Recap: Joint Distribution Function

- Given X, Y rv on same probability space, the joint distribution function is
  $F_{XY} : \Re^2 \to \Re$
  $$F_{XY}(x, y) = P[X \leq x, Y \leq y]$$
- 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 $F : \Re^2 \to \Re$ satisfying the above would be a joint distribution function.

Recap: Joint Probability mass function

- $X \in \{x_1, x_2, \cdots\}$, $Y \in \{y_1, y_2, \cdots\}$
- The joint pmf: $f_{XY}(x, y) = P[X = x, Y = y]$.
- The joint pmf satisfies:
  - A1 $f_{XY}(x, y) \geq 0, \forall x, y$ and non-zero only for x_i, y_j pairs
  - A2 $\sum_i \sum_j f_{XY}(x_i, y_j) = 1$
- Given the joint pmf, we can get the joint df as

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

- Any $f_{XY} : \Re^2 \to [0,\ 1]$ satisfying $A1, A2$ above is a joint pmf. (Because then F_{XY} satisfies all properties of df).
- Given the joint pmf, we can (in principle) compute the probability of any event involving the two discrete random variables.

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

Recap: joint density

- Two cont rv X, Y have a joint density f_{XY} if

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

- The joint density f_{XY} satisfies the following
  1. $f_{XY}(x, y) \geq 0, \;\; \forall x, y$
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x', y')\; dy'\; dx' = 1$
- Any function $f_{XY} : \Re^2 \rightarrow \Re$ satisfying the above two is a joint density function. (Then the above F_{XY} can be shown to be a joint df).
- We also have

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

  and, in general,

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

Recap Marginals

- Marginal distribution functions of X, Y are

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

- X, Y discrete with joint pmf f_{XY}. The marginal pmfs are

$$f_X(x) = \sum_y f_{XY}(x, y); \quad f_Y(y) = \sum_x f_{XY}(x, y)$$

- If X, Y have joint pdf f_{XY} then the marginal pdf are

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

Recap Conditional distribution

- Let: $X \in \{x_1, x_2, \cdots\}$ and $Y \in \{y_1, y_2, \cdots\}$. 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.
- The pmf corresponding to this df is called conditional pmf

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

Recap Bayes rule for discrete rv's

- 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

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

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

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

- Also gives us Bayes rule for discrete rv

$$f_{X|Y}(x_i|y_j) = \frac{f_{Y|X}(y_j|x_i) f_X(x_i)}{\sum_i f_{Y|X}(y_j|x_i) f_X(x_i)}$$

Recap: Conditional densities

- Let X, Y have joint density f_{XY}.
- The conditional df of X given Y is

$$F_{X|Y}(x|y) = \lim_{\delta \to 0} P[X \leq x | Y \in [y,\ y + \delta]]$$

- This exists if $f_Y(y) > 0$ and then it has a density:

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(x'|y)\ dx'$$

- This conditional density is given by

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

- We (once again) have the useful identity

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

Recap: Bayes rule for continuous rv

- We have the identity

$$f_{XY}(x,y) = f_{X|Y}(x|y)\ f_Y(y)$$

- This gives us continuous analogue of total probability rule:

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

- This also gives us Bayes rule for continuous rv

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} \\ &= \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\ dx} \end{aligned}$$

- Now, let X be a continuous rv and let Y be discrete rv.
- We can define $F_{X|Y}$ as

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

This is well defined for all values that y takes. (We consider only those y)

- Since X is continuous rv, this df would have a density

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(x'|y) \, dx'$$

- Hence we can write

$$\begin{aligned} P[X \leq x, Y = y] &= F_{X|Y}(x|y) P[Y = y] \\ &= \int_{-\infty}^{x} f_{X|Y}(x'|y) \; f_Y(y) \, dx' \end{aligned}$$

- We now get

$$\begin{aligned} F_X(x) &= P[X \leq x] = \sum_y P[X \leq x, Y = y] \\ &= \sum_y \int_{-\infty}^{x} f_{X|Y}(x'|y)\ f_Y(y)\ dx' \\ &= \int_{-\infty}^{x} \sum_y f_{X|Y}(x'|y)\ f_Y(y)\ dx' \end{aligned}$$

- This gives us

$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y)$$

- This is another version of total probability rule.
- Earlier we derived this when X, Y are discrete.
- The formula is true even when X is continuous
  Only difference is we need to take f_X as the density of X.

