Probabilistic Inference

Outline

I. Probability for continuous variables

II. Inference by enumeration

* Figures are either from the textbook site or by the instructor.

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

The *cumulative distribution function* of a continuous random variable X is

$$D(x) = P(X \leq x)$$

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

The *cumulative distribution function* of a continuous random variable X is

$$\boxed{D(x) = P(X \le x)}$$

Its derivative is the *probability density function (p.d.f.)*:

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

The *cumulative distribution function* of a continuous random variable X is

$$D(x) = P(X \leq x)$$

Its derivative is the *probability density function (p.d.f.)*:

$$p(x) = \frac{d}{dx} D(x)$$

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

The *cumulative distribution function* of a continuous random variable X is

$$D(x) = P(X \leq x)$$

Its derivative is the *probability density function (p.d.f.)*:

$$p(x) = \frac{d}{dx} D(x)$$

$= \lim dx \rightarrow 0 \quad P\ X \leq x + dx -$
$\leq P(X \leq x)dx$

$\lim dx \rightarrow 0 \quad P\ X$
$\leq x + dx\ - P(X$

I. Probability Density Function

A *continuous random variable* has its outcome take on a continuous set of values.

- the high temperature tomorrow
- the reading from a noisy force sensor
- the actual arrival time of a flight, etc.

The *cumulative distribution function* of a continuous random variable X is

$$D(x) = P(X \le x)$$

Its derivative is the *probability density function (p.d.f.)*:

$$p(x) = \frac{d}{dx} D(x)$$

$= \lim dx \to 0 \quad P\ X \le x + dx -$
$\le P(X \le x) dx$

$\lim dx \to 0 \ P\ X \le x + dx - P(X$

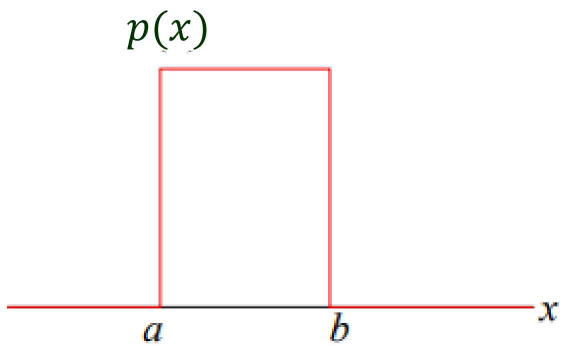
$$= \lim_{dx \to 0} \frac{P(x \le X \le x + dx)}{dx}$$

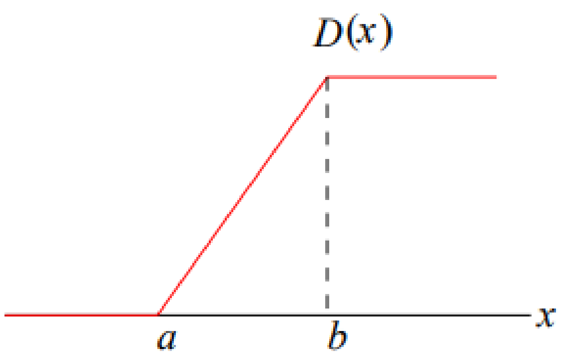
Uniform Distribution

A *uniform distribution* has constant p.d.f.

Example Continuous uniform distribution over the interval $[a, b]$.

$$p(x) = \begin{cases} 0, & \text{for } x < a, \\ \frac{1}{b-a}, & \text{for } a \le x < b, \\ 0, & \text{for } x > a, \end{cases} \qquad D(x) = \begin{cases} 0, & \text{for } x < a, \\ \frac{x-a}{b-a}, & \text{for } a \le x < b, \\ 1, & \text{for } x > b, \end{cases}$$





Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

- n experiments

Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

- n experiments
- m different values: $x_1, \ldots, x_2$

Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

- n experiments
- m different values: $x_1, \ldots, x_2$
- For $1 \leq i \leq m$, the value x_i occurs n_i times. // $n = n_1 + \cdots + n_m$

Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

- n experiments
- m different values: $x_1, \ldots, x_2$
- For $1 \le i \le m$, the value x_i occurs n_i times. // $n = n_1 + \cdots + n_m$

$$\boxed{E(X) = \frac{1}{n} \sum_{i=1}^{m} x_i \cdot n_i}$$

Expected Value

The *expected value* (or *mean*) $E(X)$ of a random variable X is its average value over a large number of chance experiments.

