Lecture 2 – Probability Distributions

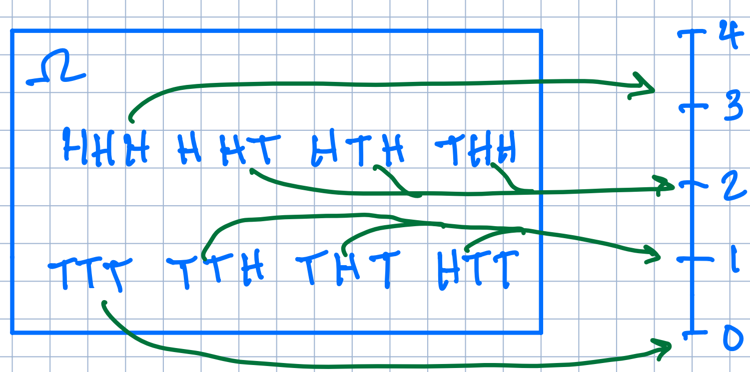
A. Random Variables

It is often inconvenient to directly deal with probability spaces and event sets. Random variables help us solve this problem.

Random variables are a mapping (i.e. a function) for assigning certain points in an event space Ω numerical values. This facilitates defining probabilities.

Example: Consider an event space Ω for a sequence of three coin tosses.

X = number of heads



$X(\omega)$ is a random variable

$$X(\omega) := \begin{cases} 0, & \omega = TTT \\ 1, & \omega \in \{TTH, THT, HTT\} \\ 2, & \omega \in \{HHT, HTH, THH\} \\ 3, & \omega = HHH \end{cases}$$

$P(X=0) = 1/8$

$P(X=1) = 3/8$

$P(X=2) = 3/8$

$P(X=3) = 1/8$

$$P(X=a) = P(\omega \in \Omega : X(\omega) = a)$$

This is a PMF

Random variables generally come in two varieties: discrete and continuous.

discrete: the values of $X(\omega)$ are distinct real numbers, usually integers

continuous: the values of $X(\omega)$ are distributed across the real number line

B. Probability distribution functions

Now that we have random variables, we can define functions of those R.V.s that correspond to probabilities.

(i) Probability mass function (pmf)

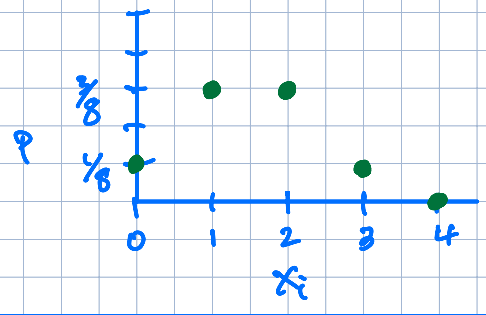
If X is discrete, the function we define is a probability mass function (pmf):

$$p(x_i) := P(X = x_i)$$ probability that RV. X equals x_i

A pmf must satisfy the properties:

$$0 \le p(x_i) \le 1 \quad , \quad \sum_i p(x_i) = 1$$

Example: PMF for the coin toss example above



(ii) Probability density functions (pdf)

For a continuous RV, we instead have a probability density function (pdf):

$$P(a \leq X \leq b) = \int_a^b f(t)\,dt$$

A pdf must satisfy these properties:

$$f(x) \geq 0 \quad , \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

"Paradox" for pdfs: with continuous variables, the probability at $x=a$ is always zero:

$$P(x=a) = \int_a^a f(x)\,dx = 0$$

This is a little disconcerting. In reality, we are always interested in some range:

$$P(a \leq x \leq a+\varepsilon) = \int_a^{a+\varepsilon} f(x)\,dx \neq 0$$

(iii) Cumulative distribution function (cdf)

Another type of function is also useful for continuous RVs, called the cumulative distribution function (cdf) or sometimes just the "distribution" function:

$$P(X \leq x) := F(x)$$

The cdf obeys these properties:

$$F(-\infty)=0\,,\quad F(+\infty)=1$$

$$\text{if } x_1 < x_2 \text{ then } F(x_1) \le F(x_2)$$ F is monotonic

$$P(a < x \le b) = F(b) - F(a)$$

Aside: there is no function called the "probability distribution function." This is ambiguous. I prefer to use this term for all of them. Instead, use pdf (density), cdf, or pmf where appropriate.

(iv) Relationships between the pmf, pdf, and cdf

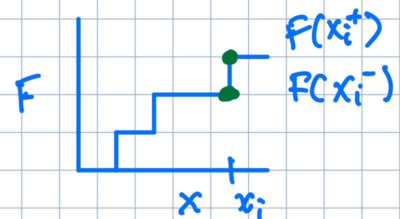
It is helpful to understand the relationships between these different functions.

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad f(x) = \frac{dF}{dx}$$ continuous

$$F(x) = \sum_{i}^{x_i \le x} p(x_i) \qquad p(x_i) = F(x_i^+) - F(x_i^-)$$ discrete

$$f(x) = \sum_i p(x_i)\,\delta(x-x_i)$$

Dirac delta function



C. Examples of pmfs, pdfs, and cdfs

There are many examples of important distributions that arise in stat thermo and in the physical sciences more generally. See the accompanying Jupyter notebook for several plots and some discussion.

Discrete Examples (pmfs)

- Uniform
- Poisson
- Bernoulli
- Geometric
- Binomial

Continuous Examples (pdfs & cdfs)

- uniform
- exponential
- Gaussian / normal