1 Brownian Motion Paths Hitting Lower Barrier

Fix $h > 0$. We will evaluate a standard Brownian motion Y starting at zero at times $t_j = jh$, $j = 1, 2, \dots$. Given a sequence of lower barrier values d_j we say that Y *dies at time* t_j if $Y(t_j) < d_j$.

Now we compute the probability that $Y(t)$ is still alive at time $t = t_j$. This is the event

$$E_j = \big[Y(t_k) \geq d_k, \, \forall k \leq j\big]$$

Set $p_j := P[E_j]$. Observing that $E_{j+1} \subseteq E_j$ and using the Markov property we obtain the recursion

$$\begin{aligned}
p_{j+1} &= P(E_{j+1}) = P(E_{j+1} \cap E_j) = P(E_j)P(E_{j+1}|E_j) \\
&= p_j P\big[Y(t_{j+1}) \geq d_{j+1} \,\big|\, E_j\big] \\
&= p_j P\big[Y(t_{j+1}) \geq d_{j+1} \,\big|\, Y(t_j) \geq d_j\big] = p_j H(j) \quad \text{with} \\
H(j) &:= P\big[Y(t_{j+1}) \geq d_{j+1} \,\big|\, Y(t_j) \geq d_j\big]
\end{aligned}$$

To compute $H(j)$ above, set $t = t_j$ and $h = t_{j+1} - t_j$ so that $t_{j+1} = t + h$ and we have

$$\begin{aligned}
H(j) &= P\big[Y(t+h) \geq d_{j+1} \,\big|\, Y(t) \geq d_j\big] \\
&= P\big[Y(t+h) - Y(t) \geq d_{j+1} - Y(t) \,\big|\, Y(t) \geq d_j\big] \\
&= \frac{1}{P(Y(t) \geq d_j)} \int_{d_j}^{\infty} P\big[Y(t+h) - Y(t) \geq d_{j+1} - y \,\big|\, Y(t) = y\big] P_{Y(t)}(dy) \\
&= \frac{1}{P(Y(t) \geq d_j)} \int_{d_j}^{\infty} P\big[\sqrt{h}N(0,1) \geq d_{j+1} - y\big] P_{Y(t)}(dy) \\
&= \frac{1}{F(-d_j/\sqrt{t})} \int_{d_j}^{\infty} F\left(\frac{y - d_{j+1}}{\sqrt{h}}\right) P_{Y(t)}(dy)
\end{aligned}$$

where $F(x) = P(N(0,1) \leq x)$ and we have used that the increment $Y(t+h) - Y(t)$ is independent of $Y(t)$ and distributed as $N(0,h) = \sqrt{h}N(0,1)$. This yields the recursion

$$p_{j+1} = p_j \times H(j) \quad \text{where} \tag{1}$$

$$H(j) = \frac{1}{F(-d_j/\sqrt{t_j})} \int_{d_j}^{\infty} F\left(\frac{y - d_{j+1}}{\sqrt{h}}\right) P_{Y(t_j)}(dy), \tag{2}$$

with starting condition (note $t_1 = h$)

$$p_1 = P[Y(t_1) \geq d_1] = F(-d_1/\sqrt{h}). \tag{3}$$

$H(j)$ as in (2) is implemented as fcn_H, see file R/BM.R.

Using the bivariate normal cumulative distribution function $F_C(I)$ where C is the covariance matrix of the distribution and I a two dimensional rectange, we can also compute the conditional probability $H(j)$ as

$$
\begin{aligned}
H(j) = P\left(Y(t_{j+1}) \geq d_{j+1} \,\middle|\, Y(t_j) \geq d_j\right) &= \frac{F_{C_j}(I_j)}{P(Y(t_j) \geq d_j)} \\
&= \frac{F_{C_j}(I_j)}{F(-d_j/\sqrt{t_j})}, \quad \text{where} \qquad (4)\\
C_j = Cov(Y(t_j), Y(t_{j+1})) = \begin{pmatrix} t_j & t_j \\ t_j & t_{j+1} \end{pmatrix} & \quad \text{and } I_j = [d_j, +\infty] \times [d_{j+1}, +\infty]. \qquad (5)
\end{aligned}
$$

This is implemented as fcn_H1 in file R/BM.R. The R-package mvtnorm provides the required multinormal distribution function.

We check that the functions fcn_H and fcn_H1 yield the same result, see test_H_function in file R/Tests.R.

The problem: In the code (function runSimulation in file R/BM.R) we run a simulation of Brownian paths and monitor the probability q_j of death at time t_j (by counting death events). For $j > 1$ death at time t_j is the event $E_{j-1} \setminus E_j$ and since $E_j \subseteq E_{j-1}$ it follows that

$$q_j = P(E_{j-1} \setminus E_j) = p_{j-1} - p_j \qquad (6)$$

with $p_0 = P(E_0) = 1$ (being alive at the start of the simulation). We observe a drastic difference between theoretical and realized death probabilities.

1.1 Simulation

We will simulate paths $Y_j = Y(t_j)$ starting from $Y_0 = 0$ and keep count how many of these paths die at each time step t_j. This gives us an empirical probability q_j^e of death at time t_j which we will compare with the theoretical value q_j above.

There will be a deviation due to sample variation and we will report a p-value associated with this deviation. To do this we write the empirical probability q_j^e as the sample mean

$$q_j^e = (X_1 + X_2 + \cdots + X_n)/n \qquad (7)$$

where X_k is the indicator variable defined as

$$X_k = \begin{cases} 1 & \text{if } Y(t) \text{dies at time } t = t_j \text{ in path } k \\ 0 & \text{else} \end{cases}$$

Then the $X_k = 0, 1$ are IID with $P(X_k = 1) = q_j$ from which it follows that $E[X_k] = q_j$ and $Var(X_k) = q_j(1 - q_j)$. By the central Limit Theorem the sample mean q_j^e in (7) is approximately normally distributed with mean $E[q_j^e] = q_j$ and variance

$$Var(q_j^e) = q_j(1 - q_j)/n$$

and because the sample size n in our simulation is extremely large (e.g. 4 million) this approximation is highly accurate.

Setting $\sigma_j := \sqrt{Var(q_j^e)}$ the variable $(q_j^e - q_j)/\sigma_j$ is approximately standard normal leading to a very accurate two sided p-value for a deviation δ of q_j^e from the theoretical value q_j as

$$\begin{aligned} pValue(\delta) &:= P\left(|q_j^e - q_j| \geq \delta\right) \\ &= P\left(|q_j^e - q_j|/\sigma_j \geq \delta/\sigma_j\right) \\ &= 2 * P\left(N(0,1) \leq -\delta/\sigma_j\right) = 2 * F\left(-\frac{\delta\sqrt{n}}{\sqrt{q_j(1-q_j}}\right). \end{aligned} \tag{8}$$

In our simulation this is extremely close to zero for $j \geq 4$ from which it follows that there is something wrong with our theory or simulation code. But what is it?