- n experiments
- m different values: $x_1, \ldots, x_2$
- For $1 \leq i \leq m$, the value x_i occurs n_i times. // $n = n_1 + \cdots + n_m$

$$E(X) = \frac{1}{n}\sum_{i=1}^{m} x_i \cdot n_i$$

Example Roll a die an infinite number of times. Each number appears 1/6 of the time.

$$E(X) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{6} i \cdot \frac{n}{6} = \frac{7}{2}$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \leq i \leq m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \le i \le m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\mathrm{var}(X) = E\big((X - E(X))^2\big)$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \le i \le m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\text{var}(X) = E\left((X - E(X))^2\right) \quad = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \leq i \leq m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\text{var}(X) = E\left((X - E(X))^2\right) \quad = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

$$= E((X - \mu)(X - \mu))$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \le i \le m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\text{var}(X) = E\big((X - E(X))^2\big) = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

$$= E\big((X - \mu)(X - \mu)\big)$$

$$= E(X^2 - 2X\mu + \mu^2)$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \le i \le m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\text{var}(X) = E\big((X - E(X))^2\big) \quad = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

$$= E((X-\mu)(X-\mu))$$

$$= E(X^2 - 2X\mu + \mu^2)$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \le i \le m$.

The mean μ of X is

$$\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\text{var}(X) = E\big((X - E(X))^2\big) \quad = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

$$= E((X - \mu)(X - \mu))$$

$$= E(X^2 - 2X\mu + \mu^2)$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2$$

Variance and Standard Deviation

Suppose p_i is the probability of the value x_i, for $1 \leq i \leq m$.

The mean μ of X is

$$\boxed{\mu = E(X) = p_1 x_1 + \cdots + p_m x_m.}$$

The *variance* of X measures the dispersion of the values $x_1, \ldots, x_n$ around μ.

$$\boxed{\text{var}(X) = E\big((X - E(X))^2\big)} \quad = \sum_{i=1}^{m} p_i (x_i - \mu)^2$$

$$= E((X - \mu)(X - \mu))$$

$$= E(X^2 - 2X\mu + \mu^2)$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2$$

The *standard deviation* of X is $\sigma = \sqrt{\text{var}(X)}$.

Mean and Variance of a Continuous Variable

If X is a continuous variable with p.d.f. $p(X)$, its expected value is

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

Mean and Variance of a Continuous Variable

If X is a continuous variable with p.d.f. $p(X)$, its expected value is

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

Its variance is given as

$$\mathrm{var}(X) = \int_{-\infty}^{\infty} p(x)(x-\mu)^2\, dx$$

Gaussian Distribution

A continuous random variable X with mean μ and variance σ^2 has *Gaussian distribution* (or *normal distribution*) if its p.d.f. is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/\left(2\sigma^2\right)}$$

