Reinforcement Learning in Finance

Week 1: Reinforcement Learning

Lesson 2: MDP for BS Model

Igor Halperin

NYU Tandon School of Engineering, 2017

Corporate defaults: The Merton model

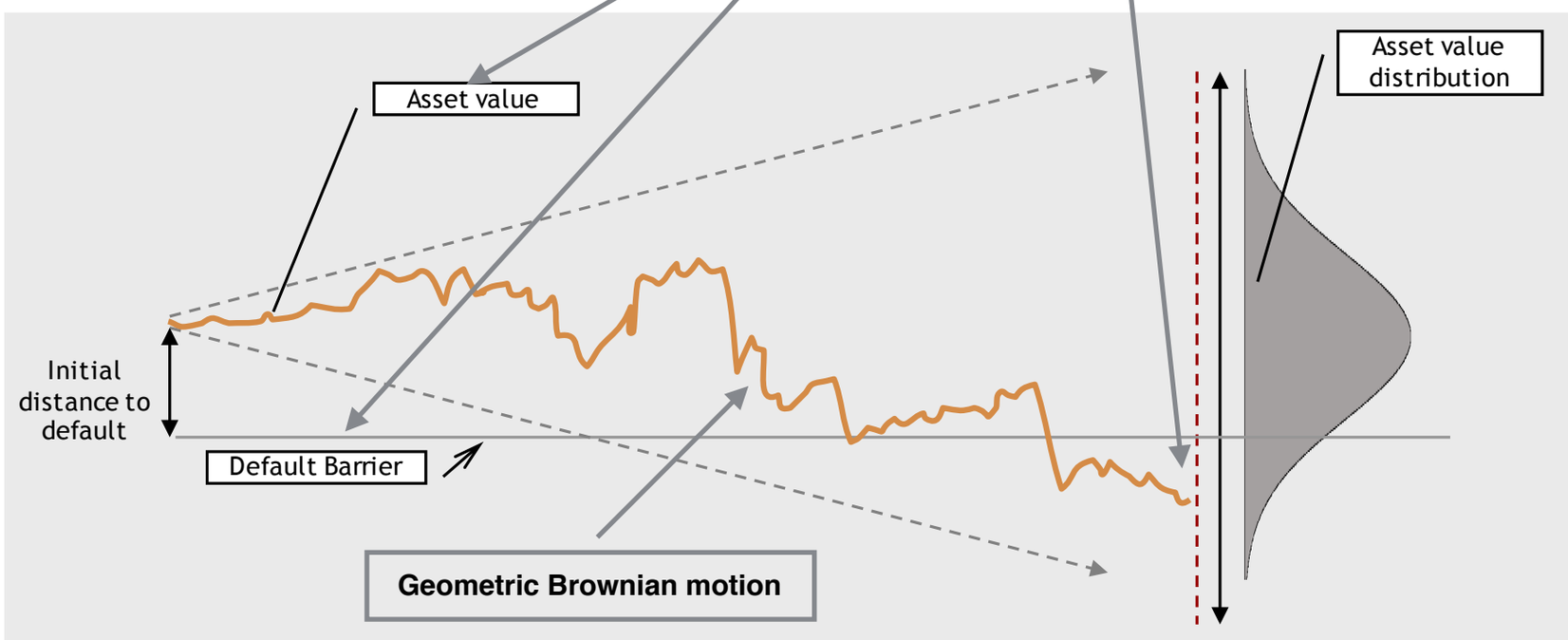
The **Merton model** of corporate defaults (1974-present) is the most popular modeling framework, used as a benchmark for many studies.

A firm is run by equity holders. At time T, they pay the face value of the debt D if the firm (asset) value is larger than D, and keep the remaining amount. If the firm value at time T is less than D, bond holders take over, and recover a "recovery" value V_T, while equity holders get nothing:

$$B_T = \min(V_T, D) = D - \max(D - V_T, 0)$$

$$S_T = \max(V_T - D, 0)$$

Limited liability of equity holders



Merton model as a structural default model

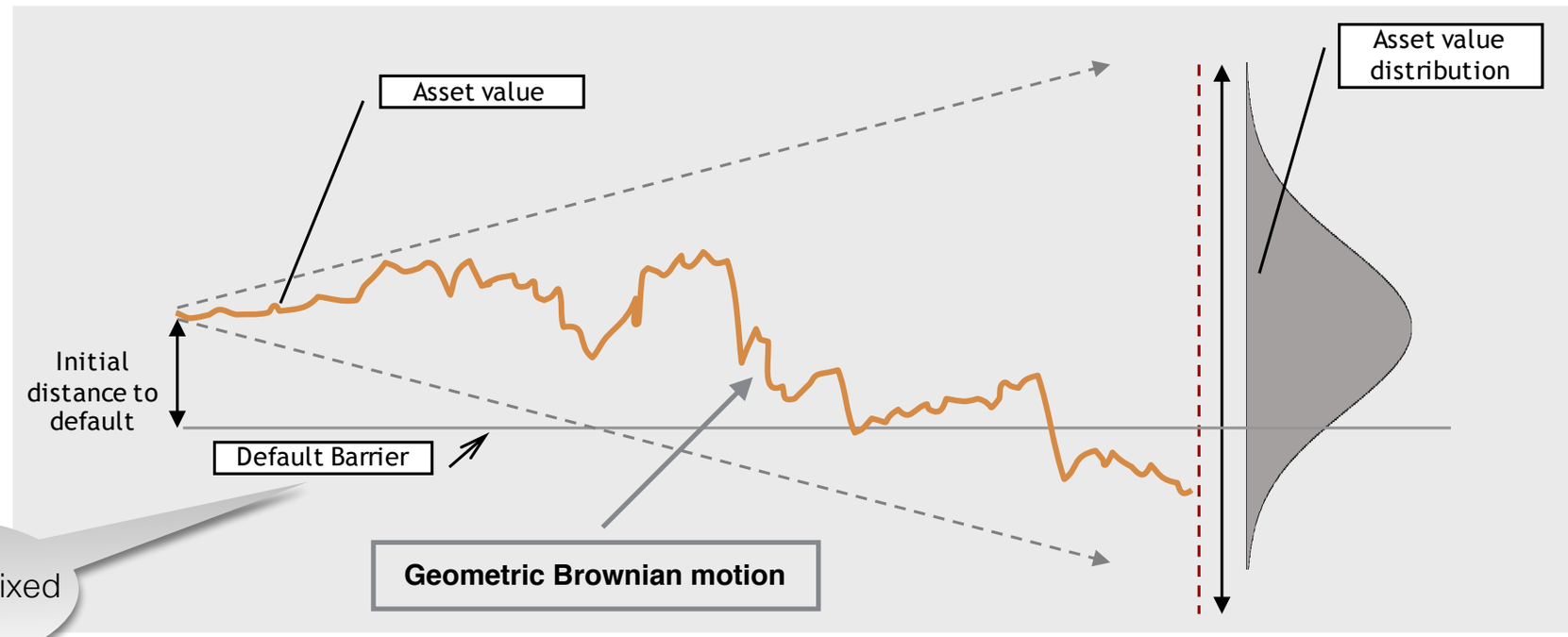
Default probability in the Merton model:

Probabilistic structural model

$$\Pr(default) = \mathbb{E}\left[\mathbb{I}_{V_T<D}\right] = \Pr(V_T < D) = N(-d_2)$$

$$d_2 = \frac{\log\frac{V_t}{D} + \left(r - \frac{\sigma_V^2}{2}\right)(T-t)}{\sigma_V\sqrt{(T-t)}}$$

Depends only on the Assets/Debt ratio and asset volatility



The barrier is fixed and static

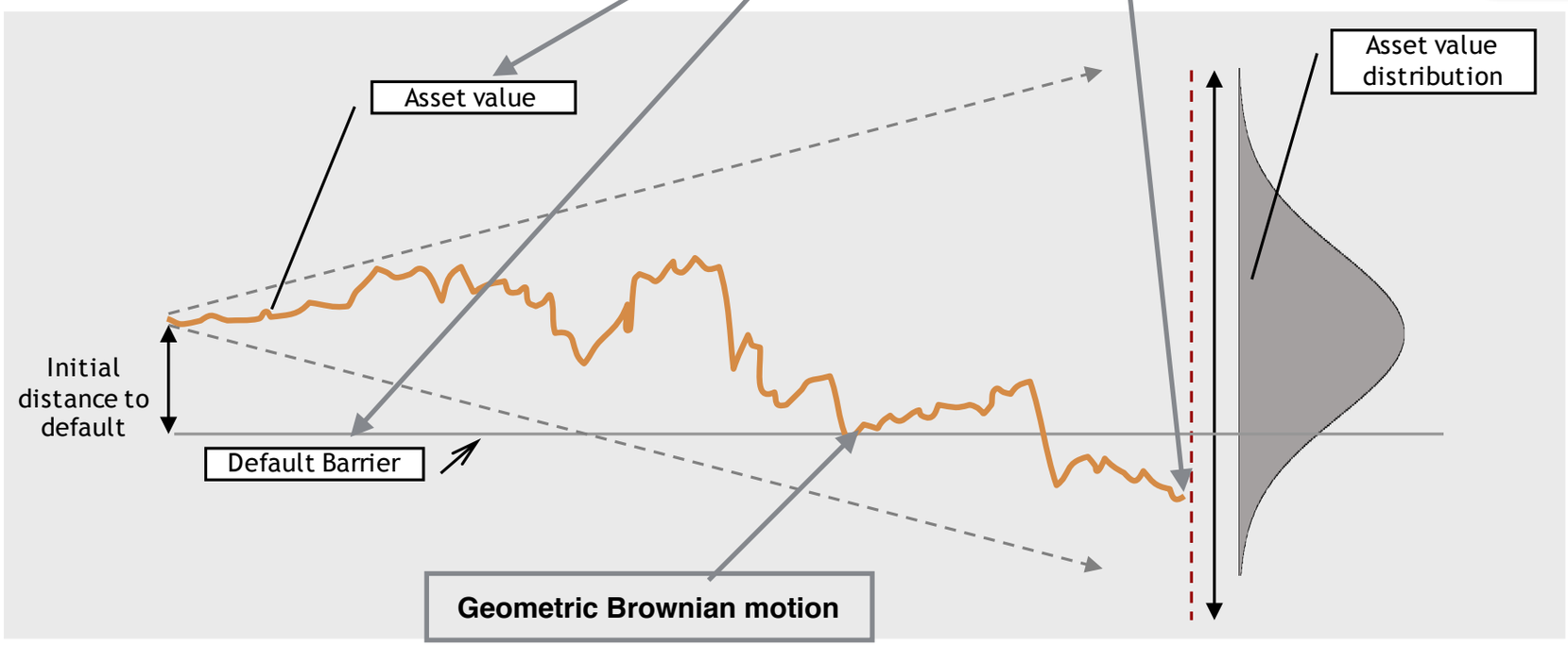
Option pricing: Black-Scholes-Merton

The equity holder offers you to buy her **future** (to be received at time T) stock **now**.
Currently the firm costs $V_0 = \$1\text{m}$, and has debt $D = \$600\text{K}$.
How much should you pay **now**?

$$B_T = \min(V_T, D) = D - \max(D - V_T, 0)$$

$$S_T = \max(V_T - D,\ 0)$$

Limited liability of equity holders



Stock option: replace the firm value by the stock price!
Stock option is a contract to get a stock for a pre-determined price at some future time

Stock options

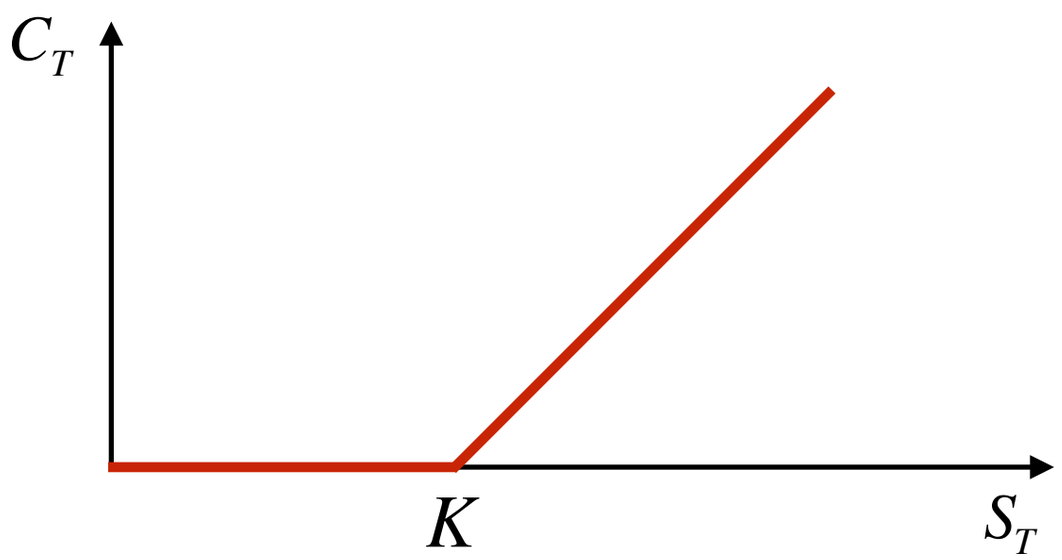
European call option on a stock is a contract to get a stock S_T for a pre-determined price (strike) K at some future time T
E.g. the current stock price is $S_0 = \$100$, and the strike $K = \$110$.
How much should you pay **now**?

