Probability Applications in Quantitative Trading

MATHEMATICAL PROBABILITY METHODS

- PRINCETON QUANTITATIVE TRADERS

Review on Random Variables

Definition and Role

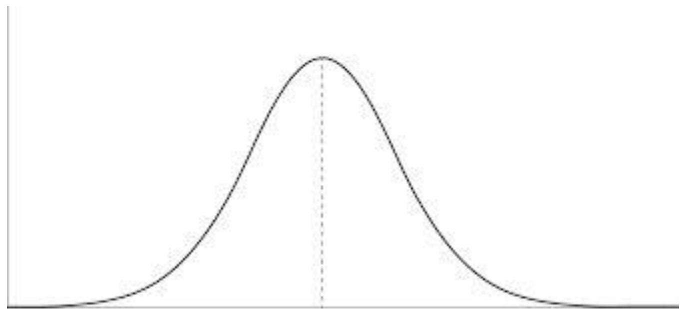
Random variables represent numerical outcomes of stochastic processes key to probability theory.

Modeling Financial Data

Random variables model uncertain financial phenomena like asset returns and price movements.

Example - Stock daily returns can be modeled as normally distributed variables with mean and variance.

- Random Variables: $X \sim N(\mu, \sigma^2)$
- Expectation: $E[X] = \int x\, f(x)\, dx$
- Variance: $Var(X) = E[(X - E[X])^2]$

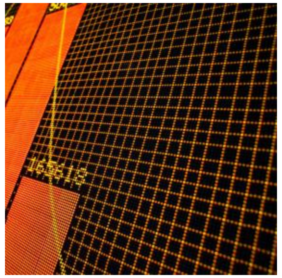


PROBABILITY CONCEPTS

Martingales in Finance

Martingale Definition

A martingale is a stochastic process where the expected future value equals the current value, indicating no drift.

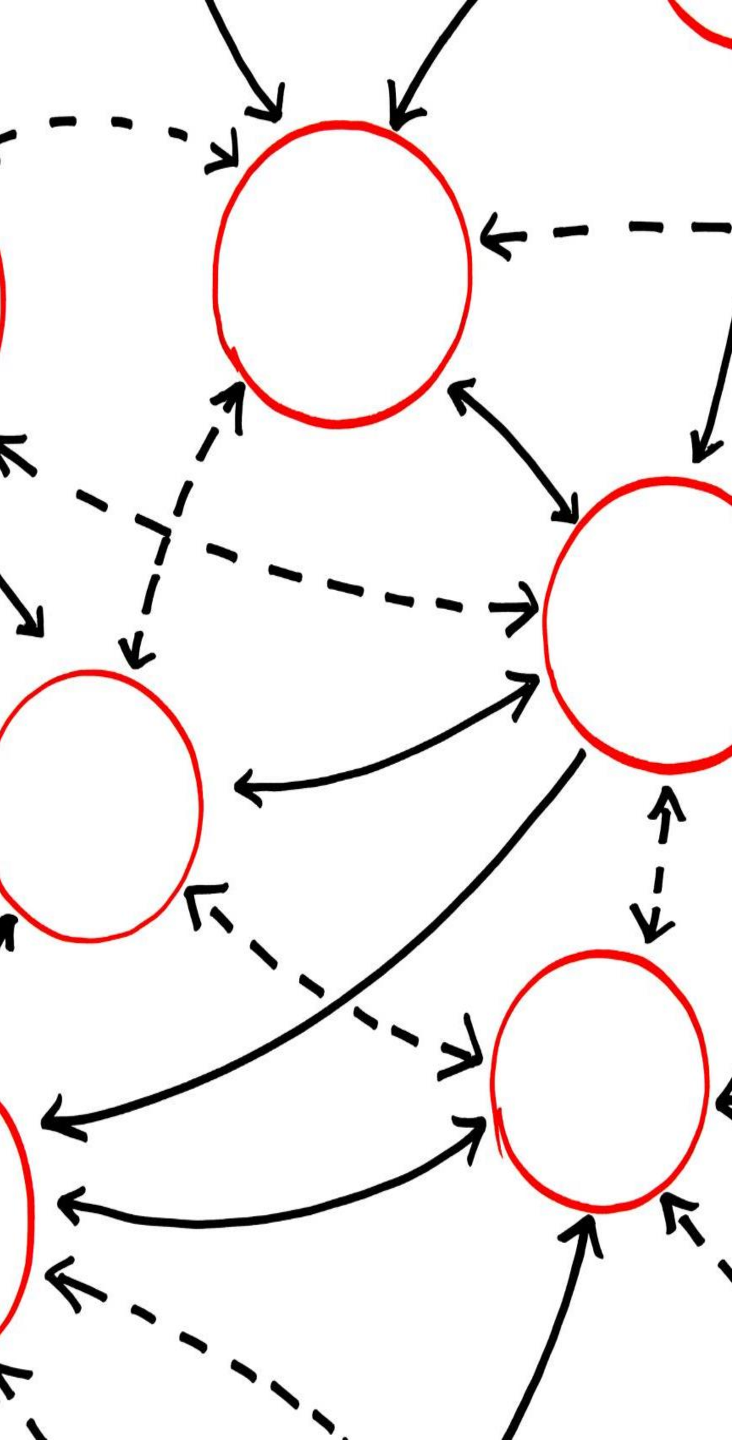


Financial Application

Martingales model asset prices under no arbitrage, forming a foundation for derivative pricing and risk-neutral valuation.

Market Efficiency and Fairness

Martingale properties help ensure market efficiency and fair betting systems by aligning expected payoffs with current prices.



Stochastic Calculus and Ito's Lemma

Stochastic Calculus Basics

Stochastic calculus generalizes calculus to functions influenced by random processes like Brownian motion.

Ito's Lemma Formula

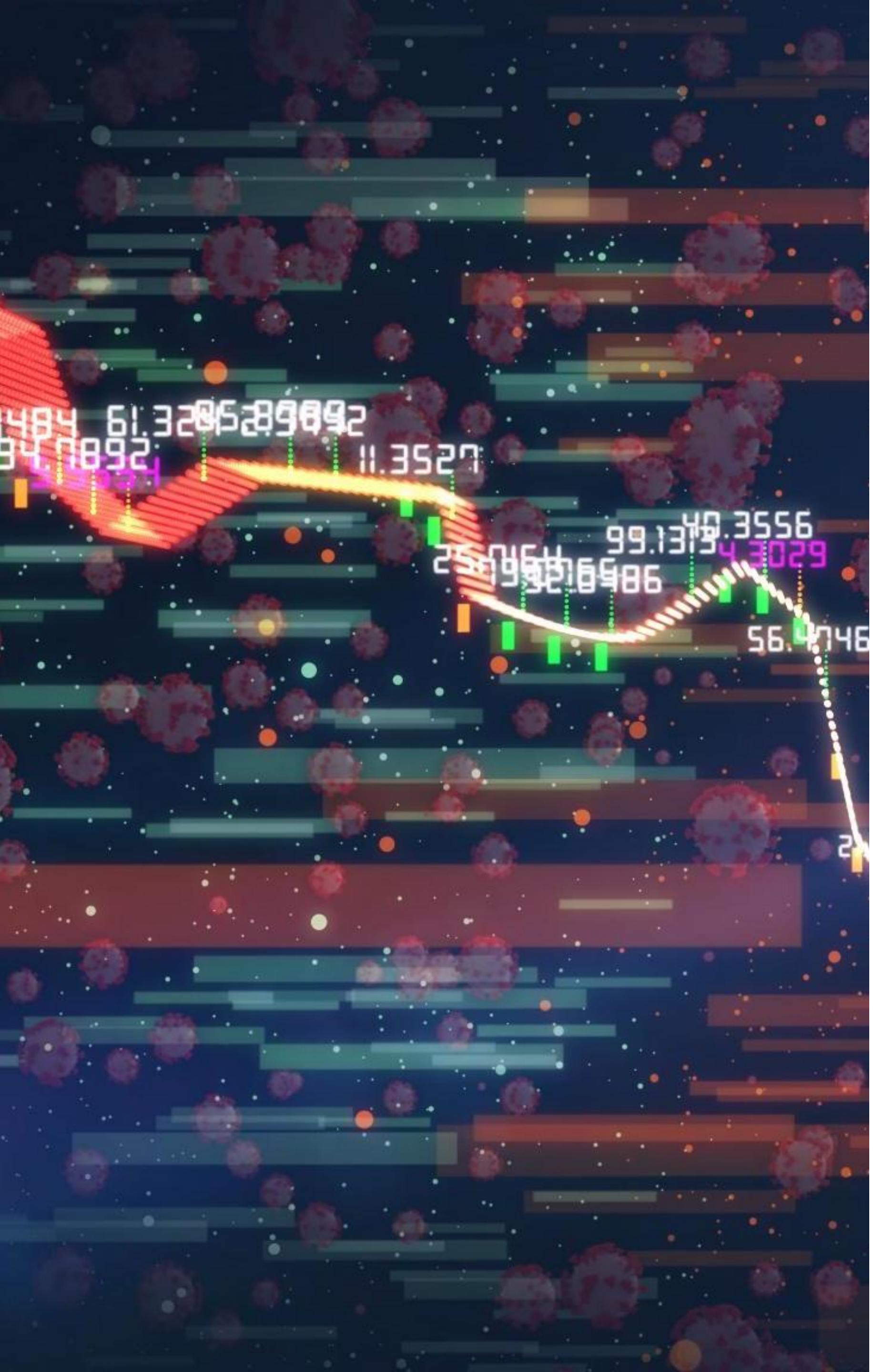
Ito's Lemma provides a differential rule for functions of stochastic variables involving drift and diffusion terms.

Applications in Finance

Ito's Lemma underpins the Black-Scholes equation and models derivative pricing and hedging strategies.

Visualizing Stochastic Paths

Visuals include stochastic paths with tangent approximations showing randomness impact on functions.



Modeling Asset Prices with SDEs

Stochastic Differential Equations

SDEs model asset price changes continuously with drift and volatility components.

Geometric Brownian Motion Model

GBM models asset prices combining deterministic trend and stochastic fluctuations.

Applications in Finance

GBM underpins Black-Scholes option pricing and quantitative trading strategies.

Visualizing Price Paths

Simulated GBM paths help traders understand price behavior and volatility impact.

MATHEMATICAL PROOFS AND INEQUALITIES

PROOF

Markov's Inequality. Let X be a non-negative random variable and let $a > 0$. Then

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Step 1: Apply Markov to a squared deviation. Let Y be any random variable with mean $\mu = E[Y]$. For any $\varepsilon > 0$,

$$\Pr\left(|Y - \mu| \geq \varepsilon\right) = \Pr\left((Y - \mu)^2 \geq \varepsilon^2\right).$$

Since $(Y - \mu)^2$ is non-negative, we can apply Markov's inequality:

$$\Pr\left((Y - \mu)^2 \geq \varepsilon^2\right) \leq \frac{E\left[(Y - \mu)^2\right]}{\varepsilon^2}.$$

Step 2: Recognize variance. Recall that $E\left[(Y - \mu)^2\right] = \mathrm{Var}(Y)$. Thus,

$$\Pr\left(|Y - \mu| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2}.$$

Chebyshev's Inequality. We have derived the classic form:

$$\Pr\left(|Y - \mu| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2}.$$

Chebyshev's Inequality – Risk Managment

Chebyshev's Inequality Concept

This inequality bounds the probability a variable deviates from its mean by k standard deviations without distribution assumptions.

Mathematical Derivation

The proof uses Markov's Inequality applied to squared deviation, linking variance to probability bounds.

Applications in Trading

Used in quantitative trading to estimate extreme event probabilities and to set risk limits and strategies.

Derivation of Hoeffding's Inequality

Markov's Inequality. For any non-negative random variable X and $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Step 1: Exponential Markov bound. Let $S_n = \sum_{i=1}^{n}(X_i - E[X_i])$. For any $h > 0$ and $t > 0$,

$$\Pr(S_n \geq t) = \Pr(e^{hS_n} \geq e^{ht}) \leq \frac{E[e^{hS_n}]}{e^{ht}}.$$

Step 2: Independence and factorization. If $X_1, \ldots, X_n$ are independent,

$$E[e^{hS_n}] = \prod_{i=1}^{n} E\left[e^{h(X_i - E[X_i])}\right].$$

Step 3: Hoeffding's Lemma. If a random variable Y satisfies $Y \in [a, b]$ and $E[Y] = 0$, then

$$E[e^{hY}] \leq \exp\left(\frac{h^2(b-a)^2}{8}\right).$$

Step 4: Apply Hoeffding's Lemma to each term. For each X_i,

$$E\left[e^{h(X_i - E[X_i])}\right] \leq \exp\left(\frac{h^2(b_i - a_i)^2}{8}\right).$$

Thus,

$$E[e^{hS_n}] \leq \exp\left(\frac{h^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right).$$

Step 5: Combine bounds. So

$$\Pr(S_n \geq t) \leq \exp\left(\frac{h^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 - ht\right).$$

Step 6: Optimize over h. Choosing

$$h = \frac{4t}{\sum_{i=1}^{n}(b_i - a_i)^2}$$

minimizes the exponent, yielding

$$\Pr(S_n \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Two-sided bound. By symmetry,

$$\Pr\left(|S_n| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

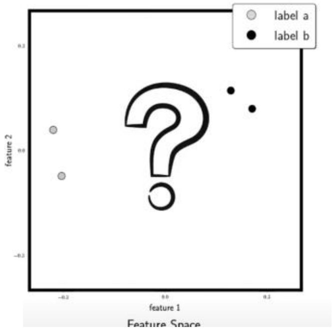
Hoeffding's Inequality. We conclude:

$$\Pr\left(\left|\sum_{i=1}^{n}(X_i - E[X_i])\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

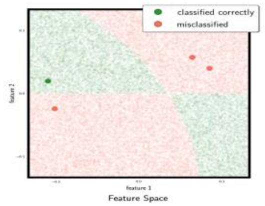
From Markov to Hoeffding Inequality

Application in Quantitative Trading

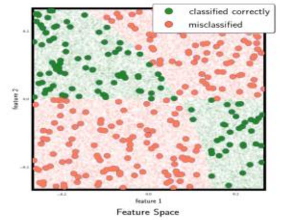
- Provides an upper bound on the probability that the sum of bounded independent random variables deviate from its expected value by a threshold



Hoeffding Inequality gives us probability of incorrect classification given n samples



$\mathbf{E_{train}(h)} = \mathbf{0.7500}$ Training Error
$\mathbf{E_{true}(h)} = \mathbf{0.6302}$ True Error



$\mathbf{m = 321}$ Sample Size
$\mathbf{E_{train}(h)} = \mathbf{0.6449}$ Training Error
$\mathbf{E_{true}(h)} = \mathbf{0.6302}$ True Error

$\mathbf{m = 100}$ Sample size
$\boxed{\varepsilon = \mathbf{0.1} \text{ Approximately close}}$

$$\mathbb{P}(|\mathbf{E_{train}(h)} - \mathbf{E_{true}(h)}| > \varepsilon) \leq \mathbf{2}\ \mathbf{exp}^{-\mathbf{2m}\varepsilon^2}$$