Joint Distribution of Two or More Random Variables

- Sometimes more than one measurement in the form of random variable is taken on each member of the sample space. In cases like this there will be a few random variables defined on the same probability space and we would like to explore their joint distribution.

- Joint behavior of 2 random variable (continuous or discrete), \( X \) and \( Y \) determined by their joint cumulative distribution function

\[
F_{X,Y}(x, y) = P(X \leq x, Y \leq y).
\]

- \( n \) – dimensional case

\[
F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).
\]
Discrete case

- Suppose $X$, $Y$ are discrete random variables defined on the same probability space.

- The joint probability mass function of 2 discrete random variables $X$ and $Y$ is the function $p_{X,Y}(x,y)$ defined for all pairs of real numbers $x$ and $y$ by
  \[
  p_{X,Y}(x,y) = P(X = x \text{ and } Y = y)
  \]

- For a joint pmf $p_{X,Y}(x,y)$ we must have: $p_{X,Y}(x,y) \geq 0$ for all values of $x,y$ and
  \[
  \sum_{x} \sum_{y} p_{X,Y}(x,y) = 1
  \]

- For any region $A$ in the $xy$ plane,
  \[
  P[(X,Y) \in A] = \sum_{A} \sum_{X,Y} p_{X,Y}(x,y)
  \]
Example for illustration

• Toss a coin 3 times. Define, $X$: number of heads on 1st toss, $Y$: total number of heads.
• The sample space is $\Omega = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$.
• We display the joint distribution of $X$ and $Y$ in the following table

Can we recover the probability mass function for $X$ and $Y$ from the joint table?

• To find the probability mass function of $X$ we sum the appropriate rows of the table of the joint probability function.
• Similarly, to find the mass function for $Y$ we sum the appropriate columns.
Marginal Probability Function

• The *marginal* probability mass function for $X$ is
  
  $$p_X(x) = \sum_y p_{X,Y}(x, y)$$

• The *marginal* probability mass function for $Y$ is
  
  $$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

• Case of several discrete random variables is analogous.
Example

• Roll a die twice. Let $X$: number of 1’s and $Y$: total of the 2 dice. There is no available form of the joint mass function for $X$, $Y$. We display the joint distribution of $X$ and $Y$ with the following table.

• The marginal probability mass function of $X$ and $Y$ are

• Find $P(X \leq 1 \text{ and } Y \leq 4)$
Conditional Probability Distribution

• Given the joint pmf of $X$ and $Y$, we want to find

$$P(X = x \mid Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}$$

and

$$P(Y = y \mid X = x) = \frac{P(X = x \text{ and } Y = y)}{P(X = x)}$$

• Back to the example on slide 5,

$$P(Y = 2 \mid X = 1) = 0$$

$$P(Y = 3 \mid X = 1) = \frac{\frac{2}{36}}{\frac{10}{36}} = \frac{1}{5}$$

$$\vdots$$

$$P(Y = 12 \mid X = 1) = 0$$

• These 11 probabilities give the conditional pmf of $Y$ given $X = 1$. 
### Definition

- For $X, Y$ discrete random variables with joint pmf $p_{X,Y}(x,y)$ and marginal mass functions $p_X(x)$ and $p_Y(y)$. If $x$ is a number such that $p_X(x) > 0$, then the *conditional pmf* of $Y$ given $X = x$ is

$$p_{Y|X}(y \mid x) = p_{Y|X}(y \mid X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

- Is this a valid pmf?

- Similarly, the conditional pmf of $X$ given $Y = y$ is

$$p_{X|Y}(x \mid y) = p_{X|Y}(x \mid Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$
The Joint Distribution of two Continuous R.V’s

• **Definition**
  Random variables $X$ and $Y$ are (jointly) **continuous** if there is a non-negative function $f_{X,Y}(x,y)$ such that
  \[
P((X,Y) \in A) = \iint_A f_{X,Y}(x,y) \, dx \, dy
  \]
  for any “reasonable” 2-dimensional set $A$.

• $f_{X,Y}(x,y)$ is called a joint density function for $(X, Y)$.

• In particular, if $A = \{(X, Y): X \leq x, Y \leq y\}$, the joint CDF of $X,Y$ is
  \[
  F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) \, du \, dv
  \]

• From Fundamental Theorem of Calculus we have
  \[
  f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = \frac{\partial^2}{\partial y \partial x} F_{X,Y}(x,y)
  \]
Properties of joint density function

- \( f_{X,Y}(x, y) \geq 0 \) for all \( x, y \in \mathbb{R} \)

- It’s integral over \( \mathbb{R}^2 \) is

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = 1
\]
Example

• Consider the following bivariate density function

\[ f_{X,Y}(x, y) = \begin{cases} \frac{12}{7}(x^2 + xy) & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

• Check if it’s a valid density function.

• Compute \( P(X > Y) \).
Marginal Density Functions

• The marginal density of $X$ is then

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

• Similarly the marginal density of $Y$ is

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

• Consider the following bivariate density function

\[ f_{X,Y}(x, y) = \begin{cases} 
6xy^2 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\
0 & \text{otherwise}
\end{cases} \]

• Check if it is a valid density function.

• Find the marginal densities of \( X \) and \( Y \).
Example

Consider the joint density

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

where \( \lambda \) is a positive parameter.

- Check if it is a valid density.

- Find the marginal densities of \( X \) and \( Y \).
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.

