1 Mean and Variance via Monte Carlo Simulation

We ask what happens when we define a random variable

$$Z_n = X_1 + X_2 + X_3 + \cdots + X_n$$

where $X_1, \ldots X_n$ are independent of each other. What does the distribution of Z_n look like given that we know $X_1, \ldots, X_n$? Z_n could represent the price of a financial asset after n timesteps, a diffusing particle, a sample mean and much more…

Exercise 1. What other distributions could be modelled by such a sum of independent variables?

We begin by studying the mean and variance of the random variable Z_n numerically.

The mean $\mu(X)$ of a random variable X is given by

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\, P(\omega)$$

where $\omega \in \Omega$ is a member of the event space Ω, $X: \Omega \to X(\Omega)$ is a random variable and $P(\omega)$ is the probability associated to the event ω. The sum only makes sense in the discrete case. In general, the above would be an integral

Exercise 2. For those of you who know about integrals, what is the analagous definition using an integral? When is each definition appropriate?

Exercise 3. What does the mean represent intuitively?

The variance $\mathrm{var}(X)$ of a random variable X is given by

$$\mathrm{var}(X) = \mu((X - \mu(X))^2) = \sum_{\omega \in \Omega} (X(\omega) - \mu(X))^2\, P(\omega)$$

Exercise 4. Plot mean and variance as a function of the timestep n. How does the mean and variance depend on the number of timesteps?

Exercise 5. Define and fit a model for the time dependence of the variance.

In fact, on can say more about Z_n. It turns out Z_n approaches a normal distributed random variable as $n \to \infty$. This result is the content of the central limit.

Theorem 1. (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be independent random variables with finite mean and variance. Then the probability density $p(z)$ of $Z_n = X_1 + X_2 + \cdots + X_n$ converges to*

$$p(z) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{(z - \mu)^2}{\sigma^2}\right)$$

as $n \to \infty$ where μ and σ^2 are the mean and variance of Z_n.

Exercise 6. Given the Central Limit Theorem, what kind of real world distributions might you expect to be normally distributed?

Exercise 7. Explore numerically whether the theorem holds. In particular, can you simulate some counter examples which are not normally distributed?

To find out how the mean and variance scale with n we can use simpler arguments. The following series of exercises will have you show how mean and variance depend on the means and variance of the individual X_i variables. To get warmed up, consider

Exercise 8. Pick a jump distribution and calculate the mean and variance for it. Hint: you'll find the following useful: $\mu(X)=\sum_x x\,p(x)$, where $p(x)$ is the probability the random variable takes on the value x.

The mean has an important property: it is linear. That is,

$$\mu(a\,X+b\,Y)=a\,\mu(X)+b\,\mu(Y)$$

where a,b are real numbers and X,Y are random variables.

Exercise 9. Prove this.

Exercise 10. Show that $\mu(Z_n)=\sum_{i=1}^{n}\mu(X_n)$ hint: use induction.

It can be shown that

$$\mathrm{var}(X)=\mu(X^2)-\mu(X)^2=\sum_x p(x)\,x^2-\left(\sum_x p(x)\,x\right)^2$$

Exercise 11. Prove this. Hint: expand $\mu((X-\mu)^2)=\mu(X^2-2\,\mu\,X+\mu^2)$ and use linearity of mean.

Exercise 12. Show that $\mathrm{var}(a\,X)=a^2\,\mathrm{var}(X)$ where a is a real number.

Finally, we are now in a position to treat the variance of a sum. It can be shown that independence of two random variables implies that the covariance is zero. This means

$$X,Y\text{ independent}\Longrightarrow\mathrm{cov}(X,Y)=\mu((X-\mu(X))\,(Y-\mu(Y)))=\mu(X\,Y)-\mu(X)\,\mu(Y)=0$$

Exercise 13. Using independence of X and Y, derive an expression for $\mathrm{var}(X+Y)$.

By induction you can now get the final expression for the variance.

Exercise 14. What is $\mathrm{var}(Z_n)$ as a function of the variances $\mathrm{var}(X_i)$?

2 Kelly Betting

You are a gambler. The amount of money you have to gamble with is x. The casino is willing to offer you the following bet. You may bet any fraction f of your bankroll x. With probability p the casino will pay you b times what you bet, so $b\,f\,x$. With probability $1-p$ the casino wins your bet and you lose $x\,f$. In this section you'll investigate what fraction f is optimal in the sense that it maximises mean returns.

Exercise 15. Implement the function *meanreturns*, which calculates the mean return over N trajectories and nt bets, given parameters f,b,p and the starting bankroll x_0.

Exercise 16. Investigate what fraction is optimal.

This problem can in fact be solved analytically (meaning on paper by hand). The result is the formula

$$f^*=p-\frac{1-p}{b}$$

where f^* is the optimal fraction.