Smoothing Linear Bernoulli Factory Functions

Divakar Kumar

OXWASP, University of Warwick, UK

Supervisor: Dr Krzysztof Latuszynski, University of Warwick, UK

Abstract

Given a coin that land heads with unknown probability p, how to construct an $f(p)-$coin for a given function f? 'Bernoulli Factory' is central to few areas of computational statistics and was first raised by von Neumann as a problem of sampling from an unbiased coin. A generic method developed by Nacu and Peres for the elbow function $f(p) = \min\{2p, 1-\epsilon\}$, relies on construction of lower and upper Bernstein polynomial approximations which is then used to address the problem of sampling from all real analytic functions. However, since elbow function f is not smooth, the Bernstein polynomial approximations converge slowly.

In many scenarios there is an upper bound on the inputs of the function. This allows to tweak f and consider a smooth version instead, allowing for fast Bernstein polynomial approximations. The aim of this report is to find such a smooth modification that has better convergence rate.

---

1. Introduction

The Bernoulli factory is an algorithm that takes input of i.i.d $p-$ coin tosses with unknown success probability $p \in S \subset [0,1]$ and outputs an $f(p)-$ coin toss, where $f : S \to [0,1]$ is known. The algorithm outputs a

Head or a *Tail* for $f(p)-$ coin depending on the outcome of first N tosses of $p-$ coin, where N is the stopping time of the algorithm. Necessary and sufficient conditions to make a draws from the $f(p)-$coin was derived by Keane and O'Brien. They prove that one can draw from $f(p)-$ coin if and only if f is constant or f is continuous and satisfies for some $n \geq 1$,

$$\min\{f(p), 1 - f(p)\} \geq \min\{p, 1 - p\}^n \tag{1}$$

That is why a Bernoulli factory algorithm cease to exist to make a draw from a $2p$ coin for $p \in [0, \frac{1}{2}]$ since the relation is not satisfied p closer to $\frac{1}{2}$. However for some $\epsilon > 0$, Bernoulli factory exists to draw from such coin for $p \in [0, \frac{1}{2} - \epsilon]$.

First problem concerning Bernoulli factory dates back to Von Neumann's problem of drawing from a constant factory function $f(p) = \frac{1}{2}$. Which is same as drawing from a unbiased coin based on the draws from a coin with unknown success probability p. The solution to this problem is rather straightforward. It needs tossing the $p-$ coin twice and declaring a head whenever HT occurs and a tail if TH occurs and continue tossing the $p-$ coin in the draws of HH and TT. This algorithm of making a draws for the unbiased coin works since HT and TH have equal probabilities of occurrences. Thus the stopping time of the algorithm lies in the set $\{2, 4, 6...\}$.

Nacu and Peres proposed a detailed algorithm to generate a single draw from the factory function $f(p) = 2p$ via the construction of Bernstein polynomials. However the algorithm requires working with the sets of exponential sizes. Hence the practical application of Nacu and Peres is rather limited.

2. Motivating Applications

2.1. MCMC for jump diffusion process

[Goncalves, Roberts, and K. Latuszynski, *Unpublished Work*] explains the MCMC algorithm for jump diffusion process with stochastic jump rate. This method has interesting application of Bernoulli factory problems. Jump diffusion process for $t \in [0, T]$ with stochastic jump rate is given by following model:

$$\gamma_t \sim Ornstein - Uhlenbeck(\theta_1) \tag{2}$$

$$\lambda_t = \exp(\gamma_t) \tag{3}$$

$$J_t \sim PoissonProcess(\lambda_t, d\Delta) \tag{4}$$

$$dV_t = \mu(V_{t-}, \theta_2)dt + \sigma(V_{t-}, \theta_2)dB_t + dJ_t \tag{5}$$

Here γ_t is unobeserved, mean reverting $Ornstein - Uhlenbeck$ process. J_t is a poisson process with intensity λ_t and distribution of the jump given by $d\Delta$. Process dV_t has drift μ, diffusion σ and depends on the jump size of the poisson process. Specification of full posterior for unobeserved processes will require drawing Gibbs samples that would alternate between $((J_t, V_t)|\cdot), (\lambda_t|\cdot), (\theta_1|\cdot)$, and $(\theta_2|\cdot)$. Since λ_t depends on γ_t thus updating $(\lambda_t|\cdot)$ would require first computation of $p(\gamma_t|\cdot)$. This is given as follows:

