

Population sizing

- Correct size of the population important:
 - too small: premature convergence to sub-optimal solutions
 - too large: computational inefficient
- Here we focus on the Counting-Ones problem, but the model can be extended to more complex functions
- We also focus on incremental tournament selection ($s = 2$), but again extensions are possible
- Key question: how does the optimal population size scales with the complexity of the problem, ie. the length of the string ?

Selection error

- Tournament selection ($s = 2$): two strings compete to become member of the parent pool:
 - $s_1 : 1100011100$, fitness = 5
 - $s_2 : 0100111101$, fitness = 6
 - string s_2 is selected
- Looking at this competition at the schema level (order-1 sufficient since we focus one Counting-Ones):
 - partition $f*****$: schema $0*****$ wins from schema $1***** \Rightarrow$ selection decision error
 - partitions $****f*****$ and $*****f$: schema $****1*****$ wins from schema $****0****$, and schema $*****1$ wins from schema $*****0 \Rightarrow$ correct selection decisions

– other partitions: nothing changes

- How many selection decision errors can we afford to make before the optimal bit-value within a partition is completely lost in the population = premature convergence
- Population sizing is basically a statistical decision making problem

Binomial distribution

- Random variable X is binomial distributed if it has the discrete probability density: $x \in \{0, \dots, \ell\}$

$$f(x) = P(X = x) = \binom{\ell}{x} p^x (1-p)^{\ell-x}$$

- The probability distribution function corresponding to the binomial density is:

$$F_X(x) = P(X \leq x) = \sum_{k=0}^x \binom{\ell}{k} p^k (1-p)^{\ell-k}$$

mean: $\mu = \ell p$

variance: $\sigma^2 = \ell p(1-p)$

Normal distribution

- Random variable X is normal distributed if it has the discrete probability density: $x \in \mathfrak{R}$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- The probability distribution function corresponding to the normal density is (mean μ , variance σ^2):

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

- Standard normal variable: $X \sim \mathbb{N}(0, 1)$

$$\Phi(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{y^2}{2}} dy$$

and $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$

Normal approximation to the binomial distribution

- Assume random variable X_ℓ is binomial distributed with parameters ℓ and p .

If $\ell \rightarrow \infty$ the probability distribution function of the standardized random variable $X_\ell^* = \frac{X_\ell - \ell p}{\sqrt{\ell p(1-p)}}$ approaches the probability distribution function Φ of the standard normal distribution:

$$P(X_\ell^* \leq \frac{x - \ell p}{\sqrt{\ell p(1-p)}}) \approx \Phi\left(\frac{x - \ell p}{\sqrt{\ell p(1-p)}}\right)$$

or

$$P(X_\ell \leq x) \approx \Phi\left(\frac{x - \ell p}{\sqrt{\ell p(1-p)}}\right)$$

The approximation is acceptable when minimum $(\ell p, \ell(1-p)) \geq 5$

- note: $p(a \leq X_\ell \leq b) \approx \Phi\left(\frac{b - \ell p}{\sqrt{\ell p(1-p)}}\right) - \Phi\left(\frac{a - \ell p}{\sqrt{\ell p(1-p)}}\right)$

Probability selection decision error

- Schemata fitness $f(H_1 : ***1*****)$ and $f(H_2 : ***0*****)$ binomial distributed
→ approximating with normal distribution $\mathbb{N}(\mu, \sigma^2)$:

$$\mu_{H_1} = 1 + (\ell - 1)p, \quad \sigma_{H_1}^2 = (\ell - 1)p(1 - p)$$

$$\mu_{H_2} = (\ell - 1)p, \quad \sigma_{H_2}^2 = (\ell - 1)p(1 - p)$$

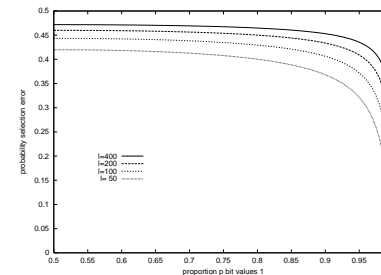
- Distribution of the fitness difference of the best schema and the worst schema $f(H_1) - f(H_2)$ is also normal distributed:

$$\mu_{H_1 - H_2} = 1, \quad \sigma_{H_1 - H_2}^2 = 2(\ell - 1)p(1 - p)$$

- Probability selection error is equal to the probability that the best schema is represented by a string with fitness less than the representative of the worst schema, which is also equal to the

