STATISTICS

Module 2 *of*
MATHEMATICS AND STATISTICS

Marina Santacroce

marina.santacroce@unibocconi.it
Department of Decision Sciences, Bocconi University

L7. Random variables.

Outline

Random variables: introduction

In many scenarios, we are interested in the value of some numerical quantity determined by the outcome as opposed to the actual outcome.

- We are interested in the sum of two dice and not in the separate values of each die;
- when we flip a coin n times, we want to know the number of tails and not the exact sequence;
- when we flip the coin until a tail occurs, we are interested in the number of times the coin has been flipped;
- when n balls are randomly chosen without replacement from an urn with m red and $N - m$ blue balls and you win 1$ for each red ball, we are interested only in the gain i.e. in the number of red balls selected;
- when your car has several things that need to be fixed and you just care of the total cost of the repair and not on the cost related to the different damages.

In each of the previous cases we can think to a (real-valued) "function" of the outcome called $\Rightarrow$ *random variable*.

Most of these examples are related to well known *random variables*, which we will study soon.

Example

Suppose Ugo has a subject to pass and each time he sits the exam the probability of passing is p. If he fails, he sits the exam again. We suppose these trials are independent and we denote by X the number of exams he sits before passing the subject.
We have

- $\mathbb{P}(X = 1) = \mathbb{P}(\text{Pass}) = p$,
- $\mathbb{P}(X = 2) = \mathbb{P}(\text{Fail, Pass}) = (1 - p)p$,
- ....
- $\mathbb{P}(X = n) = \mathbb{P}(\text{Fail,..., Fail (n-1 times), Pass}) = (1 - p)^{n-1}p$.

One can check that

$$\sum_{k:1}^{\infty} \mathbb{P}(X = k) = \sum_{k:1}^{\infty} (1 - p)^{k-1} p = 1$$

Probability distribution function

Def. Consider a sample space S, a family of events and a probability $\mathbb{P}$, a *random variable* X is a function $X : S \to \mathbb{R}$ such that, for any $B \subset \mathbb{R}$, $\{X \in B\} = \{s : X(s) \in B\}$ is an event [1].

We associate to the event $\{X \in B\}$ its probability $\Rightarrow$ $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$.

- $\mathbb{P}_X(\cdot)$ is a *probability measure on* $\mathbb{R}$ and is called *law* or *distribution* of X.

Def. The *probability distribution function* of a random variable X is a function $F : \mathbb{R} \to [0; 1]$ defined by

$$F(x) = \mathbb{P}(X \leq x), \qquad -\infty < x < \infty.$$ [2]

The next properties follows from the definition:

Prop. The probability distribution function F of a random variable X is

1. non decreasing;
2. right-continuous, i.e. if $\{x_n, n \geq 1\}$ is a decreasing sequence $x_n \downarrow x$, then $\lim_{n\to\infty} F(x_n) = F(x)$;
3. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

[1] B is to be in the Borel σ-algebra of $\mathbb{R}$

[2] $F(x)$ is the law of X computed on the sets $B = (-\infty; x]$

Discrete Random Variables

Def. A random variable X that can take on *at most* a countable number of possible values is said to be **discrete**.

- The probability mass function of a discrete random variable X is denoted by
$$p(a) = \mathbb{P}(X = a)$$
- If X can take on the values $\{x_1, x_2, ...\}$ then
$$p(x_i) \geq 0 \quad \text{and} \quad \sum_{i:1}^{\infty} p(x_i) = 1.$$

