Independence of random variables

- **Definition**
  Random variables $X$ and $Y$ are independent if their joint distribution function factors into the product of their marginal distribution functions
  \[ F_{X,Y}(x,y) = F_X(x)F_Y(y) \]

- **Theorem**
  Suppose $X$ and $Y$ are jointly continuous random variables. $X$ and $Y$ are independent if and only if given any two densities for $X$ and $Y$ their product is the joint density for the pair $(X,Y)$ i.e.
  \[ f_{X,Y}(x,y) = f_X(x)f_Y(y) \]
  **Proof:**

- If $X$ and $Y$ are independent random variables and $Z = g(X)$, $W = h(Y)$ then $Z$, $W$ are also independent.
Example

• Suppose $X$ and $Y$ are discrete random variables whose values are the non-negative integers and their joint probability function is

$$p_{x,y}(x, y) = \frac{1}{x! y!} \lambda^x \mu^y e^{-(\lambda+\mu)} \quad x, y = 0, 1, 2, \ldots$$

Are $X$ and $Y$ independent? What are their marginal distributions?

• Factorization is enough for independence, but we need to be careful of constant terms for factors to be marginal probability functions.
Example and Important Comment

- The joint density for $X, Y$ is given by
  \[ f_{X,Y}(x, y) = \begin{cases} 
  4(x + y^2) & x, y > 0, \ x + y \leq 1 \\
  0 & \text{otherwise} 
  \end{cases} \]

- Are $X, Y$ independent?

- Independence requires that the set of points where the joint density is positive must be the Cartesian product of the set of points where the marginal densities are positive i.e. the set of points where $f_{X,Y}(x,y) > 0$ must be (possibly infinite) rectangles.
Conditional densities

- If $X, Y$ jointly distributed continuous random variables, the *conditional density function* of $Y \mid X$ is defined to be
  
  $$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

  if $f_X(x) > 0$ and 0 otherwise.

- If $X, Y$ are independent then $f_{Y \mid X}(y \mid x) = f_Y(y)$.

- Also,

  $$f_{X,Y}(x, y) = f_{Y \mid X}(y \mid x)f_X(x)$$

  Integrating both sides over $x$ we get

  $$f_Y(y) = \int_{-\infty}^{\infty} f_{Y \mid X}(y \mid x)f_X(x)dx$$

- This is a useful application of the law of total probability for the continuous case.
Example

• Consider the joint density

\[
f_{X,Y}(x, y) = \begin{cases} 
\lambda^2 e^{-\lambda y} & 0 \leq x \leq y \\
0 & \text{otherwise}
\end{cases}
\]

• Find the conditional density of \(X\) given \(Y\) and the conditional density of \(Y\) given \(X\).
Properties of Expectations Involving Joint Distributions

• For random variables $X, Y$ and constants $a, b \in R$

  $$E(aX + bY) = aE(X) + bE(Y)$$

  **Proof:**

• For independent random variables $X, Y$

  $$E(XY) = E(X)E(Y)$$

  whenever these expectations exist.

  **Proof:**
Covariance

• Recall: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \ E[(X-E(X))(Y-E(Y))]$

• Definition
  For random variables $X, Y$ with $E(X), E(Y) < \infty$, the covariance of $X$ and $Y$ is
  \[
  \text{Cov}(X, Y) = E[(X-E(X))(Y-E(Y))]
  \]

• Covariance measures whether or not $X-E(X)$ and $Y-E(Y)$ have the same sign.

• Claim: \[
  \text{Cov}(X, Y) = E(XY) - E(X)E(Y)
  \]
  Proof:

• Note: If $X, Y$ independent then $E(XY) = E(X)E(Y)$, and $\text{Cov}(X, Y) = 0$. 
Example

- Suppose $X, Y$ are discrete random variables with probability function given by

<table>
<thead>
<tr>
<th>$y$</th>
<th>$x$</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>$p_X(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1/8</td>
<td>0</td>
<td>1/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p_Y(y)$

- Find $\text{Cov}(X,Y)$. Are $X, Y$ independent?
Important Facts

- Independence of $X, Y$ implies $\text{Cov}(X,Y) = 0$ but NOT vice versa.