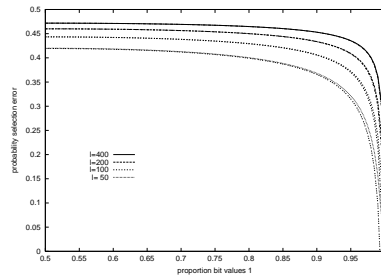
probability that the fitness difference is negative:

$$\begin{aligned} P[SelErr] &= P(F_{H_1 - H_2} < 0) \\ &= P\left(\frac{F_{H_1 - H_2} - \mu_{H_1 - H_2}}{\sigma_{H_1 - H_2}} < \frac{-\mu_{H_1 - H_2}}{\sigma_{H_1 - H_2}}\right) \\ &= \Phi\left(\frac{-1}{\sqrt{2(\ell - 1)p(1 - p)}}\right) \end{aligned}$$



- Approximation by first two terms of power series expansion for the normal distribution:

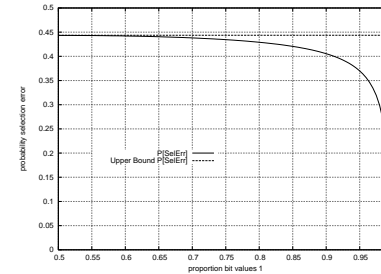
$$P[\text{SelErr}] \approx \frac{1}{2} - \frac{1}{2\sqrt{\pi(\ell-1)p(1-p)}}$$



- Selection error is upper bounded by:

$$P[\text{SelErr}] \leq \frac{1}{2} - \frac{1}{\sqrt{\pi\ell}}$$

this is a conservative estimate of the selection error that ignores the reduction in error probability when the proportion of optimal bit values $p(t)$ increases.



GA population sizing

- Selection viewed as decision making process within partitions: schemata competition
- When best schema loses competition we have a selection decision error
- How many decision errors can we afford to make given a certain population size ?
- Answer given by Gambler's ruin model: within each partition a random walk is played

Random walks

Random walks are mathematical models used to predict the outcome of certain stochastic processes. Consider the following random walk:

1. a one-dimensional, discrete space of size $N + 1$: $[0, 1, \dots, N - 1, N]$
2. a particle somewhere in the above space: position $x \in \{0, \dots, N\}$
3. the particle can move one step to the right with probability p , and one step to the left with probability $1 - p$
4. when the particle reaches the boundaries ($x = 0$, or $x = N$) the random walk ends
5. call $P_N(x)$ (resp. $P_0(x)$) the probability that the particle is absorbed by the boundary $x = N$ (resp. $x = 0$) when it is currently at position x

- Difference equation:

$$P_N(x) = pP_N(x+1) + (1-p)P_N(x-1)$$

with boundary conditions: $P_N(N) = 1$, and $P_N(0) = 0$

- Solving the above equation gives the probability that the particle - starting from position x_0 - is absorbed by the $x = N$ boundary:

$$P_N(x_0) = \frac{1 - \left(\frac{1-p}{p}\right)^{x_0}}{1 - \left(\frac{1-p}{p}\right)^N}$$

- when $p = 1 - p = 0.5$ we get $P_N(x_0) = \frac{x_0}{N}$
note also: $P_0(x_0) = 1 - P_N(x_0)$

Gambler's ruin model

The previous random walk is called the Gambler's ruin model describing a gambler betting against a casino:

1. the position x represents the amount of money a gambler possesses
2. the size N represents the total amount of money of the gambler and the casino
3. p (resp. $1 - p$) is the probability the gambler wins (resp. loses) a bet, and he subsequently gains (resp. loses) one unit of money
4. $P_0(x_0)$ (resp. $P_N(x_0)$) is the probability the gambler is ruined (resp. breaks the casino) when starting with x_0 amount of money

Population sizing & Gambler's ruin

Mapping GA concepts in Gambler's ruin model:

1. The number of optimal bit values '1' in the population at a certain position - this is, for the partition considered - corresponds to the position x in the Gambler's ruin model
2. The boundaries $x = N$ (resp. $x = 0$) correspond to all bit values in the population at the partition being equal to the bit value '1' (resp. '0')
3. When the boundaries are reached the random walk ends: either the population is filled with ones at the partition or filled with zeroes (recall that we do not take mutation into account here)
4. The probability that the amount of optimal string values in the population at the partition is increased by one is equal to the

probability that the selection decision making is correct for that partition. This corresponds to the probability that the particle moves to the right (p)

5. Desired convergence corresponds to the particle reaching the $x = N$ boundary. Premature convergence - this is, losing the optimal bit value in the population - corresponds to the particle reaching the $x = 0$ boundary.

