Improved Runtime Bounds for the Univariate Marginal Distribution Algorithm via Anti-Concentration

Phan Trung Hai Nguyen & Per Kristian Lehre
{PXN683, P.K.Lehre}@cs.bham.ac.uk

July 18th, 2017

School of Computer Science
University of Birmingham
Birmingham B15 2TT
United Kingdom

Outlines

Background

- Estimation of Distribution Algorithms (EDAs)
- Univariate Marginal Distribution Algorithm (UMDA)

Useful tools

- Level-based theorem
- Anti-concentration bound
- Feige's inequality

Our result

- Improved upper bound on runtime of UMDA on OneMax

Conclusion

Estimation of Distribution Algorithms (EDAs)

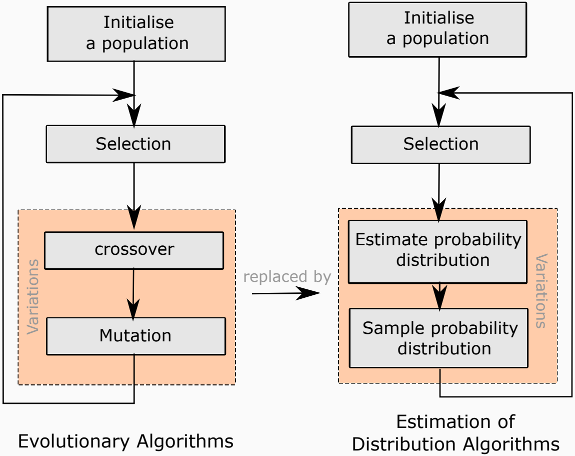


Figure 1: Comparison between EAs and EDAs (Shakya 2006)

Univariate Marginal Distribution Algorithm (UMDA)

Solution representation: $(X_1, \ldots, X_n) \in \{0,1\}^n$.

Probabilistic model at generation t: $\mathcal{M}_t := (p_t^1, p_t^2, \ldots, p_t^n)$.

Marginal probabilities: $p_t^i := \Pr(X_i = 1)$.

Algorithm 1: UMDA

begin
- initial model $\mathcal{M}_0 = (1/2, 1/2, \ldots, 1/2)$
- **repeat**
  - Sample P_t of λ individuals using joint probability

$$\Pr(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} \Pr(X_i).$$

  - Select the μ best individuals according to fitness.
  - Update the probabilistic model

$$p_{t+1}^i := \frac{1}{\mu}\sum_{j=1}^{\mu} X_i^{[j]} \in \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$$

  where $X_i^{[j]}$ is the i-th bit of the j-th individual.
- **until** *termination condition fulfilled*

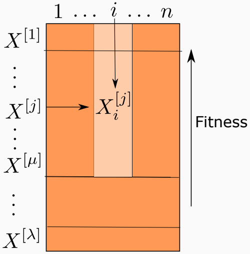


Figure 2: UMDA illustration

Previous Work

	(1+1) EA	UMDA	
LeadingOnes	$\Theta(n^2)$	$\mathcal{O}(n\lambda \log \lambda + n^2)$	(Dang & Lehre 2015)[1]
BVLeadingOnes	$\Theta(n \log n)$	∞ w.o.p.	(Chen et al. 2010)
		$\mathcal{O}(n\lambda)$	(Chen et al. 2010)[2]
SubString	$2^{\Omega(n)}$ w.o.p.	$\mathcal{O}(n\lambda)$	(Chen et al. 2009)[4]
OneMax	$\Theta(n \log n)$	$\mathcal{O}(n\lambda \log \lambda)$	(Dang & Lehre 2015)[3]
		$\Omega(\mu\sqrt{n} + n \log n)$	(Krejca & Witt 2017)[3]

[1] $\lambda = \Omega(\log n)$
[2] $\lambda = \omega(n^2)$
[3] $\lambda = (1 + \Theta(1))\mu$

Open Question

Runtime analysis of UMDA on OneMax (latest results):

- Upper bound: $\mathcal{O}(n \log n \log \log n)$ by Dang & Lehre (Dang & Lehre 2015).
- Lower bound: $\Omega(\mu\sqrt{n} + n \log n)$ by Krejca & Witt (Krejca & Witt 2017).

