today:
  * finish prob. review
  * parametric models

Bayes rule:

\[ p(x|y) = \frac{p(y|x)p(x)}{p(y)} \]

by def:

\[ p(y|x) = \frac{p(x,y)}{p(x)} \]

\[ p(x,y) = p(y|x)p(x) \]

"product rule"

Chain rule:

\[ p(x_1, \ldots, x_n) = p(x_n | x_{1:n-1}) p(x_{1:n-1}) \]

\[ p(x_{1:n-1}) = p(x_{n-1} | x_{1:n-2}) p(x_{1:n-2}) \]

\[ \vdots \]

\[ = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \]

Convention: \(1 \cdot 0 = 0\), \(p(x_1 | x_{ \varnothing }) = p(x_1)\)

Always true?

Can be simplified when making conditional indep.

\[ \text{assumptions} \]

(\text{from a directed graphical model})

Conditional independence

\(X\) is \(\text{cond. indep.}\) of \(Y\) given \(Z\)

\[ X \perp Y | Z \]

\[ \iff p(x,y|z) = p(x|z)p(y|z) \quad \forall x \in \mathcal{X} \]

\[ \forall y \in \mathcal{Y} \]

\[ \forall z \in \mathcal{Z} \quad \text{st. } p(z) \neq 0 \]

Exercise to the reader: prove that if \(X \perp Y | Z\)

then \(p(x|y,z) = p(x|z)\)

[conditional analog of \(X \perp Y \Rightarrow p(x) = p(z)\)]

Example:

\(Z = \text{indicator whether mother carries a genetic disease}\)

\(X = \text{son has disease}\)

\(X \perp Y | Z\)
Example:

\( z \) - uxoroor unless mother carries a genetic disease

\[
\begin{align*}
X &= \text{son 1 has disease} \\
Y &= \text{son 2 has disease}
\end{align*}
\]

\( X \perp Y \mid Z \)

Here, \( X \) is not "marginally independent" of \( Y \) (i.e., \( X \perp Y \)).

\[ P(x, y, z) = P(x | z) P(y | z) P(z) \]

Example of pairwise independence RV's that are not mutually independent:

\( X = \text{coin flip as \# 1, 2} \)

\[
\begin{align*}
Y &= \text{coin flip} \\
\{X \perp Y\}
\end{align*}
\]

Define \( Z \equiv X \times Y \)

\( Z \perp X \) \hspace{1cm} Z \perp Y \)

\( Z \not\perp (X, Y) \)

Other notation:

\( \mathbb{E}[x] \equiv \int_{x \in \text{RV}} x p(x) \) (for discrete RV)

\[
\mathbb{E}[x] \equiv \int_{x} x p(x) \text{ d}x \quad \text{cts. RV}
\]

\( \mathbb{E}[x+y] \) is a linear operator \( \Rightarrow \mathbb{E}[aX + bY] = a \mathbb{E}[X] + b \mathbb{E}[Y] \)

\( \text{scalar} \)

Variance:

\[
\text{Var}[x] \equiv \mathbb{E} \left[ (x - \mathbb{E}[x])^2 \right]
\]

\[
= \mathbb{E}[x^2] - (\mathbb{E}[x])^2
\]
parametric models:

Let a family of distributions set of valid parameters

\[ \mathcal{F}_\Theta = \{ p(x;\theta) \mid \theta \in \Theta \} \]

possible pmf/pdfs depending on parameter \( \theta \)

base of notation: \{ p(x;\theta) \mid \theta \in \Theta \}

"beta" notation \{ p(X;\theta) \mid \theta \in \Theta \}

Notation: \( X \sim \text{Bern}(\theta) \)

"R.V. \( X \) is distributed as a Bernoulli dist."

\[ p(x;\theta) = \text{Bern}(x;\theta) \]

\( \theta \) : pmf

Bernoulli: prob. of a coin-flip \( \Omega = \{0,1\} \)

\[ \Pr(X=1 \mid \theta) = \theta \Rightarrow \theta = \{0,1\} \]

\[ p(x;\theta) = \theta^x (1-\theta)^{1-x} = \text{Bern}(x;\theta) \]

\[ \mathbb{E}[X] = \theta \]

\[ \text{Var}[X] = \theta(1-\theta) \]

\[ \text{Var}[X;\theta] = \text{Var}[\text{Bern}(X;\theta)] \]

\[ \text{Var}[X] = \sigma^2 \]
**binomial distribution**: model n independent coin flips

\[ X \sim \text{Bin}(n, \theta) \]

\[ \Pr(X = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \]

\[ \text{mean: } \mu = n \theta \]

\[ \text{variance: } \sigma^2 = n \theta (1-\theta) \]

\[ \text{other distributions: } \text{Poisson}(\lambda), \quad \chi^2, \quad t, \quad F, \quad \text{etc.} \]

\[ \text{statistical concepts: } \text{prob. theory, } \text{well defined} \]
Statistical concepts

- Model
- Data
- Statistics
- Well-defined
- Ill-defined
- "Inverse problem"

**Example:** Model n independent coin flips

\( \text{prob. theory } \rightarrow \text{prob. k heads in a row} \)

Statistics: I have observed k heads, what is \( \theta \)?

n-k tails

**Frequentist vs. Bayesian:**

- **Semantic of prob.:** meaning of a prob.?
  a) (traditional) Frequentist semantic

  \[ P(X=x) \text{ represents the } \text{limiting frequency of observing } X=x \]
  \[ \text{If I could repeat } \# \text{ of trials experiments} \]

  b) Bayesian (subjective) semantic

  \[ P(X=x) \text{ encodes an agent's "belief" that } X=x \]

  laws of prob. characterize a "rational" way to combine "beliefs" and "evidence" [observations]

  \[ \Rightarrow \text{ has motivation in terms of gambling, utility/decision theory, etc.} \]

**Subjectively:**

Bayesian approach:

- Very simple philosophically;
  - Treat all uncertain quantities as R.V.
  - I.e. encode all knowledge about the system ("beliefs") as a prior on probability model
  and then use law of prob. (and Bayes rule) to get answers.
for a discrete RV, suppose \( \mathbb{P}(X = x) = \theta \)

\[ \Rightarrow \mathbb{P}(X \neq x) = 1 - \theta \]

\[ B \overset{d}{=} \sum_{i=1}^{n} X_i = 2 \]

\[ \Rightarrow B \overset{d}{=} \text{Bin}(n) \quad \text{RV} \]

indicator function \( I_A(u) = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{otherwise} \end{cases} \)

repeat i.i.d. experiments i.e. \( B \) is \( \text{Bin}(n) \)

by LLN \( \frac{1}{n} \sum_{i=1}^{n} B_i \xrightarrow{a.s.} \mathbb{E}[B] = \theta \)

by CLT \( \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} B_i - \theta \right) \xrightarrow{d} N(0, \sigma^2) \)