

# SFS for single mutations

```

f[S_, x_] := (1 - E^(-S (1 - x))) / (x (1 - x) (1 - E^(-S)))

Q[n_, i_, x_] := n! / (i! (n - i)!) x^i (1 - x)^(n - i)

SFSn[n_] := Table[1/i, {i, 1, n - 1}]

SFSnE[n_] := Table[1/i, {i, 1, n}]

SFS[n_, S_] :=
Module[{SFS, SFSn},

SFS = Table[NIntegrate[f[S, x] Q[n, i, x], {x, 0, 1}], {i, 1, n - 1}];
SFSn = Table[1/i, {i, 1, n - 1}];
TableForm [{SFS, SFS/SFSn, SFS - SFSn}]]

SFSpositive[n_, S_] :=
Module[{SFS, SFSnorm },

SFS = Table[NIntegrate[f[S, x] Q[n, i, x], {x, 0, 1}], {i, 1, n - 1}];
SFSnorm = SFS/Apply[Plus, SFS];
SFS]

SFSpositiveE[n_, S_] := Table[NIntegrate[f[S, x] Q[n, i, x], {x, 0, 1}], {i, 1, n}]

D0[S_] := S / (1 - E^-S)

```

# John' s approximations for one sided gamma

```
H2[meanS_ ,  $\beta$ _ , x_] := ( $\beta$ /meanS ) ^ $\beta$  (Zeta[ $\beta$ ,  $\beta$ /meanS + x] - Zeta[ $\beta$ ,  $\beta$ /meanS + 1])
```

```
P2[n_ , i_ , meanS_ ,  $\beta$ _] := NIntegrate[
  H2[meanS ,  $\beta$  , x] Q[n, i, x] / (x (1 - x)), {x, 0, 1}]
```

```
D2[meanS_ ,  $\beta$ _] :=  $\beta$  ( $\beta$ /meanS ) ^ $\beta$  Zeta[1 +  $\beta$ , 1 +  $\beta$ /meanS ]
```

```
SelectedSFS2[n_ , meanS_ ,  $\beta$ _] := Table[P2[n, i, meanS ,  $\beta$ ], {i, 1, n - 1}]
```

```
SelectedSFS2E[n_ , meanS_ ,  $\beta$ _] := Table[P2[n, i, meanS ,  $\beta$ ], {i, 1, n}]
```

```
D2[0.0001, 0.1]
```

```
0.99995
```

# d anb b statistics

## Infinite sites

```

DCEmethod [n_, pA_, Sa_, meanS_ ,  $\beta$ _] :=
Module[{SFSSs, SFSn, dNdS,  $\alpha$ True, normSFSSs , normSFSn , foldedSFSSs, foldedSFSn,
  pN, pS,  $\alpha$ original, Nabove, Sabove, blig, fneu, dstr,  $\alpha$ PopDrowser,  $\alpha$ Sussex,  $\alpha$ Fay,
  trueWeakDel, trueStrDel, trueNeu, dNdSNEU, pNBelow, pSBelow, dstr2, c},

  SFSn = pA*SFSpositive[n, Sa] + (1-pA) SelectedSFS2[n, meanS ,  $\beta$ ];
  SFSSs = Table[1/i, {i, 1, n-1}];
  pN = Apply[Plus, SFSn] // N;
  pS = Apply[Plus, SFSSs] // N;

  foldedSFSSs = Drop[SFSSs+Reverse[SFSSs], -Floor[(n-1)*0.5]];
  (*FOLDED SFS-MAF*)
  foldedSFSn = Drop[SFSn+Reverse[SFSn], -Floor[(n-1)*0.5]];
  (*FOLDED SFS-MAF*)

  foldedSFSn = foldedSFSn/pN;
  foldedSFSSs = foldedSFSSs/pS;
  dNdS = pA*D0[Sa] + (1-pA)*D2[meanS ,  $\beta$ ];
  dNdSNEU = (1-pA)*D2[meanS ,  $\beta$ ];
   $\alpha$ True = pA*D0[Sa]/dNdS;
   $\alpha$ original = 1-pN/pS*1/dNdS;

  (*DGRP original WAY*)
  pNBelow = 1-Apply[Plus, Drop[foldedSFSn, Ceiling[0.025*n]]];
  pSBelow = 1-Apply[Plus, Drop[foldedSFSSs, Ceiling[0.025*n]]];
  blig = (pNBelow-pSBelow)*(pN/pS);
  fneu = (pN/pS)*(1-blig);
  dstr = 1-(pN/pS);
  dstr2 = 1-(pN/(pS+1));

  {dstr, dstr2, 1-(pNBelow-pSBelow), (pNBelow-pSBelow)}]