The upper and lower bounds are still different by $\Theta(\log \log n)$.

Open Question: Could this gap be closed?

Theorem

The expected optimisation time of UMDA with

- $c \log n \leq \mu \leq c' \sqrt{n}$ for some constants $c, c' > 0$,
- $\lambda \geq a\mu$ for some constant $a \geq 13e$

on OneMax is

$$\mathcal{O}(n\lambda).$$

Intuition: UMDA can optimise OneMax within $\mathcal{O}(n)$ generations.

UMDA **on** ONEMAX ($n = 1000$)

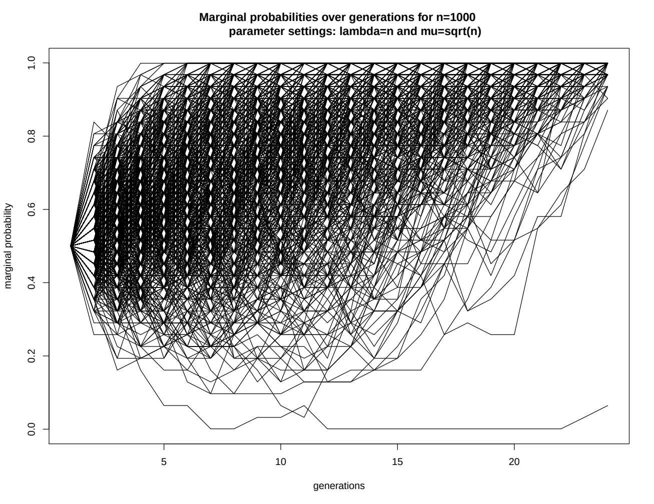
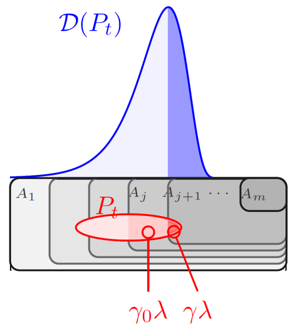


Figure 3: Marginal probabilities of UMDA on ONEMAX for $n = 1000$

Level-Based Theorem[4]

Provide **upper bounds** on the expected runtime of some population-based algorithms on a wide range of optimisation problems.



$y \sim \mathcal{D}(P_t)$

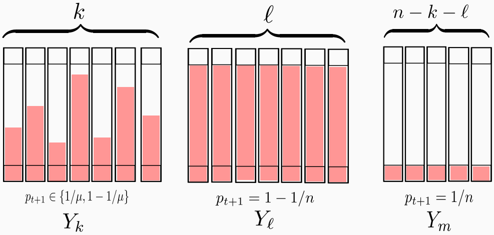
- **(G1)** $j \in [m-1]$, if $|P_t \cap A_{\geq j}| \geq \gamma_0 \lambda$ then
$$\Pr\left(y \in A_{\geq j+1}\right) \geq z_j.$$
- **(G2)** similar to (G1) and $|P_t \cap A_{\geq j+1}| \geq \gamma\lambda$ then
$$\Pr\left(y \in A_{\geq j+1}\right) \geq (1+\delta)\,\gamma.$$
- **(G3)** and the population size $\lambda \in \mathbb{N}$ satisfies
$$\lambda \geq \left(\frac{4}{\gamma_0 \delta^2}\right) \ln\left(\frac{128m}{z_* \delta^2}\right)$$
where $z_* := \min_{j \in [m-1]}\{z_j\}$, then expected runtime
$$\mathbb{E}\left[T\right] \leq \left(\frac{8}{\delta^2}\right) \sum_{j=1}^{m-1} \left[\lambda \ln\left(\frac{6\delta\lambda}{4 + z_j \delta\lambda}\right) + \frac{1}{z_j}\right].$$

[4] See Corus et al. (2016) for more details on the theorem.

Proof idea (UMDA with margins)

- **Level definition:** $A_j := \left\{x \in \{0,1\}^n \mid \textsc{OneMax}(x) = j-1\right\}$.]
- Let A_j denote the current level.
- We choose $\gamma_0 := \mu/\lambda \Rightarrow$ all μ best individuals have $\geq j-1$ ones.
- Upgrade $\Rightarrow$ Sample an offspring with $\geq j$ ones, i.e. $\Pr(Y \geq j)$.
- Verify condition (G1): $\Pr(Y \geq j) \geq z_j$.
- Verify condition (G2): $\Pr(Y \geq j) \geq (1+\delta)\gamma$.

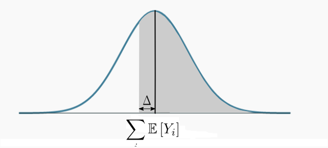
Verify condition (G2)



The probability of sampling j one-bits:

$$\begin{aligned}
\Pr(Y \geq j) &\geq \Pr(Y_k + Y_m \geq j - \ell) \cdot \Pr(Y_\ell = \ell) \\
&\geq \Pr(Z > j - \ell - 1) \cdot \left(1 - \frac{1}{n}\right)^{\ell} && (Z := Y_k + Y_m) \\
&\geq \Pr\left(Z > \mathbb{E}[Z] - \frac{\gamma}{\gamma_0}\right) \cdot \frac{1}{e} && \left(\mathbb{E}[Z] \geq j - \ell - 1 + \frac{\gamma}{\gamma_0}\right) \\
&\geq \quad \ldots \\
&\geq \quad \ldots \\
&\geq (1+\delta)\gamma.
\end{aligned}$$

Feige's Inequality



Theorem (Feige (2004))

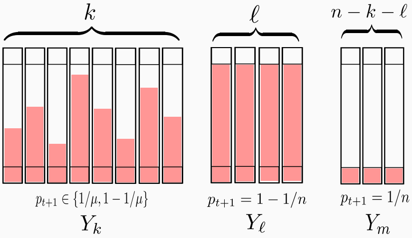
Given n independent r.v. $Y_1, \ldots, Y_n \in [0,1]$, then for all $\Delta > 0$

$$\Pr\left(\sum_{i=1}^{n} Y_i > \sum_{i=1}^{n} \mathbb{E}[Y_i] - \Delta\right) \geq \min\left\{\frac{1}{13}, \frac{\Delta}{1+\Delta}\right\}.$$

The probability of sampling j one-bits now becomes

$$\begin{aligned}\Pr(Y \geq j) &\geq \Pr\left(Z > \mathbb{E}[Z] - \frac{\gamma}{\gamma_0}\right) \cdot \frac{1}{e} \\ &\geq \min\left\{\frac{1}{13}, \frac{\gamma/\gamma_0}{1+\gamma/\gamma_0}\right\} \cdot \frac{1}{e} \\ &\geq \frac{\gamma}{13\gamma_0} \cdot \frac{1}{e} \geq (1+\delta)\gamma.\end{aligned}$$

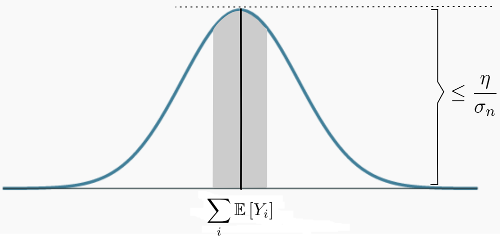
Verify (G1): k is sufficiently large



The upgrade probability in this case:

$$\begin{aligned}
\Pr(Y_k \geq j-\ell) &= \Pr\left(Y_k \geq j-\ell-1\right) - \Pr\left(Y_k = j-\ell-1\right) \\
&= \Pr\left(Y_k \geq \mathbb{E}\left[Y_k\right]\right) - \Pr(Y_k = j-\ell-1) \\
&\geq \quad \ldots \\
&\geq \quad \ldots
\end{aligned}$$