- When X, Y are discrete we have

$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y) \ (P[X = x] = \sum_y P[X = x|Y = y] P[Y = y]$$

- When X is continuous and Y is discrete, we defined $f_{X|Y}(x|y)$ to be the density corresponding to $F_{X|Y}(x|y) = P[X \leq x|Y = y]$
- Then we once again get

$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y)$$

  Now, f_X is density (and not a mass function).
- Suppose $Y \in \{1, 2, 3\}$ and $f_Y(i) = \lambda_i$.
  Let $f_{X|Y}(x|i) = f_i(x)$. Then

$$f_X(x) = \lambda_1 f_1(x) + \lambda_2 f_2(x) + \lambda_3 f_3(x)$$

  Called a mixture density model

- Continuing with X continuous rv and Y discrete. We have

$$F_{X|Y}(x|y) = P[X \leq x | Y = y] = \int_{-\infty}^{x} f_{X|Y}(x'|y)\, dx'$$

- We also have

$$P[X \leq x, Y = y] = \int_{-\infty}^{x} f_{X|Y}(x'|y)\, f_Y(y)\, dx'$$

- Hence we can define a 'joint density'

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

- This is a kind of mixed density and mass function.
- We will not be using such 'joint densities' here

- Continuing with X continuous rv and Y discrete
- Can we define $f_{Y|X}(y|x)$?
- Since Y is discrete, this (conditional) mass function is

$$f_{Y|X}(y|x) = P[Y = y|X = x]$$

But the conditioning event has zero prob
We now know how to handle it

$$f_{Y|X}(y|x) = \lim_{\delta \to 0} P[Y = y|X \in [x, x + \delta]]$$

- For simplifying this we note the following:

$$P[X \leq x, Y = y] = \int_{-\infty}^{x} f_{X|Y}(x'|y) \; f_Y(y) \; dx'$$

$$\Rightarrow \quad P[X \in [x, x{+}\delta], Y = y] = \int_{x}^{x+\delta} f_{X|Y}(x'|y) \; f_Y(y) \; dx'$$

- We have

$$
\begin{aligned}
f_{Y|X}(y|x) &= \lim_{\delta \to 0} P[Y = y | X \in [x, x+\delta]] \\
&= \lim_{\delta \to 0} \frac{P[Y = y, X \in [x, x+\delta]]}{P[X \in [x, x+\delta]]} \\
&= \lim_{\delta \to 0} \frac{\int_x^{x+\delta} \ f_{X|Y}(x'|y) \ f_Y(y) \ dx'}{\int_x^{x+\delta} \ f_X(x') \ dx'} \\
&= \lim_{\delta \to 0} \frac{f_{X|Y}(x|y)\delta \ f_Y(y)}{f_X(x) \ \delta} \\
&= \frac{f_{X|Y}(x|y) \ f_Y(y)}{f_X(x)}
\end{aligned}
$$

- This gives us further versions of total probability rule and Bayes rule.

- First let us look at the total probability rule possibilities
- When X is continuous rv and Y is discrete rv, we derived

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

  Note that f_Y is mass fn, f_X is density and so on.
- Since $f_{X|Y}$ is a density (corresponding to $F_{X|Y}$),

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\ dx = 1$$

- Hence we get

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

- Earlier we derived the same formula when X, Y have a joint density.

- Let us review all the total probability formulas

$$\textbf{1. } f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y)$$

- We first derived this when X, Y are discrete.
- But now we proved this holds when Y is discrete
  If X is continuous the $f_X, f_{X|Y}$ are densities; If X is also discrete they are mass functions

$$\textbf{2. } f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\, dx$$

- We first proved it when X, Y have a joint density
  We now know it holds also when X is cont and Y is discrete. In that case f_Y is a mass function

- When X is continuous rv and Y is discrete rv, we derived

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

- This once again gives rise to Bayes rule:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\ f_Y(y)}{f_X(x)} \quad f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

- Earlier we showed this hold when X, Y are both discrete or both continuous.
- Thus Bayes rule holds in all four possible scenarios
- Only difference is we need to interpret f_X or $f_{X|Y}$ as mass functions when X is discrete and as densities when X is a continuous rv
- In general, one refers to these always as densities since the actual meaning would be clear from context.

Example

- Consider a communication system. The transmitter puts out 0 or 5 volts for the bits of 0 and 1, and, volage measured by the receiver is the sent voltage plus noise added by the channel.
- We assume noise has Gaussian density with mean zero and variance σ^2.
- We may want the probability that the sent bit is 1 when measured voltage at the receiver is x to decide what is sent.
- Let X be the measured voltage and let Y be sent bit.
- We want to calculate $f_{Y|X}(1|x)$.
- We want to use the Bayes rule to calculate this

- We need $f_{X|Y}$. What does our model say?
- $f_{X|Y}(x|1)$ is Gaussian with mean 5 and variance σ^2 and $f_{X|Y}(x|0)$ is Gaussian with mean zero and variance σ^2

$$P[Y=1|X=x] = f_{Y|X}(1|x) = \frac{f_{X|Y}(x|1)\ f_Y(1)}{f_X(x)}$$

- We need $f_Y(1), f_Y(0)$. Let us take them to be same.
- In practice we only want to know whether $f_{Y|X}(1|x) > f_{Y|X}(0|x)$
- Then we do not need to calculate $f_X(x)$.
  We only need ratio of $f_{Y|X}(1|x)$ and $f_{Y|X}(0|x)$.

- The ratio of the two probabilities is

$$\begin{aligned}\frac{f_{Y|X}(1|x)}{f_{Y|X}(0|x)} &= \frac{f_{X|Y}(x|1)\ f_Y(1)}{f_{X|Y}(x|0)\ f_Y(0)} \\ &= \frac{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-5)^2}}{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-0)^2}} \\ &= e^{-0.5\sigma^{-2}(x^2-10x+25-x^2)} \\ &= e^{0.5\sigma^{-2}(10x-25)}\end{aligned}$$

- We are only interested in whether the above is greater than 1 or not.
- The ratio is greater than 1 if $10x > 25$ or $x > 2.5$
- So, if $X > 2.5$ we will conclude bit 1 is sent. Intuitively obvious!

- We did not calculate $f_X(x)$ in the above.
- We can calculate it if we want.
- Using total probability rule

$$
\begin{aligned}
f_X(x) &= \sum_y f_{X|Y}(x|y) f_Y(y) \\
&= f_{X|Y}(x|1) f_Y(1) + f_{X|Y}(x|0) f_Y(0) \\
&= \frac{1}{2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2\sigma^2}} + \frac{1}{2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}
\end{aligned}
$$

- It is a mixture density

- As we saw, given the joint distribution we can calculate all the marginals.
- However, there can be many joint distributions with the same marginals.
- Let F_1, F_2 be one dimensional df's of continuous rv's with f_1, f_2 being the corresponding densities.
  Define a function $f : \Re^2 \to \Re$ by

  $$f(x, y) = f_1(x) f_2(y) \left[1 + \alpha(2F_1(x) - 1)(2F_2(y) - 1)\right]$$

  where $\alpha \in (-1, 1)$.
- First note that $f(x, y) \geq 0, \ \forall \alpha \in (-1, 1)$.
  For different α we get different functions.
- We first show that $f(x, y)$ is a joint density.
- For this, we note the following

$$\int_{-\infty}^{\infty} f_1(x)\ F_1(x)\ dx = \left.\frac{(F_1(x))^2}{2}\right|_{-\infty}^{\infty} = \frac{1}{2}$$

