

# Recap: Random Variables

- ▶ Given a probability space  $(\Omega, \mathcal{F}, P)$ , a random variable is a real-valued function on  $\Omega$ .
- ▶ It essentially results in an induced probability space

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}, P_X)$$

where  $\mathcal{B}$  is the Borel  $\sigma$ -algebra and

$$P_X(B) = P[X \in B] = P(\{\omega \in \Omega : X(\omega) \in B\})$$

# Recap: Distribution function of a random variable

- ▶ Let  $X$  be a random variable. Its distribution function,  $F_X : \mathfrak{R} \rightarrow \mathfrak{R}$ , is defined by

$$F_X(x) = P[X \leq x] = P(\{\omega \in \Omega : X(\omega) \leq x\})$$

- ▶ The distribution function,  $F_X$ , completely specifies the probability measure,  $P_X$ .

# Recap: Properties of distribution function

- ▶ The distribution function satisfies
  1.  $0 \leq F_X(x) \leq 1, \forall x$
  2.  $F_X(-\infty) = 0; F_X(\infty) = 1$
  3.  $F_X$  is non-decreasing:  $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
  4.  $F_X$  is right continuous and has left-hand limits.
- ▶ Any real-valued function of a real variable satisfying the above four properties would be a distribution function of some random variable.
- ▶ We also have
$$F_X(x^+) - F_X(x^-) = F_X(x) - F_X(x^-) = P[X = x]$$
$$P[a < X \leq b] = F_X(b) - F_X(a).$$

# Recap: Discrete Random Variable

- ▶ A random variable  $X$  is said to be discrete if it takes only finitely many or countably infinitely many distinct values.
- ▶ Let  $X \in \{x_1, x_2, \dots\}$
- ▶ Its distribution function,  $F_X$  is a stair-case function with jump discontinuities at each  $x_i$  and the magnitude of the jump at  $x_i$  is equal to  $P[X = x_i]$

## Recap: probability mass function

- ▶ Let  $X \in \{x_1, x_2, \dots\}$ .
- ▶ The probability mass function (pmf) of  $X$  is defined by

$$f_X(x_i) = P[X = x_i]; \quad f_X(x) = 0, \quad \text{for all other } x$$

- ▶ It satisfies
  1.  $f_X(x) \geq 0, \forall x$  and  $f_X(x) = 0$  if  $x \neq x_i$  for some  $i$
  2.  $\sum_i f_X(x_i) = 1$

- ▶ We have

$$F_X(x) = \sum_{i: x_i \leq x} f_X(x_i)$$
$$f_X(x) = F_X(x) - F_X(x^-)$$

- ▶ We can calculate the probability of any event as

$$P[X \in B] = \sum_{\substack{i: \\ x_i \in B}} f_X(x_i)$$

## Recap: continuous random variable

- ▶  $X$  is said to be a continuous random variable if there exists a function  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  satisfying

$$F_X(x) = \int_{-\infty}^x f_X(x) \, dx$$

The  $f_X$  is called the probability density function.

- ▶ Same as saying  $F_X$  is absolutely continuous.
- ▶ Since  $F_X$  is continuous here, we have

$$P[X = x] = F_X(x) - F_X(x^-) = 0, \quad \forall x$$

- ▶ A continuous rv takes uncountably many distinct values. However, not every rv that takes uncountably many values is a continuous rv

## Recap: probability density function

- ▶ The pdf of a continuous rv is defined to be the  $f_X$  that satisfies

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad \forall x$$

- ▶ It satisfies

1.  $f_X(x) \geq 0, \quad \forall x$

2.  $\int_{-\infty}^{\infty} f_X(t) dt = 1$

- ▶ We can, in principle, compute probability of any event as

$$P[X \in B] = \int_B f_X(t) dt, \quad \forall B \in \mathcal{B}$$

- ▶ In particular,

$$P[a \leq X \leq b] = \int_a^b f_X(t) dt$$

# Recap: Function of a random variable

- ▶ If  $X$  is a random variable and  $g : \mathfrak{R} \rightarrow \mathfrak{R}$  is a function, then  $Y = g(X)$  is a random variable.
- ▶ More formally,  $Y$  is a random variable if  $g$  is a Borel measurable function.
- ▶ We can determine distribution of  $Y$  given the function  $g$  and the distribution of  $X$



# Recap

- ▶ Let  $X$  be a rv and let  $Y = g(X)$ .
- ▶ The distribution function of  $Y$  is given by

$$\begin{aligned}F_Y(y) &= P[g(X) \leq y] \\&= P[X \in \{z : g(z) \leq y\}]\end{aligned}$$

- ▶ This probability can be obtained from distribution of  $X$ .

