2. Random Variables

FRM Part 1: Quantitative Analysis

Devere Anthony Weaver

---

1 Introduction

For the FRM, this section covers fairly basic and very specific concepts concerning random variables, although it is *incredibly* helpful to already have some sort of background covering the more theoretical aspects of it since the GARP books gloss over much of the detail.

---

2 Definition of a Random Variable

Definition. *Random Variable*
*For some given sample space $\mathcal{S}$ of some experiment, a **random variable** is any rule that associates a number with each outcome in $\mathcal{S}$. It's a function whose domain is $\mathcal{S}$ and whose range is some subset of real numbers $\mathbb{R}$.*

NOTE: The GARP texts use ω to represent an event and Ω to represent a sample space.

---

3 Discrete Random Variables

Definition. *Bernoulli Random Variable*
*A **Bernoulli random variable** is any random variable whose only possible values are 0 and 1.*

Bernoulli random variables are frequently encountered when measuring binary random events (e.g. the default of a loan).

Definition. *Discrete Random Variable*
*A **discrete random variable** is a random variable whose possible values constitute either a finite set or a countably infinite set.*

3.1 Probability Distributions for Discrete Random Variables

Definition. *Probability Mass Function*
*The **probability mass function** (pmf) of a discrete random variable is defined $\forall\, x$,*

$$p(x) = P(X = x) = P(\forall\, s \in Si : X(s) = x)$$

.

In other words, the pmf is used to compute the probability that a discrete random variable takes on a certain value. It describes how the mass of 1 is distributed at various points along the axes of possible value of the random variable.

3.2 A Parameter of a Probability Distribution

Suppose $p(x)$ (a pmf) depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. This value is called a **parameter** of the probability distribution.

The collection of all probability distributions for different values of the parameter is called a family of probability distributions. An as example, take the pmf of a Bernoulli random variable, given by

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1. \end{cases}$$

This pmf form allows it to take on different values for the parameter α.

The *family of geometric distributions* is given by

$$p(x) = (1-p)^{x-1}p, \;\; x = 1, 2, 3 \ldots$$

3.3 Cumulative Distribution Function (cdf)

Definition. *Cumulative Distribution Function (cdf)*
The cdf $F(x)$ of a discrete random variable X with pmf $p(x)$ is defined $\forall$ x

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y).$$

In other words, the cdf measures the total probability of observing a value less than or equal to the input value x. The cdf measures total probability, thus it must be an increasing monotonic function in x.

Going back to the Bernoulli random variable, the cdf is a simple step function.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases}$$

One important property of a cdf is that if we can also use it to find the probability that a random variable takes on some value. We do this by computing the cdf at two points and subtracting them.

So, $\forall$ a, where $a \leq b$,

$$P(a \leq x \leq b) = F(b) - F(a-).$$

This means to find the probability that our random variable lies between these two points, a and b, we evaluate the cdf at both points and subtract, where the second term is the limits of $F(x)$ as x approaches from the left.

3.4 Expected Values of a Discrete Random Variable

Definition. *Let X be a discrete random variable with a set of possible value $\mathcal{D}$ and pmf $p(x)$. The* ***expected value*** *or* ***mean value*** *of X, denoted by $E(X)$ or μ_x, is given by*

$$E(X) = \mu_x = \mu = \sum_{\forall x \in \mathcal{D}} xp(x).$$

The expectation will exist if and only if $\sum |x|p(x) < \infty$.

Observe that the expectation is weighted, thus is might not match the arithmetic mean of the possible values.

For an unbounded set of possible values of X, the expected value may or may not exist, it depends on if the infinite series converges or diverges.

It is also possible to find the expected value of any function $h(x)$,

$$E[h(x)] = \sum_{\forall x \in \mathcal{D}} h(x)p(x).$$

So, we just multiply the function's value at x and the pmf at x.

The expectation operator is linear. For a random variable X, and constants, a and b,

$$E(aX + b) = aE(x) + b.$$

3.5 A Financial Example using Expectation

As an example of using computing expectation related to a financial problem, let's consider a call option.

Recall, a call option is a derivative contract with a payoff that is a nonlinear function of the future price of the underlying asset.

The payoff of the call option (at the expiry date) is

$$c(S) = max(S - K, 0),$$

where S is the value of the underlying at expiry and K is the strike price.

If the price is below the strike price at expiry, the call option pays zero. Valuing this call option involves computing the expectation of a function (i.e. the payoff) of the price (i.e. a random variable).

So, suppose the following is the pmf of the value of the underlying that gives the probability of different values of the call option.

$$p(s) = \begin{cases} 0.2 & s = 20 \\ 0.5 & s = 50 \\ 0.4 & s = 30. \end{cases}$$

Now, if the strike price is $K = 40$, the the payoff of the call option can be written as

$$c(S) = max(S - 40, 0).$$

Thus, the expected payoff of the call option involves computing the expected value of the call option function using the probability of each value.

$$E[c(S)] = .2(max(20 - 40, 0) + .5(max(50 - 40, 0) + 0.3(max(100 - 40, 0) = 23.$$

Therefore, the expected value of the call option by expiry 23.

3.6 Moments and Moment Generating Functions of DRVs

The expected values of integer powers of X and $(X - \mu)$ are often referred to as *moments*.

Definition. *The kth moment of a random variable X is $E(x^k)$. The kth moment about the mean (or kth central moment) of X is $E[(X - \mu)^k]$.*

It's possible to compute any kth moment; however, we're often only concerned with a few. Specifically for our types of quantitative analysis, we're only concerned with

- mean (expected value) = the first moment

$$\mu = \sum_{x \in \mathcal{D}} xp(x).$$

- variance = the second moment

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x \in \mathcal{D}} (x - \mu)^2 p(x).$$

- skewness = the third moment (measures departure from symmetry)

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

- kurtosis - the fourth moment (indicates a degree of central ”peakedness”, fatness of outer tails)

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

Skewness measures asymmetry in a distribution and negative values indicate skew to the left and positive values indicate skew to the right.

In financial analysis terms, a random variable with a kurtosis greater than 3 are described as being heavy-tailed/fat-tailed. This is incredibly common in financial return distributions.

Definition. *The **moment generating function** (mgf) of a discrete random variable X is defined to be*

$$\mu_x(t) = E(e^{tx}) = \sum_{x \in \mathcal{D}} e^{tx} p(x).$$

One thing to note about mgfs is that the mgf at $t = 0$ is always 1. Another important property of mgfs is that they uniquely specify a probability distribution.

---

4 Continuous Random Variables

Definition. *Continuous Random Variables A random variable is **continuous** if*

- *its set of possible values consists either of all numbers in a single interval on the number line or all numbers in a disjoint union of such intervals*
- *no possible value of the variable has positive probability*

4.1 Probability Density Functions and Cumulative Distribution Functions

Definition. *Let X be a continuous random variable. The probability distribution of a **probability density function** $f(x)$ is given by*

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

This gives us the probability that X takes on a value in the interval $[a, b]$. The computed probability is the area above the interval and under the graph of the density function.

The pdf must be greater than or equal to 0 for all values in its support and integrating from negative infinity to positive infinity must equal 1.

Definition. *The **cumulative distribution function** F(x) for a continuous random variable X is defined as*

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy.$$

Again, we can use the cdf to compute probabilities by subtraction.

For the continuous case, we can the pdf by taking the first derivative of the cdf. Likewise, we can obtain the cdf by integrating over the pdf.

4.2 Expected Values and Moment Generating Functions

Definition. *The **expected value** or **mean value** of a continuous random variable X with pdf $f(x)$ is given by*

$$\mu = \mu_x = E(X) = \int_{-\infty}^{\infty} x f(d) dx.$$

In practice, the limits of integration are specified by the support of the pdf.

Similar to discrete random variables, we can also find the expected value of function with a continuous random variable.

Basically, everything else is analogous to the discrete case, it's just that we instead integrate instead of sum.

---

5 Quantiles

Definition. *Let p be a number such that $0 < p < 1$. Then the **(100p)th percentile** or **pth quantile** of a distribution of a continuous random variable X is defined as*

$$p = F(\eta_p) = \int_{-\infty}^{\eta_p} f(y) dy.$$

Assuming we can find the inverse of cdf $F(x)$, we can also write

$$\eta_p = F^{-1}(p).$$