Some Concepts of Probability (Review)

Volker Tresp

Summer 2014

Summary

- Conditional probability:

$$P(y|x) = \frac{P(x,y)}{P(x)} \quad \text{with} \quad P(x) > 0$$

- Product rule

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

- Chain rule

$$P(x_1,\ldots,x_M) = P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)\ldots P(x_M|x_1,\ldots,x_{M-1})$$

- Bayes theorem

$$P(y|x) = \frac{P(x,y)}{P(x)} = \frac{P(x|y)P(y)}{P(x)} \quad P(x) > 0$$

- Marginal distribution

$$P(x) = \sum_y P(x, y)$$

- Independent random variables

$$P(x, y) = P(x)P(y|x) = P(x)P(y)$$

Discrete Random Variables

- A **random variable** $X(c)$ is a variable (more precisely a function), whose value depends on the result of a random process

- Examples:

  – c is a coin toss and $X(c) = 1$ if the result is head

  – c is a person, randomly selected from the University of Munich. $X(c)$ is the height of that person

- A **discrete random variable** X can only assume a countable number of states. Thus $X = x$ with $x \in \{x_1, x_2, \ldots\}$

Discrete Random Variables (2)

- A probability distribution specifies with which probability a random variable assumes a particular state
- A probability distribution of X can be defined via a **probability function** $f(x)$:

$$P(X = x) = P(\{c : X(c) = x\}) = f(x)$$

- $f(x)$ is the probability function and x is a realisation of X
- One often writes

$$f(x) = P_X(x) = P(x)$$

Elementary / Atomic Events

- In statistics, one attempts to derive the probabilities from data (machine learning)
- In probability one assumes either that some probabilities are known, or that they can be derived from some atomic events
- **Atomic event**: using some basic assumptions (symmetry, neutrality of nature, fair coin, ...) one assumes the probabilities for some elementary events

Example: Toss of a Fair Coin

- Atomic events: $c = \{h, t\}$
- The probability of each elementary event is $1/2$
- $X(c)$ is a random variable that is equal to one if the result is head and is zero otherwise
- $P(X = 1) = 0.5$

Example: Two Tosses of a Fair Coin

- Atomic event: results of two tosses

- $c = \{(hh), (ht), (th), (tt)\}$

- The probability for each elementary event is $1/4$

- $X(c)$ is a random variable that is equal to one if the result of the first is head

- $P(X = 1) = 1/4P(h, h) + 1/4P(h, t) = 1/2$

Multivariate Probability Distributions

- Define two random variables $X(c)$ and $Y(c)$. A **multivariate distribution** is defined as:

$$P(x, y) = P(X = x, Y = y) = P(X = x \wedge Y = y)$$

$$= P(\{c : X(c) = x \wedge Y(c) = y\})$$

Example: Two Coin Tosses

- $X(c)$ is a random variable that is equal to one if the result of the first toss is head
- $Y(c)$ is a random variable that is equal to one if the result of the second toss is head
- $P(X = 1) = 1/2$, $P(Y = 1) = 1/2$,

$$P(X = 1, Y = 1) = P(X = 1, Y = 1) = P(X = 1 \wedge Y = 1)$$

$$= P(\{c : X(c) = 1 \wedge Y(c) = 1\}) = 1/4$$

Special Cases

- If two random variables are independent, then $P(X, Y) = P(X)P(Y)$. Example: the two random variables in the previous example
- Some states of random variables can be mutually exclusive:
  - $Z(c)$ is a random variable that is equal to one if the result of the first toss is tail $P(X = 1, Z = 1) = 0$
- Mutual exclusive and collectively exhaustive: if two random variables are mutually exclusive and, in addition, $P(X = 1) + P(Z = 1) = 1$. Examples: everyone has exactly one height

Which Random Variables?

- It should be clear from the discussion that the definition of random variables in a domain is up to the researcher, although there is often a "natural" choice (height of a person, income of a person, age of a person, ...)
- Example: In a dice experiment, we can define binary random variables for each of the six results $X_1, \ldots, X_6$ or we define only one random variable with six states $X = x$, with $x \in \{1, 2, 3, 4, 5, 6\}$

Conditional Distribution

- I am interested in the probability distribution of the random variable Y but consider only atomic events, where $X = x$. Example: I am interested in the probability distribution of height, but only for teenagers

- *Definition* of a **conditional probability distribution**

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} \text{ with } P(X = x) > 0$$

- The distribution is identical to the one for the unconditional case, only that I have to divide by $P(X = x)$ (re-normalize)

Product Rule and Chain Rule

- It follows: **product rule**

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

- and **chain rule**

$$P(x_1, \ldots, x_M) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots P(x_M|x_1, \ldots, x_{M-1})$$

Bayes Theorem

- **Bayes Theorem**

$$P(y|x) = \frac{P(x,y)}{P(x)} = \frac{P(x|y)P(y)}{P(x)} \quad P(x) > 0$$

Disjunction

- Recall that $P(X = x, Y = y) = P(X = x \wedge Y = y)$ is the probability for a **conjunction**

- We get for the **disjunction**

$$P(X = x \ \vee \ Y = y) =$$

$$P(X = x, Y = y) + P(X = x, Y \neq y) + P(X \neq x, Y = y) =$$

$$[P(X = x, Y = y) + P(X = x, Y \neq y)] + [P(X = x, Y = y) + P(X \neq x, Y = y)]$$

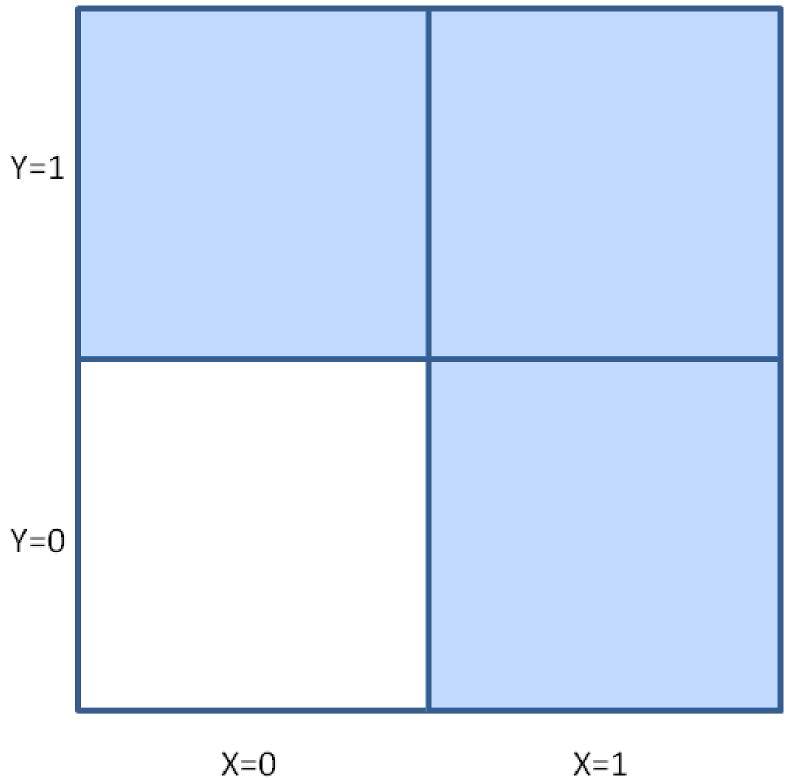
$$-P(X = x, Y = y)$$

$$= P(X = x) + P(Y = y) - P(X = x, Y = y)$$

  is the probability for a disjunction

- If states are **mutually exclusive**, $P(X = x, Y = y) = 0$, and

$$P(X = x \vee Y = y) = P(X = x) + P(Y = y)$$



Disjunction

Marginal Distribution

- The **marginal distribution** can be calculated from a joint disctribution as:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Marginalization and Conditioning: Basis for Probabilistic Inference

- $P(I, F, S)$ where $I = 1$ stands for influenza, $F = 1$ stands for fever, $S = 1$ stands for sneezing
- What is the probability for influenza, when the patient is sneezing, but temperature is unknown?
- Thus I need (conditioning) $P(I = 1|S = 1) = P(I = 1, S = 1)/P(S = 1)$
- I calculate via marginalization

$$P(I = 1, S = 1) = \sum_f P(I = 1, F = f, S = 1)$$

$$P(S = 1) = \sum_i P(I = i, S = 1)$$

Independent Random Variables

- **Independence**: two random variables are independent, if,

$$P(x, y) = P(x)P(y|x) = P(x)P(y)$$

- It follows for independent random variables,

$$P(X = x \vee Y = y) = P(X = x) + P(Y = y) - P(X = x)P(Y = y)$$

Expected Values

- **Expected value**

$$E(X) = E_{P(x)}(X) = \sum_i x_i P(X = x_i)$$

Variance

- The **Variance** of a random variable is:

$$var(X) = \sum_i (x_i - E(X))^2 P(X = x_i)$$

Covariance

- **Covariance**:

$$cov(X,Y) = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))P(X = x_i, Y = y_j)$$

- **Covariance matrix**:

$$\Sigma_{[XY],[XY]} = \begin{pmatrix} var(X) & cov(X,Y) \\ cov(Y,X) & var(Y) \end{pmatrix}$$

Correlation

- Useful identity:

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

where $E(XY)$ is the **correlation**.

Correlation coefficient (confusing naming!) is

$$r = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

Continuous Random Variables

- **Probability density**

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$$

- Thus

$$P(a < x < b) = \int_a^b f(x)dx$$

- The **distribution function** is

$$F(x) = \int_{-\infty}^{x} f(x)dx = P(X \leq x)$$

Expectations for Continuous Variables

- Expected value

$$E(X) = E_{P(x)}(X) = \int xP(x)dx$$

- Variance

$$var(X) = \int (x - E(x))^2 P(x)dx$$

- Covariance:

$$cov(X, Y) = \int (x - E(X))(y - E(Y))P(x, y)dxdy$$