$$p(\gamma_t|\cdot) = p(\gamma_t|J_t) \propto p(\gamma_t) \exp\left\{ -\int_0^T e^{\gamma_t} dt + \sum_{j=1}^{N_j} \gamma_{t_j} \right\} \tag{6}$$

$$= P(\gamma_t) C_\gamma \exp\left\{ -\int_0^T e^{\gamma_t} dt \right\} \tag{7}$$

$$= p(\gamma) C_\gamma I(\gamma) \tag{8}$$

Here $C_\gamma = \exp\left\{\sum_{j=1}^{N_j} \gamma_{t_j}\right\}$, $I(\gamma) = \exp\left\{-\int_0^T e^{\gamma_t} dt\right\}$. In the above expression if we consider $p(\gamma)$ as the proposal distribution in the Metropolis step, then the acceptance rate is given by

$$\alpha(\gamma^{(i)}, \gamma^{(i+1)}) = \min\{1, C_{\gamma^{(i)},\gamma^{(i+1)}} I(\gamma^{(i)}, \gamma^{(i+1)})\} \tag{9}$$

where $C_{\gamma^{(i)},\gamma^{(i+1)}}$ is known constant and there is a method to generate events with probability $I(\gamma^{(i)}, \gamma^{(i+1)})$. Thus the proposal takes the following form:

$$f(p) = \min\{1, Cp\} \tag{10}$$

Above proposal function take the form of elbow function and samples from it can be drawn via Bernoulli factory algorithms.

2.2. MCMC for Markov switching diffusion

Another interesting application of Bernoulli factory problems is due to [K. Latuszynski, J. Palczewski, G.O. Roberts, *Unpublished Work*]. Here Y_t for $t \in [0, T]$ is a continuous time Markov chain defined on the finite state space $\{1, 2, 3, ..., m\}$ with the intensity matrix Λ. The model is given below

$$\begin{aligned} Y_t &\sim \text{Continuous time Markov chain on}\{1, 2, ..., m\} \\ dV_t &\sim \mu(\theta, V_t, Y_t)dt + \sigma(\theta, V_t)\gamma(\theta, Y_t)dB_t \end{aligned}$$

To make inferences on the unobserved process Y_t, following needs to be computed

$$p(Y_t|\cdot) = p(Y_t|\Lambda)G(\theta, Y_t, V_t) \tag{11}$$

Here $G(\theta, Y_t, V_t)$ is found using Girsanov's theorem. In the above expression if we consider $p(Y_t|\Lambda)$ as the proposal density then acceptance rate in the Metropolis step can be given by

$$\alpha(Y^{(i)}, Y^{(i+1)}) = \min\{1, C_{\gamma^{(i)},\gamma^{(i+1)}} \bar{G}(\theta, V_t, (Y^{(i)}, Y^{(i+1)}))\} \tag{12}$$

Here $C_{\gamma^{(i)},\gamma^{(i+1)}}$ is a known constant and simulations can be drawn on events with probability $\bar{G}(\theta, V_t, (Y^{(i)}, Y^{(i+1)}))$ which depends on $G(\theta, Y_t, V_t)$. Thus the proposal takes the following form:

$$f(p) = \min\{1, Cp\} \tag{13}$$

Above proposal function take the form of elbow function again and samples from it can be drawn via Bernoulli factory algorithms.

2.3. Perfect sampling for Markov chains

Another motivating application of Bernoulli factory can be seen in [Flegal and Herbei, 2005]. For this we consider $\{X_n\}_{n\geq 0}-$ an ergodic Markov chain with transition kernel P whose limiting distribution is given by π. Under mild *minorization* condition on $\{X_n\}_{n\geq 0}$, transition kernel P can be represented as

$$P(x, \cdot) = s(x)v(\cdot) + (1 - s(x))R(x, \cdot) \tag{14}$$

Transition kernel P can be seen as mixture of two Markov transition kernels with respective probabilities $s(x)$ and $1 - s(x)$. Thus to sample according to P is first flipping a coin with probability $s(x)$ then deciding whether to sample from $v(\cdot)$ or $R(x, \cdot)$. Let τ be the first time when necessity to sample according to $v(\cdot)$ arises, then p_n is defined on the set of natural numbers via

$$p_n = \frac{P(\tau > n)}{E(\tau)} \quad \text{for} \quad n \in \{1, 2, 3, ...\} \tag{15}$$

Thus under *minorization* condition on X, for any n and measurable set A, if $Q_n(A) = P(X_n \in A|\tau > n)$, π can be decomposed as follows:

$$\pi(\cdot) = \sum_{n=1}^{\infty} p_n Q_n(\cdot) \tag{16}$$

Suppose there exists a probability distribution $d : N \to [0, 1]$ such that for some $M > 0$, $P(\tau > n) \leq Md(n)$. Thus we can write

$$p_n = \frac{P(\tau > n)}{E(\tau)} = \frac{P(\tau > n)}{E(\tau)d(n)} d(n) \tag{17}$$

Thus we can reject $d(n)$ with probability $\frac{P(\tau>n)}{E(\tau)d(n)}$. Since $P(\tau > n) \leq Md(n)$, following can be used to reject $d(n)$

$$\frac{P(\tau > n)}{Md(n)} = \frac{1}{Md(n)} P(\tau > n) \tag{18}$$

$$= Cp < 1 \tag{19}$$

Here $C = 1/(Md(n))$ and $p = P(\tau > n)$. Simulations can be drawn easily from events with probability p. Thus it takes the form of Bernoulli factory problems.

3. Converging Polynomial Envelopes

In this section I am going to state the result which promises the existence of upper and lower converging polynomials for the factory function. But first we define the concept of Bernstein polynomails which are used to contruct the upper and lower envelopes for factory function.

Definition: Bernstein Polynomials for a real valued function f defined over $[0,1]$ is given by

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} \tag{20}$$

Nacu and Peres [1] devised the necessary and sufficient conditions on the existence of the upper and lower Bernstein approximations for the factory function. The result is stated below.

Proposition [Nacu & Peres, 2005] : If there exist an algorithm which can simulate a factory function f on a set $S \subset (0,1)$ if $\forall n \geq 1$ there exist polynomials

$$g_n(x,y) = \sum_{k=0}^{n} a\,(n,k)\binom{n}{k} x^k (1-y)^{n-k} \tag{21}$$

$$h_n(x,y) = \sum_{k=0}^{n} b\,(n,k)\binom{n}{k} x^k (1-y)^{n-k} \tag{22}$$

satisfying the following properties:

1. $0 \leq a\,(n,k) \leq b\,(n,k) \leq 1$.
2. $a\,(n,k)\binom{n}{k}$ and $b\,(n,k)\binom{n}{k}$ are integers.
3. $\lim_{n\to\infty} g_n(p,1-p) = f(p) = \lim_{n\to\infty} h_n(p,1-p) \quad \forall \quad p \in S$.
4. $\forall \quad m < n$, the following conditions on $a(n,k)$ and $b(n,k)$ are satisfied:

$$a(n,k) \geq \sum_{i=0}^{k} \frac{\binom{n-m}{k-i}\binom{m}{i}}{\binom{n}{k}} a(m,i)$$

$$b(n,k) \leq \sum_{i=0}^{k} \frac{\binom{n-m}{k-i}\binom{m}{i}}{\binom{n}{k}} b(m,i)$$

Converesely if there exists such upper and lower polynomials satisfying above four conditions then there exists an algorithm to simulate from f on S such that the ruuning time of the algorithm N satisfies

$$P(N > n) = h_n(p, 1-p) - g_n(p, 1-p)$$

■

The sequence of lower polynomials $g_n(p, 1-p)$ are increasing and upper polynomials $h_n(p, 1-p)$ are decreasing. If factory function f is continuous then Bernstein polynomials converges uniformly on unit interval $[0, 1]$. If the function f linear on some interval the Bernstein polynomials have exponential convergence. I present the result concerning the rate of convergence for function having different characteristics. First lets see hypergeometric distribution and its moments which is needed for lemma concerning rate of convergence.

