Random Variables
and
Random Vectors

Random variables
• Samples from a random variable are real numbers
  – A random variable is associated with a probability distribution
    over these real values
  – Two types of random variables
    • Discrete
      – Only finitely many possible values for the random variable:
        \(X \in \{a_1, a_2, \ldots, a_n\}\)
      – (Could also have a countable infinity of possible values)
      – E.g., the random variable could take any positive integer value
      – Each possible value has a finite probability of occurring.
    • Continuous
      – Infinitely many possible values for the random variable
      – E.g., \(X \in \text{Real numbers}\)

Discrete random variables
• Discrete random variables have a pmf (probability mass function), \(P\)
  \(P(X = a) = P(a)\)
• Example: Coin flip
  \(X = 0\) if heads
  \(X = 1\) if tails
  – What is the pmf of this random variable?

Continuous random variables
• Continuous random variables have a pdf (probability density function), \(p\)
• Example: Uniform distribution
  \(p(1.3) = ?\)  \(p(2.4) = ?\)
  What is the probability that \(X = 1.3\) exactly:
  \(P(X = 1.3) = ?\)
  Probability corresponds to area under the pdf.
  \(P(1 < X < 1.5) = \int_{1}^{1.5} p(x) dx = 0.25\)

Good Review Materials

http://www.imageprocessingbook.com/DIP2E/dip2e_downloads/review_material_downloads.htm
• (Gonzales & Woods review materials)
• Chapt. 1: Linear Algebra Review
• Chapt. 2: Probability, Random Variables, Random Vectors
Continuous random variables

• What is the total area under any pdf? 
  \[ \int_{-\infty}^{\infty} p(x) dx = 1 \]

• Example continuous random variable: Human heights

<table>
<thead>
<tr>
<th>p(h)</th>
<th>p(h)</th>
<th>p(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>h</td>
<td>h</td>
</tr>
</tbody>
</table>

Random variables

• How much change do you have on you?
  – Asking a person (chosen at random) that question can be thought of as sampling from a random variable.

• Is the random variable “Amount of change people carry” discrete or continuous?

Random variables: Mean & Variance

• These formulas can be used to find the mean and variance of a random variable when its true probability distribution is known.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Discrete r.v.</th>
<th>Continuous r.v.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \mu )</td>
<td>( \mu = \sum a_i P(a_i) )</td>
<td>( \mu = \int_{-\infty}^{\infty} x \cdot p(x) dx )</td>
</tr>
<tr>
<td>Variance ( \text{Var}(X) )</td>
<td>( \sum (a_i - \mu)^2 P(a_i) )</td>
<td>( \int_{-\infty}^{\infty} (x - \mu)^2 \cdot p(x) dx )</td>
</tr>
</tbody>
</table>

The Gaussian distribution

\[ p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]

\( X \sim \text{N}(\mu, \sigma^2) \)

An important type of random variable

Estimating the Mean & Variance

– After sampling from a random variable \( n \) times, these formulas can be used to estimate the mean and variance of the random variable.

• Samples \( x_1, x_2, x_3, \ldots, x_n \)

Estimated mean:

\[ m = \frac{1}{n} \sum_{i=1}^{n} x_i \]

Estimated variance:

\[ \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2 \]

← maximum likelihood estimate

\[ \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2 \]

← unbiased estimate
Finding mean, variance in Matlab

- Samples \( x = [x_1, x_2, \ldots, x_n] \)
- Mean
  \[
  m = \left( \frac{1}{n} \right) \sum x
  \]
- Variance
  \[
  \sigma^2 = \frac{1}{n} [x_1 - m \ x_2 - m \ \ldots \ x_n - m]
  \]
  \[
  \begin{bmatrix}
  x_1 - m \\
  x_2 - m \\
  \vdots \\
  x_n - m
  \end{bmatrix}
  \]

Method 1:
\[
\triangleright v = (\frac{1}{n}) \ast (x - m) \ast (x - m)'
\]
Method 2:
\[
\triangleright z = x - m \\
\triangleright v = (\frac{1}{n}) \ast z \ast z'
\]

Example continuous random variable

- People’s heights (made up)
  - Gaussian
    \( \mu = 67, \sigma^2 = 20 \)
  - What if you went to a planet where heights Gaussian
    \( \mu = 75, \sigma^2 = 5 \)
  - How would they be different from us?

Example continuous random variable

- Time people woke up this morning
  - Gaussian
    \( \mu = 9, \sigma^2 = 1 \)

Random vectors

- An \( n \)-dimensional random vector consists of \( n \) random variables that are all associated with the same events.
- Example 2-D random vector:
  \[
  V = \begin{bmatrix}
  X \\
  Y
  \end{bmatrix}
  \]
  where \( X \) is random variable of human heights
  \( Y \) is random variable of wake-up times
- Sample \( n \) times from \( V \).
  \[
  \begin{bmatrix}
  v_1 & v_2 & \ldots & v_n \\
  x_1 & x_2 & \ldots & x_n \\
  y_1 & y_2 & \ldots & y_n
  \end{bmatrix}
  \]

Let’s collect some samples and graph them:

\( y \) (wake-up times)
\( x \) (heights)

Mean of a random vector

- Estimating the mean of a random vector
  - \( n \) samples from \( V \)
  \[
  \begin{bmatrix}
  v_1 & v_2 & \ldots & v_n \\
  x_1 & x_2 & \ldots & x_n \\
  y_1 & y_2 & \ldots & y_n
  \end{bmatrix}
  \]

\[
\text{Mean } m = \frac{1}{n} \sum v = \frac{1}{n} \sum x = \frac{[m_x]}{[m_y]}
\]

- To estimate mean of \( V \) in Matlab
  \[
  \triangleright (\frac{1}{n}) \ast \sum (v, 2)
  \]
Random vector

- Example 2-D random vector:
  \[
  \mathbf{V} = \begin{bmatrix} X \\ Y \end{bmatrix}
  \]
  where \( X \) is random variable of human **heights**  
  \( Y \) is random variable of human **weights**

- Sample \( n \) times from \( \mathbf{V} \):
  \[
  \mathbf{v}_1 \mathbf{v}_2 \ldots \mathbf{v}_n
  \]
- What will graph look like?

Covariance of two random variables

- Height and wake-up time are uncorrelated, but height and weight are correlated.

- Covariance
  \[
  \text{Cov}(X, Y) = 0 \quad \text{for} \ X = \text{height}, \ Y = \text{wake-up times}
  \]
  \[
  \text{Cov}(X, Y) > 0 \quad \text{for} \ X = \text{height}, \ Y = \text{weight}
  \]
  - Definition:
    \[
    \text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]
    \]
  If \( \text{Cov}(X, Y) < 0 \) for two random variables \( X, Y \), what would a scatterplot of samples from \( X, Y \) look like?

Estimating covariance from samples

- Sample \( n \) times:
  \[
  \begin{bmatrix} x_1 \ x_2 \ldots \ x_n \\ y_1 \ y_2 \ldots \ y_n \end{bmatrix}
  \]
  \[
  \text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_x)(y_i - m_y)
  \]
  ← maximum likelihood estimate
  \[
  \text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m_x)(y_i - m_y)
  \]
  ← unbiased estimate
- \( \text{Cov}(X, X) = \text{Var}(X) \)
- How are \( \text{Cov}(X, Y) \) and \( \text{Cov}(Y, X) \) related?
  \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)

Estimating covariance in Matlab

- **Samples**
  \[
  x = [x_1 \ x_2 \ldots \ x_n] \\
  y = [y_1 \ y_2 \ldots \ y_n]
  \]
- **Means**
  \[
  m_x \leftarrow \bar{x} \quad m_y \leftarrow \bar{y}
  \]
- **Covariance**
  \[
  \text{Cov}(X, Y) = \frac{1}{n} [x_1 - m_x \ x_2 - m_x \ldots \ x_n - m_x] \cdot [y_1 - m_y \ y_2 - m_y \ldots \ y_n - m_y]
  \]
  - Method 1:
    \[
    \mathbf{v} = (1/n) \cdot (x - \bar{x}) \cdot (y - \bar{y})'
    \]
  - Method 2:
    \[
    \mathbf{w} = x - \bar{x} \\
    \mathbf{z} = y - \bar{y} \\
    \mathbf{v} = (1/n) \cdot \mathbf{w} \cdot \mathbf{z}'
    \]

Covariance matrix of a \( D \)-dimensional random vector

- In 2 dimensions
  \[
  \text{Cov}(\mathbf{V}) = E[(\mathbf{V} - \mu)(\mathbf{V} - \mu)']
  \]
  \[
  = E\left[\begin{bmatrix} X - \mu_X \\ Y - \mu_Y \end{bmatrix}[X - \mu_X \ Y - \mu_Y]'ight] = \begin{bmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{bmatrix}
  \]
- In \( D \) dimensions
  \[
  \text{Cov}(\mathbf{V}) = E[(\mathbf{V} - \mu)(\mathbf{V} - \mu)']
  \]
- When is a covariance matrix symmetric?
  A. always, B. sometimes, or C. never

Example covariance matrix

- People’s heights (made up)
  \( X \sim N(67, 20) \)
- Time people woke up this morning
  \( Y \sim N(9, 1) \)
- What is the covariance matrix of
  \[
  \mathbf{V} = \begin{bmatrix} X \\ Y \end{bmatrix}
  \]
  \[
  \begin{bmatrix} 20 & 0 \\ 0 & 1 \end{bmatrix}
  \]
Estimating the covariance matrix from samples (including Matlab code)

- Sample \( n \) times and find mean of samples
  \[
  \mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}
  \implies \mathbf{m} = \frac{1}{n} \sum \mathbf{v} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}
  \]

  - Find the covariance matrix
  \[
  \text{Cov} = \frac{1}{n} \begin{bmatrix}
  x_1 - m_1 & y_1 - m_2 \\
  x_2 - m_1 & y_2 - m_2 \\
  \vdots & \vdots \\
  x_n - m_1 & y_n - m_2
  \end{bmatrix}
  \]

  \[
  \begin{align*}
  & \gg m = (1/n) * \text{sum}(v,2) \\
  & \gg z = v - \text{repmat}(m,1,n) \\
  & \gg v = (1/n)*z*z' \\
  \end{align*}
  \]

Gaussian distribution in \( D \) dimensions

- 1-dimensional Gaussian is completely determined by its mean, \( \mu \), and variance, \( \sigma^2 \):
  \[
  X \sim \mathcal{N}(\mu, \sigma^2) \implies p(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
  \]

- \( D \)-dimensional Gaussian is completely determined by its mean, \( \mu \), and covariance matrix, \( \Sigma \):
  \[
  X \sim \mathcal{N}(\mu, \Sigma) \implies p(x) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}
  \]

- What happens when \( D = 1 \) in the Gaussian formula?