$$f(x,y) = f_1(x) f_2(y) \left[1 + \alpha(2F_1(x) - 1)(2F_2(y) - 1)\right]$$

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\; dx\; dy &= \int_{-\infty}^{\infty} f_1(x)\; dx \int_{-\infty}^{\infty} f_2(y)\; dy \\
+\alpha \int_{-\infty}^{\infty} (2 f_1(x) F_1(x) - f_1(x))\; dx &\int_{-\infty}^{\infty} (2 f_2(y) F_2(y) - f_2(y))\; dy \\
&= 1
\end{aligned}$$

because $2 \int_{-\infty}^{\infty} f_1(x)\; F_1(x)\; dx = 1$. This also shows

$$\int_{-\infty}^{\infty} f(x,y) dx = f_2(y); \quad \int_{-\infty}^{\infty} f(x,y) dy = f_1(x)$$

- Thus infinitely many joint distributions can all have the same marginals.
- So, in general, the marginals cannot determine the joint distribution.
- An important special case where this is possible is that of independent random variables

Independent Random Variables

- Two random variable X, Y are said to be independent if for all Borel sets B_1, B_2, the events $[X \in B_1]$ and $[Y \in B_2]$ are independent.
- If X, Y are independent then

$$P[X \in B_1, Y \in B_2] = P[X \in B_1]\ P[Y \in B_2], \ \ \forall B_1, B_2 \in \mathcal{B}$$

- In particular

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

- **Theorem**: X, Y are independent if and only if $F_{XY}(x, y) = F_X(x) F_Y(y)$.

- Suppose X, Y are independent discrete rv's

$$f_{XY}(x, y) = P[X = x, Y = y] = P[X = x]P[Y = y] = f_X(x)f_Y(y)$$

  The joint mass function is a product of marginals.

- Suppose $f_{XY}(x, y) = f_X(x)f_Y(y)$. Then

$$\begin{aligned} F_{XY}(x, y) &= \sum_{x_i \leq x, y_j \leq y} f_{XY}(x_i, y_j) = \sum_{x_i \leq x, y_j \leq y} f_X(x_i) f_Y(y_j) \\ &= \sum_{x_i \leq x} f_X(x_i) \sum_{y_j \leq y} f_Y(y_j) = F_X(x)F_Y(y) \end{aligned}$$

- So, X, Y are independent if and only if $f_{XY}(x, y) = f_X(x)f_Y(y)$

- Let X, Y be independent continuous rv

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

- This implies joint density is product of marginals.
- Now, suppose $f_{XY}(x,y) = f_X(x)f_Y(y)$

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

- So, X, Y are independent if and only if $f_{XY}(x,y) = f_X(x)f_Y(y)$

- Let X, Y be independent.
- Then $P[X \in B_1 | Y \in B_2] = P[X \in B_1]$.
- Hence, we get $F_{X|Y}(x|y) = F_X(x)$.
- This also implies $f_{X|Y}(x|y) = f_X(x)$.
- This is true for all the four possibilities of X, Y being continuous/discrete.

More than two rv

- Everything we have done so far is easily extended to multiple random variables.
- Let X, Y, Z be rv on the same probability space.
- We define joint distribution function by

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

- If all three are discrete then the joint mass function is

$$f_{XYZ}(x, y, z) = P[X = x, Y = y, Z = z]$$

- If they are continuous , they have a joint density if

$$F_{XYZ}(x, y, z) = \int_{-\infty}^{z} \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XYZ}(x', y', z') \, dx' \, dy' \, dz'$$

- Easy to see that joint mass function satisfies
  1. $f_{XYZ}(x, y, z) \geq 0$ and is non-zero only for countably many tuples.
  2. $\sum_{x,y,z} \; f_{XYZ}(x, y, z) = 1$
- Similarly the joint density satisfies
  1. $f_{XYZ}(x, y, z) \geq 0$
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \; f_{XYZ}(x, y, z) \; dx \; dy \; dz = 1$
- These are straight-forward generalizations
- The properties of joint distribution function such as it being non-decreasing in each argument etc are easily seen to hold here too.
- Generalizing the special property of the df (relating to probability of cylindrical sets) is a little more complicated. (An exercise for you!)
- We specify multiple random variables either through joint mass function or joint density function.

