Computer vision: models, learning and inference

Chapter 2
Introduction to probability

Please send errata to s.prince@cs.ucl.ac.uk
Random variables

- A random variable $x$ denotes a quantity that is uncertain.
- May be result of experiment (flipping a coin) or a real world measurements (measuring temperature).
- If observe several instances of $x$ we get different values.
- Some values occur more than others and this information is captured by a probability distribution.
Discrete Random Variables

![Histogram of face value of biased die](image)

- Rain
- Drizzle
- Cloud
- Snow
- Sleet
- Sun
- Wind
Joint Probability

• Consider two random variables $x$ and $y$
• If we observe multiple paired instances, then some combinations of outcomes are more likely than others
• This is captured in the joint probability distribution
• Written as $Pr(x, y)$
• Can read $Pr(x, y)$ as “probability of $x$ and $y$”
Joint Probability

a)  

b)  

c)  

d)  

e)  

f)  

Computer vision: models, learning and inference. ©2011 Simon J.D. Prince
Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables.

\[ Pr(x) = \int Pr(x, y) \, dy \]

\[ Pr(y) = \int Pr(x, y) \, dx \]
Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

\[ Pr(x) = \sum_y Pr(x, y) \]
\[ Pr(y) = \sum_x Pr(x, y) \]
Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables.

\[
Pr(x) = \sum_y Pr(x, y)
\]

\[
Pr(y) = \int Pr(x, y) \, dx
\]
Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables.

\[ Pr(x) = \int Pr(x, y) \, dy \]
\[ Pr(y) = \int Pr(x, y) \, dx \]

Works in higher dimensions as well – leaves joint distribution between whatever variables are left.

\[ Pr(x, y) = \sum_w \int Pr(w, x, y, z) \, dz \]
Conditional Probability

- Conditional probability of $x$ given that $y=y_1$ is relative propensity of variable $x$ to take different outcomes given that $y$ is fixed to be equal to $y_1$.

- Written as $\Pr(x | y=y_1)$
Conditional Probability

- Conditional probability can be extracted from joint probability
- Extract appropriate slice and normalize

\[
Pr(x|y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*)dx} = \frac{Pr(x, y = y^*)}{Pr(y = y^*)}
\]
Conditional Probability

\[ Pr(x|y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*) \, dx} = \frac{Pr(x, y = y^*)}{Pr(y = y^*)} \]

• More usually written in compact form

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

• Can be re-arranged to give

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

\[ Pr(x, y) = Pr(y|x)Pr(x) \]
Conditional Probability

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

• This idea can be extended to more than two variables

\[
Pr(w, x, y, z) = Pr(w, x, y|z)Pr(z) \\
= Pr(w, x|y, z)Pr(y|z)Pr(z) \\
= Pr(w|x, y, z)Pr(x|y, z)Pr(y|z)Pr(z)
\]
Bayes’ Rule

From before:

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

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

Combining:

\[ Pr(y|x)Pr(x) = Pr(x|y)Pr(y) \]

Re-arranging:

\[
Pr(y|x) = \frac{Pr(x|y)Pr(y)}{Pr(x)}
= \frac{Pr(x|y)Pr(y)}{\int Pr(x, y) \, dy}
= \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y) \, dy}
\]
Bayes’ Rule Terminology

Likelihood – propensity for observing a certain value of $x$ given a certain value of $y$

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y) \, dy}$$

Prior – what we know about $y$ before seeing $x$

Posterior – what we know about $y$ after seeing $x$

Evidence – a constant to ensure that the left hand side is a valid distribution
Independence

If two variables $x$ and $y$ are independent then variable $x$ tells us nothing about variable $y$ (and vice-versa)

\[
Pr(x \mid y) = Pr(x) \\
Pr(y \mid x) = Pr(y)
\]
Independence

• If two variables $x$ and $y$ are independent then variable $x$ tells us nothing about variable $y$ (and vice-versa)

$$Pr(x|y) = Pr(x)$$
$$Pr(y|x) = Pr(y)$$
Independence

- When variables are independent, the joint factorizes into a product of the marginals:

\[
Pr(x, y) = Pr(x|y)Pr(y) = Pr(x)Pr(y)
\]
Expectation

Expectation tell us the expected or average value of some function $f[x]$ taking into account the distribution of $x$.

Definition:

$$E[f[x]] = \sum_x f[x]Pr(x)$$

$$E[f[x]] = \int f[x]Pr(x) \, dx$$
Expectation

Expectation tells us the expected or average value of some function $f(x)$ taking into account the distribution of $x$.

Definition in two dimensions:

$$E[f(x, y)] = \int \int f(x, y) Pr(x, y) \, dx \, dy$$
Expectation: Common Cases

\[ E[f[x]] = \int f[x] Pr(x) \, dx \]

<table>
<thead>
<tr>
<th>Function ( f[\bullet] )</th>
<th>Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>mean, ( \mu_x )</td>
</tr>
<tr>
<td>( x^k )</td>
<td>( k^{th} ) moment about zero</td>
</tr>
<tr>
<td>( (x - \mu_x)^k )</td>
<td>( k^{th} ) moment about the mean</td>
</tr>
<tr>
<td>( (x - \mu_x)^2 )</td>
<td>variance</td>
</tr>
<tr>
<td>( (x - \mu_x)^3 )</td>
<td>skew</td>
</tr>
<tr>
<td>( (x - \mu_x)^4 )</td>
<td>kurtosis</td>
</tr>
<tr>
<td>( (x - \mu_x)(y - \mu_y) )</td>
<td>covariance of ( x ) and ( y )</td>
</tr>
</tbody>
</table>
Rule 1:

Expected value of a constant is the constant

\[
E[\kappa] = \kappa
\]
Expectation: Rules

\[ E[f(X)] = \int f(x) Pr(X = x) \, dx \]

Rule 2:

Expected value of constant times function is constant times expected value of function

\[ E[kf(x)] = kE[f(x)] \]
Expectation: Rules

\[ E[f[X]] = \int f(x) Pr(X = x) dx \]

Rule 3:

Expectation of sum of functions is sum of expectation of functions

\[ E[f[x] + g[x]] = E[f[x]] + E[g[x]] \]
Rule 4:

Expectation of product of functions in variables $x$ and $y$ is product of expectations of functions if $x$ and $y$ are independent.

$$E[f(X)] = \int f(x) Pr(X = x) \, dx$$

$$E[f(x)g(y)] = E[f(x)]E[g(y)] \quad \text{if } x, y \text{ independent}$$
Conclusions

• Rules of probability are compact and simple

• Concepts of marginalization, joint and conditional probability, Bayes rule and expectation underpin all of the models in this book

• One remaining concept – conditional expectation – discussed later