Gaussian Distribution

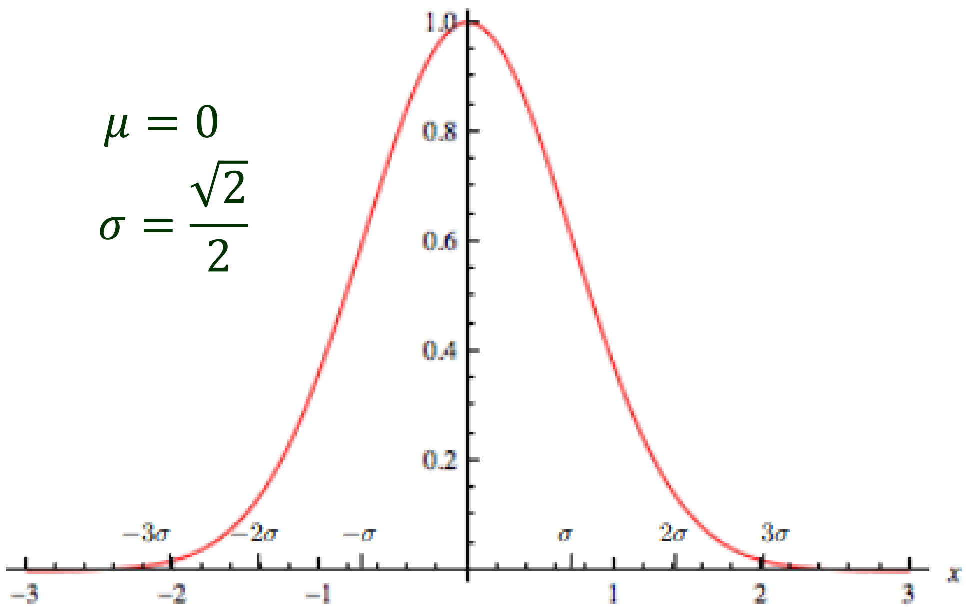
A continuous random variable X with mean μ and variance σ^2 has *Gaussian distribution* (or *normal distribution*) if its p.d.f. is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$$\sqrt{\pi}\, p(x) = e^{-x^2}$$

$$\mu = 0$$

$$\sigma = \frac{\sqrt{2}}{2}$$

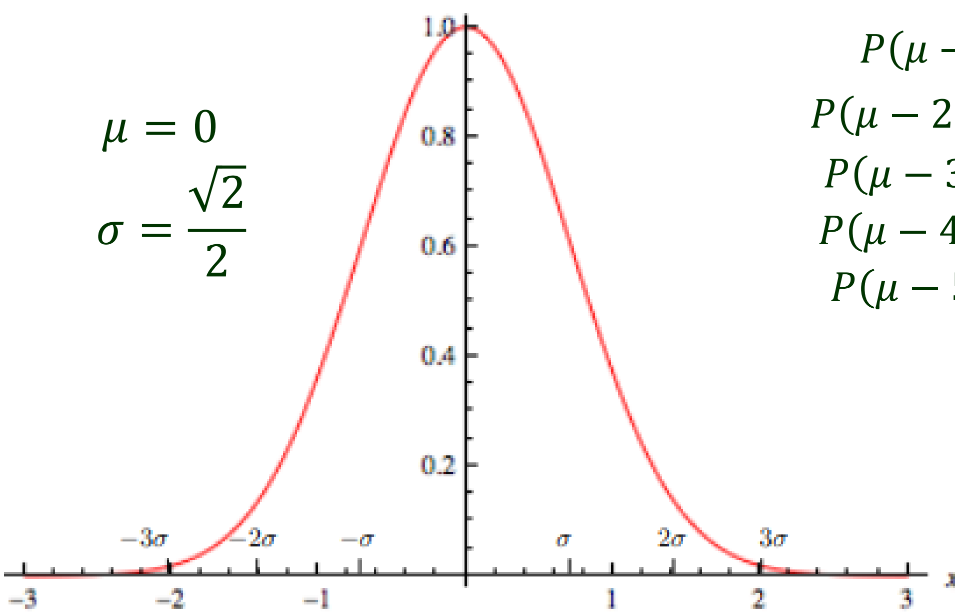


Gaussian Distribution

A continuous random variable X with mean μ and variance σ^2 has *Gaussian distribution* (or *normal distribution*) if its p.d.f. is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$\sqrt{\pi}\, p(x) = e^{-x^2}$