Recalling the probability for a selection decision error (using tournament selection, $s = 2$):

$$P[SelErr] \leq \frac{1}{2} - \frac{1}{\sqrt{\pi\ell}}$$

we can compute the probability of convergence to the optimal bit

value at a certain partition as:

$$\begin{aligned}
 P[OptConv] &= \frac{1 - \left(\frac{P[SelErr]}{1 - P[SelErr]}\right)^{N/2}}{1 - \left(\frac{P[SelErr]}{1 - P[SelErr]}\right)^N} \\
 &\approx 1 - \left(\frac{P[SelErr]}{1 - P[SelErr]}\right)^{N/2} \\
 &\approx 1 - \left(\frac{\frac{1}{2} - \frac{1}{\sqrt{\pi\ell}}}{\frac{1}{2} + \frac{1}{\sqrt{\pi\ell}}}\right)^{N/2} \\
 &\approx 1 - \left(\frac{\sqrt{\pi\ell} - 2}{\sqrt{\pi\ell} + 2}\right)^{N/2}
 \end{aligned}$$

Note: the second step - the approximation - is obtained by observing that the denominator approaches 1 much more rapidly as the numerator since $P[SelErr] < 1 - P[SelErr]$

- Taking the logs:

$$\frac{N}{2} \ln \frac{\sqrt{\pi\ell} - 2}{\sqrt{\pi\ell} + 2} \approx \ln(1 - P[OptConv])$$

- Using the Taylor series approximation:

$$\begin{aligned}
 \ln \frac{x-2}{x+2} &= \ln(x-2) - \ln(x+2) \\
 &\approx \left(\ln x - \frac{2}{x} - \frac{2}{x^2}\right) - \left(\ln x + \frac{2}{x} - \frac{2}{x^2}\right) \\
 &\approx -\frac{4}{x}
 \end{aligned}$$

we get:

$$\frac{N}{2} \frac{-4}{\sqrt{\pi\ell}} \approx \ln(1 - P[OptConv])$$

- Finally, we obtain the result for the critical population size as:

$$N \approx \ln(1 - P[OptConv]) \frac{\sqrt{\pi\ell}}{-2}$$

= the minimal required population size scales as the square root of the problem complexity !

- Alternatively, we can compute the probability that at a certain partition the optimal bit value will be found as:

$$P[OptConv] \approx 1 - e^{\frac{-2N}{\sqrt{\pi\ell}}}$$

- The number of optimal bits F in the entire string of length ℓ is binomially distributed:

$$P(F = x) = \binom{\ell}{x} P[OptConv]^x (1 - P[OptConv])^{\ell-x}$$

with mean: $\mu = \ell P[OptConv]$

and variance: $\sigma^2 = \ell P[OptConv](1 - P[OptConv])$,

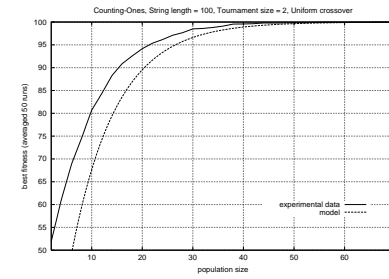
- The probability the optimal string will be reached is:

$$P[OptimalString] = P[OptConv]^\ell.$$

Experimental validation

- Counting-Ones problem, string length $\ell = 100$, uniform crossover, generational tournament selection ($s = 2$).
- Expected fitness of the string after convergence:

$$E[F] \approx 100 \left(1 - e^{-\frac{2N}{\sqrt{100\pi}}}\right)$$



- The probability of reaching the optimal string:

$$P[\text{OptimalString}] = \left(1 - e^{-\frac{2N}{\sqrt{100\pi}}}\right)^{100}$$