Definition: If X is a random variable having hypergeometric distribution $H(2n, k, n)$ if its mass function is given by

$$P(X = i) = \frac{\binom{n}{i}\binom{n}{k-i}}{\binom{2n}{k}} \quad \text{for} \quad i = 0, 1, ..., n \tag{23}$$

with the mean and variance of $\frac{X}{n}$ given by

$$E\left(\frac{X}{n}\right) = \frac{k}{2n} \tag{24}$$

$$\text{Var}\left(\frac{X}{n}\right) = \frac{k(2n-k)}{4(2n-1)n^2} \leq \frac{1}{4(2n-1)} \tag{25}$$

It is easy to derive the mean and variance of $H(2n, k, n)$ while the upper bound on the variance follows from the fact that $\frac{k(2n-k)}{n^2} \leq 1$ which implies

$$\frac{k(2n-k)}{4(2n-1)n^2} \leq \frac{1}{4(2n-1)} \tag{26}$$

This bound is slightly better than one used in [1] which is $\frac{1}{2n}$. Presented below is the result for the rate of convergence for functions with different characteristics

Lemma [Nacu & Peres,2005]: If X is a random variable having hypergeometric distribution $H(2n, k, n)$ and let $f : [0, 1] \to R$ be a factory function with $|f| \leq 1$, then following conditions holds:

1. If f is Lipschitz with Lipschitz constant C then
$$\left| Ef\left(\frac{X}{n}\right) - f\left(\frac{k}{2n}\right)\right| \leq \frac{C}{\sqrt{2n}}$$
2. If f is twice differentiable with bounded second derivative i.e. $|f''| \leq C$, then
$$\left| Ef\left(\frac{X}{n}\right) - f\left(\frac{k}{2n}\right)\right| \leq \frac{C}{8(2n-1)}$$
3. For some $x > 0$ if f is linear on $[\frac{k}{2n} - x, \frac{k}{2n} + x]$, then

$$\left| Ef\left(\frac{X}{n}\right) - f\left(\frac{k}{2n}\right)\right| \leq (2|C| + 4)\exp(-2x^2 n)$$

In above lemma, it should be noticed that right hand side bound for the twice differentiable functions is different from the one given in [1]. This is due to (26). I am reformulating the proof in [1] by considering (26)

Proof of Lemma part 2: Since f is twice differentiable, error bounds on second order Taylor expansion of the function follows the following relation:

$$\left| f\left(\frac{X}{n}\right) - f\left(\frac{k}{2n}\right) - \left(\frac{X}{n} - \frac{k}{2n}\right) f'\left(\frac{k}{2n}\right)\right| \leq \frac{1}{2}\left(\frac{X}{n} - \frac{k}{2n}\right)^2 \sup|f''|$$

Since $E\left(\left(\frac{X}{n}-\frac{k}{2n}\right)f'\left(\frac{k}{2n}\right)\right)=0$, this implies that

$$E\left[f\left(\frac{X}{n}\right)-f\left(\frac{k}{2n}\right)-\left(\frac{X}{n}-\frac{k}{2n}\right)f'\left(\frac{k}{2n}\right)\right]=E\left[f\left(\frac{X}{n}\right)-f\left(\frac{k}{2n}\right)\right]$$