Anti-Concentration Bound



Theorem (Baillon et al. (2016))

Given n independent Bernoulli r.v. $Y_1, \ldots, Y_n$ with success probabilities $p_1, \ldots, p_n$. For all n, y and p_i.

$$\Pr\left(\sum_{i=1}^{n} Y_i = y\right) \leq \frac{\eta}{\sigma_n}$$

where $\sigma_n^2 := \sum_{i=1}^{n} p_i(1-p_i)$ and $\eta \sim 0.4688$ is an absolute constant.

Verify (G1): k is sufficiently large

Theorem (Jogdeo & Samuels 1968)

Let $Y_1, Y_2, \ldots, Y_n$ be n independent Bernoulli random variables. Let $Y := \sum_{i=1}^{n} Y_i$ be the sum of these random variables. If the expectation of Y is an integer, then

$$\Pr(Y \geq \mathbb{E}[Y]) \geq \frac{1}{2}.$$

Now the upgrade probability now becomes

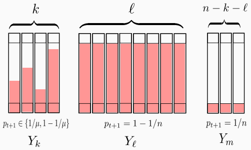
$$\begin{aligned}
\Pr(Y_k \geq j - \ell) &= \Pr(Y_k \geq j - \ell - 1) - \Pr(Y_k = j - \ell - 1) \\
&= \Pr(Y_k \geq \mathbb{E}[Y_k]) - \Pr(Y_k = j - \ell - 1) \\
&\geq \quad \frac{1}{2} \quad - \quad \frac{\eta}{\sqrt{\mathrm{Var}[Y_k]}} \\
&\geq \Omega(1).
\end{aligned}$$

Remaining cases

CASE 2: small k and large ℓ (so is j)

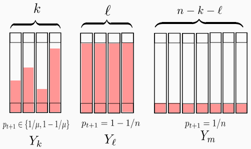
- Being very close to optimal level.
- If $j \geq n + 1 - \frac{n}{\mu}$, then according to Dang & Lehre (2015)

$$\Pr(Y \geq j) = \Omega\left(\frac{1}{\mu}\right) = \Omega\left(\frac{n-j+1}{n}\right).$$



CASE 3: Small k and not close to optimal level (i.e. not too large j).

$$\begin{aligned}\Pr(Y \geq j) &\geq \Pr(Y_k \geq j - \ell - 1) \cdot \Pr(Y_\ell = \ell) \cdot \Pr(Y_m \geq 1) \\ &\geq \frac{1}{2} \cdot \frac{1}{e} \cdot \frac{n-k-\ell}{n} \\ &\geq \Omega\left(\frac{n-j+1}{n}\right).\end{aligned}$$



Condition (G3) and Expected runtime

Combining all three cases

$$z_j := \min\left\{\Omega(1), \Omega\left(\frac{n-j+1}{n}\right)\right\} = \Omega\left(\frac{n-j+1}{n}\right).$$

(G3) satisfied iff

$$\lambda \geq \left(\frac{4}{\gamma_0\delta^2}\right)\ln\left(\frac{128m}{z_*\delta^2}\right) = \Omega(\log n).$$