```

## Sampling

```

RMSE[a_, trueValue_] := Sqrt[Mean[(a-trueValue)^2]]

```

```

Sampling [Method_, N_, n_, Ltheta_, {pA_, Sa_, meansS_ ,  $\beta$ _}] :=

Module[{SFSs, SFSn, dNdS,  $\alpha$ True, normSFSs , normSFSn , eS, eN, pN,
  pS, foldedSFSs, foldedSFSn, expectedValue, sampleEstimates ,
  sampleSFSs , sampleSFSn , summaryStats , results, expSFSn, expSFSs},

  SFSn = pA*SFSpositive[n, Sa] + (1-pA) SelectedSFS2[n, meansS ,  $\beta$ ];
  SFSs = Table[1/i, {i, 1, n-1}];
  pN = Apply[Plus, SFSn] // N;
  pS = Apply[Plus, SFSs] // N;

  (*Folding the SFS*)
  foldedSFSs = Drop[SFSs+Reverse[SFSs], -Floor[(n-1)*0.5]];
  (*FOLDED SFS-MAF*)
  foldedSFSn = Drop[SFSn+Reverse[SFSn], -Floor[(n-1)*0.5]];
  (*FOLDED SFS-MAF*)
  normSFSs = foldedSFSs/pS;
  normSFSn = foldedSFSn/pN;

  (*Codons:Every sin.site has 2 nonsin.sites*)
  eS = Ltheta*pS;
  eN = eS*(pN/pS)*2;
  dNdS = 2*(pA*D0[ $S_a$ ] + (1-pA)*D2[meansS ,  $\beta$ ]);
   $\alpha$ True = 2*pA*D0[ $S_a$ ]/dNdS;
  expSFSn = normSFSn *eN*1000000;
  expSFSs = normSFSs *eS*1000000;

  (*Infinite sites expectation*)
  expectedValue = Method[n, expSFSn, expSFSs, dNdS];

  (*Rebuilding the SFSs (folded) with infinitesegregating sites*)
  SFSs = normSFSs *eS; SFSn = normSFSn *eN;

  (*Sampling the SFS for few segregating sites*)
  sampleEstimates =
    Table[sampleSFSn = Map[Random[PoissonDistribution[ $\mu$ ]] &, SFSn];
      sampleSFSs = Map[Random[PoissonDistribution[ $\mu$ ]] &, SFSs];
      Method[n, sampleSFSn , sampleSFSs , dNdS], {N}];
  summaryStats = {expectedValue,
    Mean[sampleEstimates ], RMSE[sampleEstimates , expectedValue]};
  (*summaryStats = {sampleEstimates };*)
  summaryStats ]

```

## Statistics

```

DGRPB[n_, foldedSFSn_, foldedSFSS_, dNdS_] :=
Module[{blig, fneu, dstr, pN, pS, pNBelow, pSBelow, normSFSn , normSFSS },
  pN = Apply[Plus, foldedSFSn] // N;
  pS = Apply[Plus, foldedSFSS] // N;
  normSFSS = foldedSFSS / (pS + 1);
  normSFSn = foldedSFSn / (pN + 1);
  pN = pN / 2;
  pNBelow = 1 - Apply[Plus, Drop[normSFSn , Ceiling[0.025 * n]]];
  pSBelow = 1 - Apply[Plus, Drop[normSFSS , Ceiling[0.025 * n]]];
  blig = (pNBelow - pSBelow);
  blig]
DGRPD[n_, foldedSFSn_, foldedSFSS_, dNdS_] :=
Module[{blig, fneu, dstr, pN, pS, pNBelow, pSBelow, normSFSn , normSFSS },
  pN = Apply[Plus, foldedSFSn] // N;
  pS = Apply[Plus, foldedSFSS] // N;
  pN = pN / 2;
  dstr = 1 - (pN / (pS + 1));
  dstr]
DGRPPs[n_, foldedSFSn_, foldedSFSS_, dNdS_] :=
Module[{blig, fneu, dstr, pN, pS, pNBelow, pSBelow, normSFSn , normSFSS },
  pN = Apply[Plus, foldedSFSn] // N;
  pS = Apply[Plus, foldedSFSS] // N;
  pS]
DGRPPn[n_, foldedSFSn_, foldedSFSS_, dNdS_] :=
Module[{blig, fneu, dstr, pN, pS, pNBelow, pSBelow, normSFSn , normSFSS },
  pN = Apply[Plus, foldedSFSn] // N;
  pS = Apply[Plus, foldedSFSS] // N;
  pN]
DGRPPS[n_, foldedSFSn_, foldedSFSS_, dNdS_] :=
Module[{blig, fneu, dstr, pN, pS, pNBelow, pSBelow, normSFSn , normSFSS },
  pN = Apply[Plus, foldedSFSn] // N;
  pS = Apply[Plus, foldedSFSS] // N;
  pS]

