Statistical Methods for Economics

Andrea Tancredi
Sapienza University of Rome

March 2, 2020
Probability

- Let \( \Omega \) be a set \( \mathcal{F} \) a collection of subsets of \( \omega \). A *probability measure*, or simply a *probability* on \((\Omega, \mathcal{F})\) is a function

\[
P : \mathcal{F} \to [0, 1]
\]

such that

- To be a probability \( P \) must satisfy
  1. \( \forall A \in \mathcal{F} \ 0 \leq P(A) \leq 1 \)
  2. \( P(\Omega) = 1 \)
  3. if \( A_1 \) e \( A_2 \) are disjoint then \( P(A_1 \cup A_2) = P(A_1) + P(A_2) \)
• We can show that property 3 holds for any finite collection of disjoint sets

\[ P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \]

• It is common practice to assume that 3 hold for countable collections of disjoint sets

\[ P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \]
• In practical examples $\Omega$ is the set of outcomes of an experiment and $P$ is determined by experience, logic or judgement

• Rolling a six-sided die, the set of outcomes is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
  
  ▶ If we believe the die to be fair we may assign

  $$P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = \frac{1}{6}$$

  ▶ The laws of probability then imply, e.g.

  $$P\{1, 2\} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \quad P\{1, 2, 3\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

• Setting $P\{i\} = \frac{1}{6}$ is not automatic simply because a die a six faces, we have to believe the die to be fair. For example we could also have

  $$P\{1\} = \frac{1}{2} \quad P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = \frac{1}{10}$$
• Equivalent statements
  ▶ The probability that the die lands 1
  ▶ \( P(1) \)
  ▶ \( P([\text{the die lands } 1]) \)
  ▶ \( \mu\{1\} \)

• We also use the word distribution in place of probability measures
Properties of the probability

- \( P(A) = 1 - P(\bar{A}) \)
- If \( A \subseteq B \) \( P(A) \leq P(B) \)
- For general set \( A \) and \( B \)

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B)
\]
The game of craps

- Craps is a gambling game played with two dice

  ▶ For the dice thrower (shooter) the object of the game is to throw a 7 or an 11 on the first roll (a win) and avoid throwing a 2, 3 or 12 (a loss).

  ▶ If none of these numbers (2, 3, 7, 11 or 12) is thrown on the first throw (the Come-out roll) then a Point is established (the point is the number rolled) against which the shooter plays.

  ▶ The shooter continues to throw until one of two numbers is thrown, the Point number or a Seven.

  ▶ If the shooter rolls the Point before rolling a Seven he/she wins, however if the shooter throws a Seven before rolling the Point he/she loses.
• Ultimately we would like to calculate

$$P(\text{shooter wins})$$

but for now let’s calculate

$$P(\text{shooter wins on Come-out roll}) = P(7 \ or \ 11)$$

• The set of possible outcomes is the set of ordered pairs $$(d_1, d_2)$$ where $d_1 \in \{1, 2, 3, 4, 5, 6\}$ is the outcome of the first die and $d_2 \in \{1, 2, 3, 4, 5, 6\}$ is the outcome of the second die,

$$
\begin{align*}
(6, 6) & \quad (6, 5) & \quad (6, 4) & \quad (6, 3) & \quad (6, 2) & \quad (6, 1) \\
(5, 6) & \quad (5, 5) & \quad (5, 4) & \quad (5, 3) & \quad (5, 2) & \quad (5, 1) \\
(4, 6) & \quad (4, 5) & \quad (4, 4) & \quad (4, 3) & \quad (4, 2) & \quad (4, 1) \\
(3, 6) & \quad (3, 5) & \quad (3, 4) & \quad (3, 3) & \quad (3, 2) & \quad (3, 1) \\
(2, 6) & \quad (2, 5) & \quad (2, 4) & \quad (2, 3) & \quad (2, 2) & \quad (2, 1) \\
(1, 6) & \quad (1, 5) & \quad (1, 4) & \quad (1, 3) & \quad (1, 2) & \quad (1, 1)
\end{align*}
$$
• If the dice are fair, then the pairs are all equally likely and 
\( P(d_1, d_2) = 1/36 \) since there are 36 pairs.