- If $X, Y$ independent then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

- If $X, Y$ are NOT independent then
  $$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

- $\text{Cov}(X,X) = \text{Var}(X)$. 
Example

- Suppose $Y \sim \text{Binomial}(n, p)$. Find $\text{Var}(Y)$.
Properties of Covariance

For random variables $X, Y, Z$ and constants $a, b, c, d \in R$

- $\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$
- $\text{Cov}(X+Y, Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
Correlation

- **Definition**
  For $X, Y$ random variables the *correlation* of $X$ and $Y$ is
  \[
  \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}
  \]
  whenever $V(X), V(Y) \neq 0$ and all these quantities exists.

- **Claim:**
  \[
  \rho(aX+b, cY+d) = \rho(X, Y)
  \]
  **Proof:**

- This claim means that the correlation is scale invariant.
Theorem

• For $X, Y$ random variables, whenever the correlation $\rho(X, Y)$ exists it must satisfy

$$-1 \leq \rho(X, Y) \leq 1$$

Proof:
Interpretation of Correlation $\rho$

- $\rho(X,Y)$ is a measure of the strength and direction of the linear relationship between $X$, $Y$.

- If $X$, $Y$ have non-zero variance, then $\rho \in [-1,1]$.

- If $X$, $Y$ independent, then $\rho(X,Y) = 0$. Note, it is not the only time when $\rho(X,Y) = 0$ !!!

- $Y$ is a linearly increasing function of $X$ if and only if $\rho(X,Y) = 1$.

- $Y$ is a linearly decreasing function of $X$ if and only if $\rho(X,Y) = -1$. 
Example

• Find $\text{Var}(X - Y)$ and $\rho(X,Y)$ if $X, Y$ have the following joint density

$$f_{X,Y}(x, y) = \begin{cases} 
3x & 0 \leq y \leq x \leq 1 \\
0 & \text{otherwise}
\end{cases}$$
**Conditional Expectation**

- For $X, Y$ discrete random variables, the conditional expectation of $Y$ given $X = x$ is

$$E(Y \mid X = x) = \sum_y y \cdot p_{y\mid x}(y \mid x)$$

and the conditional variance of $Y$ given $X = x$ is

$$V(Y \mid X = x) = \sum_y \left( y - E(Y \mid X = x) \right)^2 \cdot p_{y\mid x}(y \mid x)$$

$$= E(Y^2 \mid X = x) - (E(Y \mid X = x))^2$$

where these are defined only if the sums converges absolutely.

- In general,

$$E(h(Y) \mid X = x) = \sum_y h(y) \cdot p_{y\mid x}(y \mid x)$$
• For $X, Y$ continuous random variables, the conditional expectation of $Y$ given $X = x$ is

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y \mid X}(y \mid x)dy$$

and the conditional variance of $Y$ given $X = x$ is

$$V(Y \mid X = x) = \int_{-\infty}^{\infty} [y - E(Y \mid X = x)]^2 \cdot f_{Y \mid X}(y \mid x)dy$$

$$= E(Y^2 \mid X = x) - (E(Y \mid X = x))^2$$

• In general,

$$E(h(Y) \mid X = x) = \int_{-\infty}^{\infty} h(y) \cdot f_{Y \mid X}(y \mid x)dy$$
Example

- Suppose $X, Y$ are continuous random variables with joint density function

$$f_{X,Y}(x, y) = \begin{cases} e^{-y} & y > 0, 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Find $E(X \mid Y = 2)$. 
More about Conditional Expectation

- Assume that $E(Y | X = x)$ exists for every $x$ in the range of $X$. Then, $E(Y | X)$ is a random variable. The expectation of this random variable is

$$E[E(Y | X)]$$

- **Theorem**

$$E[E(Y | X)] = E(Y)$$

This is called the “Law of Total Expectation”.