- Similarly, the conditional density of $X \mid Y$ is given by

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

if $f_Y(y) > 0$ and 0 otherwise.
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 \(X\).
Independence of Random Variables

- **Definition**
  Random variables $X$ and $Y$ are *independent* if the events $(X \in A)$ and $(Y \in B)$ are independent.

- **Theorem:** Two discrete random variables $X$ and $Y$ with joint pmf $p_{X,Y}(x,y)$ and marginal mass function $p_X(x)$ and $p_Y(y)$, are independent if and only if
  \[ p_{X,Y}(x,y) = p_X(x)p_Y(y) \]

- **Note:** If $X, Y$ are independent random variables then $P_{X|Y}(x|y) = P_X(x)$.

- **Question:** Back to the rolling die 2 times example, are $X$ and $Y$ independent?
• **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.

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

• If $X$, $Y$ are independent then

$$f_{Y|X}(y \mid x) = f_Y(y)$$
Example

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

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

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.
Expectation

• In the long run, rolling a die repeatedly what average result do you expect?

• In 6,000,000 rolls expect about 1,000,000 1’s, 1,000,000 2’s etc.

\[
\text{Average is: } \frac{1,000,000(1) + 1,000,000(2) + \cdots + 1,000,000(6)}{6,000,000} = 3.5
\]

• For a random variable \( X \), the Expectation (or expected value or mean) of \( X \) is the expected average value of \( X \) in the long run. It is also called the mean of the probability distribution of \( X \).

• Symbols: \( \mu \), \( \mu_X \), \( E(X) \) and \( EX \).
Expectation of Random Variable

• For a discrete random variable $X$ with pmf $p_X(x)$

$$E(X) = \sum_{x} x \cdot p(x)$$

whenever the sum converge absolutely (i.e. $\sum_{x} |x| \cdot p(x) < \infty$)

• For a continuous random variable $X$ with density $f_X(x)$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx$$

whenever this integral converge absolutely.
Examples

1) Roll a die. Let $X$ = outcome on 1 roll. Then $E(X) = 3.5$.

2) Bernoulli trials $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Then
   
   $$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

3) $X \sim \text{Uniform}(a, b)$. Then…

4) $X$ is a random variable with density

   $$f_X(x) = \begin{cases} 
   x^{-2} & \text{for } x > 1 \\
   0 & \text{otherwise}
   \end{cases}$$

   (i) Check if this is a valid density.
   (ii) Find $E(X)$. 
Theorem

For $g: R \rightarrow R$

- If $X$ is a discrete random variable then
  \[
  E[g(X)] = \sum_x g(x)p_X(x)
  \]

- If $X$ is a continuous random variable
  \[
  E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx
  \]

- Note, this theorem can be generalized to bivariate distributions.
Examples

1. Suppose $X \sim \text{Uniform}(0, 1)$. Let $Y = X^2$ then,

2. Suppose $X$ has the following probability mass function

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

Let $Y = e^X$, then
Properties of Expectation

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

- $E(aX + b) = aE(X) + b$
  
  Proof: Continuous case

- If $X$ is a non-negative random variable, then $E(X) = 0$ if and only if $X = 0$ with probability 1.

- If $X$ is a non-negative random variable, then $E(X) \geq 0$

- $E(a) = a$

- $E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$
Variance

- The expected value of a random variable $E(X)$ is a measure of the “center” of a distribution.

- The *variance* is a measure of how closely concentrated to center ($\mu$) the probability is. It is also called 2nd central moment.

- **Definition**
  The variance of a random variable $X$ is
  \[
  Var(X) = E[(X - E(X))^2] = E[(X - \mu)^2]
  \]

- **Claim:** $Var(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$
  Proof:

- We can use the above formula for convenience of calculation.

- The *standard deviation* of a random variable $X$ is denoted by $\sigma_X$; it is the square root of the variance i.e. $\sigma_X = \sqrt{Var(X)}$. 
Properties of Variance

For $X, Y$ random variables and are constants, then

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$

  Proof:

- $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2abE[(X - E(X))(Y - E(Y))]$

  Proof:

- $\text{Var}(X) \geq 0$

- $\text{Var}(X) = 0$ if and only if $X = E(X)$ with probability 1

- $\text{Var}(a) = 0$
Examples

1. Suppose $X \sim \text{Uniform}(0, 1)$, then $E(X) = \frac{1}{2}$ and $E(X^2) = \frac{1}{3}$ therefore

$$Var(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

2. Suppose $X \sim \text{Bernoulli}(p)$, then $E(X) = p$ and $E(X^2) = 1^2p + 0^2q = p$ therefore,

$$Var(X) = p - p^2 = p(1 - p)$$

3. Suppose $X$ has the following probability mass function

$$p_X(x) = 0.2(0.8)^{x-1} \quad x = 1, 2, \ldots$$

find the mean and variance of $X$. 
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

\[
\begin{array}{c|ccc|c}
  y & x & -1 & 0 & 1 & p_x(x) \\
  \hline
  -1 & 1/8 & 1/8 & 1/8 & \\
  0 & 1/8 & 0 & 1/8 & \\
  1 & 1/8 & 1/8 & 1/8 & \\
  \hline
  p_y(y) & & & & \\
\end{array}
\]

• Find 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

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

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

\[
\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}
\]

whenever \( \text{Var}(X), \text{Var}(Y) \neq 0 \) 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

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

$$f_{x,y}(x, y) = \begin{cases} 
3x & 0 \leq y \leq x \leq 1 \\
0 & \text{otherwise}
\end{cases}$$
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 $a > 0$

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

- Proof: