Perspectives on Stochastic Modeling

Peter W. Glynn

Stanford University

Distinguished Lecture on Operations Research

Naval Postgraduate School, June 2nd, 2017

Management Science in the 20th Century

- Data gathered only when necessary
- Computers owned by corporations/governments
- “Humans in the Loop” on all decisions

Management Science in the 21st Century

- Computation everywhere
- Data everywhere
- Real-time decision-making

Management Science in the 21st Century

... and uncertainty everywhere

$\Downarrow$

stochastic models

Today:

- I entered the field in the 1980s...
- What are the 10 lessons I have learned about stochastic modeling?

Lesson 1: “Coin-toss world” versus “Scenario world”

- “Coin-toss world” is what we learn about in a probability class
- If we observe 100 coin tosses and see 40 heads, we can predict the number of heads in the next 1000 tosses
- We are using the belief that “the past predicts the future” and the law of large numbers

Lesson 1: “Coin-toss world” versus “Scenario world”

- “Scenario world” examples
- Global Climate Change:
  - Sea level increases? How much?
  - Hurricanes?
  - Regional rainfall/temperatures?
- Political Crises:
  - Ukraine
- Demand for a new product
- No historical data...

Lesson 1: “Coin-toss world” versus “Scenario world”

- Financial Risk Management
- Beware of over-reliance on historical data!!
- e.g. The U.S. Housing Bubble in 2008
- Collateralized mortgage obligations:
  - Reduce risk by packaging mortgages from different regional housing markets
  - No nationwide decline observed from 1970’s to 2007
- Historical time series can produce an unrealistic assessment of risk

However, a simple scenario analysis would have revealed the level of exposure to housing price risk.

Lesson 1: “Coin-toss world” versus “Scenario world”

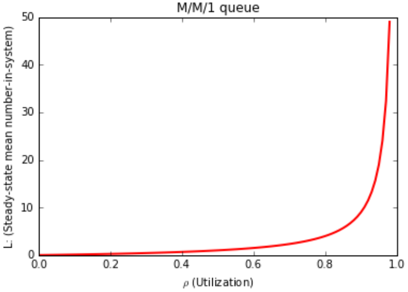
- Related questions:
- When calibrating financial models, how many trading days should be used?
- How much of the time series can be viewed as representative of future behavior?
- Challenge: Build quantitative tools capable of offering reliable insight into this question

Lesson 2: Stochastic models are often built with no data at all!

- Descriptive
- Predictive
- Prescriptive

Lesson 2: Stochastic models are often built with no data at all!

- Descriptive Model: (e.g. M/M/1 queue)



- Moral of the Story: You can't run systems at close to full utilization without affecting "Quality of Service"

Note: no data required!

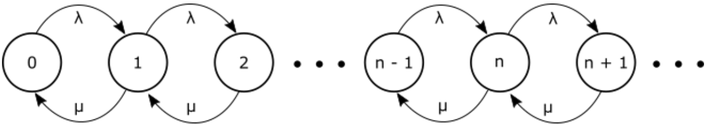
Lesson 2: Stochastic models are often built with no data at all!

- Prescriptive Model
- e.g. models for wireless networks
- Multiple channels, multiple user types
- Study the stability region

Again, no data needed

Lesson 3: Be aware that a computational model can be as useful as a "closed-form"

M/M/1 queue[1]



- The queue is initially in state i and write $p_k(t)$ for the probability of being in state k at time t:

$$p_k(t) = e^{-(\lambda+\mu)t}\left[\rho^{\frac{k-i}{2}} I_{k-i}(at) + \rho^{\frac{k-i-1}{2}} I_{k+i+1}(at) + (1-\rho)\rho^k \sum_{j=k+i+2}^{\infty} \rho^{-j/2} I_j(at)\right]$$

where $\rho = \lambda/\mu$, $a = 2\sqrt{\lambda\mu}$ and I_k is the modified Bessel function of the first kind.

[1] diagram source: grotto-networking

Lesson 3: Be aware that a computational model can be as useful as a “closed-form”

1980’s...

“Simulation is the method of last resort."

2017:

When models are being used predictively/prescriptively:

- Formulate the model from the start based on
  - The data available
  - Ease of computation

Models can be either easy or hard to simulate...

Lesson 3: Be aware that a computational model can be as useful as a "closed-form"

Hard to simulate:

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t)$$

Easy to simulate:

A model in discrete time...

Both possibilities are used in quantitative finance to describe asset prices...

Lesson 4: Diversification of risk can dramatically lower risk

Example: Auto insurance[2]

[2] picture credit: Consumerreports.org

Lesson 4: Diversification of risk can dramatically lower risk

- Auto insurance:
  - 1 million policy-holders
    Average claim size $\$1000$
    Standard deviation of claim size is $\$2000$
  - Suppose premiums are set at $\$1100$ per policy-holder
    Revenue = $\$1.1$ billion per year
  - How large a reserve to set?
  - Mean total claims: $\$1$ billion
    Standard deviation of total claims: $\$2000 \times \sqrt{1 \text{ million}}$ = $\$2$ million

Lesson 4: Diversification of risk can dramatically lower risk

- Property insurance in California
  - If it covers earthquake risk, it can not be diversified away
  - Private insurers will not cover
  - California Earthquake Authority

Lesson 4: Diversification of risk can dramatically lower risk

This diversification effect arises in many other settings:

- Warranties...
- “The square root staffing formula" for call centers:
  - The number of service reps needed beyond that required to handle the mean incoming load is small (i.e. of the order of the square root of the arrival time)
  - Averaging over many incoming customers “diversifies away" the noise and makes the fluctuations relatively smaller

Lesson 5: Randomization can be an effective tool for generating distributed algorithms

- Implementing “join the shortest” queue when there are a large number of queues
- Need to know “state” of each queue
- Needs a centralized decision-maker

Lesson 5: Randomization can be an effective tool for generating distributed algorithms

- A randomized alternative:
  - Each arriving customer chooses 2 queues at random and goes to the shorter of the 2 queues
  - Brings many of the same benefits as “standard” join-the-shortest queue

Lesson 5: Randomization can be an effective tool for generating distributed algorithms

In many settings, a small sample can carry almost as much information as a massive data set

- Used in data base settings to provide statistical answers to queries at lower computational overhead than running through entire data base
- Easier to parallelize

Lesson 6: Stochastic models can tell us what features of the data are important for calibration purposes

- Many server systems (e.g. call centers)
  - One builds a “high capacity” system because one needs to handle a high volume of incoming customers
  - This suggests studying systems where the arrival intensity n is sent to infinity:

$$n^{1/2}(Q_n(t) - q) \Rightarrow Z(t)$$

  - Important insight is that Z approaches equilibrium over roughly the typical service time

Lesson 6: Stochastic models can tell us what features of the data are important for calibration purposes

- If typical service time is 15 minutes, we need to understand the statistics of the arrival process over 15 minutes
- If arrival rate is 4000 calls per hour, we need to understand variability and correlation structure of arrivals over roughly 1000 customer arrivals
- What happens over a single customer inter-arrival is irrelevant
- Top-down statistical modeling based on accurate representation over 15 minute intervals

Similar statistical insights from systems in heavy traffic: The statistical calibration needs to focus on long time scales

Lesson 7: Rare event analysis is necessarily built on strong assumptions

- Want to design a trading strategy that generates 10% trading loss on at most 1 in 1000 trading days (relative to benchmark)

Figure: RBC Trading Floor

Lesson 7: Rare event analysis is necessarily built on strong assumptions

- An Important Modeling Principle:
- If we make n observations, we can estimate probabilities roughly of the order of $1/n$ in a "model free" way from the historical data
- Corollary: When we compute a probability of smaller order than $1/n$ from a model, the numerical value of the probability is primarily being driven by the model, not by the data

Lesson 7: Rare event analysis is necessarily built on strong assumptions

- A related insight:
- Suppose we want to “size” a distribution system so that the probability of order fulfillment taking longer than 2 days is less than 1%
- We fit a Poisson process to the demand data based on maximum likelihood (all data counts equally)
- But queueing theory tells us that it is the “left tail” of the inter-arrival distribution that matters, not the middle
- And it is the right tail of the service fulfillment times that matters...
- Predictions will be good only if we model the appropriate tails well

Lesson 8: High-dimensional data analysis is necessarily built on strong assumptions

- Risk Management:
  - Options are priced on the basis of volatility assessments implied by current "plain vanilla" option prices
- Historical data also used extensively
- Overall risk management depends on co-movement of asset prices
- Need for joint distributions

Lesson 8: High-dimensional data analysis is necessarily built on strong assumptions

- What does theory tell us?
- High-dimensional nonparametric statistical estimation requires enormous sample sizes (convergence rate $= n^{-2/(d+4)}$)
- So, we need to model the high-dimensional interactions (e.g. “factor” models)

Lesson 8: High-dimensional data analysis is necessarily built on strong assumptions

- When we have a lot of problem insight, we can model the interactions using physical/economic principles
- We can resort to simplified statistical approaches
- Fit the marginals
  Use a copula model (e.g. Gaussian)

Lesson 9

Be aware of:

Known knowns

Known unknowns

- In financial markets, correlation structure is estimated/calibrated from normal market behavior
- But correlations can change dramatically during a financial crises (fear-based markets)
- Stress testing via scenario analysis

Unknown unknowns

Lesson 10: Different types of bias in observing data

Be aware of the biases inherent in the data that is collected

e.g. clinical trial data in testing a treatment

The outcome can be:

- patient dies of treated disease
- patient survives trial period
- patient dies of other cause

When patient dies of other causes (or leaves the study for other reasons), we don't see the additional survival from the treatment.

Lesson 10: Different types of bias in observing data

e.g. online shopping

> If someone purchases at price x, we have no information about whether that consumer would have purchased at a higher price

Censored data...

Lesson 10: Different types of bias in observing data

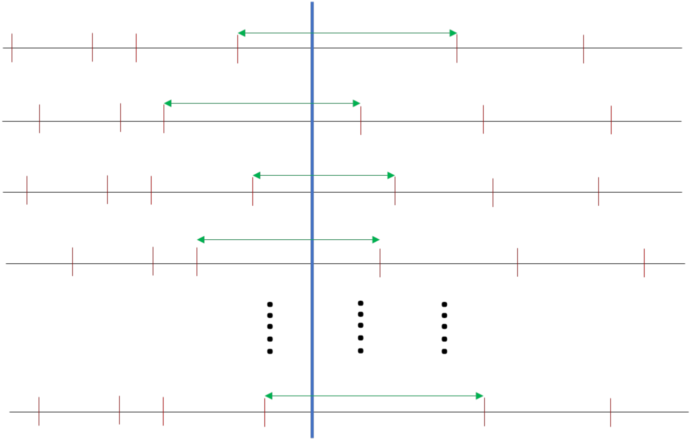
Credit Risk:

- Look at a financial product where the pay-off involves the number of firms that default on their bond obligations within a certain time period
- Model calibration involves modeling valuation of company
- Bias: The product involves only companies that have survived
- The sample involves companies that are “conditioned on not yet having gone bankrupt"
- So, naive statistical analysis leads to default probabilities that are biased upwards...

A grand jury visited a prison and polled the inmates on the length of their sentences

They compared this with the prison's own statistics on average sentence duration...

The inmates reported much longer sentences...



"Length Biasing"

Conclusions

- Uncertainty is always present when making decisions
- More and more decision tools contain embedded statistical and stochastic models
- More and more data is available to help inform decision-making and build models
- But we need to be increasingly sophisticated, as both model-builders and model-users, in how data gets used and interpreted
- This is both a challenge, and an opportunity, for OR/MS!

Thank you!!