**Proof:**
Example

• Suppose we roll a fair die; whatever number comes up we toss a coin that many times. What is the expected number of heads?
Theorem

• For random variables $X, Y$

\[ V(Y) = V[E(Y|X)] + E[V(Y|X)] \]

Proof:
Example

- Let $X \sim \text{Geometric}(p)$.
  Given $X = x$, let $Y$ have conditionally the Binomial$(x, p)$ distribution.
- Scenario: doing Bernoulli trails with success probability $p$ until 1st success so $X$ : number of trails. Then do $x$ more trails and count the number of success which is $Y$.
- Find, $E(Y), V(Y)$. 
Law of Large Numbers

• Toss a coin $n$ times.

• Suppose

$$X_i = \begin{cases} 
1 & \text{if } i^{th} \text{ toss came up } H \\
0 & \text{if } i^{th} \text{ toss came up } T 
\end{cases}$$

• $X_i$’s are Bernoulli random variables with $p = \frac{1}{2}$ and $E(X_i) = \frac{1}{2}$.

• The proportion of heads is $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

• Intuitively $\bar{X}_n$ approaches $\frac{1}{2}$ as $n \to \infty$. 
Markov’s Inequality

• If $X$ is a non-negative random variable with $E(X) < \infty$ and $a > 0$ then,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof:
Chebyshev’s Inequality

• For a random variable $X$ with $E(X) < \infty$ and $V(X) < \infty$, for any $a > 0$

\[ P\left(|X - E(X)| \geq a\right) \leq \frac{V(X)}{a^2} \]

• Proof:
Back to the Law of Large Numbers

• Interested in sequence of random variables $X_1, X_2, X_3, \ldots$ such that the random variables are independent and identically distributed (i.i.d).

  Let

  \[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \]

  Suppose $E(X_i) = \mu, V(X_i) = \sigma^2$, then

  \[
  E(\bar{X}_n) = E\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu
  \]

  and

  \[
  V(\bar{X}_n) = V\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}
  \]

• Intuitively, as $n \to \infty$, $V(\bar{X}_n) \to 0$ so $\bar{X}_n \to E(\bar{X}_n) = \mu$
• Formally, the Weak Law of Large Numbers (WLLN) states the following:

• Suppose $X_1, X_2, X_3, \ldots$ are i.i.d with $E(X_i) = \mu < \infty$, $V(X_i) = \sigma^2 < \infty$, then for any positive number $a$

$$P\left(\left|\bar{X}_n - \mu\right| \geq a\right) \to 0$$

as $n \to \infty$.

This is called Convergence in Probability.

Proof:
Example

• Flip a coin 10,000 times. Let

\[ X_i = \begin{cases} 1 & \text{if } i^{th} \text{ toss came up } H \\ 0 & \text{if } i^{th} \text{ toss came up } T \end{cases} \]

• \( E(X_i) = \frac{1}{2} \) and \( V(X_i) = \frac{1}{4} \).
• Take \( a = 0.01 \), then by Chebyshev’s Inequality

\[
P\left( \left| \frac{\bar{X}_n - 1}{2} \right| \geq 0.01 \right) \leq \frac{1}{4(10,000)} \cdot \frac{1}{(0.01)^2} = \frac{1}{4} \]

• Chebyshev Inequality gives a very weak upper bound.
• Chebyshev Inequality works regardless of the distribution of the \( X_i \)’s.
Strong Law of Large Number

• Suppose $X_1, X_2, X_3, \ldots$ are i.i.d with $E(X_i) = \mu < \infty$, then $\bar{X}_n$ converges to $\mu$ as $n \to \infty$ with probability 1. That is

$$P\left(\lim_{n \to \infty} \frac{1}{n} (X_1 + X_2 + \cdots + X_n) = \mu\right) = 1$$

• This is called convergence almost surely.