The expected optimisation time is

$$\begin{aligned}
\mathcal{O}\left(\lambda\sum_{j=1}^{n}\ln\left(\frac{1}{z_j}\right) + \sum_{j=1}^{n}\frac{1}{z_j}\right) &= \mathcal{O}\left(\lambda\sum_{j=1}^{n}\ln\left(\frac{n}{n-j+1}\right) + \sum_{j=1}^{n}\frac{n}{n-j+1}\right) \\
&= \mathcal{O}\left(\lambda\ln\prod_{j=1}^{n}\frac{n}{n-j+1} + n\sum_{k=1}^{n}\frac{1}{k}\right) \\
&= \mathcal{O}\left(\lambda\ln\frac{n^n}{n!} + n\sum_{k=1}^{n}\frac{1}{k}\right) \qquad \text{(Stirling's approximation}^{5}\text{)} \\
&= \mathcal{O}(n\lambda) + \mathcal{O}(n\log n) \\
&= \mathcal{O}(n\lambda).
\end{aligned}$$

[4]Stirling's approximation: $n! \propto n^{n+0.5}e^{-n}$

Summary

- The upper bound $\mathcal{O}(n\lambda)$ **holds** for $a\log(n) \leq \mu \leq a'\sqrt{n}$ where a, a' are some positive constants.
- The result finally **closes the gap** $\Theta(\log\log n)$ between the first upper bound $\mathcal{O}(n\log n\log\log n)$ (Dang & Lehre 2015) and recently discovered lower bound $\Omega(\mu\sqrt{n} + n\log n)$ (Krejca & Witt 2017).
- **Anti-concentration** may be applied to analyse runtime of other algorithms.

Witt (2017) independently obtained the same upper bound $\mathcal{O}(n\lambda)$ when $\lambda = (1+\Theta(1))\mu$ and $\mu \geq c\log(n)$. Our result relaxes the proportional relationship between λ and μ but convers smaller range of μ.

Thank you

Improved Runtime Bound for the Univariate Marginal Distribution Algorithm via Anti-concentration

Phan Trung Hai Nguyen & Per Kristian Lehre

{PXN683, P.K.Lehre}@cs.bham.ac.uk

References

Baillon, J.-B., Cominetti, R. & Vaisman, J. (2016), 'A sharp uniform bound for the distribution of sums of bernoulli trials', *Combinatorics, Probability and Computing* **25**(3), 352–361.

Chen, T., Lehre, P. K., Tang, K. & Yao, X. (2009), When is an estimation of distribution algorithm better than an evolutionary algorithm?, *in* 'Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2009', pp. 1470–1477.

Chen, T., Tang, K., Chen, G. & Yao, X. (2010), 'Analysis of computational time of simple estimation of distribution algorithms', *IEEE Transactions on Evolutionary Computation* **14**(1), 1–22.

Corus, D., Dang, D., Eremeev, A. V. & Lehre, P. K. (2016), 'Level-based analysis of genetic algorithms and other search processes', *CoRR* **abs/1407.7663**.
URL: *http://arxiv.org/abs/1407.7663*

Dang, D. & Lehre, P. K. (2015), Simplified runtime analysis of estimation of distribution algorithms, *in* 'Proceedings of Genetic and Evolutionary Computation Conference, GECCO'15'.

Droste, S. (2006), 'A rigorous analysis of the compact genetic algorithm for linear functions', *Natural Computing* **5**(3), 257–283.

Feige, U. (2004), On sums of independent random variables with unbounded variance, and estimating the average degree in a graph, *in* 'Proceedings of the 36th STOC', pp. 594–603.

Jogdeo, K. & Samuels, S. M. (1968), 'Monotone convergence of binomial probabilities and a generalization of ramanujan's equation', *The Annals of Mathematical Statistics* **39**(4), 1191–1195.

Krejca, M. S. & Witt, C. (2017), Lower bounds on the run time of the univariate marginal distribution algorithm on onemax, *in* 'Proceedings of Foundation of Genetic Algorithms, FOGA'17'.

Shakya, S. (2006), 'Deum: A framework for an estimation of distribution algorithm based on markov random fields'.

Witt, C. (2017), Upper bounds on the runtime of the univariate marginal distribution algorithm on onemax, *in* '(To appear) Proceedings of Genetic and Evolutionary Computation Conference, GECCO'17'.