ECE 275A: Parameter Estimation I
Random Vectors and Bayes Rule

Florian Meyer

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

0

Discrete Random Vectors

- $\boldsymbol{x} \in \mathbb{R}^n$ denotes a *random vector*
- $\boldsymbol{x}$ can take on a countable number of values in $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_I\}$
- $p_{\boldsymbol{x}}(\boldsymbol{x}_i)$, or $p(\boldsymbol{x}_i)$, is the *probability* that the random vector $\boldsymbol{x}$ takes on value $\boldsymbol{x}_i$
- $p(\cdot)$ is called *probability mass function (pmf)*
- Example: If x is the outcome of a dice roll, we have $\mathcal{X} = \{1, 2, \ldots, 6\}$ and
  $$p(x_i) = 1/6, \forall x_i \in \mathcal{X}$$

1

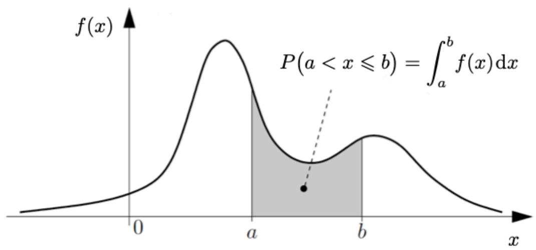
1

Continuous Random Vectors

- Random vector $\boldsymbol{x} \in \mathbb{R}^n$ takes on values in the continuum
- $f_{\boldsymbol{x}}(\boldsymbol{x})$, or $f(\boldsymbol{x})$, is its ***probability density function (pdf)***, *i.e.,*

$$P(\boldsymbol{x} \in \mathcal{R}) = \int_{\mathcal{R}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- Example:



2

2

Joint and Conditional Distributions

- $p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}, \boldsymbol{y})$ or $p(\boldsymbol{x}, \boldsymbol{y})$ is the joint pmf of random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$
- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are *independent* then

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y})$$

- $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ is the probability of $\boldsymbol{x}$ given (conditioned on) $\boldsymbol{y}$

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x}, \boldsymbol{y})/p(\boldsymbol{y}) \qquad p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})$$

- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are *independent* then

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x}) \qquad p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y})$$

- Equivalent expressions exist for the pdfs of continuous random vectors

3

3

Law of Total Probabilities, Marginals

Discrete Case

$$\sum_{\boldsymbol{x} \in \mathcal{X}} p(\boldsymbol{x}) = 1$$

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathcal{Y}} p(\boldsymbol{x}, \boldsymbol{y})$$

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathcal{Y}} p(\boldsymbol{x}|\boldsymbol{y})\, p(\boldsymbol{y})$$

Continuous Case

$$\int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$$

$$f(\boldsymbol{x}) = \int f(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

$$f(\boldsymbol{x}) = \int f(\boldsymbol{x}|\boldsymbol{y})\, f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

Bayes Rule

- Recall $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})\, p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})$

- It therefore follows that

likelihood (y is fixed) prior

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

evidence

- Equivalent expressions exist for the pdfs of continuous random vectors

Normalization

- For $\boldsymbol{y}$ observed and thus fixed

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

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

$$\propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})$$

- The constant C ensures that $p(\boldsymbol{x}|\boldsymbol{y})$ sums to one and can be calculated as

$$C = \frac{1}{\sum_{\boldsymbol{x} \in \mathcal{X}} p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})}$$

- Equivalent expressions exist for the pdfs of continuous random vectors

6

6

Conditioning

- Law of total probability

$$p(\boldsymbol{x}|\boldsymbol{z}) = \int p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z})\, \mathrm{d}\boldsymbol{y}$$
$$= \int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z})\, p(\boldsymbol{y}|\boldsymbol{z})\, \mathrm{d}\boldsymbol{y}$$

- Bayes rule with background knowledge

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

- Equivalent expressions exist for the pdfs of continuous random vectors

7

7

Conditional Independence

- Condition on $\boldsymbol{z}$, random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent if

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

- This is equivalent to

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

and

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

- Equivalent expressions exist for the pdfs of continuous random vectors

8

8

Expectation

- Expectation of a random vector $\boldsymbol{x}$

discrete case

$$\mu_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\} = \sum_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}\, p(\boldsymbol{x})$$

continuous case

$$\mu_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\} = \int \boldsymbol{x}\, f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

- Expectation of transformed random vector $g(\boldsymbol{x})$

$$\mathbb{E}\{g(\boldsymbol{x})\} = \sum_{\boldsymbol{x}\in\mathcal{X}} g(\boldsymbol{x})\, p(\boldsymbol{x}) \qquad \mathbb{E}\{g(\boldsymbol{x})\} = \int g(\boldsymbol{x})\, f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

9

Covariance

- Covariance of a random vector $\boldsymbol{x}$

discrete case

$$\mathbb{E}\{(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{x}-\boldsymbol{\mu_x})^{\mathrm{T}}\}$$
$$=\sum_{\boldsymbol{x}\in\mathcal{X}}(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{x}-\boldsymbol{\mu_x})^{\mathrm{T}}p(\boldsymbol{x})$$

continuous case

$$\mathbb{E}\{(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{x}-\boldsymbol{\mu_x})^{\mathrm{T}}\}$$
$$=\int(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{x}-\boldsymbol{\mu_x})^{\mathrm{T}}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

- Cross-Covariance of random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\mathbb{E}\{(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{y}-\boldsymbol{\mu_y})^{\mathrm{T}}\}$$
$$=\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{\boldsymbol{y}\in\mathcal{Y}}(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{y}-\boldsymbol{\mu_y})^{\mathrm{T}}p(\boldsymbol{x},\boldsymbol{y})$$

$$\mathbb{E}\{(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{y}-\boldsymbol{\mu_y})^{\mathrm{T}}\}$$
$$=\int(\boldsymbol{x}-\boldsymbol{\mu_x})(\boldsymbol{y}-\boldsymbol{\mu_y})^{\mathrm{T}}f(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}$$

10