today: finish proba. review
- parametric models
- statistical concepts

reminder:
- discrete r.v.: pmf. \( p(x) \Rightarrow \Pr[X=x] \)
- cont. r.v.: pdf. \( p(x) \Rightarrow \Pr[x-e\leq x \leq x+e] = \int p(x) \, dx \)

other prob concepts:
- expectation/mean of a r.v.:
  \[ \mathbb{E}[X] = \sum_{x} x \, p(x) \quad \text{(for discrete r.v.)} \]
  \[ \mathbb{E}[X] = \int_{-\infty}^{\infty} x \, p(x) \, dx \quad \text{(for cont. r.v.)} \]
- \( \mathbb{E}[\cdot] \) is a linear operator \( \Rightarrow \mathbb{E}[aX+bY] = a \mathbb{E}[X] + b \mathbb{E}[Y] \)

- variance:
  \[ \text{Var}[X] = \mathbb{E}[ (X-\mathbb{E}[X])^2 ] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \]

- independence:
  \[ X \text{ is independent of } Y \quad \Rightarrow \quad p(x,y) = p(x) \cdot p(y) \quad \forall x \in X \wedge y \in Y \]
"mutually independent": for R.v. $X_1, \ldots, X_n \iff p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i)$

- for events $A \subseteq B$, suppose that $P(B) \neq 0$
  
  then define $P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}$

'condition pdf':

- for a discrete r.v., define $P(X=x \mid Y=y) \triangleq \frac{P(X=x, Y=y)}{P(Y=y)}$
  
  joint pdf
  
  marginal pdf

same thing with 'conditional pdf':

subtle point: $p(x \mid y)$ is undefined when $p(y) = 0$

example: "$P(\text{having cancer} \mid \text{tumor measurement} = a)$" by defn: $p(y \mid x) = \frac{p(x, y)}{p(x)}$

$\Rightarrow p(x \mid y) = p(y \mid x) p(x)$

"product rule"

Bayes rule: $\Rightarrow$ inverting the conditioning

$p(x \mid y) = \frac{p(y \mid x) p(x)}{p(y)}$

"Bayes rule"

chain rule: $\Rightarrow$ successive application of product rule

$p(x_1, \ldots, x_n) = p(x_{1:n-1}) p(x_n \mid x_{1:n-1})$

$\Rightarrow \ldots$
\[
\prod_{i=1}^{n} p(x_i | x_{i-1:2})
\]

always true?

can be simplified using cond. indep. assumptions

\[ I \text{ directed graph model?} \]

cond. indep.: \( X \) is cond. indep. of \( Y \) given \( Z \)

notation: \( X \perp Y \mid Z \)

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

A.G.S.

\( x \in \{0, y \}
\)

\( y \in \{0, y \}
\)

\( z \in \{0, z \}
\)

\( x \in \{0, x \}
\)

Example: \( Z \) = indicator whether mother carries a genetic disease

\( X = \text{sm.1 has disease} \)

\( Y = \text{sm.2 has disease} \)

\( X \perp Y \mid Z \)

here \( X \) is not "marginally indep." of \( Y \) i.e. \( X \not\perp Y \)

but \( X \) is cond. indep. of \( Y \) given \( Z \)

exercise to the reader: prove that \( X \perp Y \mid Z \)

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

[also, \( X \perp Y \Rightarrow p(x | y) = p(x) \)]

parametric models:

family of distributions

\( \mathcal{R}_x = \{ p(x, \theta) \mid \theta \in \Theta \} \)
possible range of $\theta$ depending on $\Theta$

\[ \begin{align*}
\text{abuse of notation: } & \sum p(x; \theta) | \theta \in \Theta \\
\text{"correct" notation: } & \sum p(x; \theta, \Theta) | \theta \in \Theta
\end{align*} \]

notation: $X \sim \text{Bin}(\theta)$

\[ \text{"R.V. } X \text{ distributed as a Bin}(\theta) \text{ dist."} \]

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

\[ \text{parameter of dist.} \]

\[ \text{variable for } \theta \]

\[ \text{Bernoulli: prior of a coin flip} \quad \Pr[X = 1 | \theta] = \theta \quad \theta \in [0, 1] \]

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

\[ \text{Binomial distribution: models } n \text{ independent coin flips} \]

\[ \text{sum of } n \text{ independent } \text{Bin}(\theta) \text{ R.V.} \]

\[ \text{often denoted } p \]

\[ \text{let } X_i \sim \text{Bin}(\theta) \quad \text{independent & identically distributed} \]

\[ \text{implicitly is thinking about } X_1, \ldots, X_n \]

\[ \text{let } X = \sum X_i \quad \text{then we have } X \sim \text{Bin}(n; \theta) \quad \text{"binomial with} \]

\[ \text{from context: } \text{"support of distribution" i.e. } \theta \]

\[ \text{is usually fixed for all } \theta \]

\[ \text{examples: } \theta = \{0.1, 0.5\} \text{ for a coin flip} \]

\[ \theta = [0, 10] \text{ for gamma distribution} \]
\[ X_1, X_2, \ldots, X_n \]

\[ \mathbb{P}(X | \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \text{ for } x \in \Omega_x \]

\# ways to get k successes out of n trials

\[ (n) \times \frac{n!}{x!(n-x)!} \]

\[ \text{Exp}(\{1-\theta\}), \text{Exp}(\{1-\theta\}) \]

\[ = \frac{1}{\theta} \text{ Bern}(x | \theta) \]

Mean \( X = \sum_i \mathbb{E}[X_i] \)

\[ \mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n \theta \]

Similarly \( \text{Var}(X) = n \text{Var}(\theta) \)

Other distributions: Poisson(\(\lambda\)) \(\lambda = 0, 1, \ldots \) \(\lambda \in \mathbb{N} \) [count data]

Gaussian (\(\mu, \sigma^2\)) \(\mu = \text{mean}, \sigma^2 = \text{variance}\)

Gamma \(\Gamma(\alpha, \beta)\)

Shape scale

Other: Exponential, Cauchy, exponential, Beta, Dirichlet, etc.

Statistical concepts

\[ \text{Prob} \rightarrow \text{model} \rightarrow \text{data} \rightarrow \text{inference} \]

\[ \text{epistemology of science} \]
Example: model is n i.i.d. coin flips

Prob. theory: prob of k heads in a row

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

**Frequentist vs. Bayesian:**

 philos: meaning of a probability?

a) (traditional) frequentist semantic:

\[ P(X=x) \] represents the \textit{relative frequency} of observing \( X=x \)

\[ \text{If I could repeat } n \text{ of iid experiments} \]

b) Bayesian (subjective) semantic:

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

\[ \text{laws of probability characterize a \textit{rational} way to combine \textit{beliefs}} \]

\[ \text{and \textit{anticipate} \textit{Deductions}} \]

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

**Frequentist justification:**

*for a discrete R.V., suppose \( k \) \[ P(\Xi = x^2) = \delta \]

\[ \Rightarrow P(\Xi = x^2) = 1-\delta \]

\[ B \overset{A}{=} 1, \Xi = x^2 \text{ iid Bern}(\delta) \text{ R.V.} \]

repeat iid experiments i.e. \( B_i \text{ iid Bern}(\delta) \)

by L.L.N. (law of large numbers) \[ \frac{1}{n} \sum_{i=1}^{n} a_{B_i} \rightarrow E[B_i] = \delta \]

(terminates discussion)
\[
\text{Aside: } \sum_{i=1}^{n} b_i \sim \text{Bin}(n, \theta) \quad \mathbb{E}(\sum b_i) = \frac{\sum b_i}{n} = \theta
\]

\[
\text{Var} \left[ \sum_{i=1}^{n} b_i \right] = \frac{1}{n^2} \text{Var} \left[ \text{Bin}(n, \theta) \right] = \frac{1}{n^2} \cdot n \cdot \theta(1-\theta) = \frac{\theta(1-\theta)}{n}
\]

(see \( \text{Var}[aX] = a^2 \text{Var}[X] \))

\[
\sqrt{n} \left( \frac{\sum b_i}{n} - \theta \right) \xrightarrow{d} N(0, \theta(1-\theta))
\]

CLT (Central Limit Theorem)

\[\rightarrow \text{all works} \]

Bayesian approach:

- very simple philosophically: treat all uncertain quantities as R.V.
- \( \rightarrow \) i.e. encode all knowledge about the system ("beliefs")
- as a "prior" on probabilistic models
- and then use laws of fields (and Bayes rule) to get answers?

simplest example is based on

- we believe \( X \sim \text{Bin}(n, \theta) \)
- \( \theta \) unknown? \( \Rightarrow \) model it as a R.V.
- \( \Rightarrow \) we need a \( p(\theta) \) "prior distribution"

\[ \Omega_\theta = [0,1] \]

\[\text{Suppose we observe } X = x \text{ (result of flips)} \]

then, we can "update" our belief about \( \theta \) using Bayes rule

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

Bayes Rule
\( p(x|e) \) posterior belief

\( p(x|e) \) normalization "marginal likelihood"

\( p(e) \) observation model/likelihood

Note: \( p(x|e) \) is pdf
\( p(x|e) \) is a mixed distribution
\( p(e) \) is pdf

Example: Suppose \( p(e) \) is uniform on \([0,1]"no specific preference"

\( p(\theta|x) \propto \theta^x (1-\theta)^{n-x} \frac{1}{p(e)} \quad (x \in 0:n) \)

"proportional to"

Scaling:
\[ \int_0^1 \theta^x (1-\theta)^{n-x} d\theta = B(x+1, n-x+1) \]

\[ B(a,b) \leq \binom{a+b}{a} \frac{1}{B(a,b)} \]

Each fraction

\( p(\theta|x) \) is called a "beta distribution"

\[ B(\theta; a,b) \triangleq \frac{\theta^{a-1} (1-\theta)^{b-1}}{B(a,b)} \quad \text{support} \]