

生物資源モデリング Lecture 5

遺伝的浮動, 突然変異, 遺伝子流動のモデリング

Toshihide Kitakado

2019年11月19日

生物集団における対立遺伝子の時間的変化を題材に, 確率的シミュレーションについて述べる. ここではとくに Wright-Fisher モデルという集団遺伝学でよく使われるモデルを考える.

遺伝的浮動

いまある集団内に, N 個体の 2 倍体の生物が生息しているとする. ここで 2 倍体とは, ヒトのように 2 本の染色体が対をなした生物のことで, 各遺伝子座に 2 つの対立遺伝子がペアで存在する. したがって, N 個体の集団には $2N$ 個の対立遺伝子があることになる. どの個体も繁殖に寄与できると考えるとき, これを遺伝子プールとよぶ.

Wright-Fisher モデルでは, 離散的な世代を跨いで対立遺伝子頻度がどのように変化するかを確率的に表現する. 話の簡略のため, 対立遺伝子は A_1 または A_2 の 2 種類のみとし, 第 t 世代において遺伝子プール内に対立遺伝子 A_1 が Y_t 個あるとする. また雌雄の区別や, 選択 (淘汰), 個体の移動, そして突然変異が生じないとし, 世代を跨ぐことによって個体数に変化なく, 任意交配によって個体は総入れ替えとなると仮定する. このとき, 第 $t+1$ 世代における対立遺伝子 A_1 の数 Y_{t+1} は確率的にしか予測できないが, その確率的変動の規則性は以下のような確率分布で表現できる.

$$Pr(Y_{t+1} = n | Y_t = m) = \frac{(2N)!}{n!(2N-n)!} \left(\frac{m}{2N}\right)^n \left(1 - \frac{m}{2N}\right)^{2N-n} \quad (1)$$

```
Genetic_drift <- function(NN, TT, InitFreq, Nsim){  
  
  #Setting  
  Y <- array(NA, c(Nsim, TT))  
  Freq <- array(NA, c(Nsim, TT))  
  
  #Simulation  
  for(i in 1:Nsim){  
    Y[i,1] <- rbinom(1, 2*NN, InitFreq)  
    Freq[i,1] <- Y[i,1]/(2*NN)  
    for(t in 2:TT){  
      Y[i,t] <- rbinom(1, 2*NN, Freq[i,(t-1)])  
    }  
  }  
}
```

```

    Freq[i,t] <- Y[i,t]/(2*NN)
  }
}
Out <- list()
Out$Y <- Y
Out$Freq <- Freq

#Graphical output
layout(matrix(c(1,1,1,2),1,4,byrow=T))

#plot1
par(mar=c(6,6,2,0))
Generation <- 1:TT
plot(Generation, Freq[1,], type="l", col="gray", ylim=c(0,1), ylab="Allele frequency")
for(i in 2:Nsim) points(Freq[i,], type="l", col="gray")

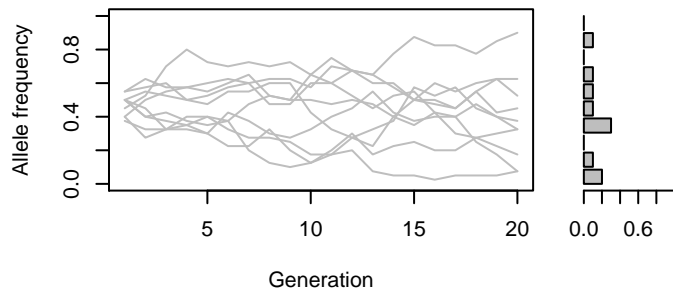
#plot2
par(mar=c(6,2,2,2))
barplot(hist(Freq[,TT], plot=F, breaks=seq(0,1,0.1))$counts/Nsim, horiz=T, xlim=c(0,1))

return(Out)
}

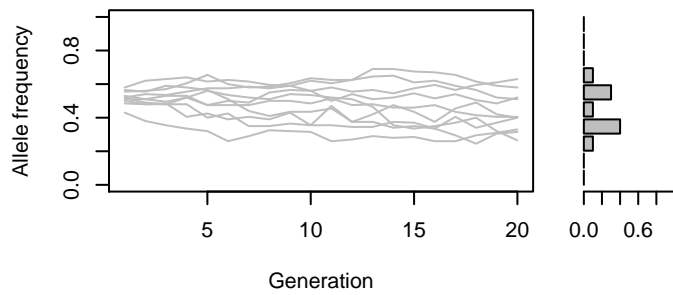
```

世代数 TT=20 (Nsim = 10)