$\mu = 0$

$\sigma = \frac{\sqrt{2}}{2}$

$P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.6826895$

$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.9544997$

$P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.9973002$

$P(\mu - 4\sigma \le X \le \mu + 4\sigma) \approx 0.9999366$

$P(\mu - 5\sigma \le X \le \mu + 5\sigma) \approx 0.9999994$

Gaussian Distribution

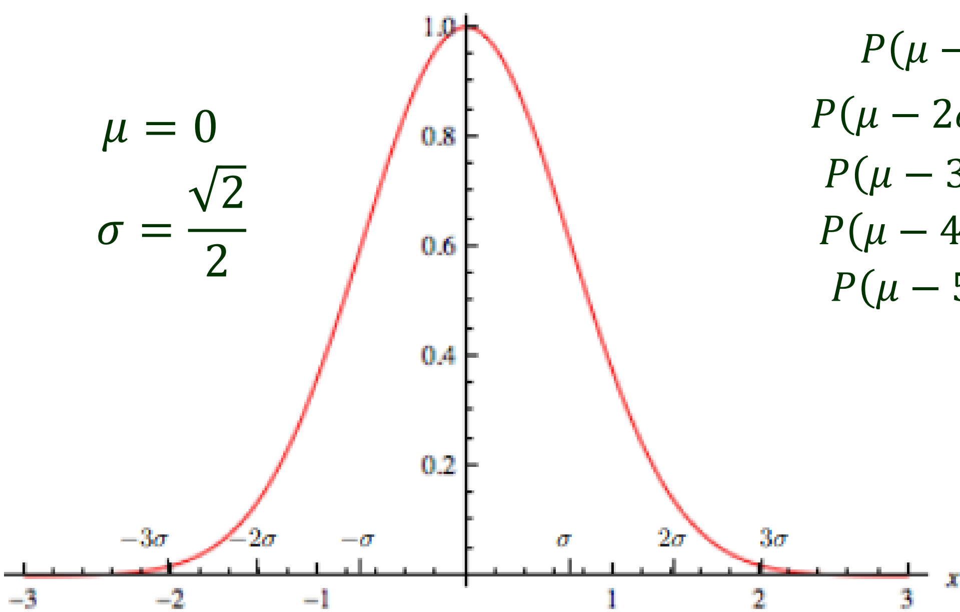
A continuous random variable X with mean μ and variance σ^2 has *Gaussian distribution* (or *normal distribution*) if its p.d.f. is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$\sqrt{\pi}\, p(x) = e^{-x^2}$

$\mu = 0$

$\sigma = \frac{\sqrt{2}}{2}$



$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.6826895$

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9544997$

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973002$

$P(\mu - 4\sigma \leq X \leq \mu + 4\sigma) \approx 0.9999366$

$P(\mu - 5\sigma \leq X \leq \mu + 5\sigma) \approx 0.9999994$

Three-sigma rule (in practice): Consider population within $(\mu - 3\sigma, \mu + 3\sigma)$.

Why Gaussian Distribution?

- It is the most important distribution because it fits many natural phenomena (e.g., human characteristics such as weight, height, body temperature, etc.)

- It is the limiting distribution of $X_1 + \cdots + X_n$ of n independent random variables $X_1, \ldots, X_n$, as $n \to \infty$, explaining a characteristic impacted by numerous independent factors (by the central limit theorem).

- It is the foundation for important methods such as least-squares, Kalman filters that are used in statistics and engineering.

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$P(\textit{cavity} \vee \textit{toothache})$

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$P(cavity \vee toothache)$

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{cavity} \vee \mathit{toothache}) = 0.108 + 0.012 + 0.072 + 0.008$$

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{cavity} \vee \mathit{toothache}) = 0.108 + 0.012 + 0.072 + 0.008$$

II. Inference by Enumeration

Probabilistic inference: compute posterior probabilities for query propositions given observed evidence.

Calculate the probability of a proposition ϕ:

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Joint distribution table:

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

$$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Marginalization

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

The unconditional (or marginal) probability of ¬*toothache*:

$P(\neg toothache)$

Marginalization

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

The unconditional (or marginal) probability of ¬*toothache*:

$P(\neg toothache)$

Marginalization

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

The unconditional (or marginal) probability of ¬*toothache*:

$$P(\neg toothache) = 0.072 + 0.144 + 0.008 + 0.576 = 0.8$$

Marginalization

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

The unconditional (or marginal) probability of ¬*toothache*:

$$P(\neg toothache) = 0.072 + 0.144 + 0.008 + 0.576 = 0.8$$

Marginalization: Sum up the probabilities for each possible value of the other variables.

$$\boldsymbol{P(Y) = \sum_z P\,(Y, Z = z)}$$

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

The unconditional (or marginal) probability of ¬*toothache*:

$$P(\neg toothache) = 0.072 + 0.144 + 0.008 + 0.576 = 0.8$$

Marginalization: Sum up the probabilities for each possible value of the other variables.

$$\boldsymbol{P(Y) = \sum_z P(Y, Z = z)}$$

Sums over all possible combinations of the values of the set of variables $\boldsymbol{Z}$.

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\begin{aligned}\boldsymbol{P}(Cavity) = \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch) \\ + \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch)\end{aligned}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\begin{aligned}\boldsymbol{P}(Cavity) = {} & \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch) \\ & + \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch)\end{aligned}$$

random variable beginning with an uppercase letter.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\boldsymbol{P}(Cavity) = \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch)$$
$$+ \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch)$$

random variable beginning with an uppercase letter.

$$= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\begin{aligned}\boldsymbol{P}(\mathit{Cavity}) &= \boldsymbol{P}(\mathit{Cavity}, \mathit{toothache}, \mathit{catch}) + \boldsymbol{P}(\mathit{Cavity}, \mathit{toothache}, \neg \mathit{catch}) \\ &\quad + \boldsymbol{P}(\mathit{Cavity}, \neg \mathit{toothache}, \mathit{catch}) + \boldsymbol{P}(\mathit{Cavity}, \neg \mathit{toothache}, \neg \mathit{catch}) \\ &= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle \\ &= \langle 0.2, 0.8 \rangle\end{aligned}$$

random variable beginning with an uppercase letter.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\begin{aligned} \boldsymbol{P}(Cavity) &= \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch) \\ &\quad + \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch) \\ &= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle \\ &= \langle 0.2, 0.8 \rangle \end{aligned}$$

random variable beginning with an uppercase letter.

Rule of *conditioning*:

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} \boldsymbol{P}(\boldsymbol{Y} \mid \boldsymbol{z}) P(\boldsymbol{z})$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\boldsymbol{P}(Cavity) = \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch)$$
$$+ \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch)$$

random variable beginning with an uppercase letter.

$$= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle$$

$$= \langle 0.2, 0.8 \rangle$$

not bold-faced

Rule of *conditioning*:

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} \boldsymbol{P}(\boldsymbol{Y} \mid \boldsymbol{z}) P(\boldsymbol{z})$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditioning

Abbreviate $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z})$ as $\boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$.

$$\boldsymbol{P}(Cavity) = \boldsymbol{P}(Cavity, toothache, catch) + \boldsymbol{P}(Cavity, toothache, \neg catch) + \boldsymbol{P}(Cavity, \neg toothache, catch) + \boldsymbol{P}(Cavity, \neg toothache, \neg catch)$$

random variable beginning with an uppercase letter.

$$= \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle$$

$$= \langle 0.2, 0.8 \rangle$$

not bold-faced

Rule of *conditioning*:

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} \underbrace{\boldsymbol{P}(\boldsymbol{Y} \mid \boldsymbol{z}) P(\boldsymbol{z})}_{= \boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(\mathit{cavity} \mid \mathit{toothache}) = \frac{P(\mathit{cavity} \wedge \mathit{toothache})}{P(\mathit{toothache})}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(\mathit{cavity} \mid \mathit{toothache}) = \frac{P(\mathit{cavity} \wedge \mathit{toothache})}{P(\mathit{toothache})}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	$\boxed{0.108}$	$\boxed{0.012}$	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(\mathit{cavity} \mid \mathit{toothache}) = \frac{P(\mathit{cavity} \wedge \mathit{toothache})}{P(\mathit{toothache})}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	$\boxed{0.108}$	$\boxed{0.012}$	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	$\boxed{0.108}$	$\boxed{0.012}$	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(\textit{cavity} \mid \textit{toothache}) = \frac{P(\textit{cavity} \wedge \textit{toothache})}{P(\textit{toothache})}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg\textit{cavity} \mid \textit{toothache}) = \frac{P(\neg\textit{cavity} \wedge \textit{toothache})}{P(\textit{toothache})}$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Computing Conditional Probabilities

1. $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $\Longrightarrow$ an expression in terms of unconditional probabilities.
2. Evaluate the expression from full joint distribution.

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.012 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)} = 0.4$$

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\mathbf{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\boldsymbol{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

$$\boldsymbol{P}(Cavity \mid toothache) = \alpha\, \boldsymbol{P}(Cavity, toothache)$$

Normalization

$$P(\mathit{cavity} \mid \mathit{toothache}) = \frac{P(\mathit{cavity} \wedge \mathit{toothache})}{P(\mathit{toothache})} = 0.6$$

$$P(\neg\mathit{cavity} \mid \mathit{toothache}) = \frac{P(\neg\mathit{cavity} \wedge \mathit{toothache})}{P(\mathit{toothache})} = 0.4$$

- $P(\mathit{toothache})$ is a normalization constant for the distribution $P(\mathit{Cavity} \mid \mathit{toothache})$ so it adds up to 1.

- We can calculate $\mathbf{P}(\mathit{Cavity} \mid \mathit{toothache})$ without $P(\mathit{toothache})$ as follows.

$$\mathbf{P}(\mathit{Cavity} \mid \mathit{toothache}) = \alpha\, \mathbf{P}(\mathit{Cavity},\ \mathit{toothache})$$

$$= \alpha\, [\mathbf{P}(\mathit{Cavity},\ \mathit{toothache}, \mathit{catch}) + \mathbf{P}(\mathit{Cavity},\ \mathit{toothache}, \neg\mathit{catch})]$$

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\mathbf{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity,\ toothache)$$

$$= \alpha\, [\mathbf{P}(Cavity,\ toothache, catch) + \mathbf{P}(Cavity,\ toothache, \neg catch)]$$

$$= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\mathbf{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

$$\begin{aligned}
\mathbf{P}(Cavity \mid toothache) &= \alpha\, \mathbf{P}(Cavity,\ toothache) \\
&= \alpha\, [\mathbf{P}(Cavity,\ toothache, catch) + \mathbf{P}(Cavity,\ toothache, \neg catch)] \\
&= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] \\
&= \alpha\, \langle 0.12, 0.008 \rangle
\end{aligned}$$

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\mathbf{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

$$\begin{aligned}
\mathbf{P}(Cavity \mid toothache) &= \alpha\, \mathbf{P}(Cavity, toothache) \\
&= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] \\
&= \alpha\, \langle 0.12, 0.008 \rangle \\
&= \langle 0.6, 0.4 \rangle
\end{aligned}$$

Normalization

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = 0.4$$

- $P(toothache)$ is a normalization constant for the distribution $P(Cavity \mid toothache)$ so it adds up to 1.

- We can calculate $\mathbf{P}(Cavity \mid toothache)$ without $P(toothache)$ as follows.

