15.455x Mathematical Methods of Quantitative Finance

Week 3: Time Series Models

Paul F. Mende
MIT Sloan School of Management

Finance at MIT
Where ingenuity drives results

Model Estimation and Identification

Finance at MIT
Where ingenuity drives results

"If your experiment needs statistics, you ought to have done a better experiment."

– *Ernest Rutherford*

But in finance, we're really quite different from laboratory sciences. Most of our data is noisy. We need statistics

Model identification and estimation

- **Model identification and estimation**
  - Which model best fits the data?
  - What are the best parameter estimates?
  - Stationarity (in models and in the so-called real world)
  - Parameter estimation and regression models
  - Exploratory data analysis
  - Scaling relationships
  - Hypothesis testing
  - Testing the Random Walk Hypothesis
  - Order identification of multiparameter models

probability distributions are don't change over time, the so– called time's translation invariance.

Statistics and estimators

- A **statistic** is a **function of the data**
  - Examples: sample mean, sample variance

$$\mu = \frac{1}{T}\sum_{t=t_1}^{t_T} r_t,$$

$$\sigma^2 = \frac{1}{T-1}\sum_{t=t_1}^{t_T}(r_t - \mu)^2$$

  - Descriptive statistics reduce a large number of observations to a few representative values

  - For well-chosen statistics, different situations that produce the same values should be comparable or equivalent in some useful sense – **whether or not** they are related to a model

Statistics and estimators

- A **estimator** is a **random variable**.
  - Function of future observations or data
  - As a random variable, it has a distribution
    - Sampling distribution
    - First and second moments easily derivable
    - Via CLT, full distribution approaches Gaussian (for well-behaved estimators) for large number of observations

- An **estimate** is a **number**
  - The definite value given by an estimator on a specified set of data

    particular set of realizations

  - Examples:

$$\widehat{\mu} = \frac{1}{T}\sum_{t=t_1}^{t_T} r_t,$$

$$\widehat{\sigma^2} = \frac{1}{T-1}\sum_{t=t_1}^{t_T}(r_t - \widehat{\mu})^2$$

Parameter estimation

- Given time series data of prices or returns, estimate parameters for drift and variance (or volatility)

$$\mu = \frac{1}{T} \sum_{t=t_1}^{t_T} r_t,$$

$$\sigma^2 = \frac{1}{T-1} \sum_{t=t_1}^{t_T} (r_t - \mu)^2$$

- These are maximum-likelihood estimators (MLE). For large *T*, convergence to "true" population parameters.

$$\mathcal{L}(r_{t_1}, r_{t_2}, \ldots | \tilde{\mu}, \sigma^2) = \prod_{t=t_1}^{t_T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r_t - \tilde{\mu})^2/2\sigma^2},$$

$$\left.\frac{\partial \log \mathcal{L}}{\partial \tilde{\mu}}\right|_{\tilde{\mu}=\hat{\mu}} = 0 = \left.\frac{\partial \log \mathcal{L}}{\partial \sigma^2}\right|_{\sigma^2=\hat{\sigma}^2}$$

Simple, eh?

- Volatilities or variances reported using standard deviations of **simple returns** are similar, but not necessarily close enough

- Relate to parameters either through change of measure or direct calculation of moments

logarithmic return

simple return $R = e^{r} - 1,$

$$m = \mathbb{E}[R] = e^{\tilde{\mu}+\sigma^2/2} - 1 = e^{\mu} - 1,$$

$$s^2 = \mathrm{Var}(R) = e^{2\tilde{\mu}+\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{2\mu}\left(e^{\sigma^2} - 1\right)$$

Simple, eh?

- For small values, use Taylor expansion to approximate

$$
\begin{aligned}
R &= e^r - 1 = r + r^2/2 + \cdots \approx r \\
m &= e^\mu - 1 = \mu(1 + \mu/2 + \cdots) \approx \mu \\
s^2 &= e^{2\mu}\left(e^{\sigma^2} - 1\right) = \left(\sigma^2 + \sigma^4/2 + \cdots\right)(1 + 2\mu + \cdots) \approx \sigma^2
\end{aligned}
$$

- Agreement to leading order, but this is only adequate when next-to-leading order terms are small. For typical values, differences are financially significant.
  - Example:

$$
\begin{aligned}
\mu &= 10\%, \quad m = 0.1 + 0.005 + \cdots = 10.52\%, \\
\sigma &= 30\%, \quad s = 0.3(1 + .1 + .0225 + \cdots) = 33.92\%
\end{aligned}
$$

What is volatility? Is it constant?

- How should the volatility parameter in the random walk model (and later, in the Black-Scholes equation) be estimated from return data?
  - Use **log returns**, not simple returns, since within model they are normally distributed.
  - Starting from parameters defining the price process,

$$P_t = P_{t-1} e^{r_t}$$

$$r_t = \mu + \sigma z_t$$

  - Although drift coefficient cancels out variance definition, it is required in estimation of volatility. (And occasionally rate of return is of interest on its own.)

Alternative estimators

- Alternative estimators attempt to use additional information to improve efficiency. For example:
  - Use intraday information from OHLC to improve over close-to-close returns
  - Account for non-trading during market closures
  - Don't accommodate intraday drift

- Parkinson, Garman-Klass:

$$\widehat{\sigma}^2_{\text{Parkinson}} = \frac{1}{4T\log 2}\sum\left(\log(H_t/L_t)\right)^2$$

$$\widehat{\sigma}^2_{\text{GK}} = \frac{1}{T}\sum\left[0.5\left(\log(H_t/L_t)\right)^2 - (2\log 2 - 1)\left(\log(C_t/O_t)\right)^2\right]$$

Hypothesis testing

1. State **null** and **alternate** hypotheses (why only two?)
2. Define a **test statistic** to see if H_0 fails.
3. *p*-value: if null were true, what is probability to observe test statistic this extreme or more?
   - small *p*-value, evidence to reject null
4. Compare to significance level

Key observation: the **test statistic** is a random variable.

- What is its distribution?

Source: xkcd

Hypothesis testing

Suppose a coin is flipped 10 times and *h* heads are observed.

- Is the coin fair or biased?
- If the coin is not known to be fair, what is the value of the bias?
- If the bias is estimated, how accurate is the estimate? How precise?

Hypotheses:

- H_0 is a **fair coin** $p=1/2$
- H_a is a **biased coin** $p\neq 1/2$

What **test statistic** should we use?

(H– T)/(H+T)

H: head, T: tail

Alternatives to the random walk

Finance at MIT

Where ingenuity drives results

Autocorrelation

If the returns are **correlated across time**, there will be non-vanishing cross-terms in the variance of the *q*-sum.

- Define autocorrelation and autocovariance coefficients, which measure the correlation and covariance of a time series **with itself**, shifted in time by *k* steps. For a covariance-stationary process, let

$$\gamma_k = \mathrm{Cov}(r_t, r_{t-k}), \text{ only depends on the time difference}$$
$$\gamma_0 = \mathrm{Var}(r_t),$$
$$\rho_k = \mathrm{Corr}(r_t, r_{t-k}) = \gamma_k / \gamma_0$$

- For instance, the variance of the two-period returns is no longer double the 1-period

$$\mathrm{Var}(r_t^{(2)}) = \mathrm{Var}(r_t + r_{t-1}) = 2\mathrm{Var}(r_t) + 2\mathrm{Cov}(r_t, r_{t-1}) = 2\mathrm{Var}(r_t)\,(1 + \rho_1)$$

Autocovariance estimation

Need T+k periods and compute

we're not using exactly the same data series

$$\hat{\gamma}_k = \frac{1}{T-1} \sum_{t=k+1}^{k+T} (r_t - \hat{\mu})(r_{t-k} - \hat{\mu}')$$

Generally expect convergence, depending on process details, as

$$\lim_{T\to\infty} \hat{\gamma}_k = \gamma_k,$$

rescale $\sqrt{T}(\hat{\gamma}_k - \gamma_k) \sim \mathcal{N}(0, \sigma_\gamma^2),$

$$\sqrt{T}(\hat{\rho}_k - \rho_k) \sim \mathcal{N}(0, 1)$$

Autocovariance observed

Serial correlation in a time series will cause the variance ratio to depart from unity

$$\mathrm{Var}(r_t + r_{t-1}) = 2\mathrm{Var}(r_t) + 2\mathrm{Cov}(r_t, r_{t-1}),$$

$$\frac{\mathrm{Var}(r_t^{(2)})}{2\mathrm{Var}(r_t)} = 1 + \rho_1 \quad \begin{cases} = 1 & \text{random walk,} \\ > 1 & \text{positive serial correlation} \\ < 1 & \text{negative serial correlation} \end{cases}$$

Positive serial correlation means that winning months tend to be followed by winners, and losing months tend to be followed by losers

Generalize to higher (finite) correlation structure:

$$\mathrm{VR}(q) = \frac{\mathrm{Var}(r_t^{(q)})}{q\mathrm{Var}(r_t)} = 1 + 2\sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \rho_k$$

Alternatives to the random walk

ARMA time series models generate lagged correlations with simple building blocks

- Autoregressive (AR) terms depend on previous lagged returns
- Moving average (MA) terms depend on previous lagged innovations

$$
\begin{aligned}
\text{RW:} \quad & r_t = \mu + \sigma z_t, \quad \text{parameters: 2} \\
\text{AR(1):} \quad & r_t = c_0 + c_1 r_{t-1} + \sigma z_t \qquad \text{parameters: 3} \\
\text{ARMA}(p,q): \quad & r_t = c_0 + c_1 r_{t-1} + \cdots + c_p r_{t-p} + \sigma z_t + \phi_1 z_{t-1} + \cdots + \phi_q z_{t-q} \quad \text{parameters: p+q+2} \\
\text{Increments:} \quad & \mathbb{E}[z_t] = 0, \quad \mathrm{Cov}(z_t, z_{t'}) = \delta_{tt'}
\end{aligned}
$$

- AR(1) example: parameters and mean reversion

$$
r_t - \mu = -\lambda(r_{t-1} - \mu) + \sigma z_t,
$$

Properties of the AR(1) model

Analyze structure of serial dependence by applying equation recursively and using stationarity

- Mean return:
$$\begin{aligned}\mathbb{E}\left[r_t - \mu\right] &= -\lambda \mathbb{E}\left[r_{t-1} - \mu\right] + \sigma \mathbb{E}\left[z_t\right] \\ &= -\lambda \mathbb{E}\left[r_t - \mu\right] \\ &= 0\end{aligned}$$

- Variance:
$$\begin{aligned}\mathrm{Var}(r_t) = \mathbb{E}\left[(r_t - \mu)^2\right] &= (-\lambda)^2 \mathbb{E}\left[(r_{t-1} - \mu)^2\right] + \sigma^2 \mathbb{E}\left[z_t^2\right] \\ &= \lambda^2 \mathrm{Var}(r_t) + \sigma^2\end{aligned}$$

$$\boxed{\mathbb{E}\left[r_t\right] = \mu, \quad \mathrm{Var}(r_t) = \frac{\sigma^2}{1 - \lambda^2}}$$

Properties of the AR(1) model

- Higher-order covariances by recursion:

$$\begin{aligned}\gamma_k &= \mathbb{E}\left[(r_t-\mu)(r_{t-k}-\mu)\right]\\ &= -\lambda\mathbb{E}\left[(r_{t-1}-\mu)(r_{t-k}-\mu)\right]\\ &= -\lambda\gamma_{k-1}\end{aligned}$$

$$\boxed{\gamma_k = (-\lambda)^k\gamma_0 = \frac{(-\lambda)^k}{1-\lambda^2}\sigma^2}$$

- The autocovariance two-point function decays exponentially as a function of the lag *k*

$$|\gamma_k| = \gamma_0 e^{-k/\kappa}$$

Estimation

- **Consistent**: estimator converges (in probability) to true value

$$\lim_{T\to\infty} \hat{\theta} = \theta$$

- **Unbiased**: expectation of estimator (as a random variable) equals true parameter value

$$\mathbb{E}[\hat{\theta}] = \theta$$

- **Asymptotically Normal**: approaches value with deviation drawn as in CLT

$$\sqrt{T}\left(\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}\right) \sim \mathcal{N}(0,1), \qquad \hat{\theta} \sim \mathcal{N}(\theta, \sigma_{\hat{\theta}}^2/T)$$

- **Efficient**: define best among set of estimators via MSE, minimum variance, etc.

Estimation

Parameter estimation in AR(1)

- Observe that the model has the structure of a typical linear regression model

$$y = \alpha + \beta x$$

$$y_i = \alpha + \beta x_i + \epsilon_i$$

- Instead of independent *x* and *y,* the left and right side variables come from the **same** time series, just **shifted** in time.

$$r_t = [\mu(1+\lambda)] + [-\lambda]\, r_{t-1} + [\sigma z_t]$$

$$r_t - \mu = -\lambda(r_{t-1} - \mu) + \sigma z_t$$

$$\Longrightarrow \quad \alpha \to \mu(1+\lambda)$$

$$\beta \to -\lambda$$

$$y_i \to r_t$$

$$x_i \to r_{t-1}$$

- Apply techniques, e.g., OLS, if errors are uncorrelated

Estimation of coefficients and variance parameter for AR(p)

least square

Applying MLE to the AR(p) model results in OLS estimation of return on p prior lagged variables

$$
\begin{aligned}
r_t &= c_0 + c_1 r_{t-1} + \cdots + c_p r_{t-p} + \sigma z_t \\
\mathbb{E}_{t-1}[r_t] &= c_0 + c_1 r_{t-1} + \cdots + c_p r_{t-p} = \bar{\mu}_t \\
\mathrm{Var}(r_t) &= \sigma^2
\end{aligned}
$$

Conditional probability

$$p(r_t | r_{t-1}, r_{t-2}, \cdots ; c_0, c_1, \cdots , c_p, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r_t - \bar{\mu}_t)^2 / 2\sigma^2}$$

$$\log \mathcal{L} = \sum_{t=p+1}^{p+T} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \big(r_t - (c_0 + c_1 r_{t-1} + \cdots) \big)^2 \right]$$

Order determination and model selection

What **value of *p*** should be used to select an appropriate AR(*p*) model?

- Order determination via PACF (partial autocorrelation function)
- Setup and estimate **sequence of models** AR(1), AR(2),..., AR(*n*)

$$
\begin{aligned}
r_t &= c_0^{(1)} + c_1^{(1)} r_{t-1} + \sigma^{(1)} z_t, \\
r_t &= c_0^{(2)} + c_1^{(2)} r_{t-1} + c_2^{(2)} r_{t-2} + \sigma^{(2)} z_t, \\
r_t &= c_0^{(n)} + c_1^{(n)} r_{t-1} + \cdots + c_n^{(n)} r_{t-n} + \sigma^{(n)} z_t,
\end{aligned}
$$

- PACF gives the coefficient of **last lagged term**
- Select *p* from model with largest number of non-negligible terms

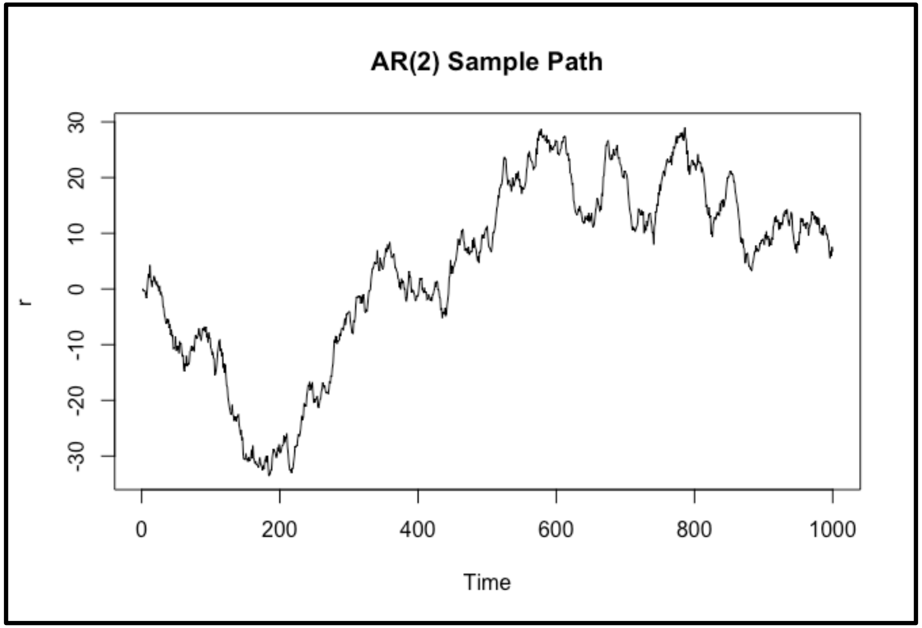
Order determination and model selection

Example: Monte Carlo data

```
c_0 <-  0.001;
c_1 <- -0.1;
c_2 <-  0.4;
Nt  <-  1000;
r   <-  matrix(0,Nt,1)
z   <- matrix(rnorm(Nt), ncol=1)
need some initial observations, which are set to 0
for (t in 3:Nt)  {
  r[t] <- c_0 + c_1*r[t-1] + c_2*r[t-2] + z[t]
}

plot(cumsum(r), type="l", main="AR(2) Sample
Path",xlab="Time",ylab="r");grid()

acf(r)
pacf(r)
```



Order determination and model selection

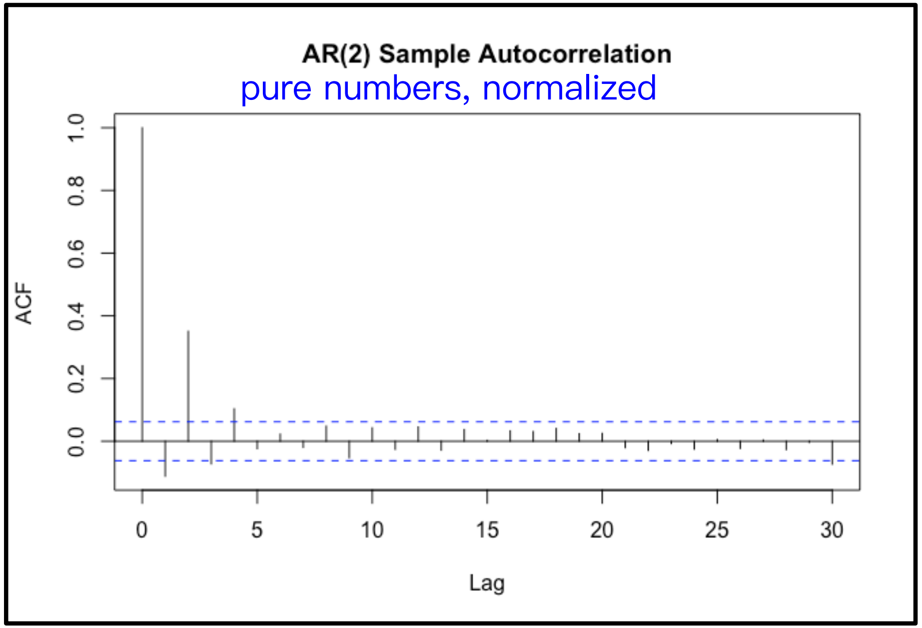
Example: Monte Carlo data

```
c_0 <-  0.001;
c_1 <- -0.1;
c_2 <-  0.4;
Nt  <-  1000;
r   <-  matrix(0,Nt,1)
z   <- matrix(rnorm(Nt), ncol=1)

for (t in 3:Nt)  {
  r[t] <- c_0 + c_1*r[t-1] + c_2*r[t-2] + z[t]
}

plot(cumsum(r), type="l", main="AR(2) Sample
Path",xlab="Time",ylab="r");grid()

acf(r)
pacf(r)
```



Order determination and model selection

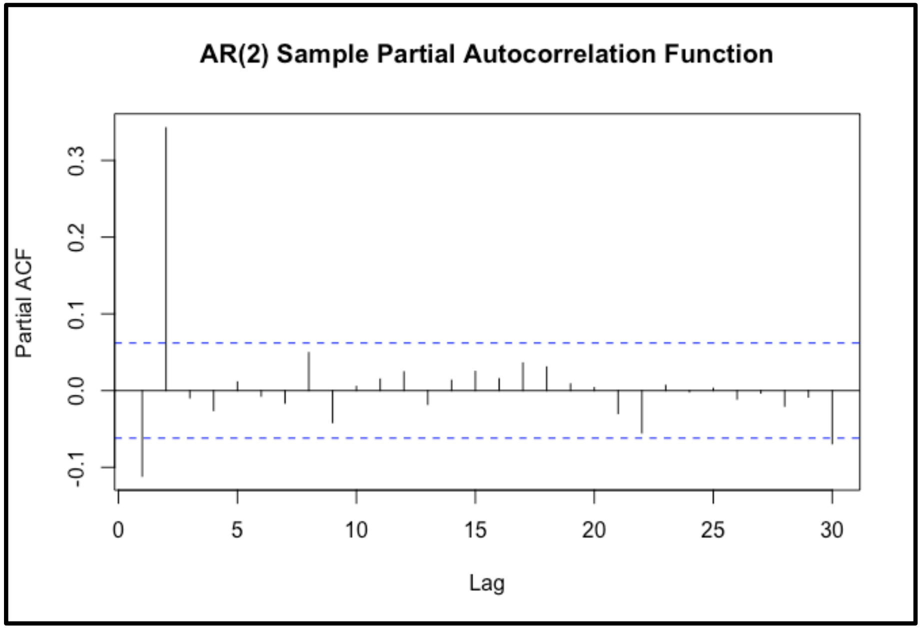
Example: Monte Carlo data

```
c_0 <-  0.001;
c_1 <- -0.1;
c_2 <-  0.4;
Nt  <-  1000;
r   <-  matrix(0,Nt,1)
z   <- matrix(rnorm(Nt), ncol=1)

for (t in 3:Nt)  {
  r[t] <- c_0 + c_1*r[t-1] + c_2*r[t-2] + z[t]
}

plot(cumsum(r), type="l", main="AR(2) Sample
Path",xlab="Time",ylab="r");grid()

acf(r)
pacf(r)
```



Order determination and model selection

- More terms and parameters can lead to better fit...or to **overfitting**
- Expand likelihood function with **penalty** for additional parameters
- AIC (Akaike Information Criterion) and BIC (Bayes Information Criterion):

$$\begin{aligned} \mathrm{AIC} &= -\frac{2}{T}\log\mathcal{L} + \frac{2N_p}{T} \\ &= \log\widehat{\sigma}^2_{(p)} + \frac{2p}{T} \\ \mathrm{BIC} &= \log\widehat{\sigma}^2_{(p)} + \frac{(\log T)p}{T} \end{aligned}$$

favor simpler models with fewer terms

Order determination and model selection

Validating the model selected:

- Residuals should be white noise process
- Test statistics for rejection of autocorrelation

$$\text{Box-Pierce:} \quad Q^*(m) = T\sum_{i=1}^{m}\hat{\rho}_i^2$$

$$\text{Ljung-Box:} \quad Q(m) = T(T+2)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2}{T-i}$$

- If returns are IID, then Q has chi-squared distribution w/ *m* degrees of freedom

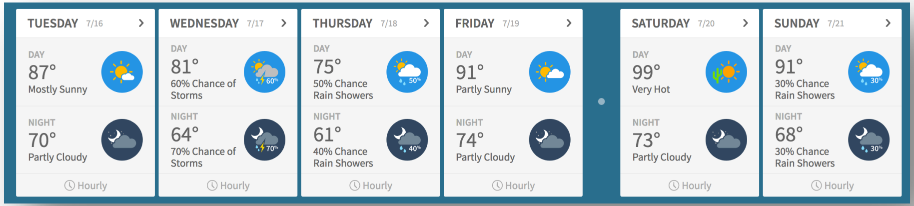
• For MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series.
• For AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process.
• Serial correlation can be identified by computing the auto correlation coefficients

Forecasting

Forecasts in time series models are **predictions** of future observations **conditioned** on information that is known at the time of the forecast

- Forecast output
  - Point forecasts: the expected value of a specific future observation
  - Probability forecasts: the likelihood of observing a specific future outcome
  - Distributions: the full probability distribution of possible outcomes at specific future time

- Horizons
  - Forecasts are framed in terms of two points in time: the **horizon** over which the forecast is made, and the time at which the forecast is evaluated
  - As time progresses, forecasts evolve based on new information and observations
  - Example: weather forecasting
    - Chance of rain on Sunday is 30%
    - Yesterday it was 0%



Forecasting

Forecasts in time series models are **predictions** of future observations **conditioned** on information that is known at the time of the forecast.

$$\boxed{Y_t \equiv \frac{R_t - \mu}{\sigma}}$$

Sigma is not the standard deviation of R, it's just the coefficient we put in front of Z in our model.

- Conditional probabilities
  - Conditioned on information set I_t
- Predict future values of the path
  - Point forecasts
  - Horizons
  - Distribution and moments
- Precision and accuracy
  - Forecast errors
  - Bias and variance

AR(1) $Y_t = z_t - \lambda Y_{t-1}$

$$Y_{t+1} = z_{t+1} - \lambda Y_t$$

$$\mathrm{E}\left[Y_{t+1} | I_t\right] = -\lambda Y_t$$

conditional expectation based on the information set available at time t

observed: not a random variable

$$Y_{t+2} = z_{t+2} - \lambda Y_{t+1}$$

$$= z_{t+2} - \lambda z_{t+1} + \lambda^2 Y_t$$

$$\mathrm{E}\left[Y_{t+2} | I_t\right] = +\lambda^2 Y_t$$

Elements of forecasting

- Nature of the process
  - Deterministic vs. stochastic
  - Stationary vs. non-stationary
  - Laws of nature vs. human behavior
- Knowledge of the process
  - Process known vs. process unknown
  - Parameters known vs. parameters unknown
  - Empirical model fit from historical data vs. theoretical model
- Model complexity
  - Univariate vs. multivariate
  - Linear vs. nonlinear
- Forecast horizon
  - One-step
  - Multi-step
  - Asymptotic t goes to infinity
- Forecast criteria
  - Observable vs. unobservable
  - Measures of forecast accuracy
  - Economic weighting of forecast accuracy

time varying volatility cannot be observed at an instant time (price can), but can be observed over a period of time

- Model evolution
  - Static parameters vs. ~~fixed~~ changing
  - Error correction when we see errors, we update our parameters
  - Regime shifts and breaks situations where model is broken

Forecasting setup

- Stochastic process up to time *t:*

$$\{\ldots x_0, x_1, \ldots, x_t, x_{t+1}, \ldots x_{t+h}\}$$

- Information set through time *t*
- Forecast horizon: all steps through *t+h*
- Conditional probabilities: forecasts conditioned on information set I_t
- Point forecasts and distributions
- Cost functions vs. confidence intervals
- Subjective information can be included in Bayesian approach.

Optimal forecast

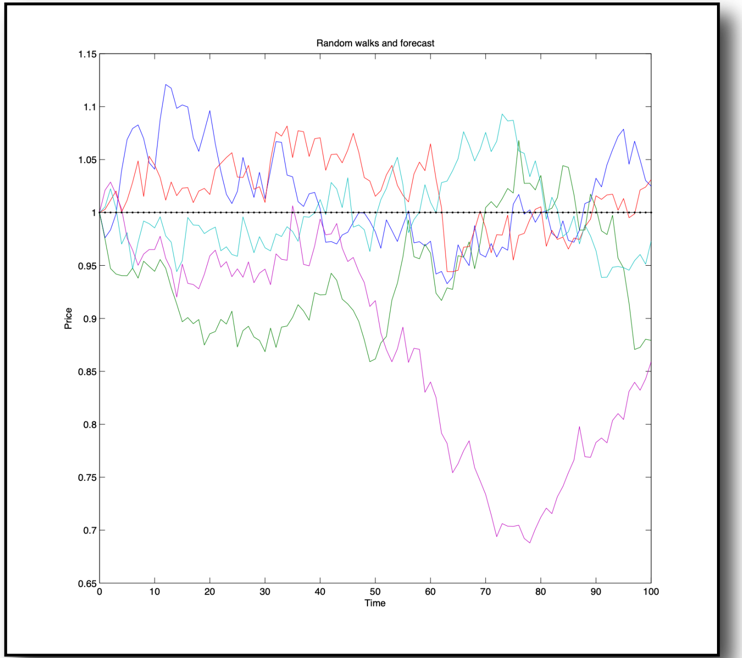
- Theorem (*cf.* Granger):
  - The optimal forecast is the conditional mean

$$f_{t,h} = \mathrm{E}\left[x_{t+h} | I_t\right]$$

  - Provided that cost function is symmetric and convex

- For a given model, use optimal forecast
- For a given set of data, find optimal model
- Example: Random walk

$$p_t = p_{t-1} + \sigma z_t$$
$$f_{t,h} = \mathrm{E}\left[p_{t+h} | I_t\right] = p_t$$



Forecast errors

- At time *t+h*, compare forecast with outcome.
  - **Forecast error** is the difference

$$e_{t+h} = x_{t+h} - f_{t,h}$$ forecast made at time t with horizon h

- Define a function of the error which quantifies the **cost**. For example, the quadratic form

scale coefficient $c(x_{t+h} - f_{t,h})^2$

is common for the usual reasons (simplicity, Taylor expansions, etc.) – but not invariant!

- Compare with loss functions (e.g., penalizing underperformance)

Non-invariance of MSFE

- The size of the forecast errors and the relative performance of different forecasts may change based on
  - Variables used
  - Parameter estimation

- Example: AR(1) process
  - Described by levels vs. differences
  - With true vs. estimated parameters
  - **One step ahead**, the forecasts differ

$$x_t = -\lambda x_{t-1} + \epsilon_t$$
$$\Delta x_t = x_t - x_{t-1}$$
$$= -(1+\lambda) x_{t-1} + \epsilon_t$$

$$f_{t,1} = \mathrm{E}\left[x_{t+1} | I_t\right] = -\lambda x_t$$
$$f'_{t,1} = \mathrm{E}\left[\Delta x_{t+1} | I_t\right] = -(1+\lambda) x_t$$

while the mean-squared forecast errors are the same

$$\mathrm{E}\left[(x_{t+1} - f_{t,1})^2\right] = \mathrm{E}\left[\epsilon_{t+1}^2\right] = \sigma^2$$

argmin MSFE = conditional mean

$$\mathrm{E}\left[(\Delta x_{t+1} - f'_{t,1})^2\right] = \mathrm{E}\left[\epsilon_{t+1}^2\right] = \sigma^2$$

Non-invariance of MSFE

- However, at **two steps ahead**, they differ.

$$
\begin{aligned}
x_{t+2} &= -\lambda x_{t+1} + \epsilon_{t+2} \\
&= \lambda^2 x_t - \lambda \epsilon_{t+1} + \epsilon_{t+2}, \\
f_{t,2} &= \mathrm{E}\left[x_{t+2} | x_t\right] = \lambda^2 x_t, \\
\mathrm{E}\left[e_{t+2}^2\right] &= \mathrm{E}\left[(\lambda \epsilon_{t+1} - \epsilon_{t+2})^2\right] = \sigma^2 (1 + \lambda^2)
\end{aligned}
$$

- In terms of the difference variable, MSFE differs (unless λ = -1/2)

$$
\begin{aligned}
f'_{t,2} &= \mathrm{E}\left[\Delta x_{t+2} | x_t\right] = \mathrm{E}\left[-(1+\lambda) x_{t+1} + \epsilon_{t+2}\right] \\
&= -(1+\lambda) \overset{f_{t+1}}{\cancel{f'_{t+1}}} = \lambda(1+\lambda) x_t, \\
\mathrm{E}\left[{e'}_{t+2}^2\right] &= \mathrm{E}\left[(-(1+\lambda)\epsilon_{t+1} + \epsilon_{t+2})^2\right] = \sigma^2 (1 + (1+\lambda)^2)
\end{aligned}
$$

Parameter uncertainty

- If parameters are uncertain, their **estimates** enter in forecast:

$$f_{t,1} = -\hat{\lambda} x_t, \quad \text{parameter might not be right}$$

$$e_{t+1} = \epsilon_{t+1} - (\lambda - \hat{\lambda}) x_t,$$

$$\mathrm{E}\left[e_{t+2}^2\right] = \sigma^2(1+\lambda^2) + x_t^2(\lambda^2 - \hat{\lambda}^2)$$

$$x_t = -\lambda x_{t-1} + \epsilon_t$$

$$\Delta x_t = x_t - x_{t-1}$$

$$= -(1+\lambda) x_{t-1} + \epsilon_t$$

- For the difference variable,

$$f'_{t,2} = \hat{\lambda}(1+\hat{\lambda}) x_t,$$

$$e'_{t+2} = -(1+\lambda)\epsilon_{t+1} + \epsilon_{t+2} + (\lambda - \hat{\lambda})(1 - \lambda - \hat{\lambda}) x_t,$$

$$\mathrm{E}\left[{e'}_{t+2}^2\right] = \sigma^2(1 + (1+\lambda)^2) + x_t^2(\lambda - \hat{\lambda})^2(1 - \lambda - \hat{\lambda})^2$$

Forecast errors and cost functions

- Thus mean-squared forecast errors differ
  - Depending on value of parameter
  - Depending on value of estimate used in forecast

- Caution required in comparing forecasts
  - Costs/error criteria
  - Non-invariance don't depend on the way we chose to label our variables to describe exactly the same thing

Random Walk on a Binomial Tree

Finance at MIT
Where ingenuity drives results

Binomial model and the random walk

The binomial tree model is an example of a random walk

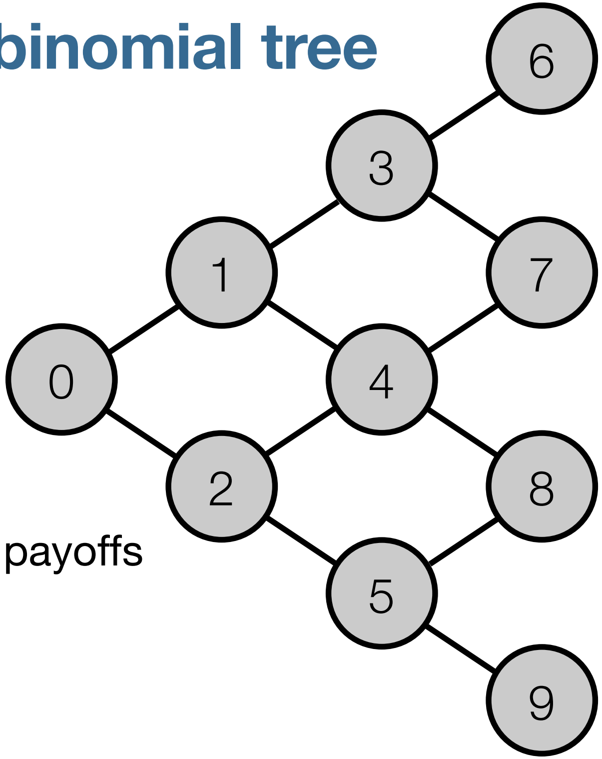
- Discrete-time stochastic process
- At each time-step, draw a new IID random variable *z* and define

$$r_t \equiv \log\left(\frac{S_t}{S_{t-1}}\right) = a + bz_t, \text{ where}$$

$$\begin{aligned} a &= (\log R_u + \log R_d)/2 \\ b &= (\log R_u - \log R_d)/2, \\ z_t &= \pm 1 \end{aligned}$$

$$\log\left(\frac{S_T}{S_0}\right) = r_1 + r_2 + \cdots r_T = aT + b\sum z_t$$

Walking and gambling on the binomial tree



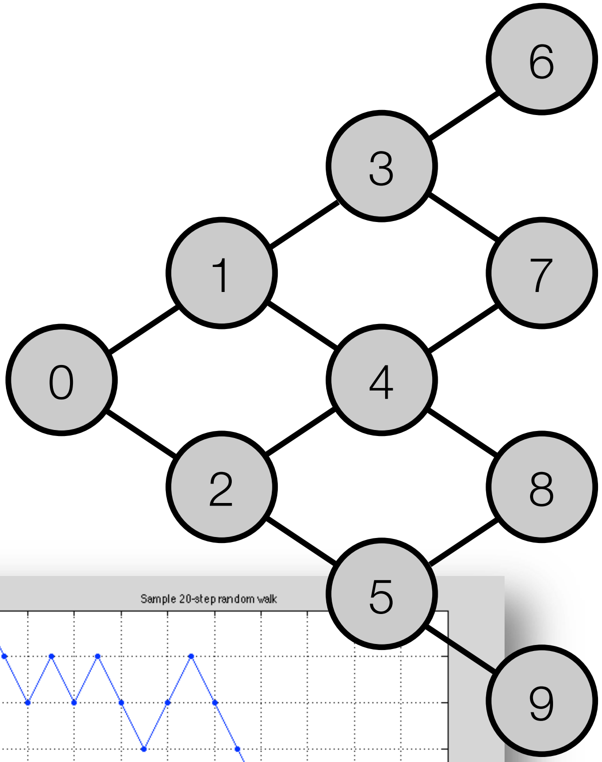
Binomial tree model

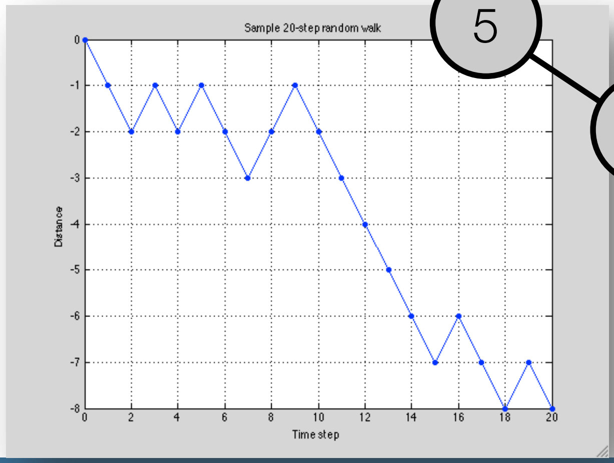
- Asset prices can move up or down. Specify
  - Probability up vs. down
  - Return up vs. down
- Analytics and pricing
  - Distribution, expected values, discounted payoffs
  - Also: boundaries, default, duration, risk,...

Binomial Tree

The random walk process is the basis of the binomial tree model of option pricing.

- Extreme simplification of dynamics.
- At each time step, there are only a discrete set of allowed prices.
- Uncertain payoffs may be state or path dependent.
- At each time step, the underlying asset price can move either up or down.
- Size of the moves defines and calibrate model.
- Assignment of probabilities defines expectations, risk, and arbitrage constraints.

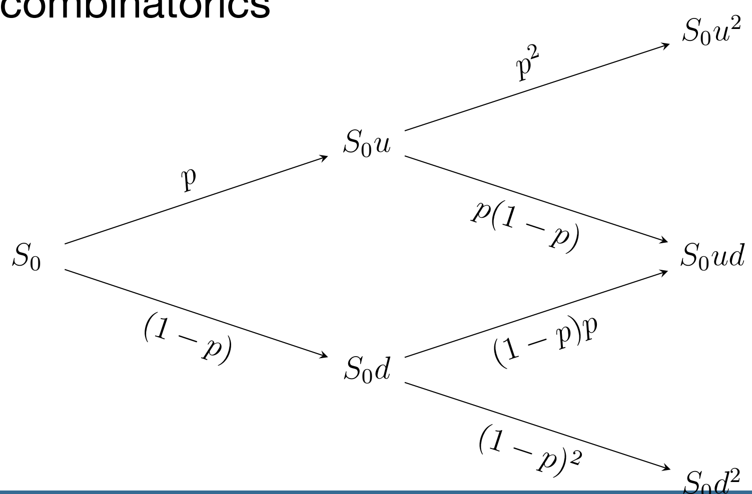




Binomial Tree

Same structure at each node

- Price multiplied by *u* with probability *p.*
- Price multiplied by *d* with probability *q.*
- Tree recombines because multiplication is commutative
- Number of paths to reach a given price level from binomial combinatorics



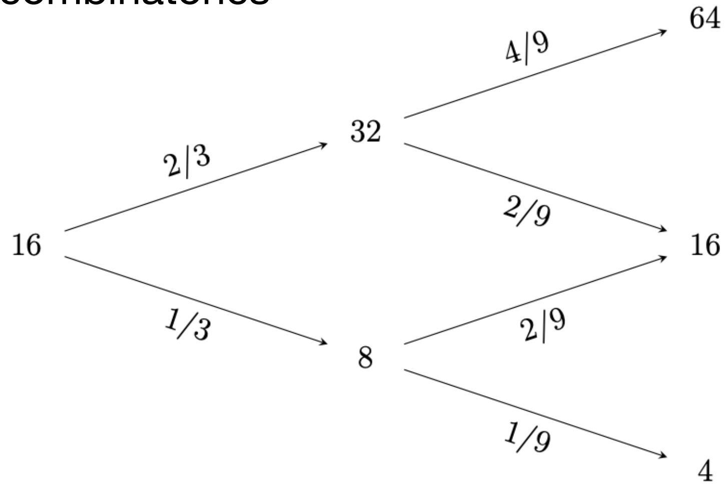
$$S_{t+1} = \begin{cases} S_t u, & \text{with probability } p \\ S_t d, & \text{with probability } q = 1-p. \end{cases}$$

$$\text{Prob}(S_t = S_0 u^k d^{t-k}) = \binom{t}{k} p^k (1-p)^{t-k}$$

Binomial Tree

Same structure at each node

- Price multiplied by *u* with probability *p.*
- Price multiplied by *d* with probability *q.*
- Tree recombines because multiplication is commutative
- Number of paths to reach a given price level from binomial combinatorics



$$S_{t+1} = \begin{cases} S_t u, & \text{with probability } p \\ S_t d, & \text{with probability } q = 1 - p. \end{cases}$$

$$\text{Prob}(S_t = S_0 u^k d^{t-k}) = \binom{t}{k} p^k (1-p)^{t-k}$$

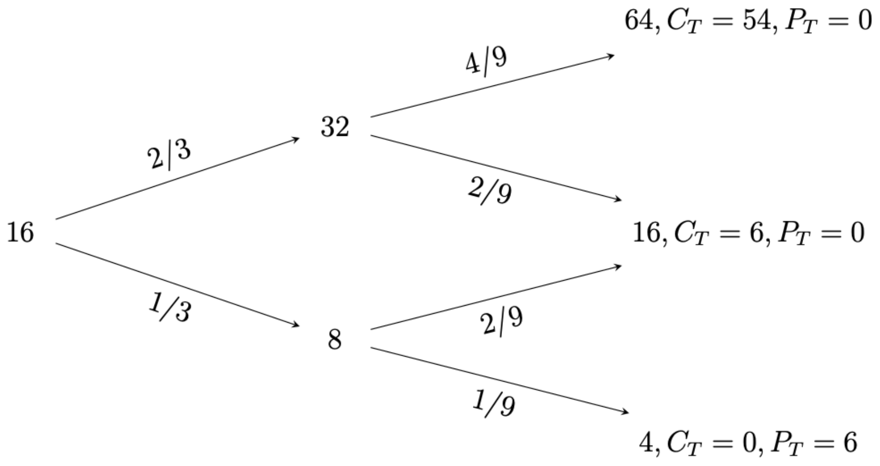
$$\boxed{\text{Example: } u = 2, d = 1/2, p = 2/3, S_0 = 16}$$

Binomial Tree

Derivative securities on the tree

- Value depends on **underlying** price and on **time**: *V(S,t)*
- Option payoffs are known at **expiration**
  - Call option: $C_T = \max(S - K, 0)$
  - Put option: $P_T = \max(0, K - S)$

$$\text{Example: } u = 2, d = 1/2, p = 2/3, \quad S_0 = 16, K = 10, T = 2$$



Binomial Tree

The tree becomes useful when we consider the future evolution of financial contracts, such as options, that are functions of the values of the nodes of the tree.

- To connect with our previous notation and calibrations, let

$$S_t = S_{t-1} e^{r_t}, \quad r_t = \begin{cases} \log u, & \text{with probability } p \\ \log d, & \text{with probability } q = 1 - p. \end{cases}$$

- Equivalently, we can formulate the return using a typical Bernoulli variable,

$$S_t = S_{t-1} e^{r_t}, \quad r_t = a + b x_t, \quad x_t = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } q = 1 - p. \end{cases}$$

Binomial Tree

To connect the model to real-world asset prices, calibrate the model parameters by identifying expectations.

$$\begin{aligned}\mu &= \mathrm{E}\left[r_t\right] = a + pb, \\ \sigma^2 &= \mathrm{Var}(r_t) = \mathrm{E}\left[(r_t - a - pb)^2\right] = b^2\,\mathrm{E}\left[(x_t - p)^2\right] \\ &= b^2 p(1-p)\end{aligned}$$

- In terms of *a,b,p* or in terms of the up/down factors, we have

$$a = \mu - \sigma\sqrt{\frac{p}{1-p}}, \qquad \log u = \mu + \sigma\sqrt{\frac{1-p}{p}},$$

$$b = \frac{\sigma}{\sqrt{p(1-p)}} \qquad \log d = \mu - \sigma\sqrt{\frac{p}{1-p}}$$

mu and sigma

- Note that this solves for two real-world values in terms of three model unknowns.

Gambler's Ruin

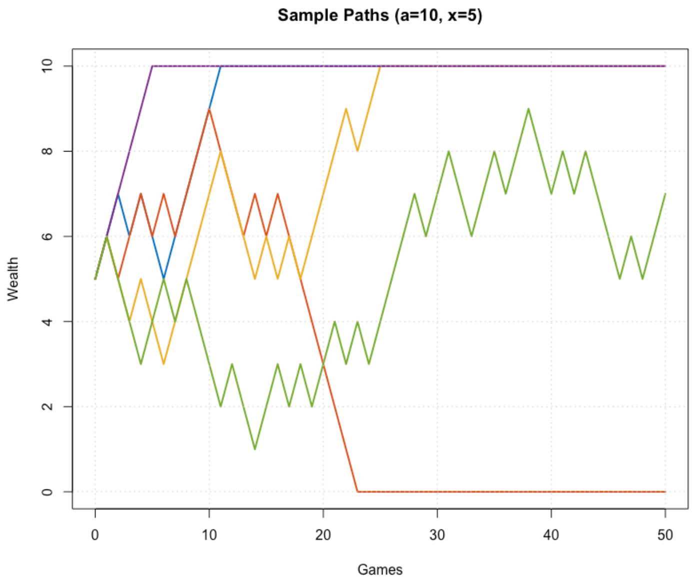
Finance at MIT

Where ingenuity drives results

Gambler's Ruin

Here is a classic problem in which the recursion takes place in space, not time. (Actually, "space" is really "money"!)

- Suppose there is a repeated set of gambles with probability of success p and of failure $q=1-p$.
- Initial capital is $x > 0$.
- You **are** permitted to quit the game.
- Stop when either
  - You break the house, accumulating all the assets, total capital a.
  - You lose your capital: **ruin** defined as $x = 0$



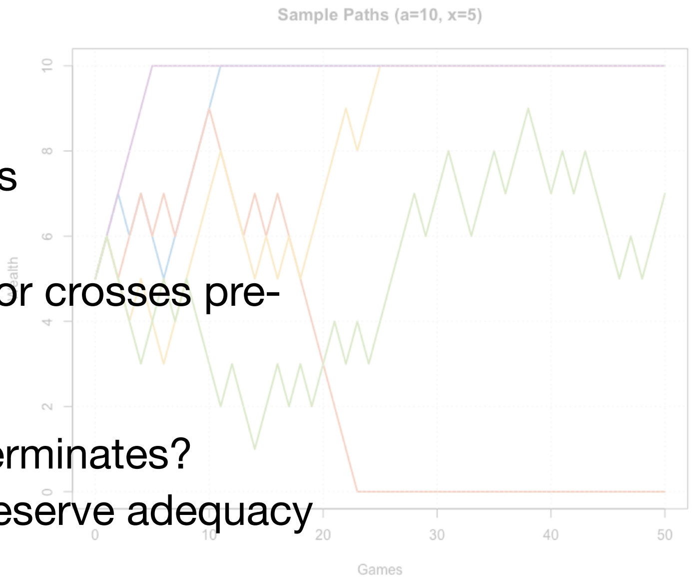
Stopping problems and boundaries

Stopping problems: process terminates, possibly optionally or conditionally

- What is the probability of stopping?
- What is the optimal stopping strategy?
- Examples: option exercise, business exit strategies

Boundary problems: process continues until path hits or crosses pre-defined values

- What is the probability of avoiding contact?
- What is the expected duration of a path before it terminates?
- Examples: credit default, stop loss limits, capital reserve adequacy



Gambler's ruin

Let Q_x denote the **probability of ruin** starting from capital x.

- It satisfies a recursion relation in x, not t, based on conditional probability. After the next game, the capital will either go up or down by one unit. Therefore

$$Q_x = pQ_{x+1} + qQ_{x-1}$$

- This second-order difference equation in general has two solutions before imposing boundary conditions. For $p \neq q$, they are

$$Q_x = 1 \text{ or } Q_x = \left(\frac{q}{p}\right)^x$$

as you can verify by substitution.

Gambler's ruin

The most general solution to the recursion is

$$Q_x = A + B\left(\frac{q}{p}\right)^x$$

where the constants *A*, *B* are fixed by the boundary conditions

$$Q_0 = 1, \quad Q_a = 0.$$

The ruin probability is therefore

$$\boxed{Q_x = \frac{(q/p)^a - (q/p)^x}{(q/p)^a - 1}}$$

Gambler's ruin

Random walk with "absorbing" boundary conditions.

- Can the sequence continue without terminating?
- Does the option to stop convey an advantage?
- Expected gain = 0 if and only if p=q.

Solution given by

$$Q_x = \frac{(q/p)^a - (q/p)^x}{(q/p)^a - 1}$$

Solution for $p = q = 1/2$ given by

$$Q_x = 1 - \frac{x}{a}$$

Gambler's ruin

Example: *a=100, x=99, p=q:*

- *99%* chance to win \$1 before losing \$99.

Example: *a=100, x=99, p=0.4:*

- 67% chance to win \$1 before losing \$99

With large initial capital, there is a significant chance of winning a relatively small amount, *a-x,* before being ruined – **even when the odds are against you**.

Gambler's ruin

What if we change the stakes? Changing from $1 bet to $*b*,

$$Q_x = pQ_{x+b} + qQ_{x-b}$$

which has solution for $p \neq q$

$$Q_x = \frac{(q/p)^{a/b} - (q/p)^{x/b}}{(q/p)^{a/b} - 1}$$

If the stakes are **increased**, the side with unfavorable odds has a **lower** chance of going broke before arriving at *a*.

the fluctuations can be to our advantage, the longer we play, the more likely we are to grind down and end up ruined. But if we take larger bets, we're more likely to have an instance where we hit our upper boundary.

Gambler's ruin

What if our appetite is unbounded?

- If *a* is infinite, then unfavorable odds lead to ruin with probability 1.
  - *Probability of return to origin*

$$Q_x = \begin{cases} 1, & \text{if } p \leq q \\ (q/p)^x, & \text{if } p > q. \end{cases}$$

What is the expected **duration** of the game?

- Example: fair game (*p=q), x=100, a=200*
  - Expected duration 10,000.
  - Grows as square when stakes are equal

$$D_x = pD_{x+1} + qD_{x-1} + 1, \quad 0 < x < a$$

$$D_x = \frac{x}{q-p} - \frac{a}{q-p}\frac{1-(q/p)^x}{1-(q/p)^a}, \quad (p \neq q)$$

$$D_x = x(a-x), \quad (p = q)$$

Gambler's ruin

MIT MANAGEMENT SLOAN SCHOOL

Summary

- Recursive, discrete-time process subject to **boundary conditions.**
- Applications:
  - Gambling
  - Insurance adequate capital to never get ruined by random processes which represent the arrival of insurable claims or losses
  - Bankruptcy
  - Credit default
  - Bet sizing
- Stopping problems defined by economic conditions
- Optimal stopping, objective functions, and strategic design
- Examples:
  - Minimize ruin probability
  - Maximize duration

References

- Books
  - Campbell, Lo, and MacKinlay (1977) – "*Econometrics of Financial Markets,*" Princeton
  - Feller, William (1968) – "*Introduction to Probability and Its Applications (3rd ed.),*" Wiley
  - Glasserman, Paul (2004) – "*Monte Carlo Methods in Financial Engineering,*" Springer
  - Tsay, Ruey S. (2010) – "*Analysis of Financial Time Series (3rd ed.),*" Wiley

- Articles
  - Fama, Eugene (1965) – "*Random Walks in Stock Market Prices,*" Financial Analysts Journal, Vol. 21, No. 5, pp. 55-59
  - Samuelson, Paul. (1965) – "*Proof that Properly Anticipated Prices Fluctuate Randomly,*" Industrial Management Review, 6:2, 41-49
  - Lo, Andrew W. & A. Craig MacKinlay (1988) – "*Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test,*" Review of Financial Studies, Vol. 1, No. 1 (Spring, 1988), pp. 41-66