the left side.

Thought experiment...

- Assume that the γ_i's are strictly concave.
- Consider being close to the bottom side.
- The value function v will equal γ_1 at the boundary.
- If the value function is smooth, then it too will be concave near this lower edge.
- Hence, the optimal strategy is to control vertically. That is, $\partial^2 v / \partial x_2^2 = 0$.
- A similar analysis applies to the left hand side.
- Posit the existance of a curve coming from the lower left corner representing a switch in policy.
- Use the *principle of smooth fit* to write down differential equation with boundary conditions.
- Solve the differential equation.

The Switching Curve

For any strictly concave function γ, let Γ denote the *increasing* function given by

$$\Gamma(x) = -\int_0^x u\gamma''(u)du.$$

The switching curve is given by

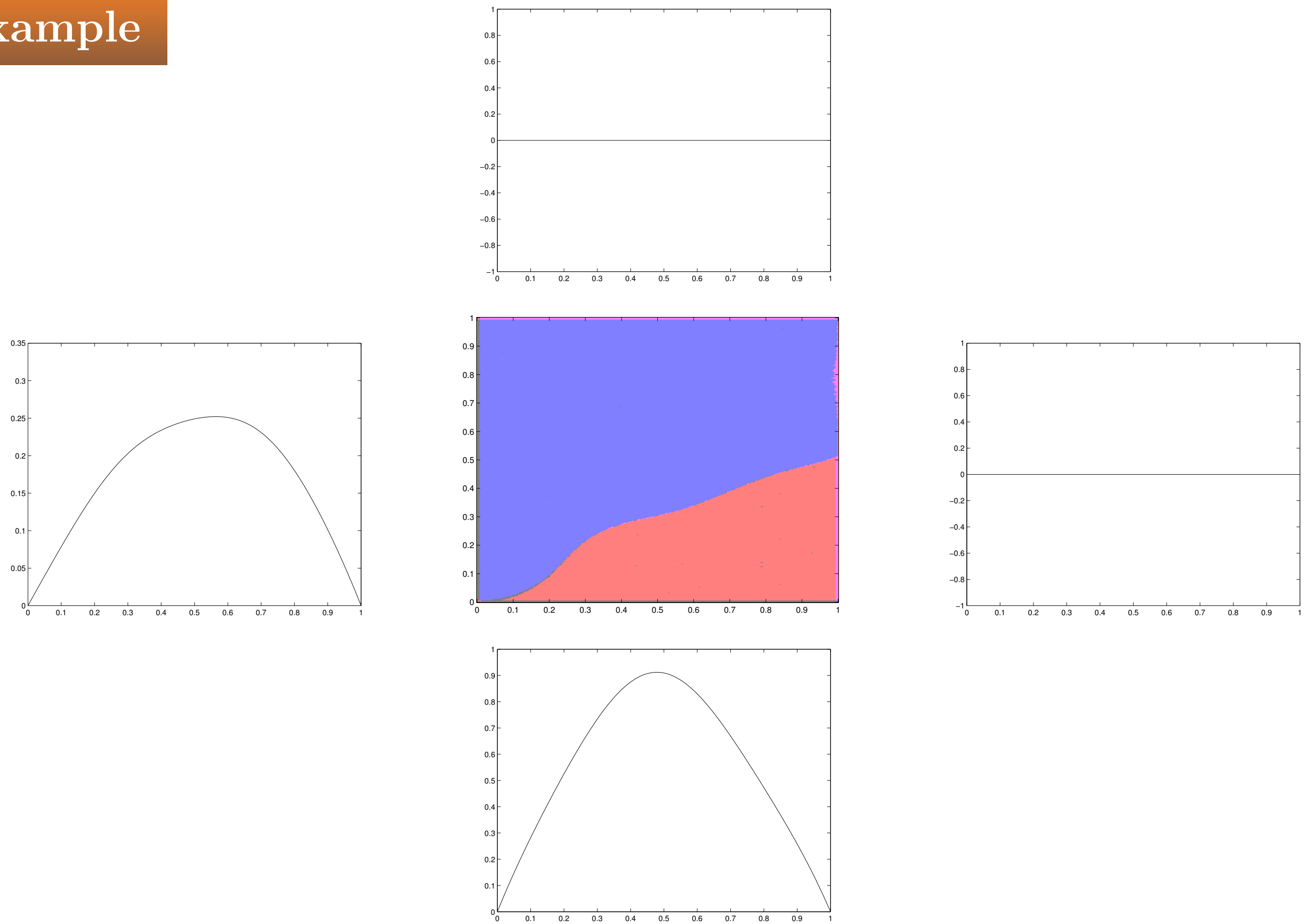
$$\Gamma_2(x_2) = \Gamma_1(x_1).$$

A similarly explicit formula exists for the value function itself.

When $\gamma_1 = \gamma_2$, the switching curve is the diagonal.
In this case, the switched process involves a *local time* process on the diagonal.

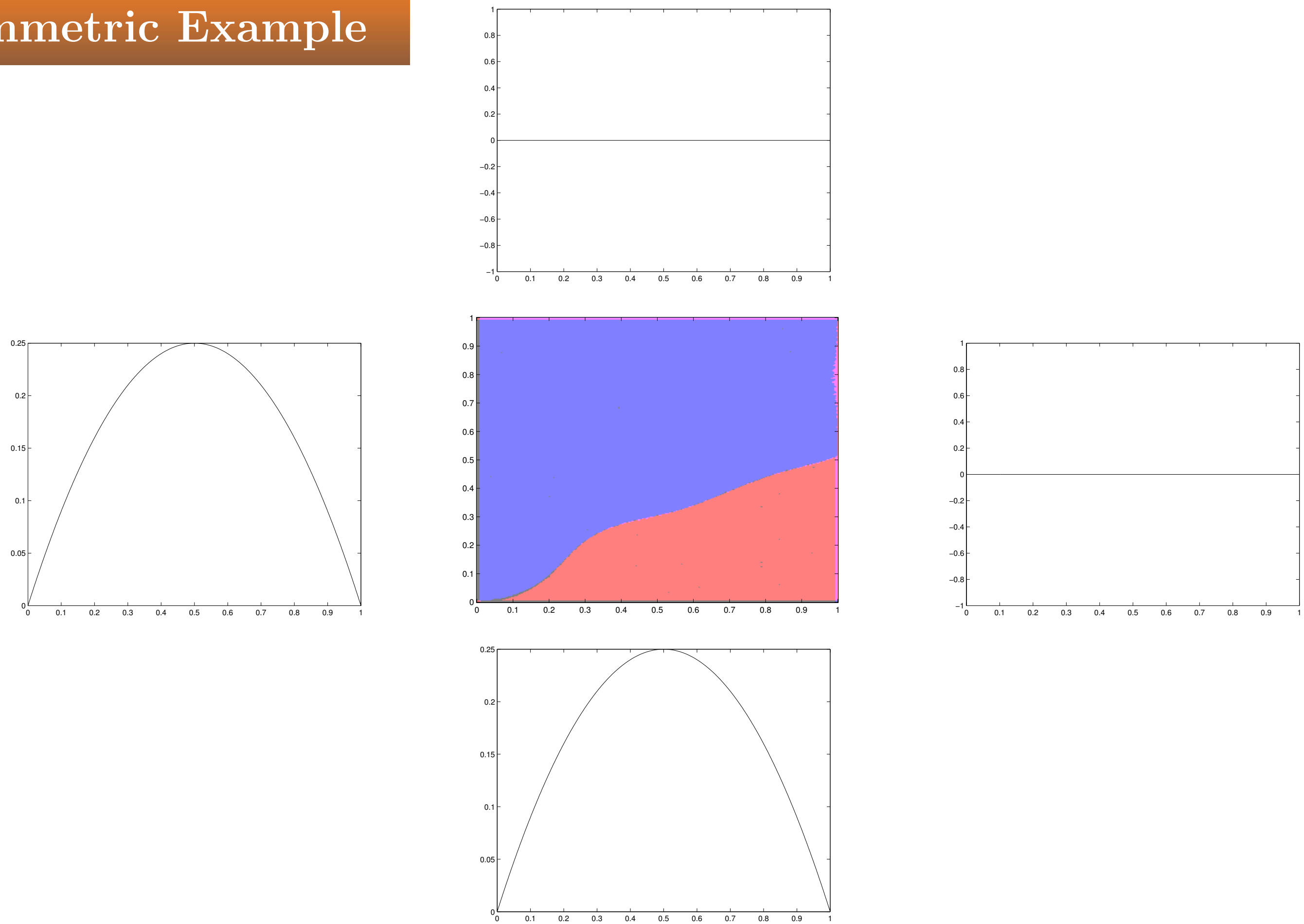
In general, the behavior of the optimally controlled process has a general local-time-like behavior along the switching curve.

An Example



Red zone: run X_2.   Blue zone: run X_1.

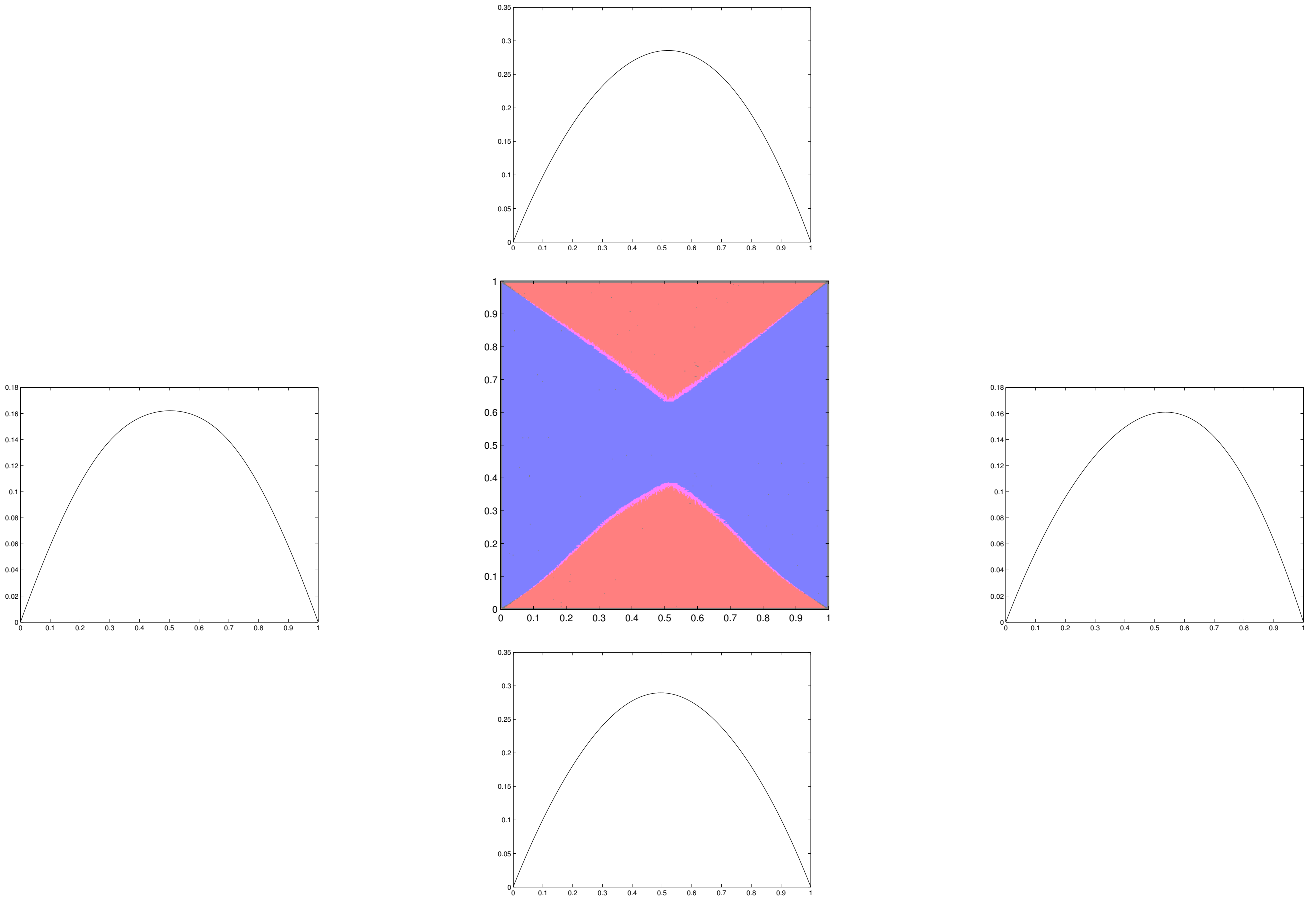
A Symmetric Example



Red zone: run X_2. Blue zone: run X_1.

Concave On All Sides

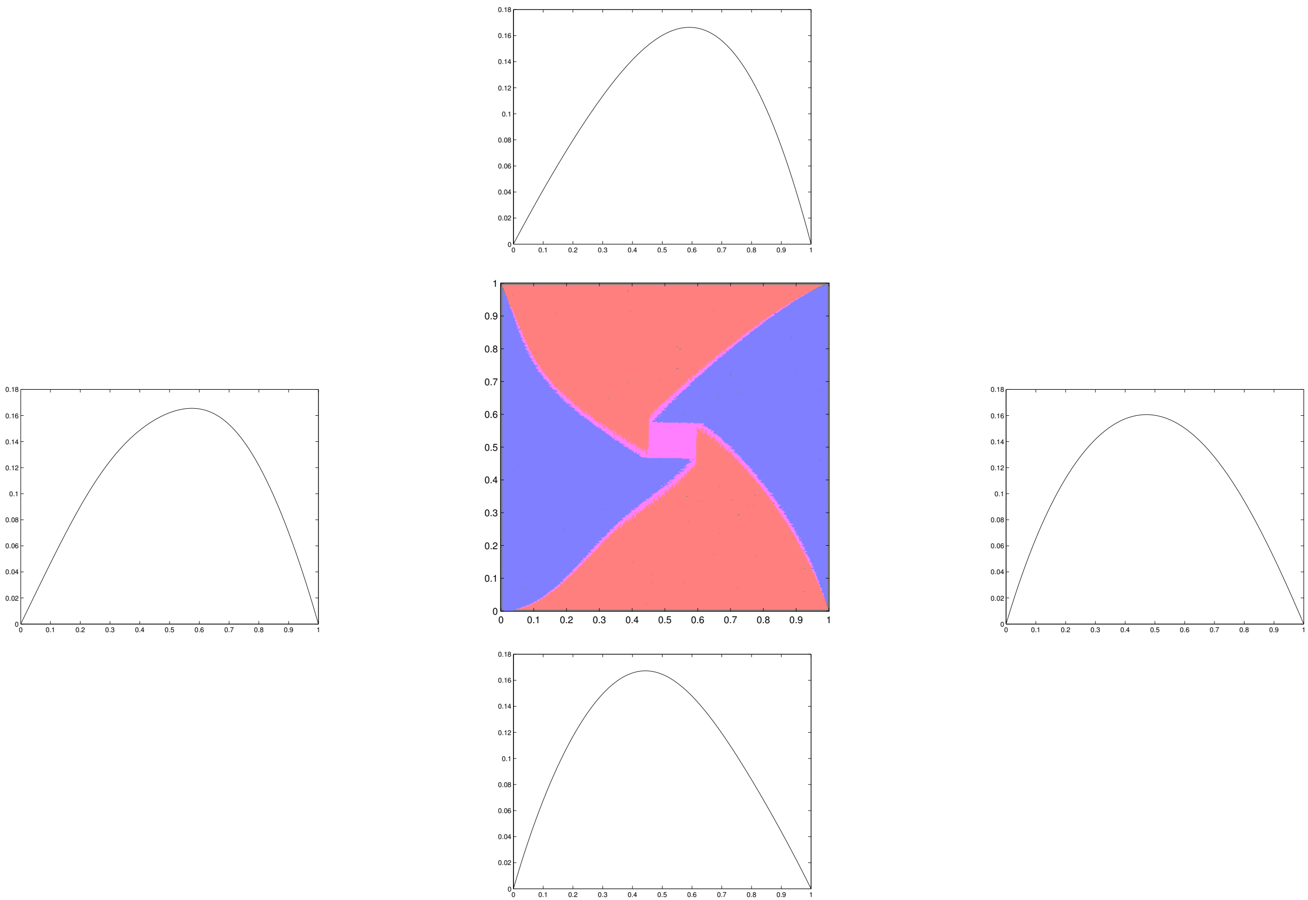
Thinking locally about the corners, one might posit that the general solution looks like this (or its transpose)...



where the switching curves are given by formulae similar to the simple two-sided case.

But Sometimes...

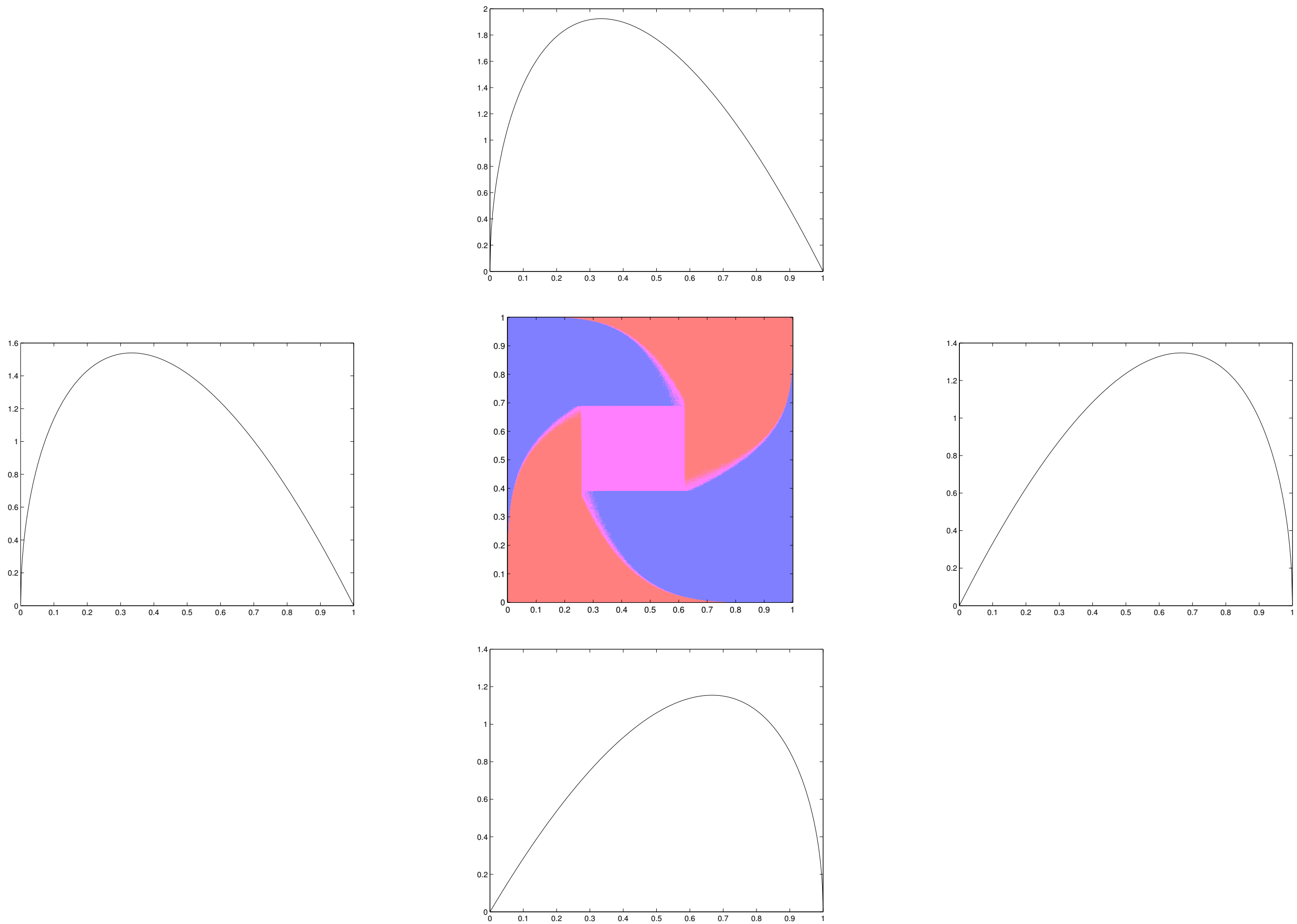
The answer (discovered by solving a *linear programming* problem), looks like this:



The central region is an *indifference zone*—run either Brownian motion.

The shape of the indifference zone is always *rectangular*.

An Example with a Large Indifference Zone



The Linear Programming Problem

$$\begin{aligned}
\text{minimize} \quad & \iint v(x_1, x_2)\, dx_1 dx_2 \\
\text{subject to} \quad & \frac{\partial^2 v}{\partial x_1^2} \leq 0 \\
& \frac{\partial^2 v}{\partial x_2^2} \leq 0 \\
& v \geq f.
\end{aligned}$$

Note: Discretize to make infinite dimensional problem into finite dimensional LP.

PS. Much faster than value iteration!

Optimal Switching References

- *Optimal Stopping and Supermartingales over Partially Ordered Sets*,
  Mandelbaum and Vanderbei,
  Z. Wahrscheinlichkeitstheorie verw. Gebiete, 57, 253–264, 1981

- *Optimal Switching Between a Pair of Brownian Motions*,
  Mandelbaum, Shepp, and Vanderbei,
  Ann. Prob., 18(3), 1010–1033, 1990.

- *Optimal Switching Among Several Brownian Motions*,
  Vanderbei,
  SIAM J. on Control and Optimization, 30, 1150–1162, 1992.

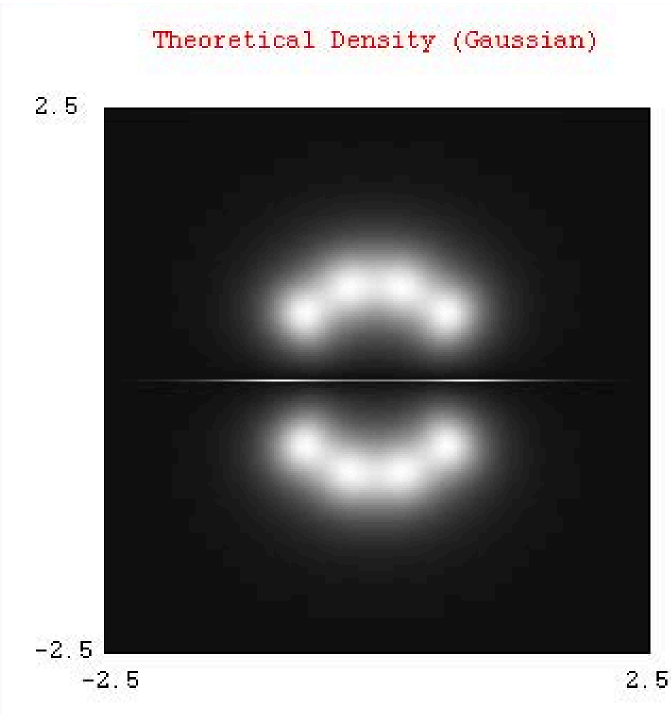
- *Brownian Bandits*,
  Mandelbaum and Vanderbei,
  Dynkin Festschrift, AMS, 1995

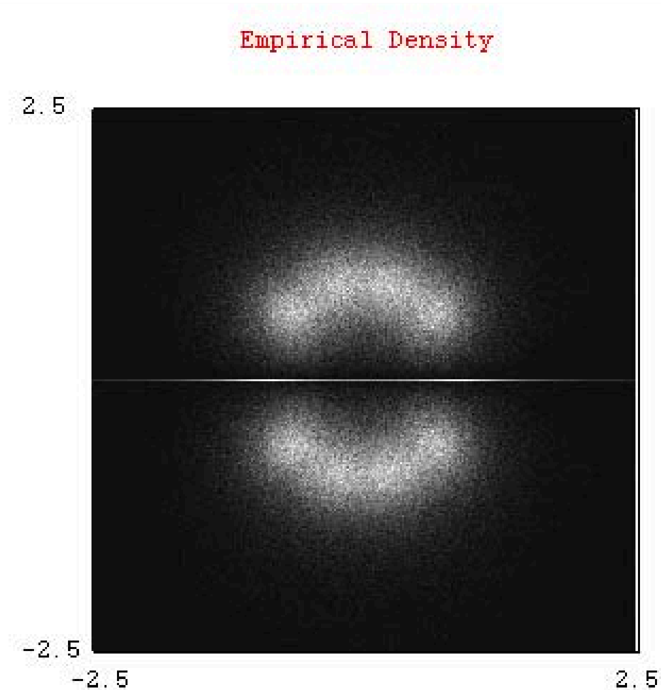
Zeros of Random Polynomials

Consider a random polynomial:

$$P_n(z) = \sum_{j=0}^{n} \eta_j z^j, \quad z \in \mathbb{C},$$

where $\eta_0, \ldots, \eta_n$ are independent standard normal random variables.
Let $\nu_n(\Omega)$ denote the number of zeros in a set Ω in the complex plane.





Theorem

For each measurable set $\Omega \subset \mathbb{C}$,

$$\mathbf{E}\nu_n(\Omega) = \int_\Omega h_n(x,y)dxdy + \int_{\Omega\cap\mathbb{R}} g_n(x)dx,$$

where

$$h_n = \frac{B_2 D_0^2 - B_0(B_1^2 + |A_1|^2) + B_1(A_0\bar{A}_1 + \bar{A}_0 A_1)}{\pi|z|^2 D_0^3},$$

and

$$g_n = \frac{(B_0 B_2 - B_1^2)^{1/2}}{\pi|z|B_0}$$

and where

$$B_k(z) = \sum_{j=0}^{n} j^k |z|^{2j}, \quad z \in \mathbb{C}, \quad k = 0, 1, 2,$$

$$A_k(z) = \sum_{j=0}^{n} j^k z^{2j}, \quad z \in \mathbb{C}, \quad k = 0, 1,$$

and

$$D_0(z) = \sqrt{B_0^2(z) - |A_0|^2(z)}.$$

Main Idea

From the *argument principle*, we get

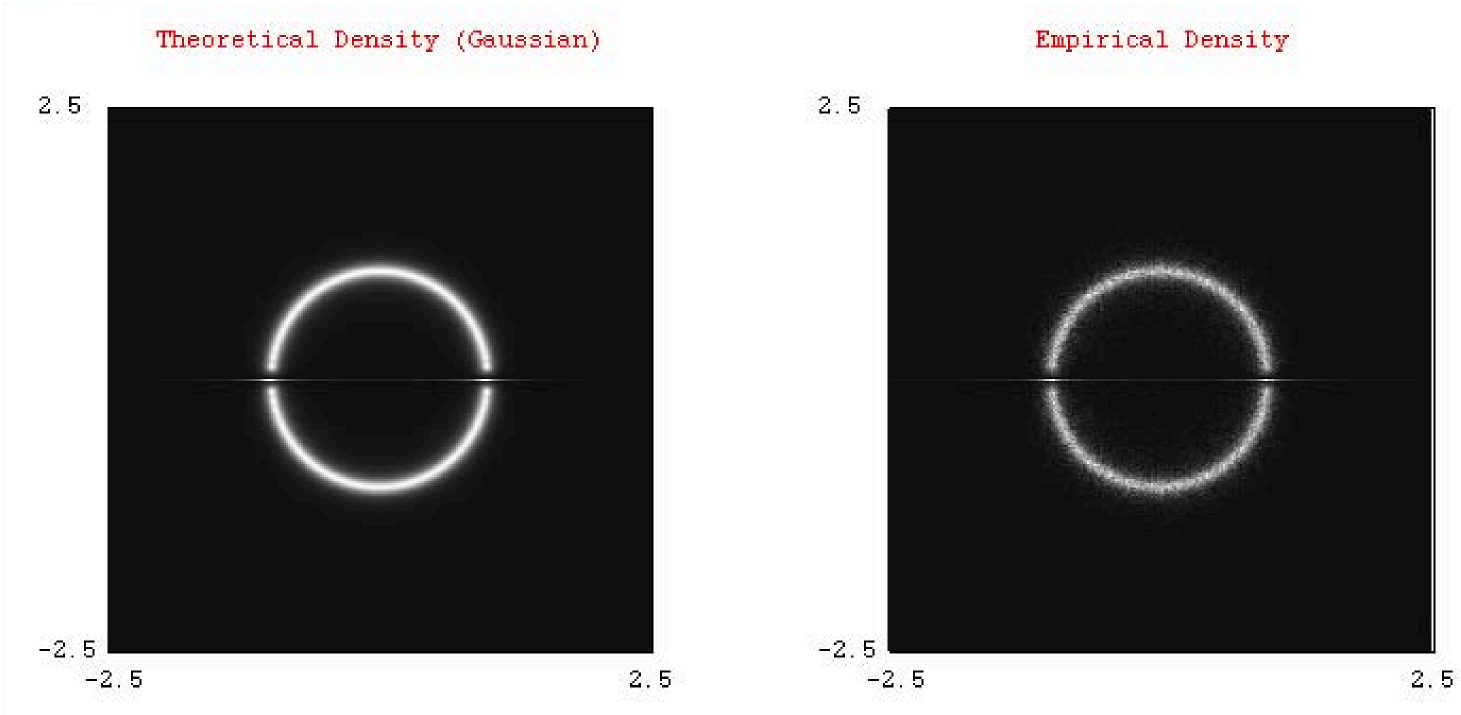
$$\nu_n(\Omega) = \frac{1}{2\pi i} \oint_{\partial\Omega} \frac{P_n'(z)}{P_n(z)} dz.$$

Taking the expectation operator inside the integral, we just need to evaluate the expected value of the ratio of two (dependent) Gaussian random variables.

The rest is tedious (and perhaps nontrivial) algebra.

Larger n

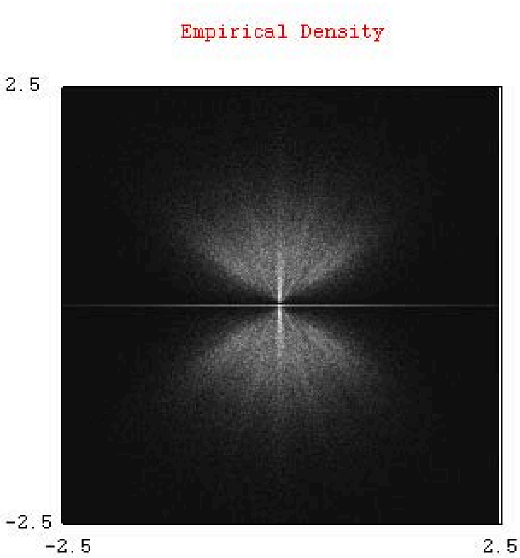
$$n = 36$$



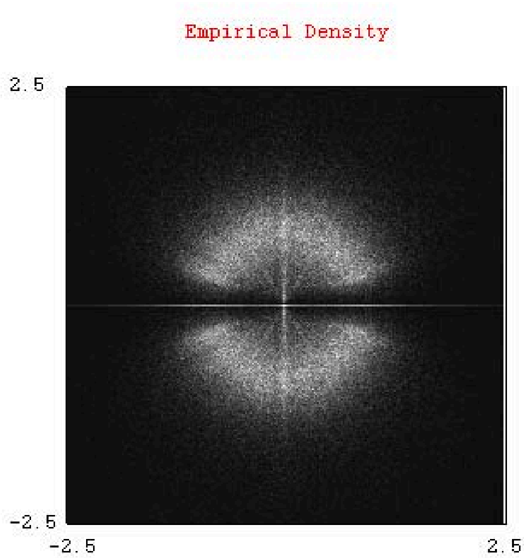
In the paper, we give an explicit formula for the limit as $n \to \infty$.

Cauchy Distribution

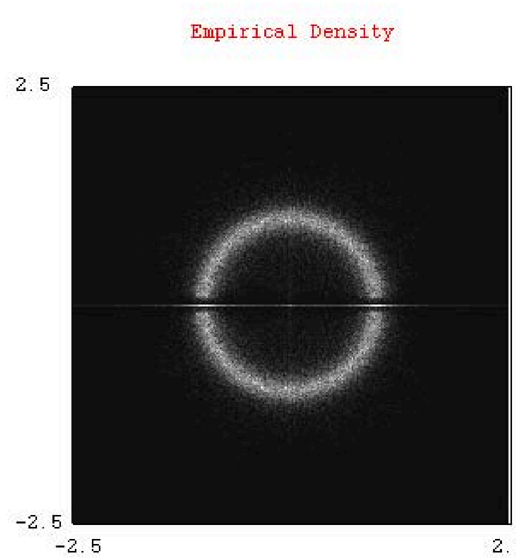
$n = 5$



$n = 10$



$n = 36$



Zeros of Random Polynomial References

- *The Complex Zeros of Random Polynomials*,
  Shepp and Vanderbei
  Transactions of the AMS, 347(11):4365-4384, 1995

The Dishonest Statistician

This is the era of *Big Data*: statistics trumps science. Cause-and-effect is passe. Correlation tells all.

Example: Consider a coin to be used in coin tossing.
It is important that the coin be *fair*: $p = q = 1/2$.

The coin in question appears to the eye to be perfectly symmetric. Physically, it seems fair (and the tosser is a well-respected member of the coin tossing community).

Such appeals to reason are not sufficient. A statistician has been hired to verify the fairness of the coin.

The statistician embarks on data collection. He instructs the tosser to toss the coin. The statistician keeps track of the number of flips, t, and the difference, X_t, between the number of heads and tails.

The Dishonest Statistician's Optimal Stopping Problem

Unfortunately, unbeknownst others, the statistician is dishonest and has a special interest in reporting that the coin is biased in favor of heads.

The only tool at the statistician's disposal is to stop the experiment at some point when heads outnumbers tails.

So, the statistician wants to solve an optimal stopping problem:

$$v(t,x) = \sup_{\tau \geq t} \mathbf{E}_{t,x}(X_\tau/\tau).$$

Of course, the statistician wants to know $v(0,0)$ but he also wants to know the strategy to achieve that "value" and to do that he needs to compute $v(t,x)$ for all t and x.

Hamilton-Jacobi-Bellman Equation

$$v(t,x) = \max \left(\frac{x}{t}, \frac{v(t+1,x+1)+v(t+1,x-1)}{2} \right)$$

Some Old Papers on the Dishonest Statistician

- L. Breiman, *"Stopping Rule Problems"*, Applied Combinatorial Mathematics, 1964.
  - Posed the problem.
- Y.S. Chow and H. Robbins, *"On Optimal Stopping Rules for S_n/n"*, Z. Wahrscheilichkeitstheorie und Verw. Gebiete, 1963.
  - There exist constants β_t such that the optimal stopping rule is to stop the first time that $X_t \geq \beta_t$.
- A. Dvoretzky, *"Existence and Properties of Certain Optimal Stopping Rules"*, Proc. 5th Berkeley Symp. Math. Statist. Prob., 1967.
  - Showed that $0.32 < \beta_t/\sqrt{t} < 4.06$.
- L.A. Shepp, *"Explicit Solutions to some Problems of Optimal Stopping"*, Annals of Mathematical Statistics, 1969.
  - Explicit formula for the limit

$$\alpha := \lim_{t\to\infty} \beta_t/\sqrt{t} = 0.83992369506\ldots$$

where

$$\alpha = (1-\alpha^2)\int_0^\infty e^{\lambda\alpha-\lambda^2/2}d\lambda$$

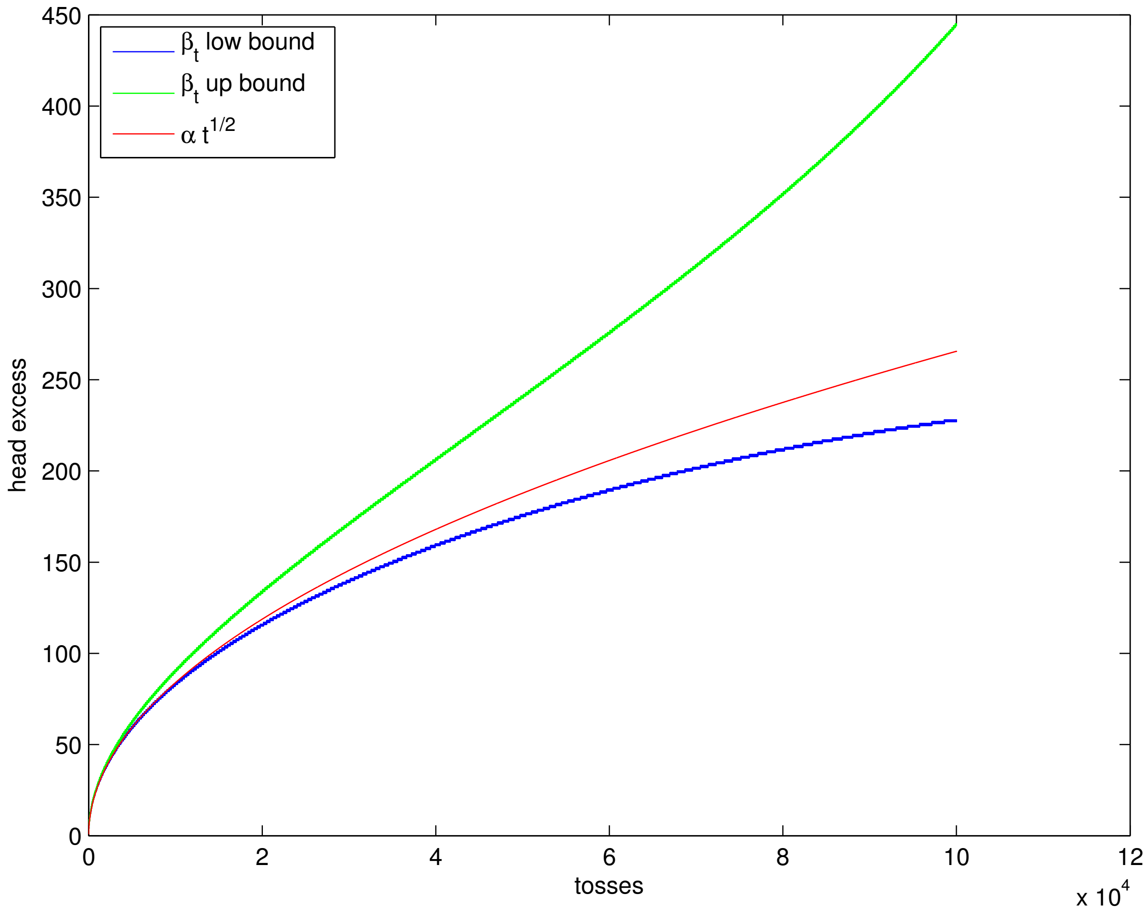
A Few New Papers

- Luis A. Medina, Doron Zeilberger,
  "An Experimental Mathematics Perspective on the Old, and still Open, Question of When To Stop?",
  in Gems in Experimental Mathematics, AMS Contemporary Mathematics series v. 517, 265–274

- O. Häggström and J. Wästlund,
  "Rigorous Computer Analysis of the Chow-Robbins Game",
  American Mathematical Monthly, 2013.

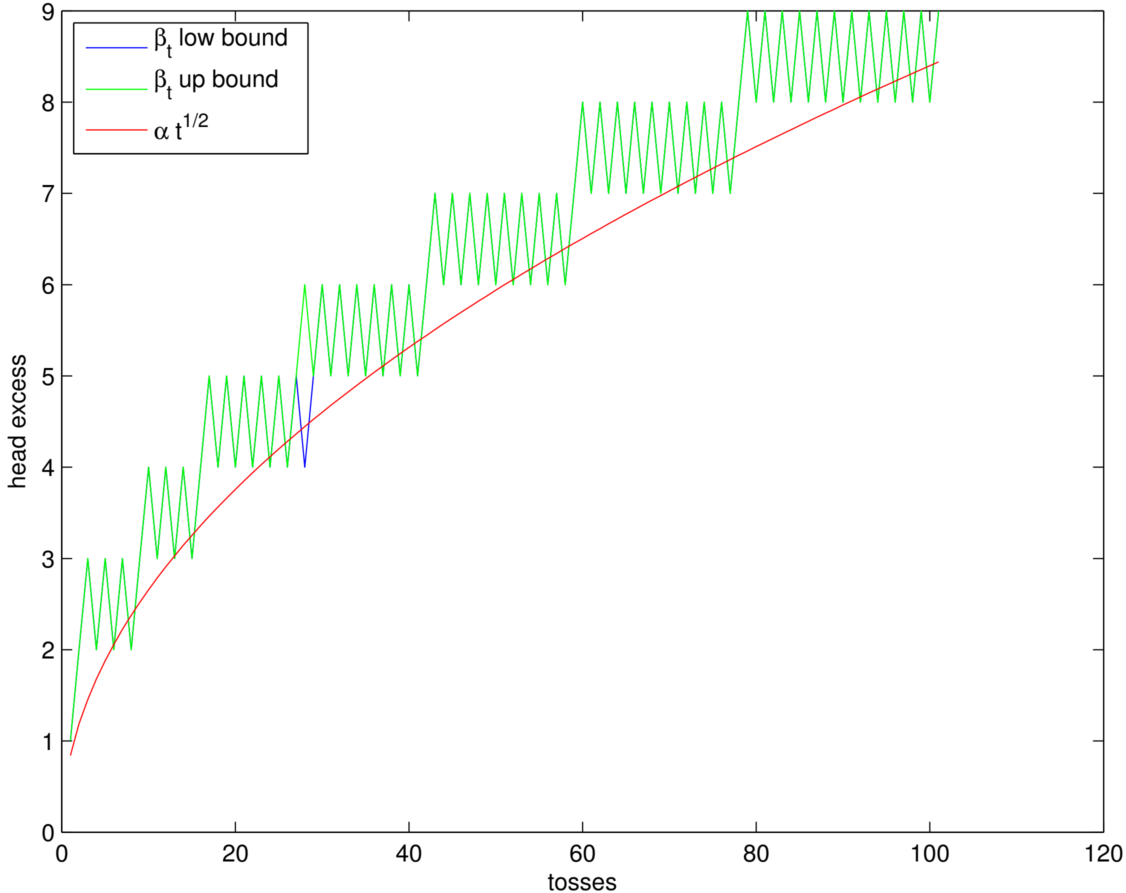
$$\max\left(\frac{x}{t},\ 0\right) \leq v(t,x) \leq \max\left(\frac{x}{t},\ 0\right) + \min\left(\frac{1}{2}\sqrt{\frac{\pi}{t}},\ \frac{1}{|x|}\right)$$

Stopping Thresholds



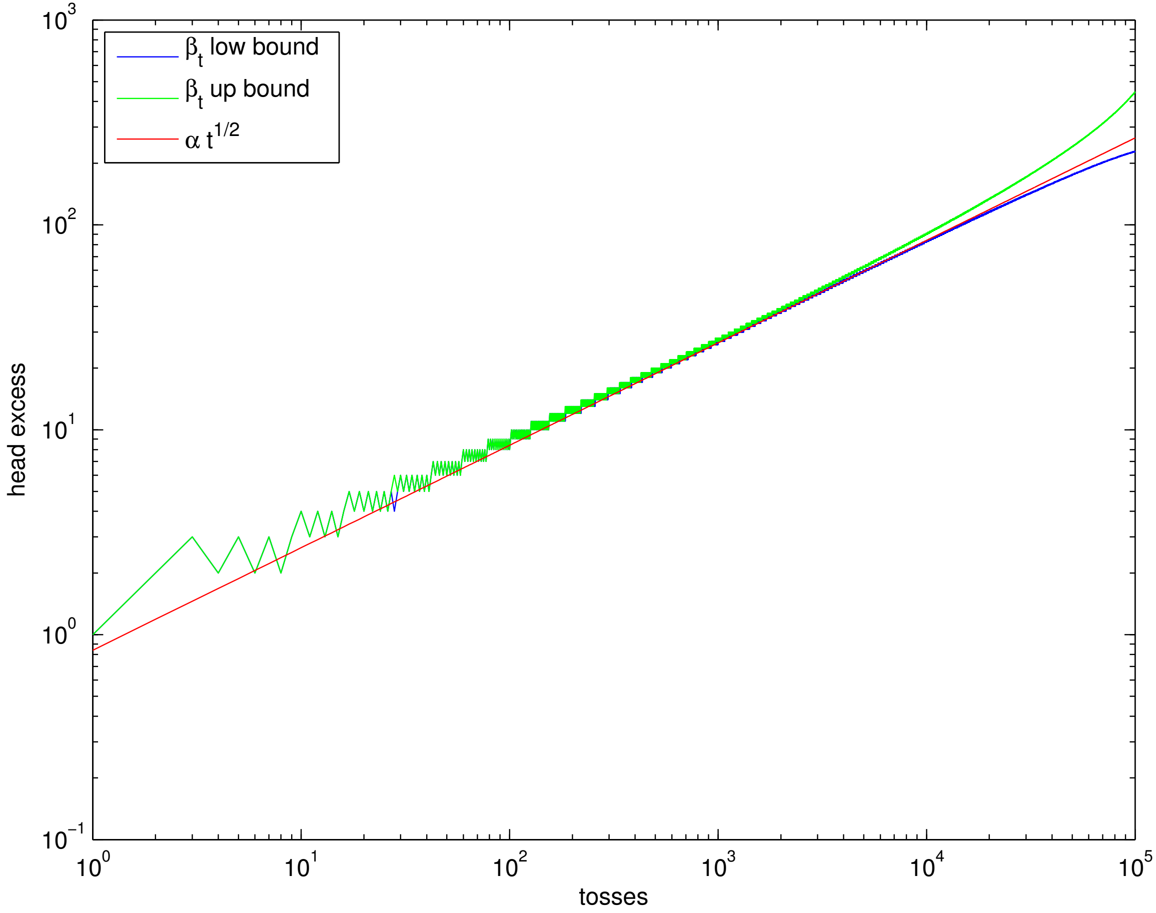
$$0.792945408 \;\leq\; \widehat{p} = (1 + v(0,0))/2 \;\leq\; 0.7929841248$$

Stopping Thresholds Near the Beginning



$$0.792945408 \;\le\; \widehat{p} \;=\; (1 + v(0,0))/2 \;\le\; 0.7929841248$$

Stopping Thresholds Log-Log Plot



$$0.792945408 \;\leq\; \widehat{p} \;=\; (1 + v(0,0))/2 \;\leq\; 0.7929841248$$

Slightly More Honest Statistician

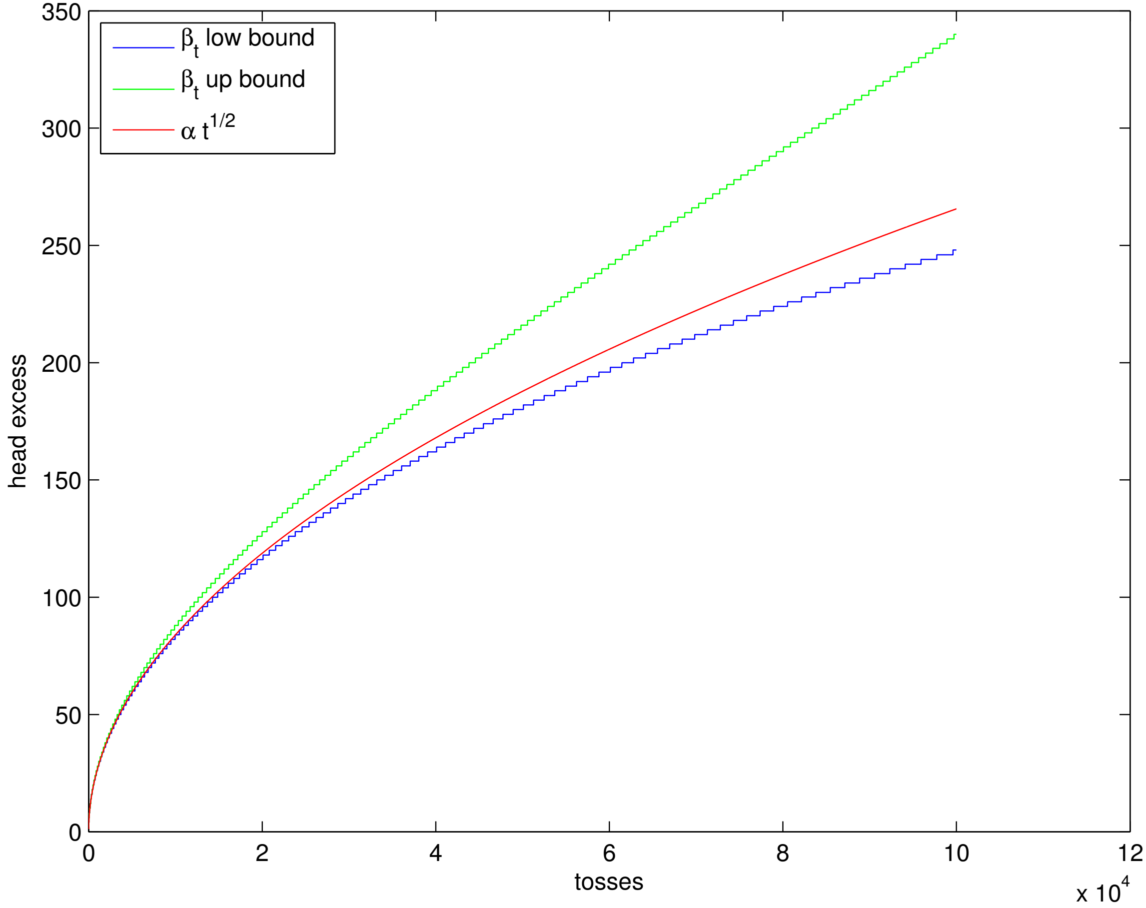
Even Number of Tosses Only

Only allowed to stop after an even number of tosses.

Modified Bellman equation:

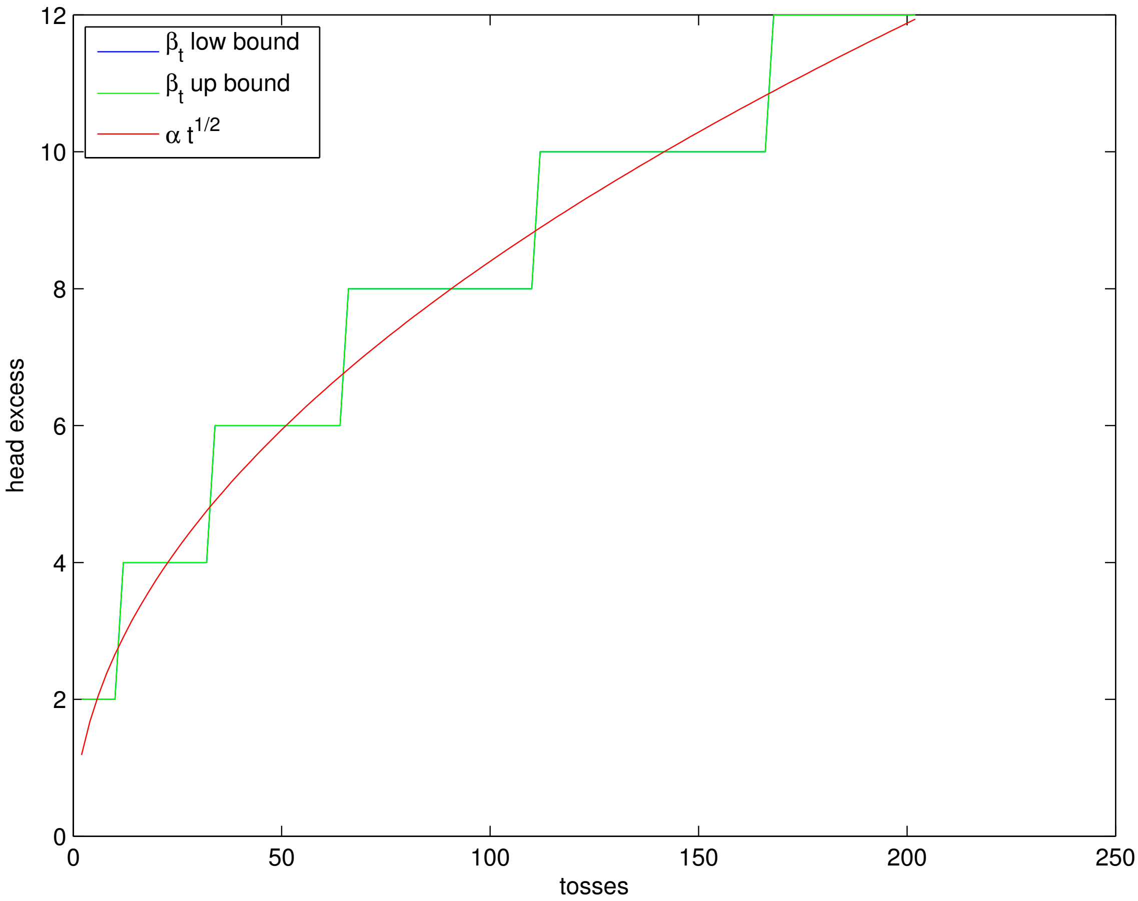
$$v(t,x) = \max\left(\frac{x}{t},\ \frac{v(t+2,x+2) + 2v(t+2,x) + v(t+2,x-2)}{4}\right).$$

Stopping Thresholds Even Number of Tosses Only



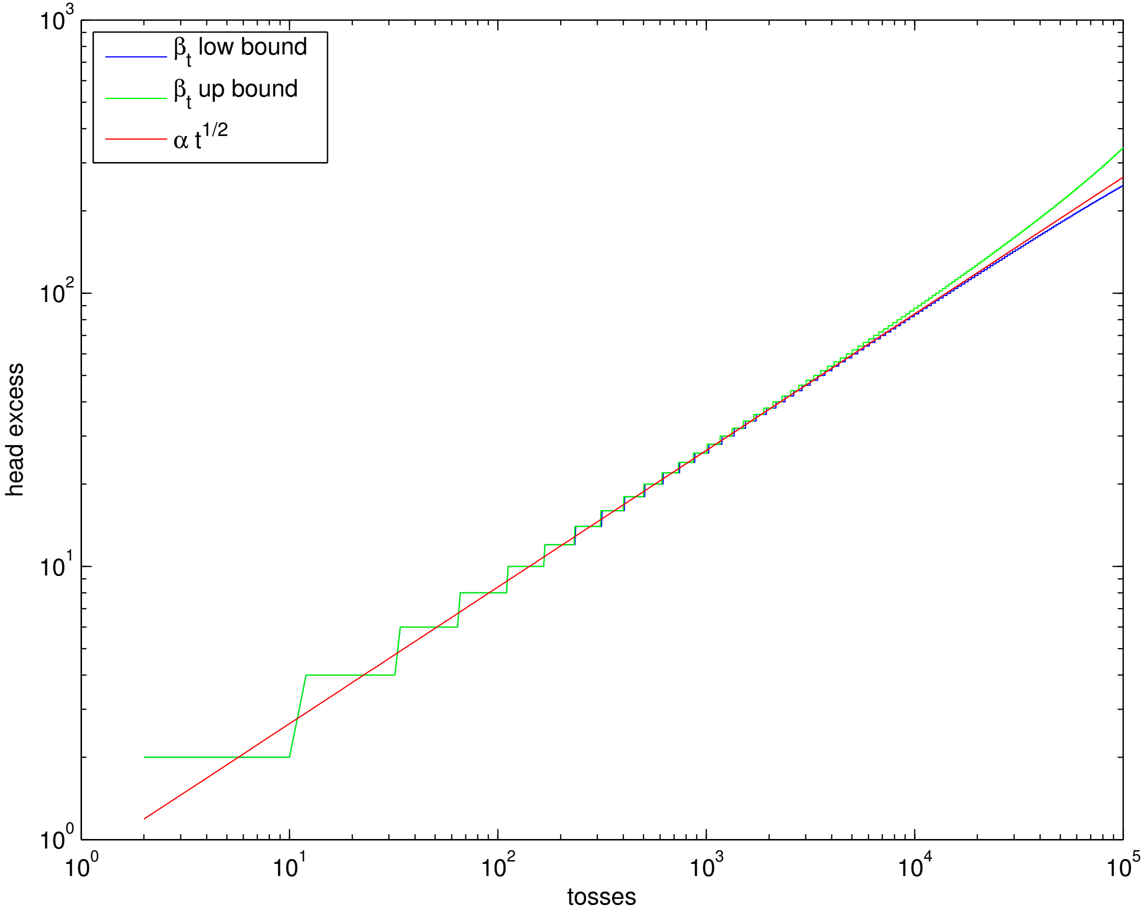
$$0.69730 \;=\; \widehat{p} \;=\; (1 + v(0,0))/2 \;=\; 0.69733$$

Stopping Thresholds Even Number of Tosses Only



$$0.69730 \;=\; \widehat{p} \;=\; (1 + v(0,0))/2 \;=\; 0.69733$$

Stopping Thresholds   Even Number of Tosses Only



$$0.69730 \;=\; \widehat{p} \;=\; (1 + v(0,0))/2 \;=\; 0.69733$$

Thank You!