```

## Analysis

```

Analysis[Method_, N_, n_, Ltheta_, ParameterList_] := Module[{x},
  x = Mean[Map[Sampling [Method, N, n, Ltheta, #] &, ParameterList ]];
  Print[x[[1]], "\t", x[[2]], "\t", x[[3]]]

```

## Finite Number of Sites

```
Analysis[DGRPPS, 100, 128, 2.5, {{0, 100, 1000, 0.1}}]
(*Number of syn segregating sites - 1 gene*)
Analysis[DGRPPS, 100, 128, 25, {{0, 100, 1000, 0.1}}]
(*Number of syn segregating sites - 10 gene*)
Analysis[DGRPPS, 100, 128, 250, {{0, 100, 1000, 0.1}}]
(*Number of syn segregating sites - 100 gene*)
```

$1.36024 \times 10^7$	13.31	$1.36024 \times 10^7$
$1.36024 \times 10^8$	136.92	$1.36024 \times 10^8$
$1.36024 \times 10^9$	1358.07	$1.36024 \times 10^9$

```
Analysis[DGRPD, 100, 128, 2.5, {{0, 100, 1000, 0.1}}]
Analysis[DGRPD, 100, 128, 2.5, {{0, 100, 1000, 0.3}}]
Analysis[DGRPD, 100, 128, 2.5, {{0, 100, 1000, 0.5}}]
Analysis[DGRPD, 100, 128, 2.5, {{0, 100, 1000, 0.7}}]
Analysis[DGRPD, 100, 128, 2.5, {{0, 100, 1000, 0.9}}]
```

0.452456	0.45179	0.235774
0.75428	0.741427	0.11903
0.859548	0.868598	0.0671182
0.90557	0.912431	0.0768415
0.928829	0.928029	0.0584566

```
Analysis[DGRPB, 100, 128, 2.5, {{0, 100, 1000, 0.1}}]
Analysis[DGRPB, 100, 128, 2.5, {{0, 100, 1000, 0.3}}]
Analysis[DGRPB, 100, 128, 2.5, {{0, 100, 1000, 0.5}}]
Analysis[DGRPB, 100, 128, 2.5, {{0, 100, 1000, 0.7}}]
Analysis[DGRPB, 100, 128, 2.5, {{0, 100, 1000, 0.9}}]
```

0.066073	0.0930499	0.179741
0.202581	0.226779	0.236654
0.326701	0.332849	0.225568
0.423654	0.459953	0.207716
0.491164	0.50526	0.196592

## Infinite Number of Sites

```
DCEmethod [128, 0.0, 100, 1000, 0.1]
DCEmethod [128, 0.0, 100, 1000, 0.3]
DCEmethod [128, 0.0, 100, 1000, 0.5]
DCEmethod [128, 0.0, 100, 1000, 0.7]
DCEmethod [128, 0.0, 100, 1000, 0.9]
{0.452164, 0.537426, 0.933445, 0.0665548}
{0.753925, 0.792222, 0.796244, 0.203756}
{0.859254, 0.881159, 0.671763, 0.328237}
{0.905336, 0.920069, 0.57466, 0.42534}
{0.928638, 0.939744, 0.507098, 0.492902}
```

## Relation between d and sample size

```
Table[Analysis[DGRPD, 1, 2^i, 2.5, {{0, 100, 1000, 0.3}}], {i, 1, 8}]
0.842085      0.875      0.0329146
0.845142      1.      0.154858
0.834568      0.625      0.209568
0.819338      0.857143      0.0378052
0.800838      0.75      0.0508378
0.779174      0.866667      0.0874925
0.75428      0.78125      0.0269701
0.726182      0.75      0.0238176
{Null, Null, Null, Null, Null, Null, Null, Null}

Table[Analysis[DGRPD, 1, 2^i, 2.5, {{0, 100, 1000, 0.9}}], {i, 1, 8}]
0.983768      1.      0.0162319
0.985053      1.      0.0149472
0.980369      0.928571      0.0517978
0.973168      0.916667      0.056501
0.962873      1.      0.0371268
0.948463      0.916667      0.031796
0.928829      0.95      0.0211708
0.903053      0.892857      0.0101957
{Null, Null, Null, Null, Null, Null, Null, Null}
```

# DFE examples

```
Plot[Evaluate@Table[PDF[GammaDistribution [ $\alpha$ , 1000], x], { $\alpha$ , {0.3, 0.9}}],  
  {x, 0, 1000}, Filling→Axis]
```