• By the probability law

\[
P(7 \text{ or } 11) = P(7 \cup 11) = P(7) + P(11)
\]

and

\[
P(7) = P((6, 1) \cup (5, 2) \cup (4, 3) \cup (3, 4) \cup (2, 5) \cup (1, 6))
\]
\[
= P(6, 1) + P(5, 2) + P(4, 3) + P(3, 4) + P(2, 5) + P(1, 6)
\]
\[
= 6/36
\]

\[
P(11) = P((6, 5) \cup (5, 6)) = P(6, 5) + P(5, 6) = 2/36
\]

so that \( P(7 \text{ or } 11) = 6/36 + 2/36 = 8/36 = 2/9 \)
Random variables

- A random variable (r.v.) is a numerical expression of the outcome of a statistical experiment or more generally the numerical manifestation of a phenomenon that can have different outcomes.

- Formally, given $\Omega$ and a probability $P$ on $\Omega$,
  
  - A r.v. $X$ is a function $X(\omega)$ defined on $\Omega$ and taking values in $\mathbb{R}$.
  
    $$X : \Omega \rightarrow \mathbb{R}$$

  - And $\forall B \subseteq \mathbb{R}$
    
    $$P(X \in B) = P(\omega \in \Omega : X(\omega) \in B)$$

- Random variables will be generally indicated with the letters $X, Y, Z...$
• A r.v. that may assume only a finite number or an infinite sequence of values is said to be discrete;

• A r.v. that may assume any value in some interval on the real number line is said to be continuous.

• For instance, a random variable representing the number of new cases of COVID-19 on one day would be discrete, while a random variable representing the weight of a person in kilograms would be continuous.

• Very often we work directly with random variables without knowing (or caring to know) the underlying probability $P$ on the space $\Omega$

• In fact we will specify (model) directly the probabilities of the outcomes of the r.v.
The (cumulative) distribution function $F_X(x)$ of $X$ is

$$F_X(x) = P(X \leq x) \quad x \in \mathbb{R}$$

Note that

- $F_X(-\infty) = 0$
- $F_X(\infty) = 1$
- $x < x' \Rightarrow F(x) \leq F(x')$
• By the distribution function and the rules of probability we can obtain other probabilities on the r.v. $X$, e.g.

$$P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$$

and since for $a < b$, $(-\infty, b] = (-\infty, a] \cup (a, b]$

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

and

$$P(a < X \leq b) = F_X(b) - F_X(a)$$
Discrete random variables

• A discrete random variable takes value only in some countable subset $D$ of $\mathbb{R}$

• Commonly this subset $D$ is a subset of the integers

• The probability that $X$ takes some given value $x$ in $D$ is

$$p(x) = P(X = x) = P(\omega \in \Omega : X(\omega) = x)$$

• The function $p(x)$ is be called *probability distribution of $X$* or probability function

• **Example:** Suppose we toss an unbiased coin 2 times in succession. What is the probability of obtaining $x$ heads ($x = 0, 1, 2$)? Let $X$ be the r.v. describing the result of such experiments. The probability function is

<table>
<thead>
<tr>
<th></th>
<th>$x$</th>
<th>$Pr(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T,T)</td>
<td>0</td>
<td>1/4</td>
</tr>
<tr>
<td>(T,H), (H,T)</td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>(H,H)</td>
<td>2</td>
<td>1/4</td>
</tr>
</tbody>
</table>
• Note that

\[ p(x) \in [0, 1] \]

and

\[ \sum_{x \in D} p(x) = 1 \]

In fact,
\[ \Omega = \bigcup_x \{ \omega : X(\omega) = x \} \text{ and } \{ \omega : X(\omega) = x \} \cap \{ \omega : X(\omega) = x' \} = \emptyset \]

otherwise we would have that \( \exists \omega : X(\omega) = x \) and \( X(\omega) = x' \) which is impossible, then

\[
1 = P(\Omega) = P\left( \bigcup_x \{ \omega : X(\omega) = x \} \right) = \sum_{x \in D} P(\{ \omega : X(\omega) = x \}) = \sum_{x \in D} p(x)
\]

• Moreover

\[ F(x) = Pr(X \leq u) = \sum_{u \leq x} p(u) \]
Cumulative distribution function for the previous example
Famous probability distributions

- \( X \sim \text{Bernoulli}(p) \) for \( 0 \leq p \leq 1 \)
  
  \[
p(1) = p \quad \text{and} \quad p(0) = 1 - p
  \]