$$C_T = \max(S_T - K, 0)$$

Option payoff at maturity T

Stock price at maturity T

Option strike K



Control question

Select all correct answers

1. Financial options are particular sort of financial derivatives whose value is derived from the value of underlying assets (e.g. stocks).
2. A European put option gives a seller of this option a right, but not an obligation, to sell a stock underlying this option, for a pre-specified price (strike) K at some future time T.
3. The profit for a buyer of a call option is equal to $\max(S_T - K, 0)$
4. The profit for a seller of a put option is equal to $\max(S_T - K, 0)$

Correct answers: 1, 2, 3

Control questions for Video 3:

1. A replicating portfolio for an option is made from another option and cash.
2. A replicating portfolio for an option is made of a stock and cash.
3. The purpose of replicating portfolio is to dynamically track option value in different state of the world, so that a 'total portfolio' made of the stock and the replication portfolio has a zero (or close to zero) value.
4. The law of one price states that securities that have the same value in all states of the world should also have the same price.

Correct answers: 2,3,4

The BSM model: lognormal stock dynamics

Lognormal process for the stock price in continuous time in the BSM model (a Geometric Brownian motion with a drift):

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

Equivalently (by Ito's lemma!):

$$d\log S_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t$$

μ - the drift, σ - volatility rate, W_t - standard Brownian motion

Discretize the stock price

Introduce a new variable :

$$X_t = -\left(\mu - \frac{\sigma^2}{2}\right)t + \log S_t$$

Then :

$$d\,X_t = -\left(\mu - \frac{\sigma^2}{2}\right)dt + d\log S_t = \sigma\, dW_t$$

Therefore X_t is a standard Brownian motion, scaled by volatility σ
If we know the value of X_t in a given scenario, then S_t is also known:

$$S_t = e^{X_t + \left(\mu - \sigma^2/2\right)t}$$

Discrete-time approximation:

$$\Delta X_t = \sigma\sqrt{\Delta t}\,\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

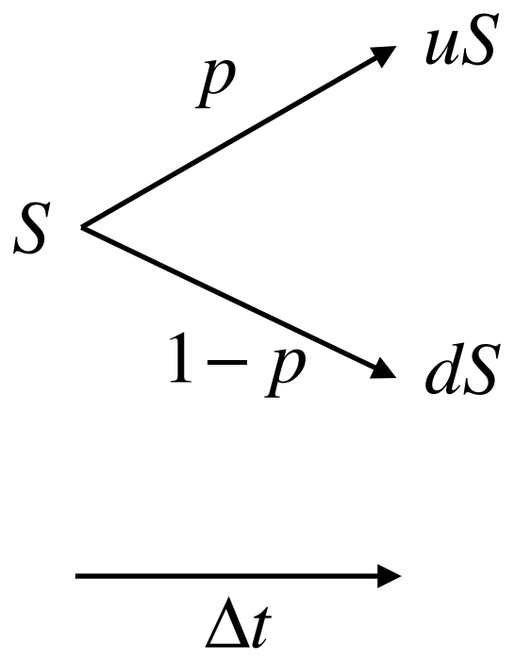
Binomial tree approximation

Discrete-time approximation for continuous-space dynamics:

$$S_t = e^{X_t + \left(\mu - \sigma^2/2\right)t}$$

$$\Delta X_t = \sigma\sqrt{\Delta t}\,\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

The binomial model gives a discrete-state approximation to this dynamics. The stock can risk to a value uS with probability p, or fall to a value dS with probability $1-p$, with $0 < d < 1 < u$



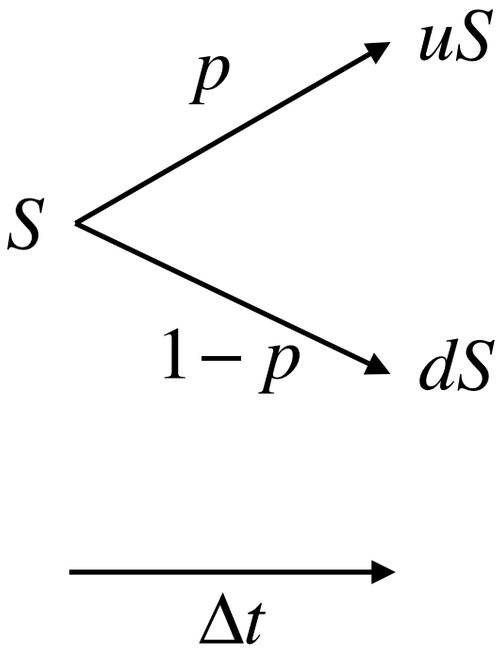
Binomial model: choice of parameters

We can choose three parameters to match the mean and standard deviation of the future stock price.

Mean $E[S] = Se^{r\Delta t} = puS + (1-p)dS$

Variance: $Var[S] = S^2 e^{2r\Delta t}\left(e^{\sigma^2 \Delta t} - 1\right) = S^2\left(pu^2 + (1-p)d^2\right)$

Have two constraints for three parameters:



$$pu + (1-p)d = e^{r\Delta t}$$

$$pu^2 + (1-p)d^2 = e^{(2r+\sigma^2)\Delta t}$$

Popular choices:

Cox, Ross, Rubinstein: $u = \frac{1}{d}$

Jarrow, Rudd: $p = \frac{1}{2}$

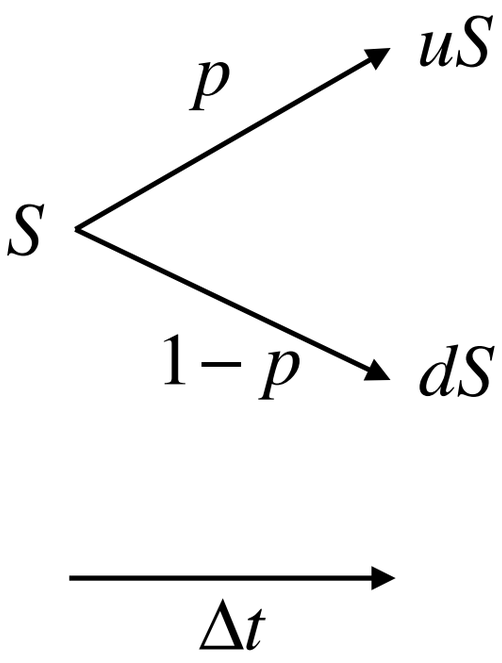
Binomial model: Cox, Ross, Rubinstein

Have two constraints for three parameters:

$$pu + (1-p)d = e^{r\Delta t}$$

$$pu^2 + (1-p)d^2 = e^{(2r+\sigma^2)\Delta t}$$

Cox, Ross, Rubinstein (CRR): $u = \dfrac{1}{d}$



$$u = e^{\sigma\sqrt{\Delta t}}$$

$$d = e^{-\sigma\sqrt{\Delta t}}$$

$$p = \frac{a-d}{u-d}$$

$$a = e^{r\Delta t}$$

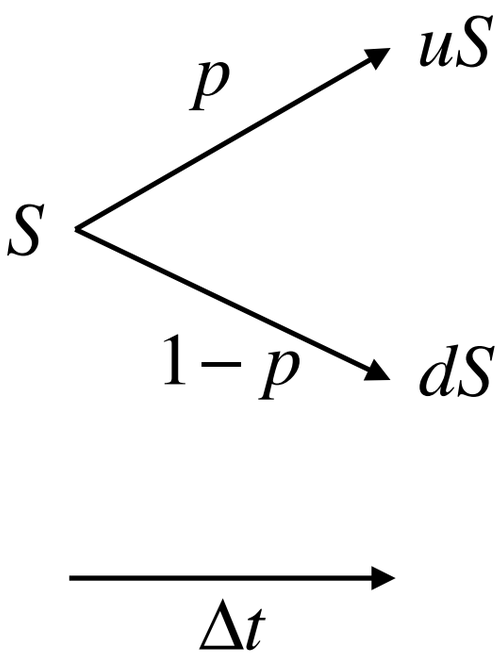
Binomial model: Jarrow-Rudd

Have two constraints for three parameters:

$$pu + (1-p)d = e^{r\Delta t}$$

$$pu^2 + (1-p)d^2 = e^{(2r+\sigma^2)\Delta t}$$

Jarrow-Rudd: $p = \frac{1}{2}$



$$u = e^{(r-\sigma^2/2)\Delta t + \sigma\sqrt{\Delta t}}$$

$$d = e^{(r-\sigma^2/2)\Delta t - \sigma\sqrt{\Delta t}}$$

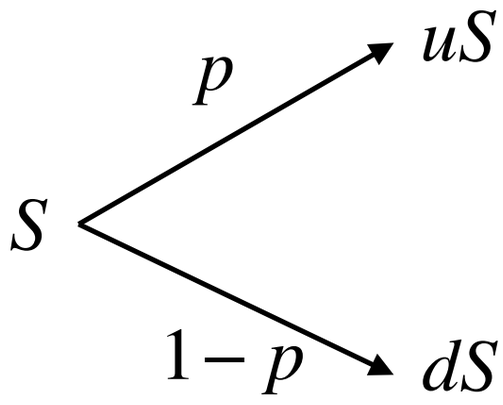
$$p = \frac{a-d}{u-d} = \frac{1}{2}$$

$$a = e^{r\Delta t}$$

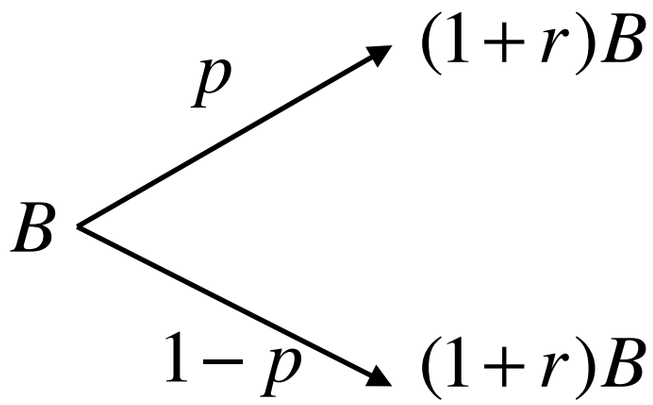
Binomial model: pricing by replication

Arbitrage pricing: price the stock by constructing a replicating portfolio of a stock and a bond.

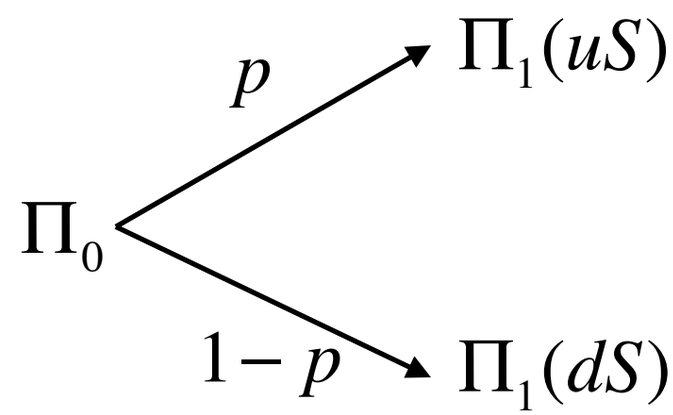
Stock



Bond



Portfolio



$$\Pi_0 = \Delta S + \theta B$$

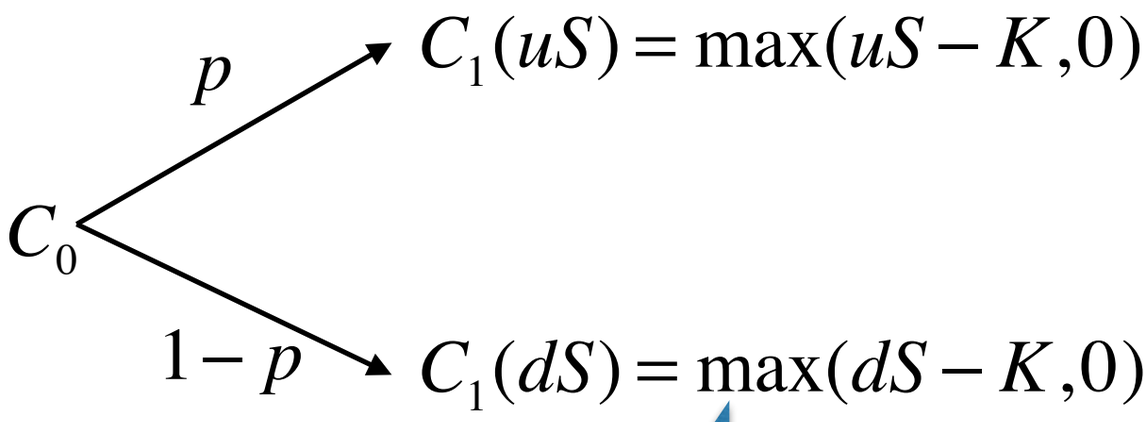
$$\Pi_1(uS) = \Delta uS + \theta(1+r)B$$

$$\Pi_1(dS) = \Delta dS + \theta(1+r)B$$

Binomial model: pricing by replication

Arbitrage pricing: price the stock by constructing a replicating portfolio of a stock and a bond.

Stock option



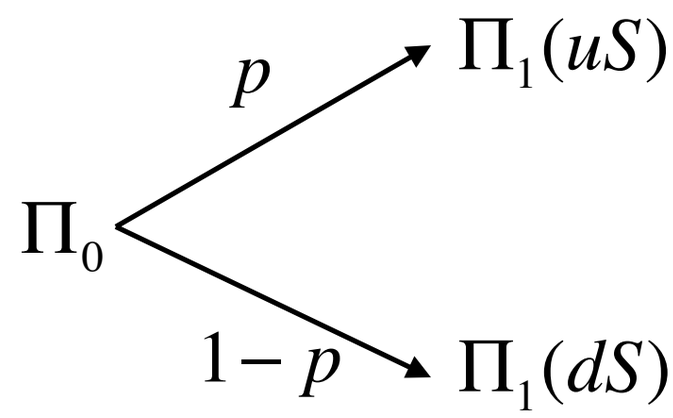
$$C_0 = \Pi_0 = \Delta S + \theta B$$

$$C_1(uS) = \Pi_1(uS) = \Delta uS + \theta(1+r)B$$

$$C_1(dS) = \Pi_1(dS) = \Delta dS + \theta(1+r)B$$

price matched in every state of the world

Portfolio



$$\Pi_0 = \Delta S + \theta B$$

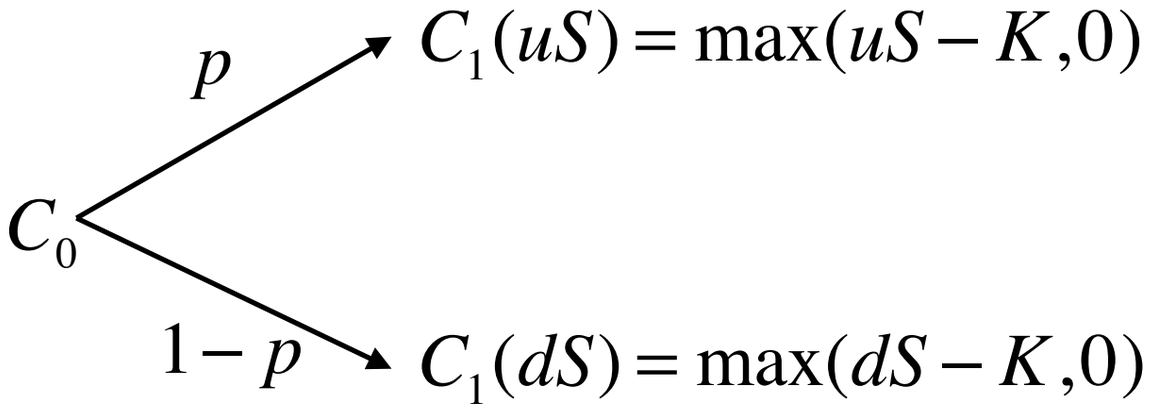
$$\Pi_1(uS) = \Delta uS + \theta(1+r)B$$

$$\Pi_1(dS) = \Delta dS + \theta(1+r)B$$

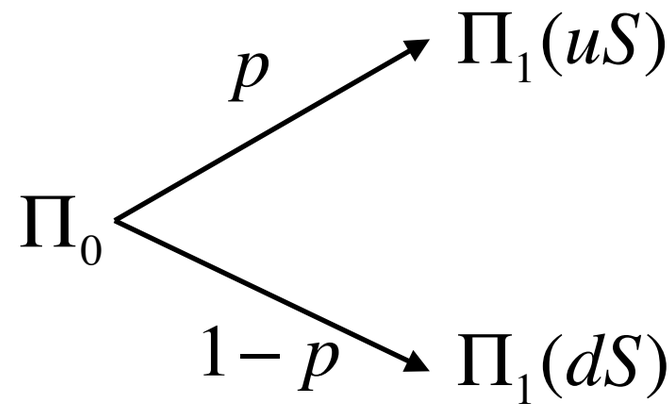
Binomial model: optimal replication

Arbitrage pricing: price the stock by constructing a replicating portfolio of a stock and a bond.

Stock option



Portfolio



$$C_0 = \Pi_0 = \Delta S + \theta B$$
$$C_1(uS) = \Pi_1(uS) = \Delta uS + \theta(1+r)B$$
$$C_1(dS) = \Pi_1(dS) = \Delta dS + \theta(1+r)B$$

$$\Delta = \frac{C_1(uS) - C_1(dS)}{uS - dS}$$
$$\theta B = \frac{uC_1(uS) - dC_1(dS)}{(1+r)(u-d)}$$

Choosing the **control** (Delta) this way completely **eliminates risk** of the option! This happens **only** for the binomial model in discrete time, and for the BSM in continuous time!

Constructing the Markov chain

Re-cap: We introduced a new variable :

$$X_t = -\left(\mu - \frac{\sigma^2}{2}\right)t + \log S_t$$

Then :

$$d\,X_t = -\left(\mu - \frac{\sigma^2}{2}\right)dt + d\log S_t = \sigma\, dW_t$$

Therefore X_t is a standard Brownian motion, scaled by volatility σ
If we know the value of X_t in a given scenario, then S_t is also known:

$$S_t = e^{X_t + \left(\mu - \sigma^2/2\right)t}$$

Discrete-time approximation:

$$\Delta X_t = \sigma\sqrt{\Delta t}\,\varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

As X_t is a stationary process without a drift (a martingale!), it stays around the current value in the sense of expectations, thus is easier to discretize!

Constructing the Markov chain

Instead of a tree, we want to discretize the dynamics to a Markov Chain (first-order Markov process with discrete states), to convert this problem to a Markov Decision Problem (MDP)

Given: initial stock price S_0
Want: create a set of representative logarithmic stock prices

$$\left[\log S_0 - I_p, \log S_0 + I_p\right]$$

$$I_p = \delta(m)\sigma\sqrt{T\,\Delta t}$$

Here (Duan and Simonato, 2001)

Δt - the time step (in years)
m - the number of discrete states (should be an odd number)
T - the option maturity (in years)
σ - stock volatility (in annualized terms)
$\delta(m)$ - a scaling factor, pick $\delta(m) = 2 + \log(\log(m))$

Grid:

$$p_i = \log S_0 + \frac{2i - m - 1}{m - 1} I_p, \quad i = 1, \ldots, m$$

Markov chain: transition probabilities

Create cells separated by values of the grid:

$$C_1 = (c_1, c_2), C_i = [c_i, c_{i+1}), i = 2,...,m$$

Here (Duan and Simonato, 2001)

$$c_1 = -\infty, c_i = \frac{p_i + p_{i-1}}{2}, i = 2,...,m, c_{m+1} = +\infty,$$

$$p_i = \log S_0 + \frac{2i - m - 1}{m - 1} I_p, \quad i = 1,...,m$$

Midpoints of cells are the values
Transition probabilities between the cells:

$$p_{ij} = N\left(\frac{c_{j+1} - p_i - (\mu - 0.5\sigma^2)\Delta t}{\sigma\sqrt{\Delta t}} \right) - N\left(\frac{c_j - p_i - (\mu - 0.5\sigma^2)\Delta t}{\sigma\sqrt{\Delta t}} \right)$$

Here $N(x)$ is the cumulative normal distribution
Alternative approach: match moments of the lognormal process (Kushner, 1990)

Option pricing is a risky business

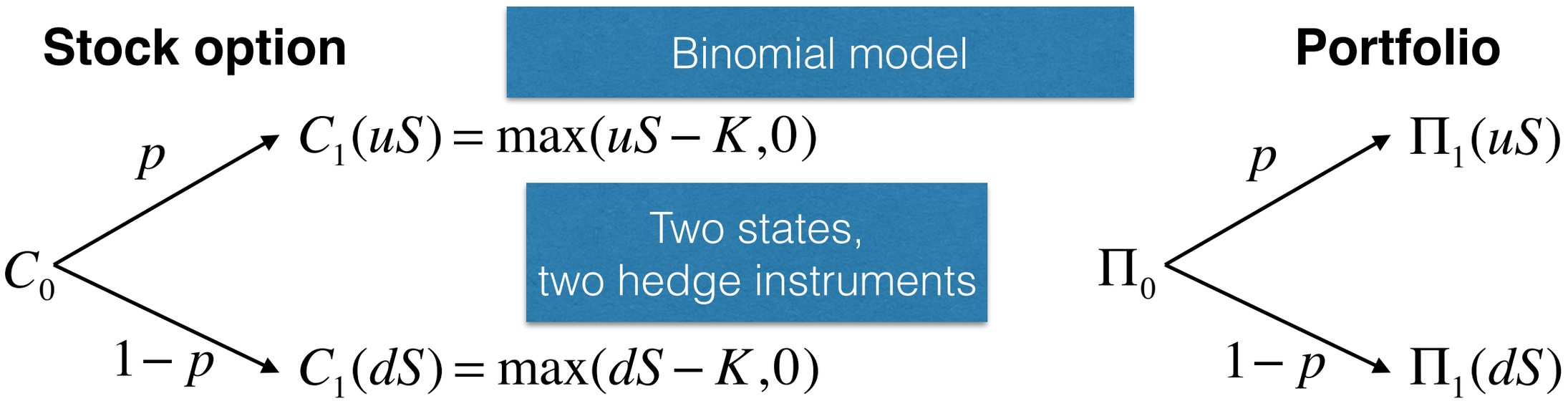
- The classical Black-Scholes-Merton model relies on **three key assumptions**:
  - Markets are **complete** (perfect replication is possible)
  - **Continuous** rebalancing of a **hedge** portfolio
  - **Zero** transaction costs

- The BSM theory results in options being redundant instruments with zero risk and unique price

- None of the above assumptions hold in practice, and option pricing and trading is **risky**!

Market incompleteness

- Complete markets:
  - Any derivative payoff can be perfectly replicated by a portfolio of basic instruments in all states of the world
- **Real markets are incomplete**!
  - Multiple factors at play, an option price is **not** just a function of a stock price!
  - **Non-zero transaction costs** prevent continuous or frequent re-hedging, but risk cannot be eliminated if re-hedging with a finite frequency
  - Market incompleteness results in a **residual risk** in derivatives
- In the Black-Scholes-Merton model (continuos-time, continuous space), risk is eliminated completely. In its discrete-time approximation, risk is completely eliminated for a binomial approximation, but remains for a multinomial approximation:



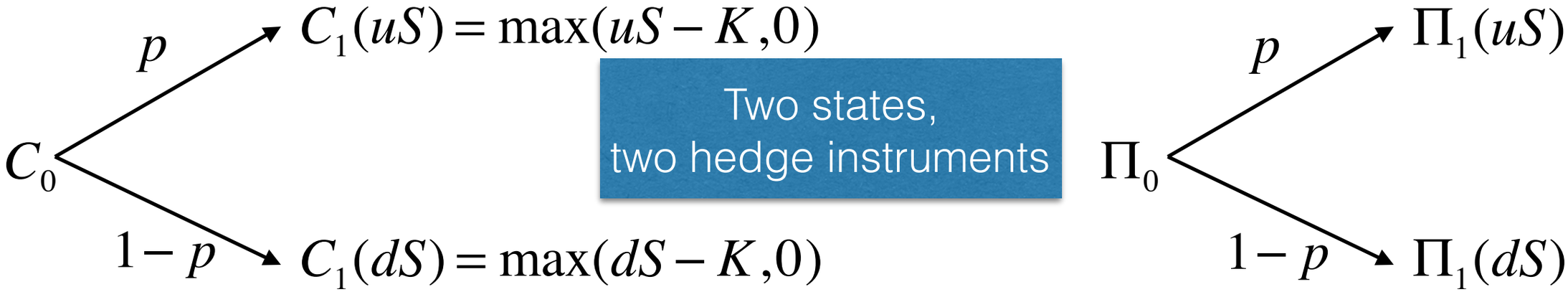
Market incompleteness in the BSM model

- In the Black-Scholes-Merton model (**continuos-time**, continuous space), risk is eliminated completely. In its **discrete-time** approximation, **risk** is completely eliminated for a binomial approximation, but **remains** for a multinomial (trinomial etc.) approximation:

Stock option

Binomial model

Portfolio

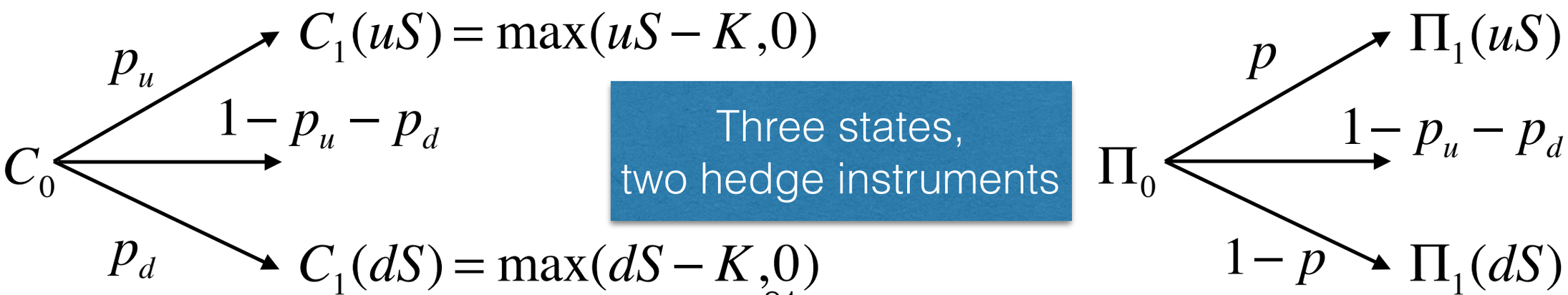


Two states,
two hedge instruments

Stock option

Trinomial model

Portfolio



Three states,
two hedge instruments

MDP approximation to the BSM model

- We want to approximate the BSM model by a MDP model:
  - Need to **discretize time**
  - Need to **discretize a state space**

- Discretization of time in the BSM model produces an ambiguity in how to hedge risk in options

- We want to tractable discrete-time dynamics that converge to the BSM model for vanishing time steps, to ensure that our results can be taken all the way to the pure (risk-free) setting of the original BSM model

- On the other hand, keeping time steps finite allows us to model more realistic market scenarios (residual risk in options) than the original BSM model or the binomial model.

- Can keep the same state dynamics (a lognormal stock price) as in the original BSM model, extensions to more complex dynamics only add computational costs but do not change the overall framework.

Hedging and pricing under market incompleteness

- In the BSM model, the option price is the price of the hedge portfolio (price of hedge). Both the hedging strategy and option price are fixed in a unique way.

- When the residual risk of a hedged option cannot be completely eliminated, the optional hedging strategy is the one that minimizes this residual risk, under a chosen metric

- As there are multiple ways to choose the measure for risk (e.g. volatility of total P&L, VaR, expected shortfall etc.), the notion of an **optimal** strategy is only meaningful when the risk measure is specified first.

- **Hedging comes ahead of pricing.** The option price can only be determined after the hedge is specified.

- Hedging, and hence the price, can be different for different dealers, depending on their risk metric, optimal hedge strategy, and risk preferences

Replicating (hedge) portfolio

- Assume we sell an option and hedge it with the stock. The number of shares we hold at time t is θ_t

- Let B_t be the amount invested in a risk-free bank account a time t

- The value of a replicating portfolio (the hedge portfolio process):

$$V_t = \theta_t S_t + B_t$$

- The option: a European option $C_t(S_t)$ with the expiration (maturity) T, and the terminal payoff $H_T(S_T)$

- At maturity T: the hedge is closed, $\theta_T = 0$, the replicating portfolio should match the derivative payoff. This fixes the terminal value that the bank account needs to have:

$$V_T = \theta_T S_T + B_T = B_T = H_T$$
$$\Rightarrow B_T = H_T$$

The present value of hedge portfolio

- We should have that at a future time T, the value of the portfolio (the same as the bank account B_T at this time) should match the option payoff in all future states of the world): $B_T = H_T$

- Now (at time $t = 0$), the exact amount B_0 needed to be put at a bank account is unknown, as it depends on the future.

- Let's require that the amount in the bank account B_t should be able to compensate the future change of the hedge (**self-financing portfolio**)

$$e^{r\Delta t} B_t + \theta_t S_{t+1} = B_{t+1} + \theta_{t+1} S_{t+1}$$

- This produces a recursive relation for B_t that can be calculated backward in time from $t = T$ to $t = 0$ given the **future** values of the hedge and the stock:

$$B_t = e^{-r\Delta t} \left(B_{t+1} + \left(\theta_{t+1} - \theta_t \right) S_{t+1} \right)$$

- This means that B_0 is **random** at $t = 0$ (measurable only at $t = T$)

The hedge portfolio process

- We should have that at a future time T, the value of the portfolio (the same as the bank account B_T at this time) should match the option payoff in all future states of the world): $B_T = H_T$

- The recursive relation for B_t that can be calculated backward in time from $t = T$ to $t = 0$ given the **future** values of the hedge and the stock:

$$B_t = e^{-r\Delta t}\left(B_{t+1} + \left(\theta_{t+1} - \theta_t\right)S_{t+1}\right)$$

- Plug the recursive relation for B_t into the formula for $V_t = \theta_t S_t + B_t$:

$$\begin{aligned} V_t &= \theta_t S_t + e^{-r\Delta t}\left(B_{t+1} + \left(\theta_{t+1} - \theta_t\right)S_{t+1}\right) \\ &= e^{-r\Delta t}\left(V_{t+1} - \theta_t\left(S_{t+1} - e^{r\Delta t}S_t\right)\right) \\ &= e^{-r\Delta t}\left(V_{t+1} - \theta_t\,\Delta S_t\right), \quad \Delta S_t \equiv S_{t+1} - e^{r\Delta t}S_t \end{aligned}$$

- This means that the hedge portfolio $V_0 = \theta_0 S_0 + B_0$ is **random** at $t = 0$ as B_0 is random (measurable only at $t = T$)

Monte Carlo Forward-Backward algorithm

- If the hedge ratio θ_t is known as a function of stock price S_t, then the distribution of the hedge portfolio value at time $t = 0$ can be estimated using a Monte Carlo simulation:

- **The Forward Pass:**
  - Simulate a set of paths of the stock price $S_0 \to S_1 \to S_2 \to \ldots \to S_T$

- **The Backward Pass:**
  - On each simulated path, set $V_T = H_T$
  - On each simulated path, use the recursive relation for V_t to compute backward in time from $t = T$:

$$V_t = e^{-r\Delta t}\left(V_{t+1} - \theta_t \, \Delta S_t\right), \quad \Delta S_t \equiv S_{t+1} - e^{r\Delta t} S_t$$

- This produces a distribution of V_0 that will be used to decide on the option price
- Note that we can re-utilize given simulated paths of the stock price to evaluate distributions obtained under different hedging scenarios
- Can use real historical data instead of simulated data, if we have enough data…

Optimal hedging

- Choose the hedge ratio θ_t at each time step, going backward, by a minimization of variance of the hedge portfolio across all Monte Carlo paths (cross-sectional analysis!) :

$$\theta_t = \arg\min_{\theta} Var\left[V_t \mid \mathcal{F}_t\right]$$

$$= \arg\min_{\theta} Var\left[V_{t+1} - \theta\,\Delta S_t \mid \mathcal{F}_t\right]$$

- Important: as the actual hedge cannot look into the future, in this calculation we condition on the currently available information $\mathcal{F}_t$ available at time t!

- As we condition on $\mathcal{F}_t$ here, variance in V_t is due to variance of B_t. Therefore, the optimal hedge minimizes the cost of hedge capital at each time step t

- The optimal hedge can be found analytically by setting the derivative to zero:

$$\theta_t^* = \frac{Cov(V_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)}$$

- The resulting penalty for mis-hedge (computed for each step on a path):

$$p_t = Var\left[V_{t+1} - \theta_t^*\ \Delta S_t \mid \mathcal{F}_t\right]$$

Fair option price

- A fair option price at time t:

$$\hat{C}_t = \mathbb{E}_t\left[V_t \mid \mathcal{F}_t\right]$$

- Can get a recursive formula for $\hat{C}_t$ by taking conditional expectation in the recursive formula for V_t and using the tower law of conditional expectations:

$$\begin{aligned}\hat{C}_t &= \mathbb{E}_t\left[e^{-r\Delta t}V_{t+1} \mid \mathcal{F}_t\right] - \theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right] \\ &= \mathbb{E}_t\left[e^{-r\Delta t}\mathbb{E}_t\left[V_{t+1} \mid \mathcal{F}_{t+1}\right] \mid \mathcal{F}_t\right] - \theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right] \\ &= \mathbb{E}_t\left[e^{-r\Delta t}\hat{C}_{t+1} \mid \mathcal{F}_t\right] - \theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right]\end{aligned}$$

- Can use the tower law of iterated expectations to equivalently express the optimal hedge in terms of $\hat{C}_t$

$$\theta_t^* = \frac{Cov(V_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)} = \frac{Cov(\hat{C}_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)}$$

The Black-Scholes limit: the hedge

- We got $\hat{C}_t = \mathbb{E}_t\left[V_t \mid \mathcal{F}_t\right]$ or alternatively

$$\hat{C}_t = \mathbb{E}_t\left[e^{-r\Delta t}\hat{C}_{t+1} \mid \mathcal{F}_t\right] - \theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right]$$

- The optimal hedge

$$\theta_t^* = \frac{Cov(V_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)} = \frac{Cov(\hat{C}_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)}$$

- The Black-Scholes limit: $\Delta t \to 0$
- In the BS limit, can use the first-order Taylor expansion

$$\hat{C}_{t+1} = C_t + \frac{\partial C_t}{\partial S_t}\Delta S_t + O\left(\Delta t\right)$$

- Substitute this into the expression for the optimal hedge:

$$\theta_t^{BS} = \lim_{\Delta t \to 0} \frac{Cov(\hat{C}_{t+1}, \Delta S_t \mid \mathcal{F}_t)}{Var(\Delta S_t \mid \mathcal{F}_t)} = \frac{\partial C_t}{\partial S_t}$$

- This means that the correct BS hedge ratio is recovered in the BS limit

The Black-Scholes limit: the price

We got $\hat{C}_t = \mathbb{E}_t\left[V_t \mid \mathcal{F}_t\right]$ or alternatively

$$\hat{C}_t = \mathbb{E}_t\left[e^{-r\Delta t}\hat{C}_{t+1} \mid \mathcal{F}_t\right] - \theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right]$$

The second term in the BS limit $\Delta t \to 0$

$$\theta_t\,\mathbb{E}_t\left[\Delta S_t \mid \mathcal{F}_t\right] \underset{dt\to 0}{\to} \theta_t^{BS} S_t(\mu - r) = (\mu - r)S_t\frac{\partial C_t}{\partial S_t}$$

To compute the first term in the BS limit, use the **second**-order Taylor expansion

$$\hat{C}_{t+1} = C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}dS_t + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}\left(dS_t\right)^2 + \ldots$$

$$= C_t + \frac{\partial C_t}{\partial t}dt + \frac{\partial C_t}{\partial S_t}S_t\left(\mu\,dt + \sigma dW_t\right) + \frac{1}{2}\frac{\partial^2 C_t}{\partial S_t^2}S_t^2\left(\sigma^2 dW_t^{\,2} + 2\mu\sigma\,dW_t\,dt\right) + \ldots$$

Substitute these into the first formula, use $\mathbb{E}_t\left[dW_t\right] = 0, \mathbb{E}_t\left[dW_t^{\,2}\right] = dt$, and simplify:

$$\frac{\partial C_t}{\partial t} + rS_t\frac{\partial C_t}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2\frac{\partial^2 C_t}{\partial S_t^2} - rC_t = 0$$

We recovered the original BS equation as the BS limit of our formulas!

Option pricing with discrete hedging

- The expected value of the hedge portfolio: $\hat{C}_0 = \mathbb{E}_{t=0}\left[V_0\right]$
- The price the option seller should request have to cover the expected value of the hedge, plus a premium for risk defined e.g. as variance $Var\left[V_0\right]$, or the Conditional Value at Risk (CVaR) $CVaR\left[V_0\right]$

$$CVaR\left[V_0\right] = \mathbb{E}_{t=0}\left[V_0 \,|\, V_0 \geq VaR_\alpha\left(V_0\right)\right]$$

- Here $VaR_\alpha\left(V_0\right)$ is the Value at Risk at the confidence level α:

$$VaR_\alpha\left(V_0\right) = K : \Pr\left[V_0 > K\right] = \alpha$$

- The option price can be chosen as (here λ is the risk aversion parameter)

$$C_0 = \mathbb{E}_{t=0}\left[V_0\right] + \lambda CVaR\left[V_0\right]$$

- The dealer's option price depends on both dealer's hedging strategy and risk tolerance!

Hedging by risk minimization

- Assume we sell an option and hedge it with the stock. The number of shares we hold between time t and $t+1$ is θ_{t+1} (set at t)

- Let $X_t = e^{-rt} S_t$ be the discounted stock price, η_t be the amount invested in a risk-free bank account, and θ_t be the hedge at time t (amount of stocks purchased at $t-1$ and held from $t-1$ to t)

- The value of a replicating portfolio (the discounted value process):

$$V_t = \theta_t X_t + \eta_t$$

- The option: a European option $C_t(S_t)$ with the expiry time (maturity) T, and the terminal payoff $H_T(S_T)$

- At maturity T: the hedge is closed, $\theta_T = 0$, the replicating portfolio should match the derivative payoff. This fixes the terminal value that the bank account needs to have:

$$V_T = \theta_T X_T + \eta_T = \eta_T = H_T$$

$$\Rightarrow \eta_T = H_T$$

Control question

Select all correct answers

1. The name “Markov Processes” first historically appeared as a result of a misspelled name “Mark-Off Processes” that was previously used for random processes that describe learning in certain types of video games, but has become a standard terminology since then.
2. The goal of (risk-neutral) Reinforcement Learning is to maximize the expected total reward by choosing an optimal policy.
3. The goal of (risk-neutral) Reinforcement Learning is to neutralize risk, i.e. make the variance of the total reward equal zero.
4. To goal of risk-sensitive Reinforcement Learning is to teach a RL agent to pick action policies that are most prone to risk of failure. Risk-sensitive RL is used e.g. by venture capitalists and other sponsors of RL research, as a quick tool to access the feasibility of new RL projects.

Correct answers: