Joint Probabilities

Frederik Mallmann-Trenn
6CCS3AIN

Joint probability distribution

- Joint probability distribution for a set of r.v.s (random variables) gives the probability of every atomic event on those r.v.s (i.e., every sample point)

  $\mathbf{P}(Cavity, Weather)$ is a 2×4 matrix of values

$$\begin{pmatrix} 0.144 & 0.02 & 0.016 & 0.02 \\ 0.576 & 0.08 & 0.064 & 0.08 \end{pmatrix}$$

  which can be interpreted as

	$W. = sunny$	$W. = rain$	$W. = cloudy$	$W. = snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution because every event is a sum of sample points.
- E.g., $P(Cavity = true\ AND\ Weather \neq cloudy)$

Inference by enumeration

(Paramount Pictures)

Inference by enumeration

- Example of a joint distribution with **three** variables (catch is something the doctor can test):

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$$

Inference by enumeration

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$\begin{aligned} P(toothache) &= 0.108 + 0.012 + 0.016 + 0.064 \\ &= 0.2 \end{aligned}$$

Inference by enumeration

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$\begin{aligned} P(cavity \vee toothache) &= 0.108 + 0.012 + 0.072 \\ &\quad + 0.008 + 0.016 + 0.064 \\ &= 0.28 \end{aligned}$$

Your turn

	$toothache$		$\neg$ $toothache$	
	$catch$	$\neg$ $catch$	$catch$	$\neg$ $catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg$ $cavity$	**.016**	**.064**	**.144**	**.576**

- Calculate $Pr(toothache \wedge cavity)$.
- I encourage you to use the KEATS forum to compare your answers!

Your turn

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- Calculate $Pr(cavity \mid toothache)$.
- I encourage you to use the KEATS forum to compare your answers!

Inference by enumeration

- Can also compute conditional probabilities:

$$
\begin{aligned}
P(\neg cavity | toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\
&= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} \\
&= 0.4
\end{aligned}
$$

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

$\boxed{P(\neg cavity \wedge toothache)}$ / $\boxed{P(toothache)}$

Notation

- Notation for conditional distributions:

  $\mathbf{P}(Cavity|toothache)$

  The outcome is a 2-dimensional vector (and so is Cavity). Write it down and compare!

Conditional probability

- Conditional or posterior probabilities

  $$P(cavity|toothache) = 0.6$$

  given that $toothache$ **is all I know**

  Recall, for convenience we write $cavity$ for $Cavity = true$ and $toothache$ for $Toothache = true$.

Conditional probability

- Conditional or posterior probabilities

  $$P(cavity|toothache) = 0.6$$

  given that $toothache$ **is all I know**

- If we know more, e.g., $cavity$ is also given, then we have

  $$P(cavity|toothache, cavity) = 1$$

  Note: the less specific belief (toothache) remains valid after more evidence arrives, but is not always useful

Conditional probability

- Conditional or posterior probabilities
  $$P(cavity|toothache) = 0.6$$
  given that $toothache$ **is all I know**

- If we know more, e.g., $cavity$ is also given, then we have
  $$P(cavity|toothache, cavity) = 1$$
  Note: the less specific belief (toothache) remains valid after more evidence arrives, but is not always useful
- New evidence may be irrelevant, allowing simplification
  $$P(cavity|toothache, \textit{your curtains are red}) = P(cavity|toothache) = 0.6$$
- This kind of inference, sanctioned by domain knowledge, is crucial.

Conditional probability

- Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) > 0$$

- Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

- A general version holds for joint distributions,

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity) \odot \mathbf{P}(Cavity)$$

(View as a $4 \cdot 2 = 8$ set of equations, not matrix multiplication - unless you know exactly what you're doing and you model everything correctly)

Chain rule

- Chain rule is derived by successive application of product rule. Concrete example:

$$\begin{aligned} P(a,b,c) &= P(a,b)P(c|b,a) \\ &= P(a)P(b|a)P(c|b,a) \end{aligned}$$

- In general,

$$\begin{aligned} \mathbf{P}(X_1,\ldots,X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\mathbf{P}(X_n|X_1,\ldots,X_{n-1}) \\ &= \mathbf{P}(X_1,\ldots,X_{n-2})\mathbf{P}(X_{n-1}|X_1,\ldots,X_{n-2})\mathbf{P}(X_n|X_1,\ldots,X_{n-1}) \\ &= \ldots \\ &= \prod_{i=1}^{n}\mathbf{P}(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$

Normalisation

We can use $\alpha = 1/P(toothache)$ to normalise (we don't need to calculate it!)

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$
$$= \alpha\left[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)\right]$$
$$= \alpha\left[\begin{pmatrix}0.108\\0.016\end{pmatrix} + \begin{pmatrix}0.012\\0.064\end{pmatrix}\right] = \alpha\begin{pmatrix}0.12,\\0.08\end{pmatrix} = \begin{pmatrix}0.6,\\0.4\end{pmatrix}$$

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

- Green boxes show step in calculation, not the desired outcome

General version

- Let $\mathbf{X}$ be all the variables.
- Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$
- Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

General version

- Then the required summation of joint entries is done by summing out the hidden variables:

$$\begin{aligned}\mathbf{P}(\mathbf{Y}|\mathbf{E}=e) &= \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=e) \\ &= \alpha\sum_h \mathbf{P}(\mathbf{Y},\mathbf{E}=e,\mathbf{H}=h)\end{aligned}$$

- The terms in the summation are joint entries because **Y**, **E**, and **H** together exhaust the set of random variables