

Probability cheatsheet

Gavin Band, [WHG GMS Programme](#) 2021

1. Probability expresses uncertainty:

$$\begin{array}{ll}
 P(X|C) = 0 & \Rightarrow \text{"X is definitely not true"} \\
 P(X|C) = 1 & \Rightarrow \text{"X is definitely true"} \\
 0 < P(X|C) < 1 & \Rightarrow \text{we are uncertain about X}
 \end{array}
 \left. \vphantom{\begin{array}{l} P(X|C) = 0 \\ P(X|C) = 1 \\ 0 < P(X|C) < 1 \end{array}} \right\} \text{Extends basic logic}$$

2. All probability is conditional:

$$P(X) \quad \text{always means} \quad P(X|C)$$

"probability of X"
"probability of X given C"

or "probability of X conditional on C"

where **C** represents **background information** (such as an assumed statistical model). C is often dropped from notation but is always there.

Examples: $P(\underbrace{\text{die}}_{\text{Model of die}} | \text{die is fair}) = 1/6$ $P(\text{transmission} | \text{model of infection})$ etc.

3. Probability is computable:

- I. $P(\underbrace{\neg A}_{\text{"not A"}} | C) = 1 - P(A|C)$ (A either occurs or doesn't.)
- II. $P(A \text{ or } B | C) = P(A|C) + P(B|C) - P(A, B|C)$ (One of them must occur, but don't double-count)
"A and B given C"
- III. $P(A, B | C) = P(A|B, C) \cdot P(B|C)$ (Split up joint probabilities into conditional ones.)
"A and B given C" "A given B and C" "B given C"

"Independence" happens in when the middle term simplifies: $P(A, B) = P(A)P(B)$

4. Practically useful tools for computation: (derived directly from the rules in 3)

LOTP: ("Law of total probability")

$$P(A|C) = \sum_X P(A, X|C)$$
 Computes probability of A by summing (or integrating) over all the possible values of another variable.

Often seen as $P(A|C) = \sum_X P(A|X, C) \cdot P(X|C)$ by applying rule III

Bayes' rule:

$$P(A|B, C) = \frac{P(B|A, C) \cdot P(A|C)}{P(B|C)}$$
 Computes P(A given B) in terms of P(B given A)

4. Probability for variables

Often X will instead denote a variable whose value we are uncertain of. It is typical to use lower-case letters like x to denote specific possible values. So we will write:

$$P(X = x|C) \quad \text{"Probability X takes value x"}$$

For a continuous variable X we must instead think of P as a density function:

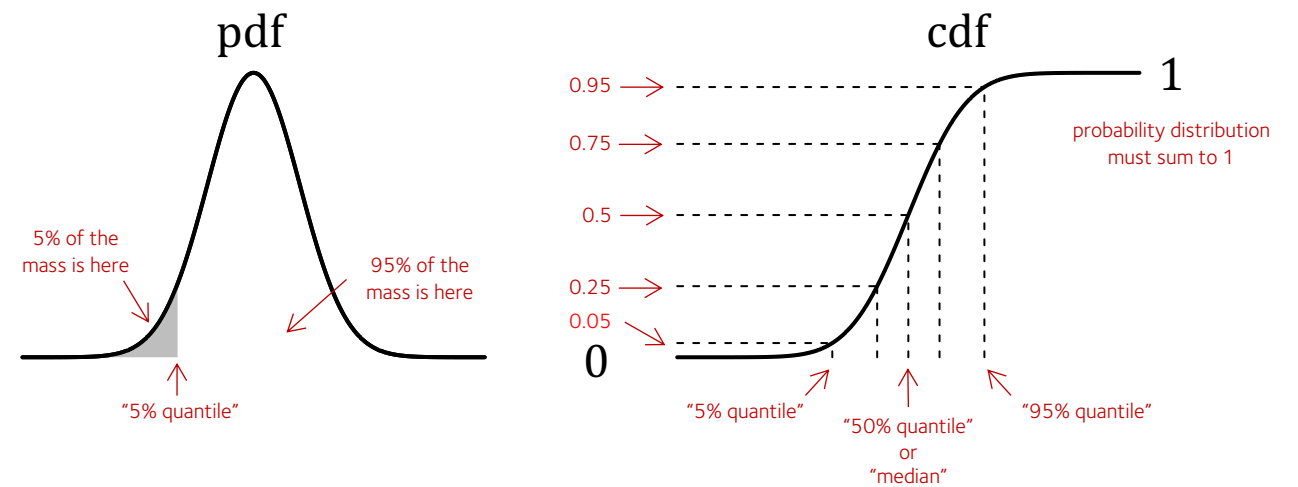
$$P(X = x|C) \quad \text{"Probability density function (pdf) evaluated at } X = x\text{"}$$

and integrate it to get an actual probability:

$$P(a \leq X \leq b|C) = \int_a^b P(X = x|C) dx \quad \text{"Probability that X is between a and b"}$$

$$P(X \leq b|C) = \int_{-\infty}^b P(X = x|C) dx \quad \text{"Cumulative distribution function (cdf) of X"}$$

Inverse of cdf is the *quantile function*.



5. Common distributions

Probabilities are often expressed in terms of a set of well-known distributions

$X \sim N(0,1)$ "X has a standard normal (or Gaussian) distribution"

$X \sim \text{binom}(N, p)$ "X has binomial distribution with N draws and success probability p "

See the separate 'distributions cheatsheet' for more.

6. Useful statistical summaries

"Mean" or "expected value" $\mu = E(x) = \sum_x \overset{\text{value}}{x} \cdot \overset{\text{pdf}}{P(X = x)}$ Average of x over all possible values

"Variance" $\text{var}(x) = E((x - \mu)^2) = \sum_x (x - \mu)^2 \cdot P(X = x)$ Average squared distance to mean (a useful measure of uncertainty)

"Entropy" $H(x) = \sum_x \frac{1}{\log P(X = x)} P(X = x)$ average information provided by x (also a measure of uncertainty)