MM Applied Cheat Sheet – AS Stats

Notations, Modelling, Rules and Reminders

Probability $\mathbb{P}(X)$; Expectation $\mathbb{E}(X)$; Variance $Var(X)$

$X ;= \cdots$: definition not property

Σ: summation operator

Π: product operator

$X \sqcup Y$: X and Y independent

$X \sim \cdots$: X follows …

$\overline{X}$: mean; $\hat{X}$: estimator (unknown)

$$\sum_{x=1}^{n} x = \frac{x(x+1)}{2}$$

$$\sum_{x=1}^{n} x^2 = \frac{x(x+1)(2x+1)}{6}$$

- Draw Venn diagram for set theory.
- Draw probability tree of difficult probabilities.
- $x \in [a, b)$ means include a exclude b in range
- $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ can be written as $\frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$

5.1 *Introduction to Continuous Random Variables*

To find probability of RV which can take uncountably infinite possible values, area can be found by definition each value not as a line on the graph (where $\int f(x) = 1$ as total prob ability has to =1) but as a strip of width δx. This way adding up all strips can give width of 1, whereas points don't have width and thus add up to 0.

Hence:

- P(s ≤ X ≤ t) = P(X ≤ t) - P(X ≤ s) = F(t) - F(s)

c.d.f. (cumulative distribution function) [area under curve]

To find probability of $X \le x$

$$F_X(x) = \mathbb{P}(X \le x)$$
$$x \to \infty: F_X(x) = 1$$
$$x \to -\infty: F_X(x) = 0$$

- Continuous RV is where X satisfies $F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(u)\,du$
  1. $f_X(x) \ge 0 \qquad \forall x \in \mathbb{R}$
  2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- Remember this is cumulative s.t. you have to add previous range's c.d.f

p.d.f. (probability density function) [equation of curve]

- Not a probability
- Is equation of curve
- Find c.d.f. by integrating p.d.f. at range b to a where $x \in [a, b)$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

At any point for which $f_X(x)$ is continuous

N.B.: for a point, p.d.f. is true as it lies on the curve (ie. equation of curve is always true for any point), but the point does not contribute to c.d.f. as it has no width.

Just the Tip™: p.d.f. is easier to find as it is the equation of the line

N.B.: for c.d.f. ranges are $a \le x < b$; for p.d.f. ranges are $x \in [a, b)$

For continuous RV with p.d.f. $f_X(x)$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

Where $\int_{-\infty}^{\infty} |x| f_X(x)\,dx$ exists

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

Where $\mu = \mathbb{E}(X)$

s.t. $Var\,(X) = \mathbb{E}(X^2) - \mu^2$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \mu^2$$