

Stats Notes

Conditional probability - Bayes Rule

- $P(A)$ - the probability of A occurring
- $P(A \cap B)$ - the intersection - the probability of both A and B occurring
- $P(A \cup B)$ - the union - the probability of A or B (or both) occurring

Note that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The probability of both A and B occurring is not necessarily just their sum, as this double counts the intersection. This is why avengers don't have double the crit rate of other classes, if you roll double 20s, it only counts as one crit. If A is a roll of 20 on one die, and B is a roll of 20 on the other, then the probability of a crit is given by $P(A \cap B)$, which is $1/20 + 1/20 - 1/20 \times 1/20 = 0.0975$

Bayes rule

Bayes rule allows us to invert the condition on a conditional probability. That is, if we know the probability of B given A , then we can infer the probability of A when B is satisfied/given/present.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Consider a diagnostic test. We can assess the test for sensitivity and specificity (see below), which describe how the test behaves in the presence of the disease. After taking the test, however, we are probably more interested in inverting this relationship - what is the probability of having the disease given the test was positive?

- A is condition (e.g a disease)
- B is observation (e.g. a positive result on a test for the disease)
- $P(A|B)$ is the probability of A given B (i.e., how likely are we to have the disease given a positive test?)
- $P(B|A)$ is the probability of B given A (i.e., how likely is the test to be positive when the disease is present)
 - this is the probability of a “true positive”
 - for disease tests this is called the **sensitivity**
- $P(B|A^c)$ is probability of B in the absence of A (i.e. how likely are we to have the disease when the test is false)* $P(B^c|A^c)$ is probability of not B in the absence of A (i.e. how likely are we to not have the disease when the test is false)

- this is the probability of a “true negative”
- for disease tests this is called the **sensitivity**

- $P(A)$ is the prevalence, or probability of the condition in the population (given no other information)

$$P(A|B) = P(B|A)P(A) / P(B) = P(B|A) P(A) / (P(B|A) P(A) + P(B|Ac) P(Ac))$$

$$P(A|B) = P(B|A) P(A) / (P(B|A) P(A) + (1- P(Bc|Ac)) (1-P(A)))$$