Also for any function g, $E(|g(X)|)\geq|Eg(X)|$, thus it follows that

$$\begin{aligned}\left|E\left[f\left(\frac{X}{n}\right)-f\left(\frac{k}{2n}\right)\right]\right| &\leq E\left|f\left(\frac{X}{n}\right)-f\left(\frac{k}{2n}\right)-\left(\frac{X}{n}-\frac{k}{2n}\right)f'\left(\frac{k}{2n}\right)\right| \\ &\leq E\left[\frac{1}{2}\left(\frac{X}{n}-\frac{k}{2n}\right)^2\sup|f''|\right] \\ &\leq \frac{C}{2}E\left(\frac{X}{n}-\frac{k}{2n}\right)^2 \\ &\leq \frac{C}{2}\mathrm{Var}\left(\frac{X}{n}\right) \\ &\leq \frac{C}{8(2n-1)}\end{aligned}$$

■

Linear Bernoulli factory $f(p)=\min\{Cp,1-\epsilon\}$ is linear on the interval $[0,\frac{1-\epsilon}{C})$, thus the Bernstein polynomials converges at the exponential rate to factory function on this interval only. Bernstein polynomials for these factory functions have $\frac{1}{\sqrt{n}}$ rate of convergence to the right of elbow since these are Lipschitz with Lipschitz constant C, while twice differentiable functions with bounded second derivative have $\frac{1}{n}$ rate of convergence. Since the function $f(p)=\min\{Cp,1-\epsilon\}$ is not differentiable, thus there is a need to make a smooth extension of the function $f(p)$ to achieve a better rate of convergence.

4. Building candidate smooth functions

4.1. Smooth extension using Normal distribution function

Here I am going to present an idea to extends our function $f(p) = \min\{Cp, 1-\epsilon\}$ after the elbow point $a = \frac{1-\epsilon}{C}$. The idea is to choose a function $g(p)$ which is concave and at the same time bounded above by 1. Together with above properties, it should have characteristics of original factory function at the elbow point. This implies

$$g(a) = Ca \tag{27}$$

$$g'(a) = C \tag{28}$$

$$g''(a) = 0 \quad \text{where} \quad a = \frac{1-\epsilon}{C} \tag{29}$$

Keeping (27) and (29) in mind and at the same time respecting the concavity property, lets us choose a function involving a normal distribution function as potential candidate after the elbow point $a = \frac{1-\epsilon}{C}$. Thus

$$g(p) = \Phi\left(\frac{p-a}{\sigma}\right) + Ca - \frac{1}{2} = \int_{-\infty}^{p} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx + Ca - \frac{1}{2}$$

Here we find that $g(p)$ is infinitely differentiable together with the property of concavity. Properties (27) and (29) are automatically satisfied since,

$$\begin{aligned}
g(a) &= \Phi\left(\frac{a-a}{\sigma}\right) + Ca - \frac{1}{2} \\
&= \Phi(0) + Ca - \frac{1}{2} \\
&= \frac{1}{2} + Ca - \frac{1}{2} \\
&= Ca \\
g''(a) &= -\frac{(a-a)}{\sigma^2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \\
&= 0
\end{aligned}$$

Property (28) can be satisfied with the choice of σ as the solution of (2) equation. Which implies,

$$\begin{aligned}
g'(a) &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\{-\frac{(a-a)^2}{\sigma^2}\} = C \\
\sigma &= \frac{1}{\sqrt{2\pi}C}
\end{aligned}$$

This candidate function faces a potential threat of exceeding the bound of 1. Since every parameter in this function has been fixed, we need to include extra parameters so that these free parameters can be tuned to bound the function by 1. Thus we make following choice of the function after the elbow point:

$$g(p) = \sum_{i=1}^{2} a_i \Phi\left(\frac{p-a}{\sigma_i}\right) + Ca - \frac{1}{2} = \sum_{i=1}^{2} a_i \int_{-\infty}^{p} \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_i}\exp\left\{-\frac{(x-a)^2}{\sigma_i^2}\right\}dx + Ca - \frac{1}{2}$$

