Distribution Representations

MLAI Lecture 2

Neil D. Lawrence

Department of Computer Science
Sheffield University

27th September 2012

Outline

Distribution Representation

Probability Density Functions

Sample Based Approximations

Distribution Representation

- We can represent probabilities as tables

y	0	1	2
$P(y)$	0.2	0.5	0.3

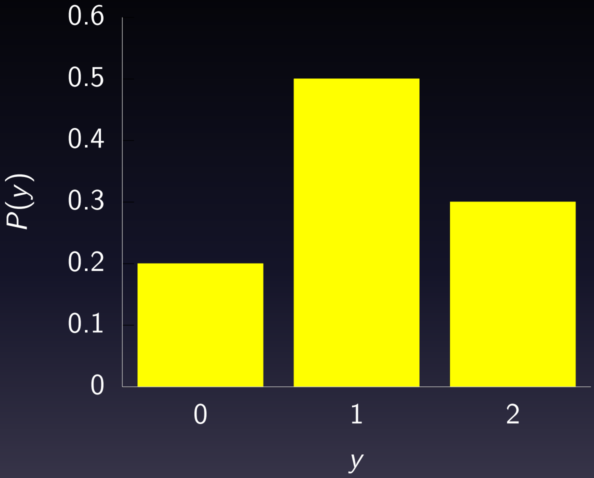

Figure: Histogram representation of the simple distribution.

Expectations of Distributions

- Writing down the entire distribution is tedious.
- Can summarise through expectations.

$$\langle f(y) \rangle_{P(y)} = \sum_{y} f(y) p(y)$$

- Consider:

y	0	1	2
$P(y)$	0.2	0.5	0.3

- We have $\langle y \rangle_{P(y)} = 0.2 \times 0 + 0.5 \times 1 + 0.3 \times 2 = 1.1$
- This is the *first moment* or mean of the distribution.

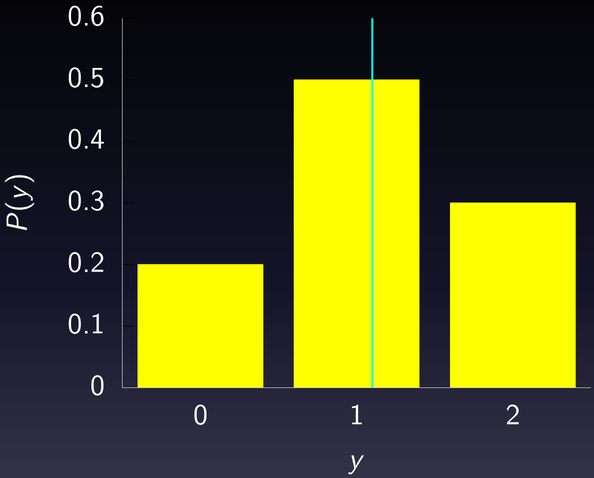

Figure: Histogram representation of the simple distribution including the expectation of y (red line), the mean of the distribution.

Variance and Standard Deviation

- Mean gives us the centre of the distribution.
- Consider:

y	0	1	2
y^2	0	1	4
$P(y)$	0.2	0.5	0.3

- *Second moment* is $\left\langle y^2 \right\rangle_{P(y)} = 0.2 \times 0 + 0.5 \times 1 + 0.3 \times 4 = 1.7$
- Variance is $\left\langle y^2 \right\rangle - \left\langle y \right\rangle^2 = 1.7 - 1.1 \times 1.1 = 0.49$
- Standard deviation is square root of variance.
- Standard deviation gives us the "width" of the distribution.

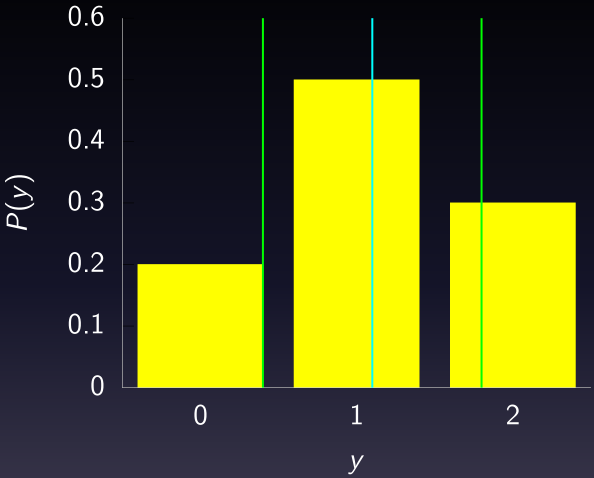

Figure: Histogram representation of the simple distribution including lines at one standard deviation from the mean of the distribution (green lines).

Outline

Distribution Representation

Probability Density Functions

Sample Based Approximations

Continuous Variables

- So far discrete values of x or y.
- For continuous models we use the *probability density function* (PDF).
- Discrete case: defined probability distributions over a discrete number of states.
- How do we represent continuous as probability?
- Student heights:
  - Develop a representation which could answer *any* question we chose to ask about a student's height.
- PDF is a positive function, integral over the region of interest is one[1].

[1] In what follows we shall use the word distribution to refer to both discrete probabilities and continuous probability density functions.

Manipulating PDFs

- Same rules for PDFs as distributions *e.g.*

$$p(y|x) = \frac{p(x|y)\,p(y)}{p(x)}$$

  where $p(x, y) = p(x|y)\,p(y)$ and for continuous variables $p(x) = \int p(x, y)\,\mathrm{d}y$.

- Expectations under a PDF

$$\langle f(x) \rangle_{p(x)} = \int f(x)\,p(x)\,\mathrm{d}x$$

  where the integral is over the region for which our PDF for x is defined.

The Gaussian Density

- Perhaps the most common probability density.

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$
$$= \mathcal{N}\left(y|\mu, \sigma^2\right)$$

- Also available in multivariate form.
- First proposed maybe by de Moivre but also used by Laplace.

Gaussian PDF I

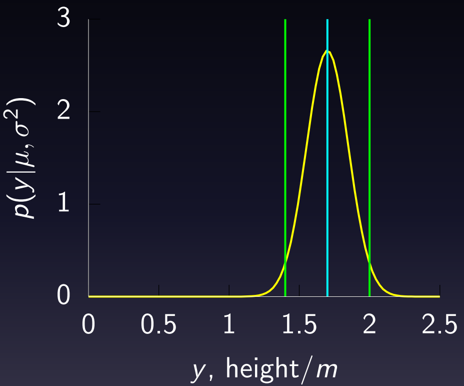

Figure: The Gaussian PDF with $\mu = 1.7$ and variance $\sigma^2 = 0.0225$. Mean shown as red line. Two standard deviations are shown as magenta. It could represent the heights of a population of students.

Cumulative Distribution Functions

- PDF doesn't represent probabilities directly
- One very common question is: what is the probability that $x < y$?
- The cumulative distribution function (CDF) represents the answer for $-\infty < x < \infty$ the CDF is given by

$$P\left(x > y\right) = \int_{-\infty}^{y} p\left(x\right) \mathrm{d}x,$$

for $0 \leq x < \infty$ then the CDF is given by

$$P\left(x > y\right) = \int_{0}^{y} p\left(x\right) \mathrm{d}x.$$

Gaussian PDF and CDF

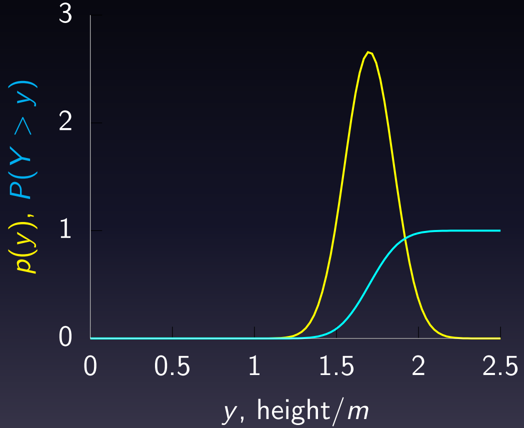

Figure: The cumulative distribution function (CDF) for the heights of computer science students. The thick curve gives the CDF and the thinner curve the associated PDF.

PDF from CDF

- The PDF can be recovered from the CDF through differentiation.

Outline

Distribution Representation

Probability Density Functions

Sample Based Approximations

Sample Based Approximations I

- It is not always possible to compute expectations directly.
- Sample based approximation

$$\left\langle f\left(y\right)\right\rangle_{P(y)} \approx \frac{1}{N}\sum_{i=1}^{N} f\left(y_i\right).$$

- Special cases of this include the *sample mean*, often denoted by $\bar{y}$, and computed as

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

Sample Mean vs True Mean

- This is an approximation to the true distribution mean

$$\langle y \rangle \approx \bar{y}.$$

- The same approximations can used for continuous PDFs, so we have

$$\begin{aligned}\langle f(y) \rangle_{p(y)} &= \int f(y)\, p(y)\, \mathrm{d}y \\ &\approx \frac{1}{N} \sum_{i=1}^{N} f(y_i),\end{aligned}$$

  where y_i are independently obtained samples from the density $p(y)$.
- Approximation gets better for increasing N and worse if the samples from $P(y)$ are *not* independent.

Expectation Computation Example

- Consider the following distribution.

y	1	2	3	4
$P(y)$	0.3	0.2	0.1	0.4

- What is the mean of the distribution?
- What is the standard deviation of the distribution?
- Are the mean and standard deviation representative of the distribution form?
- What is the expected value of $-\log P(y)$?

