Bayesian Statistics and Probabilistic Programming

Lecture 1. Probability

- Probability: interpretations.
  - frequentist: count (potentially infinite) from data
  - bayesian: degree of belief

  } check Bayes' billiard at the end.

- Statistics:
  - classical: (X, Θ), constants, variables, $\underbrace{\text{parameters}}^{\text{unknown, not random}}$ → estimators.
  - bayesian: random variables (unknown)

- Elementary conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{with } P(B) > 0$$

- $P(A|B) > P(A) \Rightarrow$ B facilitates occurence of A (and the analogous)
- $P(A|B) = P(A) \Rightarrow$ A and B are independent. $\underset{\text{def.}}{\Rightarrow}$ $P(A \cap B) = P(A)P(B)$ (not: $A \perp\!\!\!\perp B$)
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$ Bayes' rule (with $P(A), P(B) > 0$)

Pg 31. Diagram 1 does not include two independent variable. If A, it can't B.
If B, it can't A.

only diagram 2 includes independent variables.

- Sample space: (Ω) all possible results.

- Bayes formula:

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

$P(A,B) = P(A \cap B)$

$$\mu_x = \underbrace{\frac{\gamma^2}{\sigma^2 + \gamma^2}}_{\substack{\text{low var} \\ \to \text{closer to 1} \\ \text{rely on observ. } (x)}} x + \underbrace{\frac{\sigma^2}{\sigma^2 + \gamma^2}\mu}_{\substack{\text{high variance} \to \text{closer to 1} \\ \text{we rely more on population mean } (\mu)}}$$

- Beta function:

$$B(x,y) = \frac{\Gamma(}{\quad}$$

Lecture 2. Random Variables

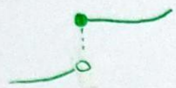
"Random.vars.r" includes supporting code (Notebook 01)

- **Probability Mass Function:** maps possible values to their probability.
  (can be shown as a function or as areas on the real line)

- **Cumulative Distribution Function:** cumsum de las probabilidades. (step function; always increasing)

(Examples of random variables in notebook 1) → bernoulli ~ tossing a coin n times

- right continuous function



discrete r.v → continuous r.v
sums ↦ integrals

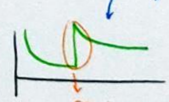
"required" better than "library"

- Laplace is not differentiable at the peak $\wedge$, while gaussian is: $\cap$

... → dots syntax: "any other parameter" → they can be passed to any other function through "..."

- **Law of Large Numbers: (LNN)** the relative frequency $\xrightarrow[\infty]{\text{tends}}$ to the probability
  - LNN for r.v. with $E < \infty$

  Cauchy distribution does not always converge in all trajectories: big jumps



The trajectories can be trapped by $\sqrt{N}$ (x a constant, eg. 1, 1.5)
↳ if we multiply the trajectories by $\sqrt{N}$ → $\underbrace{\text{the trajectories will turn into almost horizontal}}_{\text{easier to analyze.}}$

Lecture 3. Statistical Models.

↳ do not over-elaborate / over-parametrize them.

- **Inverse of** $p = \text{Prob}[X \le x]$: $Q(p) = x$
  $[0,1] \to X$