Week 6: Variance, the law of large numbers and Kelly's criterion

Expected value, variance, and Chebyshev inequality. If X is a random variable recall that the *expected value of* X, $E[X]$ is the average value of X

$$\textbf{Expected value of } X: \quad E[X] = \sum_{\alpha} \alpha\, P(X = \alpha)$$

The expected value measures only the average of X and two random variables with the same mean can have very different behavior. For example the random variable X with

$$P(X = +1) = 1/2\,, \quad P(X = -1) = 1/2$$

and the random variable Y with

$$[P(X = +100) = 1/2\,, \quad P(X = -100) \,=\, 1/2$$

have the same mean

$$E[X] = E[Y] \,=\, 0$$

To measure the "spread" of a random variable X, that is how likely it is to have value of X very far away from the mean we introduce the *variance of* X, denoted by $var(X)$. Let us consider the distance to the expected value i.e., $|X - E[X]|$. It is more convenient to look at the square of this distance $(X - E[X])^2$ to get rid of the absolute value and the variance is then given by

$$\textbf{Variance of } X: \mathrm{var}(X) \,=\, E\left[(X - E[X])^2\right]$$

We summarize some elementary properties of expected value and variance in the following

Theorem 1. *We have*

1. *For any two random variables X and Y, $E[X + Y] = E[X] + E[Y]$.*
2. *For any real number a, $E[aX] = aE[X]$.*
3. *For any real number c, $E[X + c] = E[X] + c$.*

4. *For any real number* a, $\mathrm{var}(aX) = a^2\mathrm{var}(X)$.

5. *For any real number* c, $\mathrm{var}(X + c) = \mathrm{var}(X)$.

Proof. 1. should be obvious, the sum of averages is the average of the sum. For 2. one notes that if X takes the value α with some probability then the random variable aX takes the value $a\alpha$ with the same probability. 3 is a special case of 1 if we realize that $E[a] = a$. For 4. we use 2 and we have

$$\mathrm{var}(X) \;=\; E\left[(aX - E[aX])^2\right] \;=\; E\left[a^2 (X - E[X])^2\right] \;=\; a^2 E\left[(X - E[X])^2\right] \;=\; a^2\mathrm{var}(X)\,.$$

Finally for 5. not that $X + a - E[X + a] = X - E[x]$ and so the variance does not change. ∎

Using this rule we can derive another formula for the variance.

$$\begin{aligned}
\mathrm{var}(X) &= E\left[(X - E[X])^2\right] \\
&= E[X^2 - 2XE[X] + E[X]^2 \\
&= E[X^2] + E[-2XE[X]] + E[E[X]^2] \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

So we obtain

$$\boxed{\textbf{Variance of } X \;:\; \mathrm{var}(X) \;=\; E\left[(X - E[X])^2\right] \;=\; E[X^2] - E[X]^2}$$

Example: The $0 - 1$ **random variable**. Suppose A is an event the random variable X_A is given by

$$X_A \;=\; \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

and let us write

$$p = P(A)$$

The we have

$$E[X_A] = 0 \times P(X_A = 0) + 1 \times P(X_A = 1) \;=\; 0 \times (1 - p) + 1 \times p \;=\; p\,.$$

To compute the variance note that

$$X_A - E[X_A] = \begin{cases} 1-p & \text{if } A \text{ occurs} \\ -p & \text{otherwise} \end{cases}$$

and so

$$\text{var}(X) = (-p)^2 \times P(X_A = 0) + (1-p)^2 \times P(X_A = 1) = p^2(1-p) + (1-p)^2 p = p(1-p)$$

In summary we have

The $0-1$ random variable

$$P(X=1) = p\,, \quad P(X=0) = (1-p)$$
$$E[X] = p\,, \quad \text{var}(X) = p(1-p)$$

∎

Chebyshev inequality: The Chebyshev inequality is a simple inequality which allows you to extract information about the values that X can take if you know only the mean and the variance of X.

Theorem 2. *We have*

1. **Markov inequality**. *If $X \geq 0$, i.e. X takes only nonnegative values, then for any $a > 0$ we have*
$$P(X \geq a) \leq \frac{E[X]}{\alpha}$$

2. **Chebyshev inequality**. *For any random variable X and any $\epsilon > 0$ we have*
$$P\left(|X - E[X]| \geq \epsilon\right) \leq \frac{\text{var}(X)}{\epsilon^2}$$

Proof. Let us prove first Markov inequality. Pick a positive number a. Since X takes only nonnegative values all terms in the sum giving the expectations are nonnegative we have

$$E[X] = \sum_{\alpha} \alpha P(X=\alpha) \geq \sum_{\alpha \geq a} \alpha P(X=\alpha) \geq a \sum_{\alpha \geq a} P(X=\alpha) = aP(X \geq a)$$

and thus

$$P(X \geq a) \;\leq\; \frac{E[X]}{a}\,.$$

To prove Chebyshev we will use Markov inequality and apply it to the random variable

$$Y = (X - E[X])^2$$

which is nonnegative and with expected value

$$E[Y] \;=\; E\left[(X - E[X])^2\right] = \mathrm{var}(X)\,.$$

We have then

$$\begin{aligned}
P(|X - E[X]| \geq \epsilon) \quad &= \quad P((X - E[X])^2 \geq \epsilon^2) \\
&= \quad P(Y \geq \epsilon^2) \\
&\leq \quad \frac{E[Y]}{\epsilon^2} \\
&= \quad \frac{\mathrm{var}(X)}{\epsilon^2}
\end{aligned} \tag{1}$$

■

Independence and sum of random variables: Two random variables are independent independent if the knowledge of Y does not influence the results of X and vice versa. This can be expressed in terms of conditional probabilities: the (conditional) probability that Y takes a certain value, say β, does not change if we know that X takes a value, say α. In other words

$$Y \text{ is independent of } X \text{ if } P(Y = \beta | X = \alpha) = P(Y = \beta) \text{ for all } \alpha, \beta$$

But using the definition of conditional probability we find that

$$P(Y = \beta | X = \alpha) = \frac{P(Y = \beta \cap X = \alpha)}{P(X = \alpha)} = P(Y = \beta)$$

or

$$P(Y = \beta \cap X = \alpha) \;=\; P(X = \alpha) P(Y = \beta)\,.$$

This formula is symmetric in X and Y and so if Y is independent of X then X is also independent of Y and we just say that X and Y are independent.

X and Y are **independent** if $P(Y = \beta \cap X = \alpha) = P(X = \alpha)P(Y = \beta)$ for all α, β

Theorem 3. *Suppose X and Y are independent random variable. Then we have*

1. $E[XY] = E[X]E[Y]$.
2. $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Proof. : If X and Y are independent we have

$$\begin{aligned} E[XY] &= \sum_{\alpha,\beta} \alpha\beta P(X = \alpha Y = \beta) \\ &= \sum_{\alpha,\beta} \alpha\beta P(X = \alpha)P(Y = \beta) \\ &= \sum_{\alpha} \alpha P(X = \alpha) \sum_{\beta} \beta P(Y = \beta) \\ &= E[X]E[Y] \end{aligned}$$

To compute the variance of $X + Y$ it is best to note that by Theorem 1, part 5, the variance is unchanged if we translate the the random variable. So we have for example $\mathrm{var}(X) = \mathrm{var}(X - E[X])$ and similarly for Y and $X + Y$. So without loss of generality we may assume that $E[X] = E[Y] = E[X + Y] = 0$. Then $\mathrm{var}(X) = E[X^2]$, etc...

$$\begin{aligned} \mathrm{var}(X + Y) &= E\left[(X + Y)^2\right] \\ &= E\left[X^2 + 2XY + Y^2\right] \\ &= E\left[X^2\right] + E\left[Y^2\right] + 2E\left[XY\right] \\ &= E\left[X^2\right] + E\left[Y^2\right] + 2E\left[X\right]E\left[Y\right] \quad (X, Y \text{ independent}) \\ &= E\left[X^2\right] + E\left[Y^2\right] \qquad (\text{since } E[X] = E[Y] = 0) \\ &= \mathrm{var}(X) + \mathrm{var}(Y) \end{aligned}$$

∎

The Law of Large numbers Suppose we perform an experiment and a measurement encoded in the random variable X and that we repeat this experiment n times each time in the same conditions and each time independently of each other. We thus obtain n independent copies of the random variable X which we denote

$$X_1, X_2, \cdots, X_n$$

Such a collection of random variable is called a IID sequence of random variables where IID stands for *independent and identically distributed*. This means that the random variables X_i have the same probability distribution. In particular they have all the same means and variance

$$E[X_i] = \mu\,, \quad \text{var}(X_i) = \sigma^2\,, \quad i = 1, 2, \cdots, n$$

Each time we perform the experiment n tiimes, the X_i provides a (random) measurement and if the average value

$$\frac{X_1 + \cdots + X_n}{n}$$

is called the *empirical average*. The Law of Large Numbers states for large n the empirical average is very close to the expected value μ with very high probability

Theorem 4. *Let* $X_1, \cdots, X_n$ *IID random variables with* $E[X_i] = \mu$ *and* $\text{var}(X_i)$ *for all* i*. Then we have*

$$P\left(\left|\frac{X_1 + \cdots X_n}{n} - \mu\right| \ge \epsilon\right) \le \frac{\sigma^2}{n\epsilon^2}$$

In particular the right hand side goes to 0 has $n \to \infty$.

Proof. The proof of the law of large numbers is a simple application from Chebyshev inequality to the random variable $\frac{X_1 + \cdots X_n}{n}$. Indeed by the properties of expectations we have

$$E\left[\frac{X_1 + \cdots X_n}{n}\right] = \frac{1}{n}E\left[X_1 + \cdots X_n\right] = \frac{1}{n}\left(E\left[X_1\right] + \cdots E\left[X_n\right]\right) = \frac{1}{n}n\mu = \mu$$

For the variance we use that the X_i are independent and so we have

$$\text{var}\left(\frac{X_1 + \cdots X_n}{n}\right) = \frac{1}{n^2}\text{var}\left(X_1 + \cdots X_n]\right) = \frac{1}{n^2}\left(\text{var}(X_1) + \cdots + \text{var}(X_n)\right) = \frac{\sigma^2}{n}$$

By Chebyshev inequality we obtain then

$$P\left(\left|\frac{X_1 + \cdots X_n}{n} - \mu\right| \ge \epsilon\right) \le \frac{\sigma^2}{n\epsilon^2}$$

■

Coin flip Suppose we flip a fair coin 100 times. How likely it is to obtain between 40% and 60% heads? We consider the random variable X which is 1 if the coin lands on head and 0 otherwise. We have $\mu = E[X] = 1/2$ and $\sigma^2 = \text{var}(X) = 1/4$ and by Chebyshev

$$
\begin{aligned}
P(\text{between 40 and 60 heads}) &= P\left(40 \leq X_1 + \cdots X_{100} \leq 60\right) \\
&= P\left(\frac{4}{10} \leq \frac{X_1 + \cdots X_{100}}{100} \leq \frac{6}{10}\right) \\
&= P\left(\left|\frac{X_1 + \cdots X_{100}}{100} - \frac{1}{2}\right| \leq \frac{1}{10}\right) \\
&= 1 - P\left(\left|\frac{X_1 + \cdots X_{100}}{100} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \\
&\geq 1 - \frac{1/4}{100(1/10)^2} = .75 \qquad (2)
\end{aligned}
$$

If we now flip a fair coin now 1000 obtainthe probability to obtain between 40% and 60% heads can be estimated by

$$
\begin{aligned}
P(\text{between 400 and 600 heads}) &= P\left(\left|\frac{X_1 + \cdots X_{1000}}{1000} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \\
&= 1 - P\left(\left|\frac{X_1 + \cdots X_{100}}{100} - \frac{1}{2}\right| \leq \frac{1}{10}\right) \\
&\geq 1 - \frac{1/4}{1000(1/10)^2} = .975 \qquad (3)
\end{aligned}
$$

The Monte-Carlo method: Under the name Monte-Carlo methods, one understands an algorithm which uses randomness and the LLN to compute a certain quantity which might have nothing to do with randomness. Such algorithm are becoming ubiquitous in many applications in statistics, computer science, physics and engineering. We will illustrate the ideas here with some very simple test examples.

Random number: A computer comes equipped with a *random number generator*, (usually the command **rand**, which produces a number which is uniformly distributed in $[0,1]$. We call such a number U and such a number is characterized by the fact that

$$P(U \in [a,b]) = b - a \quad \text{for any interval } [a,b] \subset [0,1]\,.$$

Every Monte-Carlo method should be in principle constructed with random number so as to be easily implementable. For example we can generate a $0-1$ random variable X with $P(X=1) = p$ and $P(X=0) = 1-p$ by using a random number. We simply set

$$X = \begin{cases} 1 & \text{if } U \leq p \\ 0 & \text{if } U > p \end{cases}$$

Then we have

$$P(X=1) \;=\; P(U \in [0,p]) \;=\; p\,.$$

An algorithm to compute the number π: To compute the number π we draw a square with side length 1 and inscribe in it a circle of radius 1/2. The area of the square of 1 while the area of the circle is $\pi/4$. To compute π we generate a random point in the square. If the point generated is inside the circle we accept it, while if it is outside we reject it. The we repeat the same experiment many times and expect by the LLN to have a proportion of accepted points equal to $\pi/4$

More precisely the algorithm now goes as follows

- Generate two random numbers U_1 and V_1, this is the same as generating a random point in the square $[0,1]\times[0,1]$.
- If $U_1^2+V_1^2 \le 1$ then set $X_1=1$ while if $U^2+V^2>1$ set $X_1=0$.
- Repeat to the two previous steps to generate $X_2, X_3, \cdots, X_n$.

We have

$$P(X_1=1) \;=\; P(U_1^2+V_1^2\le 1) \;=\; \frac{\text{area of circle}}{\text{area of the square}} \;=\; \frac{\pi/4}{1}$$

and $P(X=0)=1-\pi/4$. We have then

$$E[X] \;=\; \mu=\pi/4 \quad \mathrm{var}(X) \;=\; \sigma^2=\pi/4(1-\pi/4)$$

So using the LLN and Chebyshev we have

$$P\left(\left|\frac{X_1+\cdots X_n}{n}-\frac{\pi}{4}\right|\ge\epsilon\right) \;\le\; \frac{\pi/4(1-\pi/4)}{n\epsilon^2}$$

In order to get quantitative information suppose we want to compute π with an accuracy of $\pm 1/1000$, that is we take $\epsilon=1/000$. This is the same as computing $\pi/4$ with an accuracy of 1/4000. On the right hand side we have the variance $\pi/4(1-\pi/4)$ which is a number we don't know. But we note that the function $p(1-p)$ on $[0,1]$ has its maximum at $p=1/2$ and the maximum is 1/4 so we can obtain

$$P\left(\left|\frac{X_1+\cdots X_n}{n}-\frac{\pi}{4}\right|\ge 4/100\right) \;\le\; \frac{4,000,000}{n}$$

That we need to compute run the algorithm 80 millions times to make this probability 5/100.

The Monte-carlo method to compute the integral $\int_a^b f(x)\,dx$. We consider a function f on the interval $[a,b]$ and we wish to compute

$$I = \int_a^b f(x)\,dx\,.$$

for some bounded function f. Without loss of generality we can assume that $f \geq 0$, otherwise we replace f by $f+c$ for some constant c. Next we can also assume that $f \leq 1$ otherwise we replace f by cf for a sufficiently small c. Finally we may assume that $a=0$ and $b=1$ otherwise we make the change of variable $y = (x-a)/(b-a)$. For example suppose we want to compute the integral

$$\int_0^1 \frac{e^{sin(x^3)}}{3(1+5x^8)}\,dx$$

This cannot be done by hand and so we need a numerical method. A standard method would be to use a Riemann sum, i.e. we divide the interval $[0,1]$ in subinterval and set $x_i = \frac{i}{n}$ then we can approximate the integral by

$$\int f(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

that is we approximate the area under the graph of f by the sum of areas of rectangles of base length $1/n$ and height $f(i/n)$.

We use instead a Monte-Carlo method. We note that

$$I = \text{ Area under the graph of } f$$

and we construct a $0-1$ random variable X so that $E[X] = I$. We proceed as for computing π.

More precisely the algorithm now goes as follows

- Generate two random numbers U_1 and V_1, this is the same as generating a random point in the square $[0,1]\times[0,1]$.
- If $V_1 \leq f(U_1)$ then set $X_1 = 1$ while if $V_1 > f(U_1)$ set $X_1 = 0$.
- Repeat to the two previous steps to generate $X_2, X_3, \cdots, X_n$.

We have

$$P(X=1) = P(V \leq f(U)) = \frac{\text{Area under the graph of } f}{\text{Area of } [0,1]\times[0,1]} = I = \int_0^1 f(x)\,dx$$

and so $E[X] = I$ and $\mathrm{var}(X) = I(1-I)$.

Proportional Play and Kelly's criterion: Investing in the stock market is, in effect, making a series of bets. Contrary to bets in a casino though, one would generally believe that the stock market is on average *rising*, so we are making a series of *superfair bets*.

In order to make things concrete we imagine that we invest in the stock market by buying (idealized) *option contracts*. With such a contract you may pay a certain amount b to have the possibility to sell a certain stock at a certain price at a given date. If the market stock price is below your option price you will buy it on the market and sell it immediately at a profit. if the stock price is above your option price then you do nothing and you lose the amount you paid for your option. We will make it even simpler by assuming your option cost \$1 (this your bet), and you make a profit γ with probability p and lose your bet with probability $1-p$.

$$\boxed{P(\text{ Win } \$\gamma) \;=\; p \quad P(\text{ Lose } \$1) \;=\; q = (1-p)}$$

If the game is superfair then we assume

$$\boxed{\text{Expected gain } \mathbf{E}[\mathbf{W}] \;=\; \gamma\mathbf{p} - \mathbf{q} > \mathbf{0}}$$

Informed with our experience with the gambler's ruin we recall that is subfair game the best strategy is *bold play*, that is invest everything at each step. But if the game is superfair then our probability to win a certain amount gets larger if we make small bets. But if we buy options that are sold on a fixed time window and making very small bets won't make you much money or only very slowly. To

$$\boxed{\textbf{Proportional Play = Invest a proportion } 0 \le f \le 1 \textbf{ of your fortune at each bet.}}$$

Let us know computed your fortune after n bets. If you start with a fortune X_0 then you bet fX_0 and so your fortune at after the first bet is

$$X_1 \;=\; \begin{cases} X_0 + \gamma f X_0 & \text{if you win} \\ X_0 - \gamma X_0 & \text{if you lose} \end{cases} \;=\; \begin{cases} X_0(1+\gamma f) & \text{with probability } p \\ X_0(1-\gamma) & \text{with probability } q = 1-p \end{cases}$$

So at each step your fortune is *multiplied by a random factor* either $1+\gamma f$ or $1-f$, so we define the random variable

$Q = (1 + \gamma f)$ with probability p and $Q = (1 - \gamma)$ with probability q.

To find our fortune at time N we take

$$Q_1, Q_2, \cdots Q_N \text{ independent identically distributed random variables}$$

all with the same distribution as Q. Then we have

Fortune after N bets is $X_N = Q_N Q_{N-1} \cdots Q_2 Q_1 X_0$

Since we make a series of super fair bets we expect that our fortune to fluctuate but to increase exponentially

$$X_N \sim X_0 \exp \alpha N \, .$$

and so α represents the rate at which our fortune increases. We can compute the long run value of α by using the law of large numbers. Indeed we have

$$\alpha \sim \frac{1}{N} \log(X_N/X_0) \; = \; \frac{1}{N} \left[\log(Q_1) + \log(Q_2) + \cdots \log(Q_N)\right]$$

and so

$$\frac{1}{N} \left[\log(Q_1) + \log(Q_2) + \cdots \log(Q_N)\right] \longrightarrow E[\log(Q)] \equiv \alpha$$

and we have

$$P\left(X_0 e^{N(\alpha - \epsilon)} \le X_N \le X_0 e^{N(\alpha+\epsilon)}\right) \le \frac{\mathrm{var}(\log(Q))}{N\epsilon^2}$$

To find the best asset allocation we seek to maximize $\alpha = E[\log(Q)]$.

Optimal proportional play $\Longleftrightarrow$ maximize $E[\log(Q)]$

We have

$$E[\log(Q)] \; = \; p \log(1 + \gamma f) + q \log(1 - f)$$

Differentiating with respect to f we find

$$\begin{aligned} 0 &= \frac{d}{df} E[\log(Q)] \\ &= \frac{d}{df} p \log(1+\gamma f) + q \log(1-f) \\ &= \frac{p\gamma}{1+\gamma f} - \frac{q}{1-f} \end{aligned} \tag{4}$$

and so we have

$$p\gamma(1-f) = q(1+\gamma f)$$
$$p\gamma - q \;=\; f\gamma(p+q) \;=\; f\gamma$$

Finally we obtain

$$\textbf{Optimal f}: \qquad \mathbf{f^*} \;=\; \frac{\mathbf{p}\gamma - \mathbf{q}}{\gamma} \;=\; \frac{\textbf{Expected gain}}{\textbf{gain}} \qquad \textbf{Kelly's formula}$$

For example if your bet results in a payout of \$10 with probability 1/4 you should bet

$$f^* = \frac{10\frac{1}{4} - \frac{3}{4}}{10} = \frac{7}{40} = 0.175$$

of your fortune on each bet. Your fortune, on the long run, will grow at the rate of

$$\alpha = \frac{1}{4}\log(1 + 10\frac{7}{40}) + \frac{3}{4}\log(1 - \frac{7}{40}) \;=\; 0.047$$

Exercises:

Exercise 1: After the very confusing explanations of your professor in class about the Monty's Hall problem you decide to go home and to write a Monte-Carlo algorithm to check if the best strategy is to switch or not to switch. Explain your algorithm (your input should consist only of random numbers).

Exercise 2: Suppose that you run the Monte-Carlo algorithm to compute π 10'000 times and observe 7932 points inside the circle. What is your estimation for the value of π?

Using Chebyshev describe how accurate your estimation of π, the answer should be in the form should be in the form: based on my computation the number π belong to the interval $[a, b]$ with probability .95.

Exercise 3: I hand you a coin and make the claim that it is biased and that heads comes up with probability .48% of the times. You decide to flip the coin yourself a number of times to make sure that I was saying the truth. Use Chebyshev inequality to estimate how many times should you flip that coin to be sure that the coin is biased with a probability of .95.

Exercise 4: Suppose you start with a fortune of 1 and throw a coin n times. A very rich and not very bright friend of yours pays you $\frac{11}{5}k$ for a bet of k if the throw is heads and nothing if the throw is tail.

1. Suppose you bet everything on each throw. What is your expected gain after 20 throws? What is the probability that you win nothing after 20 throws? What is the variance of your gain after 20 throws?
2. What does Kelly's formula suggest you should do? What is your expected gain and its variance after 20 throws in that case?

Exercise 5: Consider the proportional play strategy, but now every single bet of 1 unit leads to three possible outcomes

$$P(\text{ Win } 4) = 1/2 \quad P(\text{ Lose } 1) = 1/4 \quad P(\text{ Lose } 4) = 1/4\,.$$

What is the optimal proportion f^* of your fortune you should invest?