MATHEMATICAL FOUNDATIONS FOR FINANCE

Introduction to Probability and Statistics: Part 1

Philip H. Dybvig
Washington University
Saint Louis, Missouri

Probability and Statistics

Probability is the mathematical study of randomness. Since most interesting topics in Finance involve randomness, the tools from probability theory are central to Finance, and most models in Finance are probabilistic models.

Statistics is the mathematical study of how we learn from data. It is where probability theory meets the data, and is also an application of probability theory.

For example,

- The Black-Scholes model of option pricing is a probabilistic model.
- Estimating the parameters of the Black-Scholes model and assessing the quality of the estimates is a statistical procedure.
- Testing whether the Black-Scholes model fits the data is also a statistical task.

Most of this lecture will focus on probability theory.

Random Variables, Probabilities, and Distributions

A **Random Variable** takes on a value as a function of the *state of nature* (or the state of the economy etc.). Real-valued random variables can be discrete (e.g. the number of Apple stock trades tomorrow) or continuous (the Apple stock price at the end of the day tomorrow). A random variable can also take on values in a set other than real numbers, for example, the price of IBM through the day tomorrow is a *random process* because it is a random function of time.

A **Probability Measure** has probabilities adding up to one of the different mutually exclusive possible outcomes. There can be objective probabilities (e.g. 1/2-1/2 for a coin toss),[1] subjective probabilities (your belief that Google stock will go up tomorrow with high probability, or your probability belief about the change in short Treasury rates yesterday), or sample probabilities (the monthly returns to the S&P500 index for the past 60 months, giving each month equal probability weight of 1/60).

A **Probability Distribution** gives the probabilities of the outcomes without regard to the identity of the states.

[1] notwithstanding Einstein's assertion that God "does not throw dice"

Objective Probability Distribution: Frequency Interpretation

One interpretation of objective probabilities is that in repeated independent draws of a random variable, the objective probabilities are the probabilities in a sample that gets very large. In fact, probability theory can be used to prove that if there are objective probabilities, then the sample distribution of independent draws from the probability distribution converges to the objective distribution as we add draws. This is called the **Fundamental Theorem of Statistics** because it shows that we can learn about true probabilities by looking at the data.

For example, in a repeated flip of a coin that is almost fair, objectively speaking the fraction of times we see heads might be .5000127, and the fraction of times we see tails might be .4999873 (a fair coin would have probabilities of 0.5 and 0.5). This example also illustrates the limitation of the result; if we really flipped the same coin millions or billions of times needed to form these estimates, each time the shape of the coin would change a little bit and over time the probabilities would change. And, in finance, we usually do not really see essential similar situations repeating millions of times.

Discrete Random Variables: Probability Distribution and Special Moments

The probability distribution of a random variable gives the probability π_i of each outcome x_i without worrying about which state of nature it occurs in. For the π_i's to be a proper probability distribution, we must have $(\forall i)\pi_i \geq 0$ and $\Sigma_i \pi_i = 1$.

- The **mean** or **expectation** of a discrete random variable can be computed by adding up the quantities times the probabilities:

$$\mu_x = E[x] = \sum_i \pi_i x_i.$$

- The **variance** of a discrete random variable is the expectation of the squared deviation from the mean:

$$\mathrm{var}(x) = E[(x-\mu_x)^2] = \sum_i \pi_i (x_i - \mu_x)^2 = E[x^2] - \mu_x^2 = E[x(x-\mu_x)].$$

- The **standard deviation** of a random variable is the square root of the variance: $\sigma = \sqrt{\mathrm{var}(x)}$, which has the same units as x and μ_x.

More on Moments

The mean is a measure of the middle of a distribution, and is especially important in Finance because a lot of our models are stated in terms of means. For example, the Capital Asset Pricing Model (CAPM) says that the mean excess return of a stock over the riskfree interest rate is proportional to the stock's risk as measured by its beta-coefficient. Other measures of the middle of the distribution are the mode and the median; these are not widely used in Finance.

The variance and standard deviation are both measures of volatility of a random variable around the mean. When an option trader talks about the vol of the underlying, he is talking about the standard deviation of the stock or commodity on which the option was written. This might come from the recent historical prices, the vol implicit in option prices, or the trader's own opinion.

In general, we can talk about more moments. We call $E[x^n]$ the nth moment around zero, and we call $E[(x-\mu_x)^n]$ the nth moment around the mean. In this notation, the mean is the first moment around 0 and the variance is the second moment around the mean. Higher moments tell us more about the shape of the distribution.

Simple Example

We model the return to Ameren stock over the next day as having a distribution given by the following table:

state name	value of x	probability
good state	$+3\%$	0.5
bad state	-2%	0.5

- mean return $0.5 \times .03 + 0.5 \times (-.02) = 0.5\%$
- variance of return $0.5 \times (.03 - .005)^2 + 0.5 \times (-.02 - .005)^2 = 0.000625$
- or, $0.5 \times .03^2 + 0.5 \times (-.02)^2 - 0.005^2 = 0.000625$
- standard deviation of return $\sqrt{.000625} = 2.5\%$

In-class Exercise

Model the return to Monsanto stock over the next day as having a distribution given by the following table:

state name	value of x	probability
good state	$+4\%$	$2/3$
bad state	-2%	$1/3$

Compute the mean return, variance of returns, standard deviation of returns, third moment around the mean, and fourth moment around the mean.

More Complex Example

Suppose the return x of IBM over the next day is modeled as having a distribution given by the following table:

state name	value of x	probability
good state	$+5\%$	0.5
intermediate state	-2%	0.49
bad state	-40%	0.01

- mean return $0.5 \times .05 + .49 \times (-.02) + .01 \times (-.4) = 1.12\%$
- variance of the return $0.5 \times .0025 + .49 \times .0004 + .01 \times .16 - .000025 = .00292056$
- standard deviation of the return $\sqrt{.00292056} \approx 5.4\%$
- third moment around the mean $0.5 \times (.05 - .0112)^3 + .49 \times (-.02 - .0112)^3 + .01 \times (-.4 - .0112)^3 \approx -0.000681$
- fourth moment around the mean $0.5 \times (.05 - .0112)^4 + .49 \times (-.02 - .0112)^4 + .01 \times (-.4 - .0112)^4 = 0.0002875$

Interpreting mean and standard deviation

Mean:

- a measure of the typical value of a random variable
- possible alternatives: mode, median
- mean portfolio return: reward for investing

Standard deviation:

- a measure of the noisiness of a random variable
- possible alternatives: interquartile range, mean absolute deviation
- standard deviation of a portfolio return: risk of investing
  - not a good measure of risk but easy to use
  - widely used in practice

Working with moments

Let k and k' be constants, let x and y be random variables, and let $n > 0$ be an integer..

- $E[k] = k$. $E[(k - \mu_k)^n] = 0$. $E[k^n] = k^n$.
- $E[kx] = kE[x]$. $E[(kx - \mu_{kx})^n] = k^n E[(x - \mu_x)^n]$. $E[(kx)^n] = k^n E[x^n]$.
- $E[x + y] = E[x] + E[y]$. Equivalently, $\mu_{x+y} = \mu_x + \mu_y$.
- $E[k_1 x + k_2 y] = k_1 E[x] + k_2 E[y]$.
- $\mathrm{var}(k_1 x + k_2 y) = k_1^2 \mathrm{var}(x) + 2k_1 k_2 \mathrm{cov}(x, y) + k_2^2 \mathrm{var}(y)$, where the covariance of x and y, $\mathrm{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] = E[xy - x\mu_y - y\mu_x + \mu_x \mu_y] = E[xy] - E[x]E[y] = E[(x - \mu_x)y]$. If this covariance is zero, we say x and y are uncorrelated, and a special case of this is independence of x and y (more on this later). Note, that if k is a constant, then $\mathrm{cov}(x, k) = 0$ for all x.
- $E[xy] = E[x]E[y] + \mathrm{cov}(x, y)$.

In-class Exercise: portfolio mean and variance

Let x be the random return to a broad stock index and let y be the return to a stock we expect to do well this year. Assume that

- $\mu_x = 5\%$ and $\mu_y = 10\%$
- $\text{var}(x) = .04$, $\text{var}(y) = .16$, and $\text{cov}(x, y) = .06$

- What is the mean and standard deviation of the return from putting all your money in the index? from putting all your money in the single stock? half and half?
- Which would you prefer?

Note: putting a fraction θ of the money in the single stock gives a return equal to $\theta y + (1 - \theta)x$.

Skewness and Kurtosis

When thinking intuitively about third and fourth moments, it is useful to consider dimensionless quantities generated by normalizing the moments by the appropriate power of the standard deviation. We define the skewness of x by

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = \frac{E[(x-\mu)^3]}{\sigma^3}.$$

Intuition: skewness is positive if the right tail is relatively larger and negative if the left tail is relatively larger.

We define the kurtosis of x by

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] = \frac{E[(x-\mu)^4]}{\sigma^4}.$$

Sometimes kurtosis (or "excess kurtosis") is defined by this measure less 3, since 3 is the "normal" kurtosis of any normal random variable (which follows the bell curve). Kurtosis is a measure of how fat the tail are. A lot of times we talk about stock returns as being leptokurtotic, which is a big word for positive excess kurtosis. This means that compared to a normal distribution, stock returns tend to have more big outliers.

Linear regression

Linear regression is one of the most commonly used statistical procedures. If we have two random variables, x and y, a linear regression is a fit to the equation

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

where α and β are constants and ε is an error term. The most commonly used estimates for α and β are the Ordinary Least Squares (OLS) estimates that minimize $E[\varepsilon^2]$. These estimates are given by

$$\hat{\beta} = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)}$$

and

$$\hat{\alpha} = E[y] - \hat{\beta} E[x].$$

Typically, the covariance and variance used in computing the beta-coefficient are computed from the sample distribution that puts equal weights on all the observations.

In-class exercise: linear regression

Perform the linear regression

$$R_A = \alpha + \beta R_M + \varepsilon$$

where R_A is the return on Apple stock and R_M is the return on the market index. In particular, estimate α and β using OLS, given the annual stock returns for three years of

	Apple return	Market return
year 1	45%	30%
year 2	-3%	-10%
year 3	-9%	10%

(Note: I rigged these numbers to give reasonable coefficients assuming a riskfree rate of 5%. Doing a regression like this on three yearly observations is expected to give crazy results. Even with 60 monthly observations, the results of a regression like this will be really noisy.)

Continuous random variables

However, it is often more useful to think of considering random variables that take on continuously many values. For example, we can think of stock prices quoted to the penny as taking on all values 0.01, 0.02, ... 200.00. Or, we can just think of this as a continuous random variable that can take on all real values.

For a continuous distribution, we can work with a probability density function $f(x)$ instead of the discrete probabilities π_i. Probabilities summing to 1 for the discrete and continuous cases can be written as:

$$\sum_i \pi_i = 1 \quad \int_{x=-\infty}^{\infty} f(x) = 1.$$

The formula for the mean in each case is:

$$\mu_x = \sum_i \pi_i x_i \quad \mu_x = \int_{x=-\infty}^{\infty} x f(x) dx.$$

The formula for the variance in each case is:

$$\text{var}(x) = \left(\sum_i \pi_i x^2\right) - {\mu_x}^2 \quad \text{var}(x) = \left(\int_{x=-\infty}^{\infty} x^2 f(x) dx\right) - {\mu_x}^2.$$

Example: uniform distribution

Suppose x is uniformly distributed on $[1,3]$. Then x has density $f(x) = 1/2$ for $x \in [1,3]$ and density $f(x) = 0$ everywhere else. Confirm the probabilities sum to 1 and compute the mean, variance, and standard deviation of x.[2]

$$\int_{x=1}^{3} f(x)dx = \int_{x=1}^{3} \frac{1}{2}dx = \left[\frac{x}{2}\right]_1^3 = \frac{3}{2} - \frac{1}{2} = 1.$$

$$\mu_x = \int_{x=1}^{3} xf(x)dx = \int_{x=1}^{3} x\frac{1}{2}dx = \left[\frac{x^2}{4}\right]_1^3 = \frac{3^2}{4} - \frac{1^2}{4} = 2.$$

$$\mathrm{var}(x) = \int_{x=1}^{3} x^2 f(x)dx - {\mu_x}^2 = \int_{x=1}^{3} x^2\frac{1}{2}dx - 2^2 = \left[\frac{x^3}{6}\right]_1^3 - 4 = \frac{27}{6} - \frac{1}{6} - 4 = \frac{1}{3}.$$

$$\mathrm{std}(x) = \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{3} \approx .577$$

[2] For $n \neq 0$, $dx^n/dx = nx^{n-1}$ and therefore for $m \neq -1$, $\int x^m dx = x^{m+1}/(m+1)$.