Expectation Computation Example

- Consider the following distribution.

y	1	2	3	4
$P(y)$	0.3	0.2	0.1	0.4

- What is the mean of the distribution?
- What is the standard deviation of the distribution?
- Are the mean and standard deviation representative of the distribution form?
- What is the expected value of $-\log P(y)$?

Expectation Computation Example

- Consider the following distribution.

y	1	2	3	4
$P(y)$	0.3	0.2	0.1	0.4

- What is the mean of the distribution?
- What is the standard deviation of the distribution?
- Are the mean and standard deviation representative of the distribution form?
- What is the expected value of $-\log P(y)$?

Expectation Computation Example

- Consider the following distribution.

y	1	2	3	4
$P(y)$	0.3	0.2	0.1	0.4

- What is the mean of the distribution?
- What is the standard deviation of the distribution?
- Are the mean and standard deviation representative of the distribution form?
- What is the expected value of $-\log P(y)$?

Expectations Example: Answer

- We are given that:

y	1	2	3	4
$P(y)$	0.3	0.2	0.1	0.4
y^2	1	4	9	16
$-\log(P(y))$	1.204	1.609	2.302	0.916

- Mean: $1 \times 0.3 + 2 \times 0.2 + 3 \times 0.1 + 4 \times 0.4 = 2.6$
- Second moment: $1 \times 0.2 + 4 \times 0.2 + 9 \times 0.1 + 16 \times 0.4 = 8.4$
- Variance: $8.4 - 2.6 \times 2.6 = 1.64$
- Standard deviation: $\sqrt{1.64} = 1.2806$
- Expectation $-\log(P(y))$:
  $0.3 \times 1.204 + 0.2 \times 1.609 + 0.1 \times 2.302 + 0.4 \times 0.916 = 1.280$

Sample Based Approximation Example

- You are given the following values samples of heights of students,

i	1	2	3	4	5	6
y_i	1.76	1.73	1.79	1.81	1.85	1.80

- What is the sample mean?
- What is the sample variance?
- Can you compute sample approximation expected value of $-\log P(y)$?
- Actually these "data" were sampled from a Gaussian with mean 1.7 and standard deviation 0.15. Are your estimates close to the real values? If not why not?

Sample Based Approximation Example

- You are given the following values samples of heights of students,

i	1	2	3	4	5	6
y_i	1.76	1.73	1.79	1.81	1.85	1.80

- What is the sample mean?
- What is the sample variance?
- Can you compute sample approximation expected value of $-\log P(y)$?
- Actually these "data" were sampled from a Gaussian with mean 1.7 and standard deviation 0.15. Are your estimates close to the real values? If not why not?

Sample Based Approximation Example

- You are given the following values samples of heights of students,

i	1	2	3	4	5	6
y_i	1.76	1.73	1.79	1.81	1.85	1.80

- What is the sample mean?
- What is the sample variance?
- Can you compute sample approximation expected value of $-\log P(y)$?
- Actually these "data" were sampled from a Gaussian with mean 1.7 and standard deviation 0.15. Are your estimates close to the real values? If not why not?

Sample Based Approximation Example

- You are given the following values samples of heights of students,

i	1	2	3	4	5	6
y_i	1.76	1.73	1.79	1.81	1.85	1.80

- What is the sample mean?
- What is the sample variance?
- Can you compute sample approximation expected value of $-\log P(y)$?
- Actually these "data" were sampled from a Gaussian with mean 1.7 and standard deviation 0.15. Are your estimates close to the real values? If not why not?

Sample Based Approximation Example: Answer

- We can compute:

i	1	2	3	4	5	6
y_i	1.76	1.73	1.79	1.81	1.85	1.80
y_i^2	3.0976	2.9929	3.2041	3.2761	3.4225	3.2400

- Mean: $\frac{1.76+1.73+1.79+1.81+1.85+1.80}{6} = 1.79$
- Second moment: $\frac{3.0976+2.9929+3.2041+3.2761+3.4225+3.2400}{6} = 3.2055$
- Variance: $3.2055 - 1.79 \times 1.79 = 1.43 \times 10^{-3}$
- Standard deviation: 0.0379
- No, you can't compute it. You don't have access to $P(y)$ directly.

Reading and Exercises

- Read and *understand* Bishop on:
  - Probability densities: Section 1.2.1 (Pages 17–19).
  - Expectations and Covariances: Section 1.2.2 (Pages 19–20).
  - The Gaussian density: Part of Section 1.2.4 (Pages 24–25).
- Look at exercises:
  - Exercise 1.7
  - Exercise 1.8
- Complete Exercise:
  - Exercise 1.9

References I

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006. [Google Books] .