Statistics Refresher Notes

Lucien R. Zagabe

Bayes' Theorem

Theorem (The Law of Total Probability). *Let $A_1, ..., A_k$ be a partition of Ω. Then, for any event B,*

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Theorem (Bayes' Theorem). *Let $A_1, ..., A_k$ be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each i. If $\mathbb{P}(B) > 0$ then, for each $i = 1, ..., k$,*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

Important. We call $\mathbb{P}(A_i)$ the **prior probability of** A and $\mathbb{P}(A_i|B)$ the **posterior probability of A**

Bayesian statistics

Bayesian inference differs from more traditional statistical inference by preserving uncertainty. The Bayesian world-view interprets probability as measure of believability in an event, that is, how confident we are in an event occurring.

Bayesians interpret a probability as measure of belief, or confidence, of an event occurring. A probability is a summary of an opinion. An individual who assigns a belief of 0 to an event has no confidence that the event will occur; conversely, assigning a belief of 1 implies that the individual is absolutely certain of an event occurring.

This philosophy of treating beliefs as probability is natural to humans. We employ it constantly as we interact with the world and only see partial truths, but gather evidence to form beliefs.

Consider an uncertain event, for example whether the Artctic ice cap will have disappeared by the end of the century. These are not events that can be repeated numerous times in order to define a notion of probability (i.e. Frequentist statistics). Nevertheless, we will generally have some idea, for example, of how quickly we think the polar ice is melting. If we now obtain fresh evidence, for instance from a new earth observation, we may revise our opinion on the rate of ice loss. This can all be achieve through Bayesian interpretation of probability.

Distribution Functions and Probability Functions

Definition. *The* ***cumulative distribution function****, or CDF, is the function* $F_X : \mathbb{R} \to [0,1]$ *defined by*

$$F_X(x) = \mathbb{P}(X \leq x).$$

Definition. X *is discrete if it takes countably many values* $\{x_1, x_2, ...\}$. *We define the* ***probability function*** *or* ***probability mass function*** *for* X *by* $f_x(x) = \mathbb{P}(X = x)$ ■

Definition. *A random variable* X *is continuous if there exists a function* f_X *such that* $f_X(x) \geq 0$ *for all* x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ *and for every* $a \leq b$,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

The function f_X *is called the* ***probability density function*** *(PDF). We have that,*

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

and $f_X(x) = F_X^{'}(x)$ *and all points* x *at which* F_X *is differentiable* ■

Definition. *Let* X *be a random variable with CDF* F. *The* ***inverse CDF*** *or* ***quantile function*** *is defined by*[4]

$$F^{-1}(q) = \inf\{x : F(x) > q\}$$

for $q \in [0,1]$. *If* F *is strictly increasing and continuous than* $F^{-1}(q)$ *is the unique real number* x *such that* $F(x) = q$ ■