Primer on Probability and Statistics

Goal: Quantitative characterization of uncertainty associated with the occurence of random events. (UQ: uncertainty quantificati

- frequentist (statistics) : UQ based on long sample frequencies
- Bayesian (Thomas Bayes ~1700) : UQ based on the rules of probability theory

• Discrete random variables : (events that take discrete values)

$p(A)$ denotes the probability that the event A is true.

By definition $0 \leq p(A) \leq 1$, i.e. $\begin{cases} p(A)=0 \Rightarrow \text{the event will definitely not happen} \\ p(A)=1 \Rightarrow \text{this event will definitely happen.} \end{cases}$

$\bar{A} := \{\text{not } A\}$

$p(\bar{A}) = 1 - p(A)$

A discrete random variable X is an event taking values from a discrete set $\mathcal{X}$.

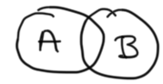
e.g. coin toss : $X = \{\text{outcome of the coin toss}\}$, $\mathcal{X} = \{\text{heads}, \text{tails}\}$

$p(x) = p(X = \text{heads}) = \frac{1}{2}$ if the coin is fair.

We denote the probability of a discrete r.v. $X = x$ as $p(X = x)$ or simply as $p(x)$.

Properties: $\begin{cases} 0 \leq p(x) \leq 1 \\ \sum_{x \in \mathcal{X}} p(x) = 1 \\ p(\{\emptyset\}) = 0 \end{cases}$ then $p(x)$ is called the probability mass function (p.m.f) of the discrete random variable X.

· Fundamental rules of probability :



Venn diagram

· Union : (A or B)

$$p(A \cup B) = p(A) + p(B) - \underbrace{p(A \cap B)}_{}^{\geq 0}$$

$$\overset{\text{union}}{\underset{\text{bound}}{\Longrightarrow}} \quad p(A \cup B) \leq p(A) + p(B)$$

· Joint : (A and B)

$$\underbrace{p(A,B)}_{\text{joint}} = p(A \cap B) = \underbrace{p(A|B)}_{}\underbrace{p(B)}_{} \qquad \text{Product rule}$$

$$= \quad \ldots \quad = \text{conditional} \times \text{marginal}$$

This directly follows from the definition of conditional probability:

$$p(A|B) := \frac{p(A,B)}{p(B)} \quad , \text{ given that } p(B) > 0$$

Given a joint distribution $p(A,B)$, we can define a marginal distribution as :

$$\begin{cases} p(A) = \sum_{b \in \mathcal{B}} p(A,B) = \sum_{b \in \mathcal{B}} p(A|B)p(B) = \sum_{b \in \mathcal{B}} p(A|B=b)p(B= \\ p(B) = \sum_{a \in \mathcal{A}} p(A,B) = \sum_{a \in \mathcal{A}} p(B|A)p(A) = \sum_{a \in \mathcal{A}} p(B|A=a)p(A \end{cases}$$

This is the sum rule or the rule of total probability.

· Bayes rule : It is a product of combining the definition of conditional probability with the product and sum rules :

$$p(A=a|B=b) = \frac{p(A=a, B=b)}{p(B=b)} = \frac{p(B=b|A=a)\,p(A=a)}{\sum_{a \in \mathcal{A}} p(B=b|A=a)\,p(A=a)}$$

· Some common discrete distributions :

– Binomial distribution : $X \sim Bin(n, \theta)$

n : # of trials

θ : probability of our desired outcome/event to occur

X : # of heads in n trials

X: # of heads in n trials

$$p(X=k) = \mathrm{Bin}(k|n,\vartheta) = \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}, \text{ where}$$

pmf

$$\binom{n}{k} := \frac{n!}{(n-k)!\,k!},$$

mean: $\mathbb{E}[X] = n\vartheta$

variance: $\mathrm{Var}[X] = n\vartheta(1-\vartheta)$

(*) This can be generalized to the so-called multinomial distribution for cases involving non-binary outcome, e.g. toss of die.

— Bernoulli distribution: Suppose we only have one trial.

Then $X \in \{\underset{\text{tails}}{0}, \underset{\text{heads}}{1}\}$ and if ϑ denotes the probability of our desired outcome the the discrete r.v. X is a Bernoulli random variable:

$$X \sim \mathrm{Ber}(\vartheta), \quad p(X=1) = \mathrm{Ber}(x|\vartheta) = \vartheta^{\mathbb{1}_{\{x=1\}}}(1-\vartheta)^{\mathbb{1}_{\{x=0\}}}$$

where $\mathbb{1}$ is the indicator function: $\mathbb{1}_{\{x=1\}} = \begin{cases} 1, & \text{if } x=1 \\ 0, & \text{if } x=0 \end{cases}$

In other words: $p(X=1) = \mathrm{Ber}(x|\vartheta) = \begin{cases} \vartheta, & \text{if } x=1 \\ 1-\vartheta, & \text{if } x=0 \end{cases}$

(*) This is a special case of the Binomial distribution with $n=1$.

(*) Can be generalized to cases with non-binary outcomes.

$\hookrightarrow$ multi-noulli distribution.

— Poisson distribution: models counts

Event/outcome space $X \in \{0, 1, 2, 3, \dots\}$

We say that a discrete random variable X has a Poisson distribution with parameter $\lambda > 0$:

$$X \sim \mathrm{Poi}(\lambda), \quad \text{pmf}: \ p(X=x) = \mathrm{Poi}(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$$

(*) This is usually used to model counts of rare events.

Independence:

- Two random variables X, Y are unconditionally independent if:

$$\underbrace{p(X,Y)}_{\text{joint}} = \underbrace{p(X)p(Y)}_{\text{product of the marginals}}, \quad X \perp Y$$

- Two random variables X, Y are conditionally independent given Z:

$$p(X,Y \mid Z) = p(X \mid Z)\, p(Y \mid Z), \quad X \perp Y \mid Z$$

- **Moments of discrete random variables:**

Goal: Quantify the statistical properties of a random variable or its associated probability distribution.

Mean : $\mu = \mathbb{E}[X] := \sum_{x \in \mathcal{X}} x\, p(x)$, $\boxed{p(x) : \text{pmf}}$

(First moment / first-order statistic)

Variance : $\mathbb{V}\mathrm{ar}[X] = \sum_{x \in \mathcal{X}} (x-\mu)^2 p(x)$

(second-moment / second-order statistic)

⋮

n-th moment / n-th-order statistic : $\sum_{x \in \mathcal{X}} (x-\mu)^n p(x)$