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity,\ toothache)$$

$$= \alpha\, [\mathbf{P}(Cavity,\ toothache, catch) + \mathbf{P}(Cavity,\ toothache, \neg catch)]$$

$$= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$

$$= \alpha\, \langle 0.12, 0.008\rangle$$

$$= \langle 0.6, 0.4\rangle$$

// $\alpha = 5 = 1/P(toothache)$.

Inference Procedure

- X: single query variable (e.g., *Cavity*).
- $\boldsymbol{E}$: evidence variables (e.g., *Toothache*).
- $\boldsymbol{e}$: their observed values.
- $\boldsymbol{Y}$: unobserved variables (e.g., *Catch*).

Inference Procedure

- X: single query variable (e.g., *Cavity*).
- $\boldsymbol{E}$: evidence variables (e.g., *Toothache*).
- $\boldsymbol{e}$: their observed values.
- $\boldsymbol{Y}$: unobserved variables (e.g., *Catch*).

$$\boldsymbol{P}(X \mid \boldsymbol{e}) \leftarrow \alpha \boldsymbol{P}(X \mid \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

♣ Summation is over all possible combinations $\boldsymbol{y}$ of values of variables in $\boldsymbol{Y}$.

♣ $\boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$ is a subset of probabilities from the full joint distribution.

♣ The full joint distribution has size exponential in # variables and is rarely computed.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

⇨ A 32-element table.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\boldsymbol{P}(Toothache, Catch, Cavity, Weather)$$

$$P(toothache, catch, cavity, cloud)$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$\Rightarrow$ A 32-element table.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

$$P(toothache, catch, cavity, cloud)$$

$$= P(cloud \mid toothache, catch, cavity)\, P(toothache, catch, cavity)$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

⇒ A 32-element table.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

$$P(toothache, catch, cavity, cloud)$$

$$= P(cloud \mid toothache, catch, cavity)\, P(toothache, catch, cavity)$$

$$\Downarrow \quad P(cloud \mid toothache, catch, cavity) = P(cloud)$$

$$= P(cloud)\, P(toothache, catch, cavity)$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

A 32-element table.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

$$P(toothache, catch, cavity, cloud)$$

$$= P(cloud \mid toothache, catch, cavity)\, P(toothache, catch, cavity)$$

$$\Downarrow \quad P(cloud \mid toothache, catch, cavity) = P(cloud)$$

$$= P(cloud)\, P(toothache, catch, cavity)$$

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\, \mathbf{P}(Weather)$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

⇒ A 32-element table.

Independence

Add a fourth variable *Weather* with domain { *sun*, *rain*, *cloud*, *snow* }.

$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$

$$P(toothache, catch, cavity, cloud)$$

$$= P(cloud \mid toothache, catch, cavity)\, P(toothache, catch, cavity)$$

$$\Downarrow \quad P(cloud \mid toothache, catch, cavity) = P(cloud)$$

$$= P(cloud\,)\, P(toothache, catch, cavity)$$

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\, \mathbf{P}(Weather)$$

8-element table + 4-element table

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$\Rightarrow$ ~~A 32-element table.~~

Independent Variables

- Two propositions a and b are *independent* if

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

Independent Variables

- Two propositions a and b are *independent* if

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

- Two variables X and Y are *independent* if

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X) \quad \text{or} \quad \boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y) \quad \text{or} \quad \boldsymbol{P}(X \wedge Y) = \boldsymbol{P}(X)P(Y)$$

Independent Variables

- Two propositions a and b are *independent* if

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

- Two variables X and Y are *independent* if

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X) \quad \text{or} \quad \boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y) \quad \text{or} \quad \boldsymbol{P}(X \wedge Y) = \boldsymbol{P}(X)P(Y)$$

- Two continuous variables X and Y are *independent* if, for all $x, y \in \mathbb{R}$,

$$P(X \leq x \wedge Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

Independent Variables

- Two propositions a and b are *independent* if

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

- Two variables X and Y are *independent* if

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X) \quad \text{or} \quad \boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y) \quad \text{or} \quad \boldsymbol{P}(X \wedge Y) = \boldsymbol{P}(X)P(Y)$$

- Two continuous variables X and Y are *independent* if, for all $x, y \in \mathbb{R}$,

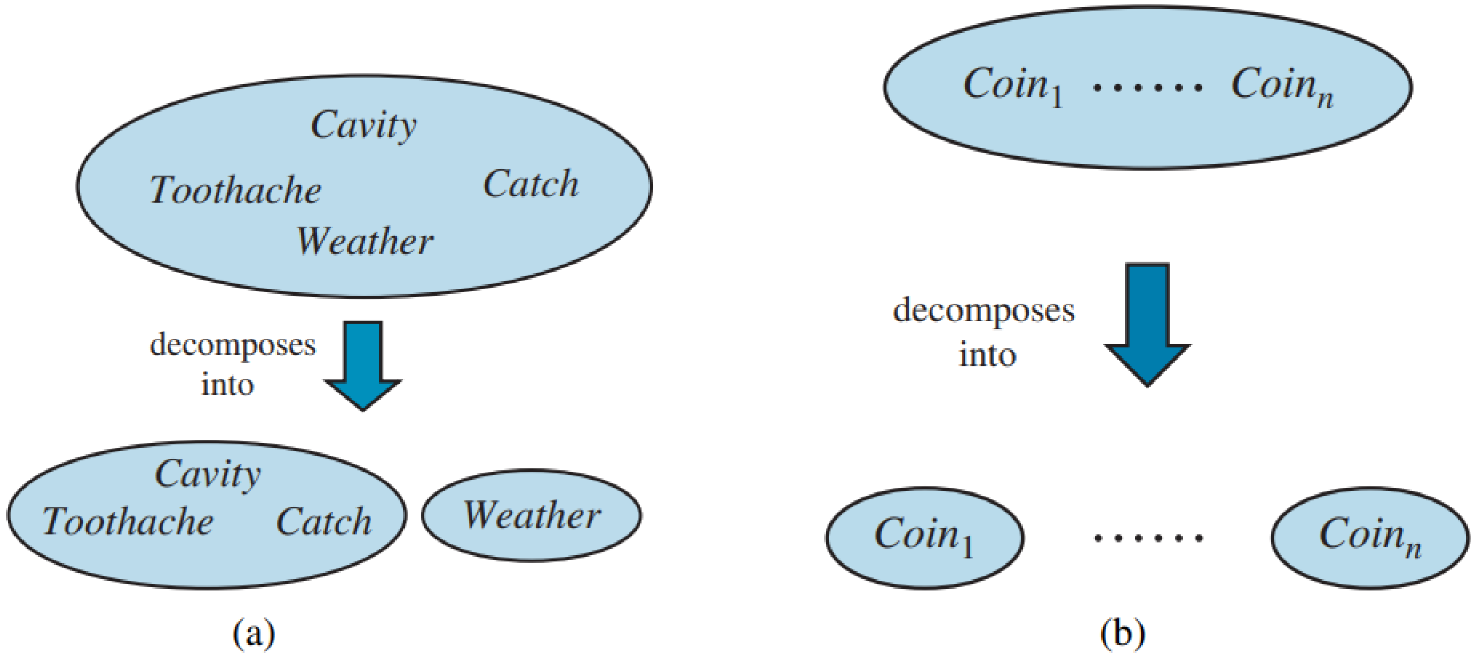
$$P(X \leq x \wedge Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

The joint probability density function $p(x, y)$ satisfies

$$p(x, y) = \int_{-\infty}^{\infty} p(x, y)\, dy \int_{-\infty}^{\infty} p(x, y)\, dx$$

Factoring a Joint Distribution

The full joint distribution can be factored into *separate* joint distributions on subsets of variables that are *independent*.



Weather and dental problems are independent.

Coin flips are independent.