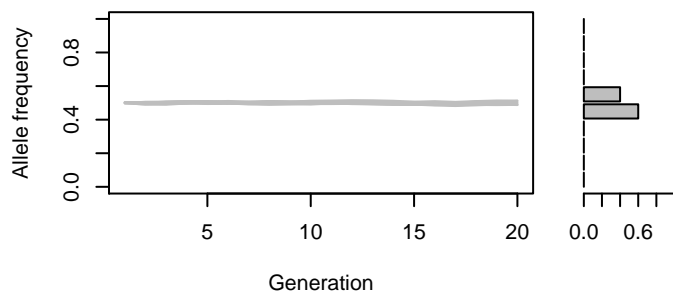
```
Res <- Genetic_drift(NN=20, TT=20, InitFreq=0.5, Nsim=10)
```



```
Res <- Genetic_drift(NN=100, TT=20, InitFreq=0.5, Nsim=10)
```

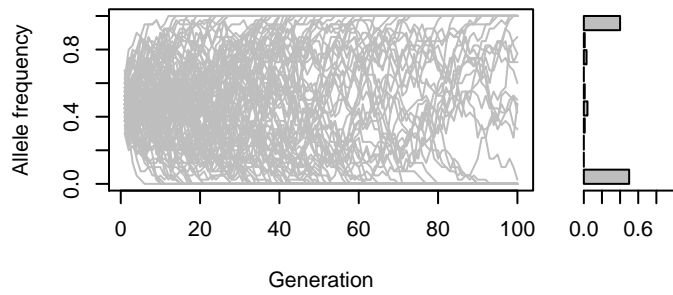


```
Res <- Genetic_drift(NN=10000, TT=20, InitFreq=0.5, Nsim=10)
```

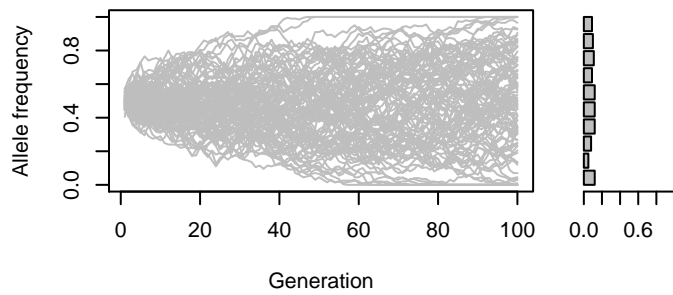


世代数 TT=100 (Nsim = 100)

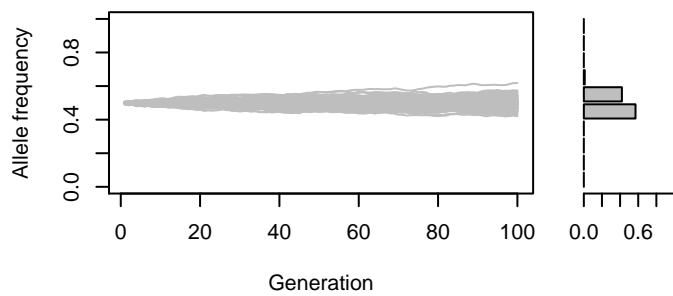
```
Res <- Genetic_drift(NN=20, TT=100, InitFreq=0.5, Nsim=100)
```



```
Res <- Genetic_drift(NN=100, TT=100, InitFreq=0.5, Nsim=100)
```

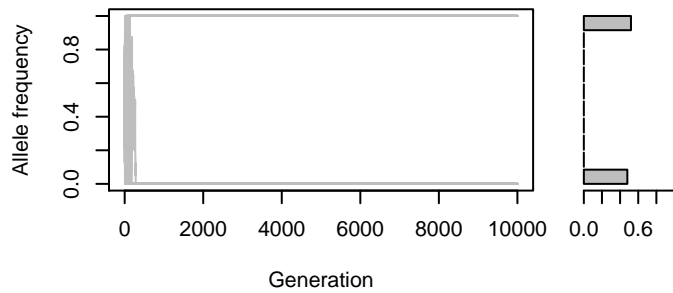


```
Res <- Genetic_drift(NN=10000, TT=100, InitFreq=0.5, Nsim=100)
```

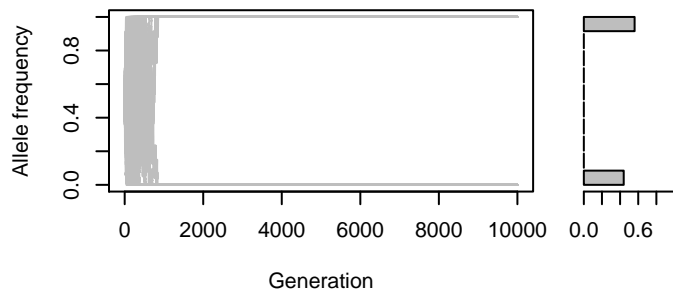


世代数 $TT = 10000$ ($N_{sim} = 100$)

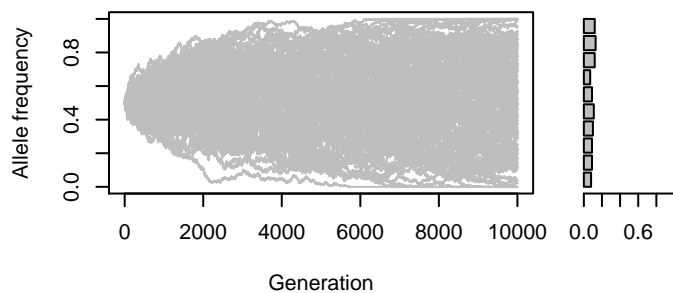
```
Res <- Genetic_drift(NN=20, TT=10000, InitFreq=0.5, Nsim=100)
```



```
Res <- Genetic_drift(NN=100, TT=10000, InitFreq=0.5, Nsim=100)
```



```
Res <- Genetic_drift(NN=10000, TT=10000, InitFreq=0.5, Nsim=100)
```



図から示されたように、世代が進むにつれてどちらかの対立遺伝子に頻度が固定されていることがわかる。集団の個体数が小さいほど遺伝的浮動が大きく、その分だけ早く対立遺伝子の固定が進むこともわかる。

突然変異

いま、ある遺伝子座において異なる2つの対立遺伝子 A_1 と A_2 があるとし、生物を過剰利用した結果、対立遺伝子がいったん A_1 に固定されたとする（すなわち、どの個体の対立遺伝子も A_1 ）。その後、個体数は遺伝的浮動が無視できる程度に回復したとする。そこから n 世代後の A_1 の頻度を p_t 、 A_2 の頻度を $q_t = 1 - p_t$ とする（ただし、 $p_0 = 1$ ）。 A_1 から A_2 への世代当たりの突然変異率を u 、また A_2 から A_1 のそれを v とするとき、頻度の世代間変化は

$$p_{t+1} = (1 - u)p_t + v(1 - p_t) \quad (2)$$

および

$$q_{t+1} = up_t + (1 - v)q_t \quad (3)$$

で表せる。式 (3) は次のように変形できる。

$$q_{t+1} - \frac{u}{u+v} = (1 - u - v) \left(q_t - \frac{u}{u+v} \right).$$

よって、

$$q_t - \frac{u}{u+v} = (1 - u - v)^t \left(q_0 - \frac{u}{u+v} \right)$$

となり、初期条件 $p_0 = 1 (q_0 = 0)$ より

$$q_t = \frac{u}{u+v} \{1 - (1 - u - v)^t\}$$

を得る。ここで、突然変異率がいずれもごく小さな値であると仮定すると、その極限は $q = \lim_{t \rightarrow \infty} q_t = u/(u+v)$ で与えられ、したがって遺伝子頻度の極限は突然変異率の比率で決まることがわかる。なお、この極限值は (3) において平衡状態 $q_{t+1} = q_t = q$ を仮定することでも求められる。

遺伝子流動

いま、互いに移住可能な二倍体生物の分集団を考える。ある遺伝子座において2つの対立遺伝子があるとし、ある時点 t の対立遺伝子 A の頻度を p_t とする。また、集団全体 (分集団の集合) における A の平均頻度を β とする。離散的な時間軸で遺伝的浮動と移住 (移住率 m) を同時に考え、 p_t の平均変化率と分散 (変動の大きさ) がそれぞれ以下のように表せるとする。

$$\begin{aligned}\Delta p &= p_{t+1} - p_t = -mp_t + m\beta, \\ \sigma_{\Delta p}^2 &= p_t(1 - p_t)/(2N_t).\end{aligned}$$

ただし、 N_t は集団の大きさとする。

このような対立遺伝子頻度の変化を連続的に捉えるとき、 p の推移の様子を表す確率密度関数 f は Fokker-Planck 方程式とよばれる拡散方程式

$$\frac{\partial}{\partial t} f(p_0, p, t) = \frac{1}{2} \frac{\partial^2}{\partial p^2} \{ \sigma_{\Delta p}^2 f(p_0, p, t) \} - \frac{\partial}{\partial p} \{ \Delta p f(p_0, p, t) \}$$

をみたすことが知られている。定常状態ではこの変化率がゼロであるから、 $\frac{\partial}{\partial t} f(p_0, p, t) = 0$ をみたす。そこで $f(p_0, p, t) = f$ と表すとき、境界条件から

$$\frac{1}{2} \frac{d}{dp} \{ \sigma_{\Delta p}^2 f \} - \{ \Delta p f \} = 0$$

が導かれる。これは線形1次微分方程式であり、これを解くと、

$$f(p|\beta, \theta) = \frac{1}{B(\theta\beta, \theta(1-\beta))} p^{\theta\beta-1} (1-p)^{\theta(1-\beta)-1}$$

を得る。これをベータ分布という。ただし、 $\theta = 4Nm$ である。ここで、 B は 2.4 で述べたベータ関数である。対立遺伝子 A の頻度 p の平均値は β 、分散は $\beta(1-\beta)/(1+\theta)$ で得られ、 θ が遺伝子流動による遺伝的差異の大きさを決めていることが分かる。