- \( X \sim \text{Geometric}(p) \) for \( 0 \leq p \leq 1 \)
  
  \[
p(x) = p(1 - p)^{x-1} \quad \text{for} \ x = 1, 2, \ldots
  \]

This r.v. represents, for example, the number of coin flips until the first head shows up (assuming independent coin flips)

![Probability distribution function for a Geometric r.v. with \( p = 0.5 \)]
\begin{itemize}
  \item $X \sim \text{Binomial}(n, p)$ for $n > 0$ and $0 \leq p \leq 1$
  
  $$p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots n$$

  The binomial r.v. represents, for example, the number of heads in $n$ independent coin flips

  Probability distribution function for a Binomial r.v. with $n = 10$ and $p = 0.5$
\end{itemize}
• $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{for } x = 0, 1, 2, \ldots$$

The Poisson r.v. often represents the number of random events, e.g. number of customers, email, COVID-19 cases etc., in some time interval

Probability distribution function for a Binomial r.v. with $n = 10$ and $p = 0.5$
Continued random variables

- A continuous r.v. can take values on an interval, either of finite or infinite length
  - Medical trials: the time until a patient experience a relapse or the time until healing
  - Economics: the income of a family
  - Health economics: the cost of a treatment

- Since the elements $x$ of a real interval are uncountable, for a continuous r.v. we must have $P(X = x) = 0$

- Formally, a r.v. $X$ is continuous if $\forall B \subset \mathbb{R}$

\[
P(X \in B) = \int_B f_X(x) \, dx
\]

for some function $f_X(x)$ that will be called probability density function or simply density
• Every density function satisfy the following two properties
  ▶ $f_X(x) \geq 0$
  ▶ $\int_{-\infty}^{\infty} f_X(x) = 1$

• In fact if $f_X(x) < 0$ on the interval $(a, b)$ then $P(X \in (a, b)) = \int_a^b f_X(x)dx < 0$ and we can’t have probabilities less than 0

• Moreover $1 = P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f_X(x)dx$

• Note that we effectively have $P(X = a) = 0 \forall a \in \mathbb{R}$
  ▶ In fact $P(X = a) = \lim_{\epsilon \to 0} P(X \in [a, a+\epsilon]) = \lim_{\epsilon \to 0} \int_a^{a+\epsilon} f_X(x)dx = 0$
Note that in physics

\[ \text{density} = \frac{\text{mass}}{\text{volume}} \]

but the concept is similar in statistics. In fact let
\( I(x) = [x - \epsilon, x + \epsilon] \) be a small interval

\[
P(X \in I(x)) = \int_{x-\epsilon}^{x+\epsilon} f_X(x) \, dx \approx f_X(x) \times \text{length } I(x)
\]

Then
\[
f_X(x) \approx \frac{P(X \in I(x))}{\text{length } I(x)}
\]

in both the case the density varies: in physics within the object in statistics along the line.
• The integral between $-\infty$ and $x$ of the density is the cumulative distribution function. In fact

$$F_X(x) = P(X \leq x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} f_X(t)dt =$$

• The derivative of the distribution function is the density

$$\frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^{x} f_X(t)dt =$$

$$\frac{d}{dx} [F(x) - F(-\infty)] = \frac{d}{dx} F(x) = f_X(x)$$
Famous continuous random variables

- Uniform: $X \sim \text{Unif}(a, b)$ for $b > a$ has density function

$$f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{for } a \leq x \leq b \\
0 & \text{otherwise}
\end{cases}$$

10 realizations from a Uniform(2,5) in R

```r
runif(n=10, min=2, max=5)
```

```r
```

The cumulative distribution function is

$$F_X(x) = \begin{cases} 
0 & \text{for } x < a \\
\frac{x-a}{b-a} & \text{for } a \leq x \leq b \\
1 & \text{for } x > b
\end{cases}$$
• Exponential: $X \sim \text{Exp}(\lambda)$ for $\lambda > 0$ has density function

$$f_X(x) = \begin{cases} 
\lambda e^{-\lambda x} & \text{for } x > 0 \\
0 & \text{otherwise}
\end{cases}$$

10 realizations from a $\text{Exp}(\lambda)$ in R

```r
rexp(n=10, rate=5)
```

```r
# [1] 0.137717767 0.315281628 0.006778736  
# [4] 1.157800351 0.160485442 0.313788585  
# [7] 0.264082466 0.016791116 0.195571822  
# [10] 0.182311110
```

The cumulative distribution function is

$$F_X(x) = \begin{cases} 
0 & \text{for } x < 0 \\
1 - e^{-\lambda x} & \text{for } x \geq 0
\end{cases}$$
Normal: \( X \sim N(\mu, \sigma^2) \) for \(-\infty < \mu < \infty\) and \(\sigma > 0\) has density function

\[
 f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma^2} \right)}
\]

10 realizations from a \( N(1, 4) \) in R

```r
rnorm(n=10,mean=2,sd=2)
```

##

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.71</td>
<td>0.01</td>
<td>2.90</td>
</tr>
<tr>
<td>4</td>
<td>5.08</td>
<td>-0.71</td>
<td>4.14</td>
</tr>
<tr>
<td>7</td>
<td>2.53</td>
<td>4.70</td>
<td>1.07</td>
</tr>
<tr>
<td>10</td>
<td>5.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The cumulative distribution function cannot be obtained analytically
Transformations of random variables

**Example.** Let $U$ be a r.v. uniformly distributed on $(0, 1)$ and define

$$Y = -\frac{1}{\lambda} \log U$$

$$F_Y(y) = P(Y \leq y) = P \left( -\frac{1}{\lambda} \log U \leq y \right) = P(\log U > -\lambda y)$$

$$= P(U > e^{-\lambda y}) = 1 - P(U \leq e^{-\lambda y}) = 1 - F_U(e^{-\lambda y})$$

Since $F_U(u) = u \; \forall u \in [0, 1]$

$$F_Y(y) = 1 - e^{-\lambda y}$$

that is $Y \sim Exp(\lambda)$
Position and scale transformation Let $X$ be a continuous r.v. with density $f_X(x)$ and cumulative distribution function $F_X(x)$. The density of $Z = a + bX$ is

$$f_Z(z) = f_X\left(\frac{z - a}{b}\right) \frac{1}{b}.$$ 

In fact

$$F_Z(z) = P(Z \leq z) = P(a + bX \leq z) = P\left(X \leq \frac{z - a}{b}\right) = F_X\left(\frac{z - a}{b}\right)$$

and the density of $Z$ is

$$f_Z(z) = F'_Z(z) = f_X\left(\frac{z - a}{b}\right) \frac{1}{b}.$$
• Example. Let $X$ be a $N(0, 1)$ and define

$$Z = X^2$$

$$F_Z(z) = P(Z < z) = P(X^2 < z) = P(-\sqrt{z} < X < \sqrt{z})$$

$$= 2[F_X(\sqrt{z}) - F_X(0)]$$

then

$$f_Z(z) = \frac{d}{dz}F_Z(z) = 2f_X(\sqrt{z})\frac{d}{dz}\sqrt{z}$$

$$= (2/\sqrt{2\pi})e^{-(1/2)(\sqrt{z})^2}(1/2)z^{-1/2}$$

$$= (1/\sqrt{2\pi})e^{-(1/2)z}z^{-1/2}$$

The density $f_Z(z)$ is the density of the $\chi^2$ random variable.
A general result. Let $X$ a random variable with density $f_X(x)$. Let $g$ be a differentiable, monotonic, invertible function and define $Z = g(X)$. Then the density of $Z$ is

$$f_Z(z) = f_X(g^{-1}(z)) \left| \frac{dg^{-1}(z)}{dz} \right|$$

In fact, if $g$ is an increasing function

$$f_Z(z) = \frac{d}{dz} P[Z \in (a, z)] = \frac{d}{dz} P[X \in (g^{-1}(a), g^{-1}(z))]$$

$$= \frac{d}{dz} \int_{g^{-1}(a)}^{g^{-1}(z)} f_X(x) dx = f_X(g^{-1}(z)) \left| \frac{dg^{-1}(t)}{dt} \right|_{t=z}$$
Let $X$ be a r.v. with density $f_X(x) = 2x$ on $[0, 1]$ and 0 otherwise. Let $Z = 1/X$. The inverse transformation is $g^{-1}(z) = 1/z$. Its derivative is $dg^{-1}(z)/dz = -1/z^2$ and

$$f_Z(z) = f_X(g^{-1}(z)) - \frac{dg^{-1}(z)}{dz} = \frac{2}{z} - \frac{1}{z^2} = \frac{2}{z^3}$$