# Recap

- ▶ Suppose  $X$  is a discrete rv with  $X \in \{x_1, x_2, \dots\}$ .
- ▶ Suppose  $Y = g(X)$ .
- ▶ Then  $Y$  is also discrete and  $Y \in \{g(x_1), g(x_2), \dots\}$ .
- ▶ We can find the pmf of  $Y$  as

$$\begin{aligned}f_Y(y) &= p[Y = y] = P[g(X) = y] \\&= P[X \in \{x_i : g(x_i) = y\}] \\&= \sum_{\substack{i: \\ g(x_i)=y}} f_X(x_i)\end{aligned}$$

# Recap

- ▶ Let  $g : \Re \rightarrow \Re$  be differentiable with  $g'(x) > 0, \forall x$  or  $g'(x) < 0, \forall x$ .
- ▶ Let  $X$  be a continuous rv and let  $Y = g(X)$ .
- ▶ Then  $Y$  is a continuous rv with pdf

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, \quad a \leq y \leq b$$

where  $a = \min(g(\infty), g(-\infty))$  and  
 $b = \max(g(\infty), g(-\infty))$

- ▶ This theorem is useful in some cases to find the densities of functions of continuous random variables

## Recap: Expectation

- ▶ Let  $X$  be a discrete rv with  $X \in \{x_1, x_2, \dots\}$ . Then

$$E[X] = \sum_i x_i f_X(x_i)$$

- ▶ If  $X$  is a continuous random variable with pdf,  $f_X$ ,

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

- ▶ Sometimes we use the following notation to denote expectation of both kinds of rv

$$E[X] = \int_{-\infty}^{\infty} x dF_X(x)$$

- ▶ We take the expectation to exist when the sum or integral above is absolutely convergent
- ▶ Note that expectation is defined for all random variables

## Recap: Expectation of a function of a random variable

- ▶ Let  $X$  be a rv and let  $Y = g(X)$ . Then,
- ▶  $EY = \int y \, dF_Y(y) = \int g(x) \, dF_X(x)$
- ▶ That is, if  $X$  is discrete, then

$$EY = \sum_j y_j f_Y(y_j) = \sum_i g(x_i) f_X(x_i)$$

- ▶ If  $X$  and  $Y$  are continuous

$$EY = \int y f_Y(y) \, dy = \int g(x) f_X(x) \, dx$$

- ▶ This is true for all rv's.

# Recap: Properties of Expectation

$$E[g(X)] = \sum_i g(x_i) f_X(x_i) \quad \text{or} \quad E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

- ▶ If  $X \geq 0$  then  $EX \geq 0$
- ▶  $E[b] = b$  where  $b$  is a constant
- ▶  $E[ag(X)] = aE[g(X)]$  where  $a$  is a constant
- ▶  $E[aX + b] = aE[X] + b$  where  $a, b$  are constants.
- ▶  $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$
- ▶  $E[(X - c)^2] \geq E[(X - EX)^2], \forall c$

# Recap: Variance of random variable



$$\text{Var}(X) = E[(X - EX)^2] = E[X^2] - (EX)^2$$

▶ Properties of Variance:

- ▶  $\text{Var}(X) \geq 0$
- ▶  $\text{Var}(X + c) = \text{Var}(X)$
- ▶  $\text{Var}(cX) = c^2 \text{Var}(X)$

# Recap: Moments of a random variable

- ▶ The  $k^{th}$  (order) moment of  $X$  is

$$m_k = E[X^k] = \int x^k dF_X(x)$$

- ▶ The  $k^{th}$  central moment of  $X$  is

$$s_k = E[(X - EX)^k] = \int (x - EX)^k dF_X(x)$$

- ▶ If moment of order  $k$  is finite then so is moment of order  $s$  for  $s < k$ .



## Recap: Moment Generating function

- ▶ The moment generating function –  $M_X : \Re \rightarrow \Re$

$$M_X(t) = Ee^{tX} = \sum_i e^{tx_i} f_X(x_i) \quad \text{or} \quad \int e^{tx} f_X(x) dx, \quad t \in \Re$$

- ▶ We say the mgf exists if  $E[e^{tX}] < \infty$  for  $t$  in some interval around zero
- ▶ If  $M_X(t)$  exists (for  $t \in [-a, a]$  for some  $a > 0$ ) then all its derivatives also exist and

$$\left. \frac{d^k M_X(t)}{dt^k} \right|_{t=0} = E[X^k]$$

# Generating function

- ▶ For  $X \in \{0, 1, 2, \dots\}$  the (probability) generating function of  $X$  is defined by

$$P_X(s) = \sum_{k=0}^{\infty} f_X(k)s^k, \quad s \in \mathbb{R}$$

- ▶ We get the pmf from it as

$$f_X(0) = P_X(0); \quad f_X(1) = \frac{P'_X(0)}{1!}; \quad f_X(2) = \frac{P''_X(0)}{2!}$$

- ▶ We can also get the moments:

$$P'_X(1) = EX, \quad P''_X(1) = E[X(X-1)]$$

## quantiles of a distribution

- ▶ Let  $p \in (0, 1)$ . The number  $x \in \Re$  that satisfies

$$P[X \leq x] \geq p \quad \text{and} \quad p[X \geq x] \geq 1 - p$$

is called the quantile of order  $p$  or the  $100p^{th}$  percentile of rv  $X$ .

- ▶ If  $x$  is quantile of order  $p$ , it satisfies

$$p \leq F_X(x) \leq p + P[X = x]$$

- ▶ For a given  $p$  there can be multiple values for  $x$  to satisfy the above.
- ▶ For  $p = 0.5$ , it is called the median.

## Recap: some moment inequalities

- ▶ **Markov inequality:** For a non-negative function,  $g$ ,

$$P[g(X) > c] \leq \frac{E[g(X)]}{c}$$

- ▶ A specific instance of this is

$$P[|X| > c] \leq \frac{E[|X|^k]}{c^k}$$

- ▶ **Chebyshev inequality**

$$P[|X - EX| > c] \leq \frac{\text{Var}(X)}{c^2}$$

- ▶ With  $EX = \mu$  and  $\text{Var}(X) = \sigma^2$ , we get

$$P[|X - \mu| > k\sigma] \leq \frac{1}{k^2}$$



## Recap: A pair of random variables

- ▶ Let  $X, Y$  be random variables on the probability space  $(\Omega, \mathcal{F}, P)$
- ▶ We can think of  $X, Y$  together as a vector-valued function mapping  $\Omega$  to  $\mathbb{R}^2$ .
- ▶ This gives rise to a new probability space  $(\mathbb{R}^2, \mathcal{B}^2, P_{XY})$  with  $P_{XY}$  given by

$$\begin{aligned} P_{XY}(B) &= P[(X, Y) \in B], \forall B \in \mathcal{B}^2 \\ &= P(\{\omega : (X(\omega), Y(\omega)) \in B\}) \end{aligned}$$

## Recap: Joint distribution function

- ▶ Let  $X, Y$  be random variables on the same probability space  $(\Omega, \mathcal{F}, P)$
- ▶ The joint distribution function of  $X, Y$  is  $F_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , defined by

$$\begin{aligned} F_{XY}(x, y) &= P[X \leq x, Y \leq y] \quad (= P_{XY}((-\infty, x] \times (-\infty, y])) \\ &= P(\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\}) \end{aligned}$$

- ▶ The joint distribution function is the probability of the intersection of the events  $[X \leq x]$  and  $[Y \leq y]$ .

# Recap: Properties of Joint Distribution Function

- ▶ Joint distribution function:  $F_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$

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

- ▶ It satisfies

1.  $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \forall x, y;$   
 $F_{XY}(\infty, \infty) = 1$
2.  $F_{XY}$  is non-decreasing in each of its arguments
3.  $F_{XY}$  is right continuous and has left-hand limits in each of its arguments
4. For all  $x_1 < x_2$  and  $y_1 < y_2$

$$F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1) \geq 0$$

- ▶ Any  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  satisfying the above would be a joint distribution function.



- ▶ Let  $X, Y$  be two discrete random variables (defined on the same probability space).
- ▶ Let  $X \in \{x_1, \dots, x_n\}$  and  $Y \in \{y_1, \dots, y_m\}$ .
- ▶ We define the joint probability mass function of  $X$  and  $Y$  as

$$f_{XY}(x_i, y_j) = P[X = x_i, Y = y_j]$$

( $f_{XY}(x, y)$  is zero for all other values of  $x, y$ )

- ▶ The  $f_{XY}$  would satisfy
  - ▶  $f_{XY}(x, y) \geq 0, \forall x, y$  and  $\sum_i \sum_j f_{XY}(x_i, y_j) = 1$
- ▶ This is a straight-forward extension of the pmf of a single discrete rv.

## Example

- ▶ Let  $\Omega = (0, 1)$  with the 'usual' probability.
- ▶ So, each  $\omega$  is a real number between 0 and 1
- ▶ Let  $X(\omega)$  be the digit in the first decimal place in  $\omega$  and let  $Y(\omega)$  be the digit in the second decimal place.
- ▶ If  $\omega = 0.2576$  then  $X(\omega) = 2$  and  $Y(\omega) = 5$
- ▶ Easy to see that  $X, Y \in \{0, 1, \dots, 9\}$ .
- ▶ We want to calculate the joint pmf of  $X$  and  $Y$

## Example

- ▶ What is the event  $[X = 4]$ ?

$$[X = 4] = \{\omega : X(\omega) = 4\} = [0.4, 0.5)$$

- ▶ What is the event  $[Y = 3]$ ?

$$[Y = 3] = [0.03, 0.04) \cup [0.13, 0.14) \cup \cdots \cup [0.93, 0.94)$$

- ▶ What is the event  $[X = 4, Y = 3]$ ?

It is the intersection of the above

$$[X = 4, Y = 3] = [0.43, 0.44)$$

- ▶ Hence the joint pmf of  $X$  and  $Y$  is

$$f_{XY}(x, y) = P[X = x, Y = y] = 0.01, \quad x, y \in \{0, 1, \dots, 9\}$$

## Example

- ▶ Consider the random experiment of rolling two dice.  
 $\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, 2, \dots, 6\}\}$
- ▶ Let  $X$  be the maximum of the two numbers and let  $Y$  be the sum of the two numbers.
- ▶ Easy to see  $X \in \{1, 2, \dots, 6\}$  and  $Y \in \{2, 3, \dots, 12\}$
- ▶ What is the event  $[X = m, Y = n]$ ? (We assume  $m, n$  are in the correct range)

$$[X = m, Y = n] = \{(\omega_1, \omega_2) \in \Omega : \max(\omega_1, \omega_2) = m, \omega_1 + \omega_2 = n\}$$

- ▶ For this to be a non-empty set, we must have  
 $m < n \leq 2m$
- ▶ Then  $[X = m, Y = n] = \{(m, n - m), (n - m, m)\}$
- ▶ Is this always true? No! What if  $n = 2m$ ?  
 $[X = 3, Y = 6] = \{(3, 3)\},$   
 $[X = 4, Y = 6] = \{(4, 2), (2, 4)\}$
- ▶ So,  $P[X = m, Y = n]$  is either  $2/36$  or  $1/36$  (assuming  $m, n$  satisfy other requirements)

## Example

- ▶ We can now write the joint pmf.
- ▶ Assume  $1 \leq m \leq 6$  and  $2 \leq n \leq 12$ . Then

$$f_{XY}(m, n) = \begin{cases} \frac{2}{36} & \text{if } m < n < 2m \\ \frac{1}{36} & \text{if } n = 2m \end{cases}$$

( $f_{XY}(m, n)$  is zero in all other cases)

- ▶ Does this satisfy requirements of joint pmf?

$$\begin{aligned} \sum_{m,n} f_{XY}(m, n) &= \sum_{m=1}^6 \sum_{n=m+1}^{2m-1} \frac{2}{36} + \sum_{m=1}^6 \frac{1}{36} \\ &= \frac{2}{36} \sum_{m=1}^6 (m-1) + \frac{1}{36} 6 \\ &= \frac{2}{36} (21 - 6) + \frac{6}{36} = 1 \end{aligned}$$

# Joint Probability mass function

- ▶ Let  $X \in \{x_1, x_2, \dots\}$  and  $Y \in \{y_1, y_2, \dots\}$  be discrete random variables.
- ▶ The joint pmf:  $f_{XY}(x, y) = P[X = x, Y = y]$ .
- ▶ The joint pmf satisfies:
  - ▶  $f_{XY}(x, y) \geq 0, \forall x, y$  and
  - ▶  $\sum_i \sum_j f_{XY}(x_i, y_j) = 1$
- ▶ Given the joint pmf, we can get the joint df as

$$F_{XY}(x, y) = \sum_{\substack{i: \\ x_i \leq x}} \sum_{\substack{j: \\ y_j \leq y}} f_{XY}(x_i, y_j)$$

- ▶ Given sets  $\{x_1, x_2, \dots\}$  and  $\{y_1, y_2, \dots\}$ .
- ▶ Suppose  $f_{XY} : \Re^2 \rightarrow [0, 1]$  be such that
  - ▶  $f_{XY}(x, y) = 0$  unless  $x = x_i$  for some  $i$  and  $y = y_j$  for some  $j$ , and
  - ▶  $\sum_i \sum_j f_{XY}(x_i, y_j) = 1$
- ▶ Then  $f_{XY}$  is a joint pmf.
- ▶ This is because, if we define

$$F_{XY}(x, y) = \sum_{\substack{i: \\ x_i \leq x}} \sum_{\substack{j: \\ y_j \leq y}} f_{XY}(x_i, y_j)$$

then  $F_{XY}$  satisfies all properties of a df.

- ▶ We normally specify a pair of discrete random variables by giving the joint pmf

- ▶ Given the joint pmf, we can (in principle) compute the probability of any event involving the two discrete random variables.

$$P[(X, Y) \in B] = \sum_{\substack{i,j: \\ (x_i, y_j) \in B}} f_{XY}(x_i, y_j)$$

- ▶ Now, events can be specified in terms of relations between the two rv's too

$$[X < Y + 2] = \{\omega : X(\omega) < Y(\omega) + 2\}$$

- ▶ Thus,

$$P[X < Y + 2] = \sum_{\substack{i,j: \\ x_i < y_j + 2}} f_{XY}(x_i, y_j)$$



- ▶ Take the example: 2 dice,  $X$  is max and  $Y$  is sum
- ▶  $f_{XY}(m, n) = 0$  unless  $m = 1, \dots, 6$  and  $n = 2, \dots, 12$ .  
For this range

$$f_{XY}(m, n) = \begin{cases} \frac{2}{36} & \text{if } m < n < 2m \\ \frac{1}{36} & \text{if } n = 2m \end{cases}$$

- ▶ Suppose we want  $P[Y = X + 2]$ .

$$\begin{aligned} P[Y = X + 2] &= \sum_{\substack{m, n: \\ n=m+2}} f_{XY}(m, n) = \sum_{m=1}^6 f_{XY}(m, m+2) \\ &= \sum_{m=2}^6 f_{XY}(m, m+2) \quad \text{since we need } m+2 \leq 2m \\ &= \frac{1}{36} + 4 \frac{2}{36} = \frac{9}{36} \end{aligned}$$

# Joint density function

- ▶ Let  $X, Y$  be two continuous rv's with df  $F_{XY}$ .
- ▶ If there exists a function  $f_{XY}$  that satisfies

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dy' dx', \quad \forall x, y$$

then we say that  $X, Y$  have a joint probability density function which is  $f_{XY}$

- ▶ Please note the difference in the definition of joint pmf and joint pdf.
- ▶ When  $X, Y$  are discrete we defined a joint pmf
- ▶ We are not saying that if  $X, Y$  are continuous rv's then a joint density exists.
- ▶ We use joint density to mean joint pdf

## properties of joint density

- ▶ The joint density (or joint pdf) of  $X, Y$  is  $f_{XY}$  that satisfies

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dy' dx', \quad \forall x, y$$

- ▶ Since  $F_{XY}$  is non-decreasing in each argument, we must have  $f_{XY}(x, y) \geq 0$ .
- ▶  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x', y') dy' dx' = 1$  is needed to ensure  $F_{XY}(\infty, \infty) = 1$ .

# properties of joint density

- ▶ The joint density  $f_{XY}$  satisfies the following
  1.  $f_{XY}(x, y) \geq 0, \forall x, y$
  2.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x', y') dy' dx' = 1$
- ▶ These are very similar to the properties of the density of a single rv

## Example: Joint Density

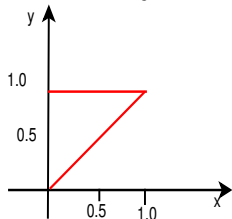
- ▶ Consider the function

$$f(x, y) = 2, \quad 0 < x < y < 1 \quad (f(x, y) = 0, \text{ otherwise})$$

- ▶ Let us show this is a density

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = \int_0^1 \int_0^y 2 \, dx \, dy = \int_0^1 2x \Big|_0^y \, dy = \int_0^1 2y \, dy = 1$$

- ▶ We can say this density is uniform over the region



## properties of joint density

- ▶ The joint density  $f_{XY}$  satisfies the following
  1.  $f_{XY}(x, y) \geq 0, \quad \forall x, y$
  2.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x', y') \, dy' \, dx' = 1$
- ▶ Any function  $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$  satisfying the above two is a joint density function.
- ▶ Given  $f_{XY}$  satisfying the above, define

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') \, dy' \, dx', \quad \forall x, y$$

- ▶ Then we can show  $F_{XY}$  is a joint distribution.

- ▶  $f_{XY}(x, y) \geq 0$  and  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x', y') dy' dx' = 1$
- ▶ Define

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dy' dx', \quad \forall x, y$$

- ▶ Then,  $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \forall x, y$  and  $F_{XY}(\infty, \infty) = 1$
- ▶ Since  $f_{XY}(x, y) \geq 0$ ,  $F_{XY}$  is non-decreasing in each argument.
- ▶ Since it is given as an integral, the above also shows that  $F_{XY}$  is continuous in each argument.
- ▶ The only property left is the special property of  $F_{XY}$  we mentioned earlier.

$$\Delta \triangleq F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1).$$

- ▶ We need to show  $\Delta \geq 0$  if  $x_1 < x_2$  and  $y_1 < y_2$ .
- ▶ We have

$$\begin{aligned} \Delta &= \int_{-\infty}^{x_2} \int_{-\infty}^{y_2} f_{XY} \, dy \, dx - \int_{-\infty}^{x_1} \int_{-\infty}^{y_2} f_{XY} \, dy \, dx \\ &\quad - \int_{-\infty}^{x_2} \int_{-\infty}^{y_1} f_{XY} \, dy \, dx + \int_{-\infty}^{x_1} \int_{-\infty}^{y_1} f_{XY} \, dy \, dx \\ &= \int_{-\infty}^{x_2} \left( \int_{-\infty}^{y_2} f_{XY} \, dy - \int_{-\infty}^{y_1} f_{XY} \, dy \right) dx \\ &\quad - \int_{-\infty}^{x_1} \left( \int_{-\infty}^{y_2} f_{XY} \, dy - \int_{-\infty}^{y_1} f_{XY} \, dy \right) dx \end{aligned}$$



- Thus we have

$$\begin{aligned}\Delta &= \int_{-\infty}^{x_2} \left( \int_{-\infty}^{y_2} f_{XY} dy - \int_{-\infty}^{y_1} f_{XY} dy \right) dx \\ &\quad - \int_{-\infty}^{x_1} \left( \int_{-\infty}^{y_2} f_{XY} dy - \int_{-\infty}^{y_1} f_{XY} dy \right) dx \\ &= \int_{-\infty}^{x_2} \int_{y_1}^{y_2} f_{XY} dy dx - \int_{-\infty}^{x_1} \int_{y_1}^{y_2} f_{XY} dy dx \\ &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY} dy dx \geq 0\end{aligned}$$

- This actually shows

$$P[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY} dy dx$$

- ▶ What we showed is the following
- ▶ Any function  $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$  that satisfies
  - ▶  $f_{XY}(x, y) \geq 0, \forall x, y$
  - ▶  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$
 is a joint density function.

- ▶ This is because now
 
$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(x, y) dx dy$$
 would satisfy all conditions for a df.

- ▶ Convenient to specify joint density (when it exists)
- ▶ We also showed

$$P[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY} dy dx$$

- ▶ In general

$$P[(X, Y) \in B] = \int_B f_{XY}(x, y) dx dy, \quad \forall B \in \mathcal{B}^2$$

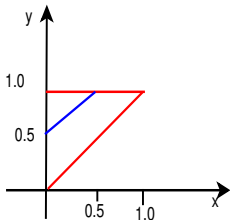
- ▶ Let us consider the example

$$f(x, y) = 2, \quad 0 < x < y < 1$$

- ▶ Suppose we want probability of  $[Y > X + 0.5]$

$$\begin{aligned} P[Y > X + 0.5] &= P[(X, Y) \in \{(x, y) : y > x + 0.5\}] \\ &= \int_{\{(x, y) : y > x + 0.5\}} f_{XY}(x, y) \, dx \, dy \\ &= \int_{0.5}^1 \int_0^{y-0.5} 2 \, dx \, dy \\ &= \int_{0.5}^1 2(y - 0.5) \, dy \\ &= 2 \left. \frac{y^2}{2} \right|_{0.5}^1 - y \Big|_{0.5}^1 = 1 - 0.25 - 1 + 0.5 = 0.25 \end{aligned}$$

- ▶ We can look at it geometrically



- ▶ The probability of the event we want is the area of the small triangle divided by that of the big triangle.

# Marginal Distributions

- ▶ Let  $X, Y$  be random variables with joint distribution function  $F_{XY}$ .
- ▶ We know  $F_{XY}(x, y) = P[X \leq x, Y \leq y]$ .
- ▶ Hence

$$F_{XY}(x, \infty) = P[X \leq x, Y \leq \infty] = P[X \leq x] = F_X(x)$$

- ▶ We define the marginal distribution functions of  $X, Y$  by

$$F_X(x) = F_{XY}(x, \infty); \quad F_Y(y) = F_{XY}(\infty, y)$$

- ▶ These are simply distribution functions of  $X$  and  $Y$  obtained from the joint distribution function.

# Marginal mass functions

- ▶ Let  $X \in \{x_1, x_2, \dots\}$  and  $Y \in \{y_1, y_2, \dots\}$
- ▶ Let  $f_{XY}$  be their joint mass function.
- ▶ Then

$$P[X = x_i] = \sum_j P[X = x_i, Y = y_j] = \sum_j f_{XY}(x_i, y_j)$$

(This is because  $[Y = y_j]$ ,  $j = 1, \dots$ , form a partition and  $P(A) = \sum_i P(AB_i)$  when  $B_i$  is a partition)

- ▶ We define the marginal mass functions of  $X$  and  $Y$  as

$$f_X(x_i) = \sum_j f_{XY}(x_i, y_j); \quad f_Y(y_j) = \sum_i f_{XY}(x_i, y_j)$$

- ▶ These are mass functions of  $X$  and  $Y$  obtained from the joint mass function

## marginal density functions

- ▶ Let  $X, Y$  be continuous rv with joint density  $f_{XY}$ .
- ▶ Then we know  $F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dy' dx'$
- ▶ Hence, we have

$$\begin{aligned} F_X(x) = F_{XY}(x, \infty) &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(x', y') dy' dx' \\ &= \int_{-\infty}^x \left( \int_{-\infty}^{\infty} f_{XY}(x', y') dy' \right) dx' \end{aligned}$$

- ▶ Since  $X$  is a continuous rv, this means

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

We call this the marginal density of  $X$ .

- ▶ Similarly, marginal density of  $Y$  is

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

- ▶ These are pdf's of  $X$  and  $Y$  obtained from the joint

## Example

- ▶ Rolling two dice,  $X$  is max,  $Y$  is sum
- ▶ We had, for  $1 \leq m \leq 6$  and  $2 \leq n \leq 12$ ,

$$f_{XY}(m, n) = \begin{cases} \frac{2}{36} & \text{if } m < n < 2m \\ \frac{1}{36} & \text{if } n = 2m \end{cases}$$

- ▶ We know,  $f_X(m) = \sum_n f_{XY}(m, n)$ ,  $m = 1, \dots, 6$ .
- ▶ Given  $m$ , for what values of  $n$ ,  $f_{XY}(m, n) > 0$  ?  
We can only have  $n = m + 1, \dots, 2m$ .
- ▶ Hence we get

$$f_X(m) = \sum_{n=m+1}^{2m} f_{XY}(m, n) = \sum_{n=m+1}^{2m-1} \frac{2}{36} + \frac{1}{36} = \frac{2}{36}(m-1) + \frac{1}{36} = \frac{2m-1}{36}$$



## Example

- ▶ Consider the joint density

$$f_{XY}(x, y) = 2, \quad 0 < x < y < 1$$

- ▶ The marginal density of  $X$  is: for  $0 < x < 1$ ,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) \, dy = \int_x^1 2 \, dy = 2(1 - x)$$

Thus,  $f_X(x) = 2(1 - x)$ ,  $0 < x < 1$

- ▶ We can easily verify this is a density

$$\int_{-\infty}^{\infty} f_X(x) \, dx = \int_0^1 2(1 - x) \, dx = (2x - x^2) \Big|_0^1 = 1$$

We have:  $f_{XY}(x, y) = 2$ ,  $0 < x < y < 1$

- ▶ We can similarly find density of  $Y$ .
- ▶ For  $0 < y < 1$ ,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_0^y 2 dx = 2y$$

- ▶ Thus,  $f_Y(y) = 2y$ ,  $0 < y < 1$  and

$$\int_0^1 2y dy = 2 \left. \frac{y^2}{2} \right|_0^1 = 1$$

- ▶ If we are given the joint df or joint pmf/joint density of  $X, Y$ , then the individual df or pmf/pdf are uniquely determined.
- ▶ However, given individual pdf of  $X$  and  $Y$ , we cannot determine the joint density. (same is true of pmf or df)
- ▶ There can be many different joint density functions all having the same marginals

# Conditional distributions

- ▶ Let  $X, Y$  be rv's on the same probability space
- ▶ We define the conditional distribution of  $X$  given  $Y$  by

$$F_{X|Y}(x|y) = P[X \leq x | Y = y]$$

(For now ignore the case of  $P[Y = y] = 0$ ).

- ▶ Note that  $F_{X|Y} : \Re^2 \rightarrow \Re$
- ▶  $F_{X|Y}(x|y)$  is a notation. We could write  $F_{X|Y}(x, y)$ .

- ▶ Conditional distribution of  $X$  given  $Y$  is

$$F_{X|Y}(x|y) = P[X \leq x | Y = y]$$

It is the conditional probability of  $[X \leq x]$  given (or conditioned on)  $[Y = y]$ .

- ▶ Consider example: rolling 2 dice,  $X$  is max,  $Y$  is sum

$$P[X \leq 4 | Y = 3] = 1; \quad P[X \leq 4 | Y = 9] = 0$$

- ▶ This is what conditional distribution captures.
- ▶ For every value of  $y$ ,  $F_{X|Y}(x|y)$  is a distribution function in the variable  $x$ .
- ▶ It defines a new distribution for  $X$  based on knowing the value of  $Y$ .

- ▶ Let:  $X \in \{x_1, x_2, \dots\}$  and  $Y \in \{y_1, y_2, \dots\}$ . Then

$$F_{X|Y}(x|y_j) = P[X \leq x|Y = y_j] = \frac{P[X \leq x, Y = y_j]}{P[Y = y_j]}$$

(We define  $F_{X|Y}(x|y)$  only when  $y = y_j$  for some  $j$ ).

- ▶ For each  $y_j$ ,  $F_{X|Y}(x|y_j)$  is a df of a discrete rv in  $x$ .
- ▶ Since  $X$  is a discrete rv, we can write the above as

$$\begin{aligned} F_{X|Y}(x|y_j) &= \frac{P[X \leq x, Y = y_j]}{P[Y = y_j]} = \frac{\sum_{i:x_i \leq x} P[X = x_i, Y = y_j]}{P[Y = y_j]} \\ &= \sum_{i:x_i \leq x} \left( \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} \right) \end{aligned}$$

# Conditional mass function

- ▶ We got

$$F_{X|Y}(x|y_j) = \sum_{i: x_i \leq x} \left( \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} \right)$$

- ▶ Since  $X$  is a discrete rv, what is inside the summation above is the pmf corresponding to the df,  $F_{X|Y}$ .
- ▶ We define the conditional mass function of  $X$  given  $Y$  as

$$f_{X|Y}(x_i|y_j) = \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} = P[X = x_i | Y = y_j]$$

- ▶ The conditional mass function is

$$f_{X|Y}(x_i|y_j) = P[X = x_i|Y = y_j] = \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)}$$

- ▶ This gives us the useful identity

$$f_{XY}(x_i, y_j) = f_{X|Y}(x_i|y_j)f_Y(y_j)$$

$$(P[X = x_i, Y = y_j] = P[X = x_i|Y = y_j]P[Y = y_j])$$

- ▶ This gives us the total probability rule for rv's

$$f_X(x_i) = \sum_j f_{XY}(x_i, y_j) = \sum_j f_{X|Y}(x_i|y_j)f_Y(y_j)$$

- ▶ This is same as

$$P[X = x_i] = \sum_j P[X = x_i|Y = y_j]P[Y = y_j]$$

$(P(A) = \sum_j P(A|B_j)P(B_j))$  when  $B_1, \dots$  form a partition)



- We have

$$f_{XY}(x_i, y_j) = f_{X|Y}(x_i|y_j)f_Y(y_j) = f_{Y|X}(y_j|x_i)f_X(x_i)$$

- This gives us Bayes rule for discrete rv's

$$\begin{aligned} f_{X|Y}(x_i|y_j) &= \frac{f_{Y|X}(y_j|x_i)f_X(x_i)}{f_Y(y_j)} \\ &= \frac{f_{Y|X}(y_j|x_i)f_X(x_i)}{\sum_i f_{XY}(x_i, y_j)} \\ &= \frac{f_{Y|X}(y_j|x_i)f_X(x_i)}{\sum_i f_{Y|X}(y_j|x_i)f_X(x_i)} \end{aligned}$$