Properties (27) and (28) puts the following constraints on the parameters:

$$\begin{aligned}
a_1 + a_2 &= 1 \\
\frac{a_1}{\sigma_1} + \frac{a_2}{\sigma_2} &= \sqrt{2\pi}C
\end{aligned}$$

while property (29) is automatically satisfied with the choice of such functions since a is the mode of both of normal densities included in above function. This choice of function gives us the flexibility with one of the a_i's and one of the σ_i's while fixes the others because of the above equations (27)-(29). Thus we can make wise choice of flexible parameters to restrict the function below one. Since $g(p)$ is defined for $p \in [a, 1]$ and it is an increasing function of p, the bound condition implies that

$$g(1) = \sum_{i=1}^{2} a_i \Phi\left(\frac{1-a}{\sigma_i}\right) + Ca - \frac{1}{2} \leq 1 \tag{30}$$

It is easy to see that,

$$\Phi\left(\frac{1-a}{\sigma_i}\right) = \frac{1}{2} + \delta_i \quad \text{for} \quad \delta_i > 0, \quad i = 1, 2$$

Plugging this in the inequality (25) implies the following:

$$\sum_{i=1}^{2} a_i \left(\frac{1}{2} + \delta_i\right) + Ca - \frac{1}{2} \leq 1$$

$$a_1\delta_1 + a_2\delta_2 \leq \epsilon \quad \text{since} \quad \epsilon = 1 - Ca$$

Last inequality follows since $a_1 + a_2 = 1$. Thus last inequality upper bounds the convex combination of δ_1 and δ_2 by ϵ. Without loss of generality we choose a_1 and σ_1. Choice of σ_1 fixes δ_1. One possible way to make a good choice of a_1 and σ_1 is by choosing both really large which makes δ_1 really small. Although above choice would make δ_2 close to $\frac{1}{2}$. But the convex combination $a_1\delta_1 + a_2\delta_2$ would bring the final value closer to δ_1 because of the arbitrary large choice of a_1. Formally, for large enough $n \in N$, choose $\delta_1 = \frac{\epsilon}{na_1}$ for pre-set a_1. If this choice of n does not satisfy the inequality $a_2\delta_2 \leq \epsilon\frac{n-1}{n}$, we move to higher values of n.

Thus the smooth version of the factory function $f(p)$ is given by

$$h(p) = \begin{cases} Cp & \text{if} \quad p < a \\ \sum_{i=1}^{2} a_i \Phi\left(\frac{p-a}{\sigma_i}\right) + Ca - \frac{1}{2} & \text{if} \quad p \geq a \end{cases} \tag{31}$$

Second derivative of smoothed function $h(p)$ is given by

$$h''(p) = \begin{cases} 0 & \text{if} \quad p < a \\ -\sum_{i=1}^{2} a_i \left(\frac{p-a}{\sigma_i^2}\right) \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_i} \exp\left\{-\frac{(p-a)^2}{2\sigma^2}\right\} & \text{if} \quad p \geq a \end{cases} \tag{32}$$

Since the function $F(x) = xe^{-x^2/2}$ defined for $x > 0$ satisfies $|F(x)| \leq e^{-1/2}$. Thus is is easy to see that second derivative of the smoothed function satisfies

$$|h''(p)| \leq \left(\sum_{i=1}^{2} \frac{a_i}{\sigma_i^2}\right) \frac{1}{\sqrt{2\pi}} e^{-1/2} \tag{33}$$

Here we get the bounds on the second derivative of the function depending on a_i's and σ_i's. The extended function is twice differentiable with bounded second derivative with bounds depending on C, This promises better rate of convergence due to part (2) of the lemma seen in the previous section.

4.2. Less smooth quadratic extension of elbow function

Here I am going to present another idea to extend the elbow function smoothly using a quadratic function after the elbow point a. The reason to choose the quadratic function is the simplicity of the function and the ease of analytical calculations to get the explicit bound on the second derivative of the function. The idea is to glue a quadratic function at the elbow point a then extend it to the point where it reaches the maximum and then glue a constant function thereafter. Further, we need to make sure that function does not exceeds the bound of 1 and it should stay concave. Thus we make the following choice of the quadratic function after the elbow point a.

$$g(p) = \begin{cases} q(p) = -\alpha(p-a)^2 + \beta(p-a) + \gamma & p \in [a, a^*] \\ q(a^*) & p > a^* \end{cases} \tag{34}$$

Here α is positive which makes sure that we have concave extension. a^* is the point of maxima of the function $q(p)$. The resultant smooth function would be less smoother than previous smooth extension we have seen. That is here we relax the second derivative matching criteria otherwise all three parameter would be fixed and we would lose control on the bounds of the function. Thus we would like the following conditions to hold:

$$g(a) = Ca \tag{35}$$

$$g'(a) = C \tag{36}$$

Above two conditions imply that $\gamma = Ca$ and $\beta = C$. Thus the extended part becomes:

$$g(p) = \begin{cases} q(p) = -\alpha(p-a)^2 + C(p-a) + Ca & p \in [a, a^*] \\ q(a^*) & p > a^* \end{cases} \tag{37}$$

In the above, point of maxima of the quadratic function satisfies $-2\alpha(a^* - a) + C = 0$, thus $(a^* - a) = C/(2\alpha)$. Since we want $g(a^*) < 1$

which implies that

$$-\alpha\left(\frac{C^2}{4\alpha^2}\right)+\frac{C^2}{2\alpha}+Ca<1 \tag{38}$$

$$\frac{C^2}{4\alpha}<1-Ca \tag{39}$$

$$\alpha>\frac{C^2}{4\epsilon} \tag{40}$$

The last inequality follows since $\epsilon = 1 - Ca$. Choosing a sensible value of α greater than $\frac{C^2}{4\epsilon}$ would glue the function at the elbow point a. Then we glue a constant function equal to maximum value $\frac{C^2}{4\alpha} + Ca$ of the quadratic function at the point of maxima a^*. Thus the smoothed extension of elbow function is given as

$$h(p)=\begin{cases} Cp & p\in[0,a) \\ -\alpha(p-a)^2+C(p-a)+Ca & p\in[a,a^*] \\ \frac{C^2}{4\alpha}+Ca & p>a^* \end{cases} \tag{41}$$

The second derivative of the above function is given by

$$h''(p)=\begin{cases} 0 & p\in[0,a) \\ -2\alpha & p\in[a,a^*] \\ 0 & p>a^* \end{cases} \tag{42}$$

Thus $|h''(p)| \leq 2\alpha$ which promises better rate of convergence due to part 2 of lemma.

5. Concluding remark

The first smoothed function defined in the last section can be used to smooth any elbow function of the kind $f(p) = \min\{Cp, 1-\epsilon\}$. Also it is easy to see that parameters of the function namely a_i's and σ_i's depend on C. Thus for really small ϵ and large values of C, the change in the derivative at the elbow point is substantial, which would result in bound on the second derivative being larger.

Although smoothing the elbow function seems a nice idea at first look since its advantage can be seen in terms of better rate of convergence, it faces a problem. In the proposition by Nacu and Peres running time of the algorithm satisfies: $P(N > n) = h_n(p, 1-p) - g_n(p, 1-p)$ which is of order $1/n$ for the upper and lower polynomails defined in [1]. Since $E(N) = \sum_n P(N > n)$. Thus the running time N of the algorithm does not have finite expectations. A sensible construction of Bernstein polynomials can help to overcome this issue.

6. References

[1] Serban Nacu , Yuval Peres, *Fast simulation of coin from old*, The Annals of Applies Probability, 2005.

[2] James M. Flegal, Radu Herbei, *Exact sampling for intractable probability distributions via a Bernoulli factory*, 2010.

[3] M. S. Keane , G.L. O'Brien, *A Bernoulli Factory*, 1994.

[4] K. Latuszynski, J. Palczewski, G.O. Roberts, *MCMC for Markov switching diffusions*, Unpublished Work

[5] F. Goncalves, G.O. Roberts, K. Latuszynski, *MCMC for jump diffusions with stochastic jump rate*, Unpublished Work