Quant Curriculum Week 2
Probability Theory and Statistics

Jonathan Qin

March 2, 2023

1 Discrete and Continuous Distributions

1.1 Motivation

We will begin by studying distribution functions widely used for quantitative modeling. Having a intuitive understanding of these functions and being able to discern the unique characteristics of a distribution is a valuable skill.

1.2 Definitions

1. Random Variable: can be discrete or continuous

2. Cumulative distribution function: $F(a) = P\{X \leq a\}$, $\int_{-\infty}^{a} f(x)\, dx$

3. Probability mass function: $p(x) = P\{X = x\}$

4. Probability distribution function: $f(x) = \frac{d}{dx} F(x)$

5. Expected value: $E[X] = \sum_{x:p(x)>0} x p(x)$, $\int_{-\infty}^{\infty} x f(x)\, dx$

6. Variance: $var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

7. Bernoulli trial: random experiment with two possible outcomes (success/failure) of consistent probability

1.3 Discrete / Continuous Random Variable Distributions

1. Uniform: probability distributions with equally likely outcomes

2. Binomial (Discrete): probability distributions arising from several Bernoulli trials

3. Normal (Gaussian): continuous probability distribution associated with the Central Limit Theorem (often used for natural and social sciences)

4. Poisson (discrete): probability distribution expressing the probability of a given number of events occurring in a fixed interval of time given a known constant mean rate and independence

5. Gamma: a family of distributions related to the beta, exponential, and chi-squared distributions

6. Geometric (discrete): probability distribution of the number of Bernoulli trials to get one success

7. Exponential: probability distribution of the time between events in a Poisson point process

8. Lognormal: random variables with the log being normally distributed

1.4 Moments of Distributions

Moments of random variables describe quantitatively the shape of their distribution. For second and higher moments, the central moment (about the mean) is typically used to provide more meaningful information about the distribution's shape. If F is a cumulative probability distribution function, the n-th moment of the distribution is given by the Riemann-Stieltjes Integral:

$$\mu'(n) = E[X^n] = \int_{-\infty}^{\infty} x^n \, dF(x)$$

Furthermore, the Moment Generating Function is an alternative way to describe the distribution of a random variable. The Moment Generating Function of random variable X, $M_X(t)$, is given by:

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) \, dx$$

The Moment Generating Function can be used to find the moments of a distribution, as $M_X^{(n)}$ expanded becomes the series:

$$M_X(t) = E[e^{tX}] = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \cdots$$

Therefore, differentiating $M_X(t)$ i times and setting $t = 0$ will give the ith moment about the origin. A problem with moment-generating functions is that unlike characteristic functions (Fourier transform series), they may not exist for certain distributions, as the integral does not necessarily converge. However, it exists for almost all common distributions.

1. First Moment: Mean / Expected Value (raw moment)

2. Second Moment: Variance, deviations from the mean (central moment)

3. Third Moment: Skewness, a measure of the asymmetry of a distribution about the mean. Right skewed is defined as a positive skewness with the tail extending towards the right side.

4. Fourth Moment: Kurtosis, a measure of the thickness of tails. Excess kurtosis is kurtosis excess of the normal distribution. In finance, kurtosis is a measure of the extent of price fluctuations (how often prices move dramatically) and can be used to model risk in Value At Risk models.

Normal Moments Example What is the moment-generating function of a standard normal distribution? What are the distribution's first, second, third, and fourth moments?

Properties of Poisson Processes Buses at a station arrive according to a Poisson process with an average arrival time of 10 minutes (0.1 / min). If you arrive at a random time, what is your expected waiting time? On average, how many minutes ago did the last bus leave?

Poisson Process has occurs at constant mean rate λ, events can not occur at the same time, and events occur independently. The Poisson distribution gives the number of events in some time interval t given expected number of events:

$$P(X = \lambda t) = \frac{\lambda t e^{-\lambda t}}{k!}$$

Meanwhile, the exponential distribution gives the expected time between events, and it is a continuous and memory-less distribution:

$$P(t) = \lambda e^{-\lambda t}$$

2 Quick Probability Problems

Problem 2.1 Tennis Match Two tennis players are playing a 3 set tennis match. The match ends when a player wins two sets. The probability of a player winning each set is constant throughout the match. If you are given money to bet on whether the game finishes in 2 or 3 sets, which outcome should you bet on?

Problem 2.2 Aces You shuffle a deck of 52 cards and begin to draw cards one by one without replacement. How many cards do you expect to draw for you to draw your first ace?

3 Quant Trader Technical: Sports Brackets Betting

This is a technical interview question (45 minutes) that was asked by a trading firm in their first round this recruiting season. The problem summarizes perfectly the core concepts of probability we covered while serving as a good introduction to Game Theory. I will present a modified version of the game, as well as an explanation and my code implementation (and solution) to the problem afterward.

3.1 Sports Bracket System

There are $n = 2^m$ teams playing in a single elimination bracket for a championship. Each team has a strength between 100 and $100n$, not necessarily unique. If team A and team B face-off, the probability of team A winning is given by:

$$P(\text{A winning}) = \frac{\text{strength(A)}}{\text{strength(A)} + \text{strength(B)}}$$

For example, if team A had strength 100 while team B had strength 200, team A would have a $\frac{1}{3}$ chance of winning.

3.2 Game Betting System

At the start of each round, odds (prices) for each remaining team will be presented for the player to bet on. The bets are structured as follows: you are able to bet on each team winning the championship for some price K. If the team does win the entire bracket, the payoff for your bet is $100. Otherwise, you win nothing. There is no limit on how many bets you may place on each team.

For $n = 2^m$ teams playing in single elimination, m rounds are needed to determine a victor, so there will be m betting rounds as well. Each round, the odds for the remaining teams will be updated, but the final payout of $100 remains the same.

For example, if team A has betting odds of $10 on Round 1, that means placing a bet on team A winning the entire bracket and the championship costs $10 with a potential payout of $100. Keep in mind that this means only "long" bets are possible - you are unable to explicitly "short" a team, other than going "long" on the opposing team.

3.3 Evaluation Metrics

At the end of the game, you are presented with your net profits, given by the total winnings less the total betting expenses. While net profits are one way to evaluate performance, more important is formulating a winning strategy that is explainable and successful long-term. It is possible to face a net loss even with a game theory optimal strategy due to chance. The next question becomes: after determining a profitable strategy, how can I minimize risk while implementing this strategy?