$\Rightarrow$ Note the mass probability function p characterizes the *law of X*

$$\Rightarrow \mathbb{P}_X(B) = \mathbb{P}(X \in B) = \sum_{x_i \in B} p(x_i)$$

The *probability distribution function* of a *discrete* random variable X is a *step function* defined by

$$F(x) = \sum_{x_i \leq x} p(x_i), \qquad -\infty < x < \infty.$$

Expected Value

Def. If X is a discrete random variable having probability mass function $p(x)$, then the expected value or expectation of X[3] is defined as

$$\mathbb{E}(X) = \sum_{x_i} x_i p(x_i)$$

Example: Assume you have a fair die and X is the number rolled then

- $p(i) = \frac{1}{6}; \quad i : 1, \dots, 6.$

$$\mathbb{E}(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

[3] For the existence we will suppose $\mathbb{E}(|X|) = \sum_{x_i} |x_i| p(x_i) < \infty$

Example: Tramway in Melbourne

In Melbourne (Australia), no one collects fare when you board a tram and no one checks systematically whether you have a ticket. Only occasionally, inspectors arrive to check tickets and give a fine to any ticketless riders.
A ticket costs 2$, a fine is 30$ and the probability of meeting inspectors 5%.
Assume you are interested in minimizing the average cost,
$\Rightarrow$ *should you buy tram tickets or not?*

Let C be the cost.

- $\Rightarrow$ If you buy a ticket each time you board a tram then the cost is C = 2$ with probability 1. $\Rightarrow \mathbb{E}(C) = 2 \times 1 = 2$$
- $\Rightarrow$ If you never buy tram tickets then the cost is
  - $C = 0$$ if you don't get caught $\Rightarrow \mathbb{P}(C = 0) = 0.95$
  - $C = 30$$ if you get caught $\Rightarrow \mathbb{P}(C = 30) = 0.05$.

  Thus,
  $$\Rightarrow \mathbb{E}(C) = 0 \times 0.95 + 30 \times 0.05 = 1.50\$$$

So as to minimize the average cost, do not buy tram tickets in Melbourne!

Example: Who wants to be a millionaire?

You have won \$64,000 and are facing a tough \$125,000 question. This means that you will receive

- \$125,000 if you answer the question correctly.
- only \$32,000 if your answer is wrong.
- \$64,000 if you decide not to answer the question.

What would be the best strategy to take?

The \$125,000 question is: *Which of the following movies did not win the Palme d'Or at the Cannes Film Festival*?

A- Parasite B- Pulp fiction

C- Underground D- Roma

If you answer, what is the expected value of the outcome? Would you answer the question or would you walk out with \$64,000?

If you have no idea and you guess, the expected value of your outcome

$$\frac{1}{4}125,000 + \frac{3}{4}32,000 = 55,250 < 64,000$$

whereas if you walk out now you have \$ 64,000.

It seems perhaps more sensible to walk out although you might like taking risks and stay in.

Example: Who wants to be a millionaire?(BIS)

Now assume you can use the 50:50 option to get rid of two possible answers, so that you are left with the two options

A- Parasite D- Roma

Suppose that you decide to answer the question. What is the expected value of the outcome? Would you answer the question or would you walk out with \$64,000?

The expected value is now

$$\frac{1}{2} \times 125,000 + \frac{1}{2} \times 32,000 = 78,500 > 64,000.$$

It might be reasonable to answer although a risk averse person will definitely not answer!

Utility

- In the previous examples, we have discussed cases where we compute the expected cost or price under different situations so as to make a decision: buy/don't buy a ticket, answer/don't answer question.
- But different people would behave in different ways in those situations. Why?

$\Rightarrow$ People have typically different utility functions and they maximize the expected utility of the cost/price, i.e.

$$\text{maximize} \qquad \mathbb{E}(U(X))$$

- a typical choice for a risk averse person is $U(x) = \log x$.
- Different people have typically different utility functions; e.g. some people will always buy a tram ticket!

How to compute the expectation $\mathbb{E}(U(X))$?

Expectation of a Function of a Random Variable

Given a discrete random variable X, $Y = g(X)$ is a *new* random variable.

From the probability mass function (pmf) of X, one can obtain the pmf of Y and use it to compute the expected value.

Example: Assume you have X taking values in $\{-2, -1, 0, 1, 2\}$ and consider $Y = X^2$. Clearly Y takes values in $\{0, 1, 4\}$ and we have

$$\begin{aligned}
\mathbb{P}(Y = 4) &= \mathbb{P}(X = 2) + \mathbb{P}(X = -2) \\
\mathbb{P}(Y = 1) &= \mathbb{P}(X = 1) + \mathbb{P}(X = -1) \\
\mathbb{P}(Y = 0) &= \mathbb{P}(X = 0)
\end{aligned}$$

The expected value of Y is

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) + 4 \cdot \mathbb{P}(Y = 4)$$

Prop. If X is a discrete random variable and g is a real valued function, the expected value of $g(X)$, if it exists, is given by

$$\mathbb{E}[g(X)] = \sum_{x_i \in \mathbb{X}} g(x_i) p(x_i).$$

Indeed,

$$\mathbb{E}(Y) = \sum_{y_j} y_j \mathbb{P}(Y = y_j) =$$

$$\sum y_j \mathbb{P}(Y = g(X) = y_j) = \sum_{y_j} y_j \sum_{x_i : g(x_i) = y_j} p_X(x_i) =$$

$$\sum_{y_j} \sum_{x_i : g(x_i) = y_j} y_j p_X(x_i) =$$

$$\sum_{x_i} g(x_i) p_X(x_i).$$

Prop. (*Linearity*)

- $\mathbb{E}[aX + b] = a\mathbb{E}(X) + b,$ where a and $b \in \mathbb{R}$
- f and g be two real valued function

$$\mathbb{E}(f(X) + g(X)) = \mathbb{E}(f(X)) + \mathbb{E}(g(X))$$

Variance

- A random variable X is entirely defined by its p.m.f. but it is very useful to be able to summarize the essential properties of it through a few suitably defined measures.

We have defined the *expectation*. Consider the following random variables

- $X = 0$ with probability 1,
- $Y = 1$ and $Y = -1$ with probability 0.5
- $Z = 100$ and $Z = -100$ with probability 0.5.

Then

$$\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$$

We observe a much greater variability in the values of Y than in X (which is deterministic) and the same is true comparing Z to Y.

In general, we would like the *expectation* to be representative of X, so it would make sense to *quantify the average deviations* from it. This is what the variance achieves.

Variance

Def. If X is a random variable with expected value $\mathbb{E}(X)$, then the *variance* of X is defined by

$$\mathrm{Var}(X) = \mathbb{E}[(X - E(X))^2].$$

We have

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Rem. $\mathrm{Var}(X) \geq 0$;

- ◇ Verify that for a and $b \in \mathbb{R} \Rightarrow \mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.
- $\mathbb{E}(X^n)$ is said n-th moment (or moment of order n) of X.
- The variance is the *centered second moment of* X.
- $\sigma_X = \sqrt{\mathrm{Var}(X)}$ is said *standard deviation* of X.
- **Rem.** (*Standardized random variable*) Let $Y = \frac{X - \mathbb{E}(X)}{\sigma_X}$.
  - ◇ Verify that Y is a random variable with null expectation and variance 1.

Moment Generating Function

Def. The moment generating function (m.g.f.) of a random variable X for a given $t \in \mathbb{R}$, is defined[4] as

$$\phi_X(t) = \mathbb{E}[e^{tX}] = \sum_{x \in \mathbb{X}} e^{tx} p_X(x)$$

Note that $M_X(0) = 1$.

Prop. A moment generating function completely determines the distribution of a random variable X.

Prop. The k-th moment of a random variable X can be found by evaluating the k-th derivative of the m.g.f. of X at $t = 0$.

$$\phi_X^{(k)}(t) = \mathbb{E}[X^k].$$

In particular, if they exist, we can compute expectation and variance

$$\mathbb{E}[X] = \phi'_X(0) \qquad \text{and} \qquad \mathrm{Var}(X) = \phi''_X(0) - (\phi'_X(0))^2.$$

[4] where there is a positive number h such that it exists for $-h < t < h$