1- probability, random variables, and basic statistics

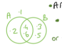
probability = measure of the likeliness that a random event will occur
↳ frequentists = defines probability as its relative frequency in many trials, objective
↳ subjectivists = degree of belief

sample space = the set of all possible outcomes of a random process
↳ $S = \{(H,H), (T,T), (H,T), (T,H)\}$ ↳ $S = [0, +\infty)$ → can be continuous

event = any collection of possible outcomes of a random process, subset of sample space
↳ getting exacty one head in tossing two coins experiment / process

prior / marginal probabilities = $P(A)$, probability of a single even occuring, independently
disjoint events = $A \cap B = \varnothing \xrightarrow{\text{then}} P(A \cup B) = P(A) + P(B)$
joint probability = $P(A \cap B)$ frequency of two events occurring together $= |A \cap B| / |U|$ → universal set
conditional probability = $P(A|B) = |A \cap B| / |B|$
conditionally independent = $X \perp Y | Z$ X and Y may not be independent in universal set, but they are independent given Z
independent events = $P(A|B) = P(A) \rightarrow P(A \cap B)/P(B) = P(A) \rightarrow P(A).P(B) = P(A \cap B) = P(A,B)$
↳ axiom = if two events are mutually independent
↳ independence is symmetric (if A ind of B ⟷ B ind of A)
* independence $\not\Rightarrow$ disjoint

example = tossing a die $P(A,B) = P(A \cap B) = \frac{2}{6}$
$A = \{\text{even results}\} \rightarrow 2, 4, 6$ $P(A) = \frac{3}{6}$
$B = \{\text{result} > 2\} \rightarrow 3, 4, 5, 6$ $P(B) = \frac{4}{6}$



- $A \cap B \neq 0 \longrightarrow$ they are not disjoint
- $P(A).P(B) = \frac{3}{6} \cdot \frac{4}{6} = \frac{1}{3} = P(A \cap B)$
or $P(A|B) = \frac{2}{6} / \frac{4}{6} = \frac{1}{2} = P(A)$
↳ they are independent

bayesian theorem = • $P(A|B) = \frac{P(A,B)}{P(B)}$ • $P(B|A) = \frac{P(A,B)}{P(A)} \Rightarrow P(A|B) = \frac{P(B|A).P(A)}{P(B)}$

the law of total probability = $P(A) = P(A|B_1).P(B_1) + P(A|B_2).P(B_2) \ldots + P(A|B_j).P(B_j)$
↳ marginalization ↳ $P(A) = \sum_j P(A|B=B_j).P(B=B_j)$ $(S = B_1 + B_2 + B_3 \ldots + B_j)$

example = a test is 95% effective in detecting when it is present, it has 1% false positive rate, 0.5% of the population has the disease, probability that a person with a positive test result actually has the disease? $P(B|A) = ?$

A = positive test result
B = having the disease
$P(A|B) = 0.95$
$P(A|\neg B) = 0.01$
$P(B) = 0.005$
$P(\neg B) = 0.995$

$P(B|A) = \frac{P(A,B)}{P(A)} = 0.323$

$P(A|B) = \frac{P(A,B)}{P(B)}$ $P(A,B) = (0.95).(0.005)$

$P(A) = P(A|B).P(B) + P(A|\neg B).P(\neg B) = (0.95).(0.005) + (0.01).(0.995)$

random variables = rule that assigns a numerical value to each outcome in sample space
probability distributions = function that maps to the probability of a value
↳ $f(x) = 0.5$ $\forall_x \in \{H, T\}$ (outcome of a coin toss)
↳ probability mass function (pmf) = prob. dist. func for discrete values
↳ $f_X(x) = P(X = x)$
↳ probability density function (pdf) = for continuous values
↳ $f_X(x) = \lim_{\Delta x \to 0} P(X = x) = 0$ $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$
↳ cumulative distribution function (cdf) = X having value up to and including x.
↳ $F_X(x) = \sum_{\forall t \leq x} f(t)$ ↳ $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$

expected value = mean of a random variable $E[X] = \mu_X$
↳ $\mu_X = \sum_x x f(x) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_k f(x_k)$ ↳ $\mu_X = \int_{-\infty}^{+\infty} x f(x)\,dx$
↳ linear combinations = $E[c_1 X + c_2 Y] = c_1 E[X] + c_2 E[Y]$
↳ product = $E[XY] = \int_X \int_Y xy\, f_{XY}(x,y)\,dx\,dy$ $f_{XY}(x,y) = f_X(y).f_Y(x) \Longleftrightarrow$ X and Y are independent
↳ joint probability

variance = $Var(X) = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2$ → differentiable (absolute error is not)
↳ square the expected deviation from the mean
standard deviation = $\sigma = \sqrt{Var(X)}$

example = a lab rat
X = number of meals eaten in a day, two meals half, one or no equally in reaming
Y = number of hours slept in a day, 10 ± 4 hours uniformly distributed

X

$f_X(x) = \begin{cases} 0.25, & x = 0 \\ 0.25, & x = 1 \\ 0.50, & x = 2 \end{cases}$

$F_X(x) = \begin{cases} 0.25, & x = 0 \\ 0.50, & x = 1 \\ 1, & x = 2 \end{cases}$

$E_X[X] = 0.(0.25) + 1.(0.25) + 2.(0.50) = 1.25$

Y

$f_Y(y) = \begin{cases} \frac{1}{8}, & 6 \leq y \leq 14 \\ 0, & \text{otherwise} \end{cases} = \frac{dF_Y(y)}{dy}$

$F_Y(y) = \begin{cases} 0, & y < 6 \\ \frac{y-6}{8}, & 6 \leq y \leq 14 \\ 1, & y > 14 \end{cases} = \int f_Y(y)\,dy$

$E_Y[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_6^{14} \frac{y}{8}\,dy = \frac{y^2}{16}\Big|_6^{14} = 10$

- expected number of hours above 8 $\rightarrow g(Y) = y - 8$ $E[g(Y)] = \int_6^{14} \frac{(y-8)}{8}\,dy = 2$
- expected number of meals per hour of sleep? (assuming independence)
$E\left[\frac{X}{Y}\right] = \int_X \int_Y \frac{x}{y} f_X(x) f_Y(y)\,dx\,dy = \underbrace{\sum_x x f_X(x)}_{1.25} \int_y \frac{1}{y} f_Y(y)\,dy = (1.25).\int_6^{14} \frac{1}{8y}\,dy = (1.25).\frac{\ln(14) - \ln(6)}{8} = 0.132$
- variance of sleep time $= E(Y^2) - \mu_Y^2 = \int_6^{14} y^2 f(Y)\,dy - 100 = \frac{y^3}{3.8}\Big|_6^{14} - 100 = 5.\overline{33}$
$\sigma = 2.309 \rightarrow$ std dev

covariance = measure of the dependence of two variables to each other
↳ $Cov(X,Y) = E[(X - \mu_X).(Y - \mu_Y)] = \underbrace{E(XY) - E(X).E(Y)}$
↳ if they are independent $E(XY) = E(X).E(Y) \Rightarrow 0 \Rightarrow Cov(X,Y) = 0$

pearson correlation = normalized measure of dependence
↳ $corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$
- $corr(X,Y) = 1 \rightarrow$ perfect positive dependence
- $corr(X,Y) = -1 \rightarrow$ perfect negative dependence

- if $corr(X,Y) = 1$ or -1, when x is known, exact value of y is known
↳ for other values y can be known as set of distribution, not exactly

↳ if X and Y are independent $\Rightarrow corr = 0$, but $corr = 0 \not\Rightarrow$ x and y independent → not known