- Now we get many different marginals:

$$F_{XY}(x,y) = F_{XYZ}(x,y,\infty); \;\; F_Z(z) = F_{XYZ}(\infty,\infty,z) \quad \text{and so on}$$

- Similarly we get

$$\begin{aligned} f_{YZ}(y,z) &= \int_{-\infty}^{\infty} f_{XYZ}(x,y,z)\,dx; \\ f_X(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)\,dy\,dz \end{aligned}$$

- Any marginal is a joint density of a subset of these rv's and we obtain it by integrating the (full) joint density with respect to the remaining variables.
- We obtain the marginal mass functions for a subset of the rv's also similarly where we sum over the remaining variables.

- We have to be a little careful in dealing with these when some random variables are discrete and others are continuous.
- Suppose X is continuous and Y, Z are discrete. We do not have any joint density or mass function as such.
- However, the joint df is always well defined.
- Suppose we want marginal joint distribution of X, Y. We know how to get F_{XY} by marginalization.
- Then we can get f_X (a density), f_Y (a mass fn), $f_{X|Y}$ (conditinal density) and $f_{Y|X}$ (conditional mass fn)
- With these we can generally calculate most quantities of interest.

- Like in case of marginals, there are different types of conditional distributions now.
- We can always define conditional distribution functions like

$$\begin{aligned} F_{XY|Z}(x, y|z) &= P[X \leq x, Y \leq y|Z = z] \\ F_{X|YZ}(x|y, z) &= P[X \leq x|Y = y, Z = z] \end{aligned}$$

- In all such cases, if the conditioning random variables are continuous, we define the above as a limit.
- For example when Z is continuous

$$F_{XY|Z}(x, y|z) = \lim_{\delta \to 0} P[X \leq x, Y \leq y|Z \in [z, z + \delta]]$$

- If X, Y, Z are all discrete then, all conditional mass functions are defined by appropriate conditional probabilities. For example,

$$f_{X|YZ}(x|y,z) = P[X = x|Y = y, Z = z]$$

- Thus the following are obvious

$$\begin{aligned} f_{XY|Z}(x,y|z) &= \frac{f_{XYZ}(x,y,z)}{f_Z(z)} \\ f_{X|YZ}(x|y,z) &= \frac{f_{XYZ}(x,y,z)}{f_{YZ}(y,z)} \\ f_{XYZ}(x,y,z) &= f_{Z|YX}(z|y,x)f_{Y|X}(y|x)f_X(x) \end{aligned}$$

- For example, the first one above follows from

$$P[X = x, Y = y|Z = z] = \frac{P[X = x, Y = y, Z = z]}{P[Z = z]}$$

- When X, Y, Z have joint density, all such relations hold for the appropriate (conditional) densities. For example,

$$
\begin{aligned}
F_{Z|XY}(z|x,y) &= \lim_{\delta\to 0} \frac{P[Z \leq z, X \in [x, x+\delta], Y \in [y, y+\delta]]}{P[X \in [x, x+\delta, Y \in [y, y+\delta]]} \\
&= \lim_{\delta\to 0} \frac{\int_{-\infty}^{z} \int_{x}^{x+\delta} \int_{y}^{y+\delta} f_{XYZ}(x', y', z') \, dy' \, dx' \, dz'}{\int_{x}^{x+\delta} \int_{y}^{y+\delta} f_{XY}(x', y') \, dy' \, dx'} \\
&= \int_{-\infty}^{z} \frac{f_{XYZ}(x, y, z')}{f_{XY}(x, y)} \, dz' = \int_{-\infty}^{z} f_{Z|XY}(z'|x, y) \, dz'
\end{aligned}
$$

- Thus we get

$$
f_{XYZ}(x, y, z) = f_{Z|XY}(z|x, y) f_{XY}(x, y) = f_{Z|XY}(z|x, y) f_{Y|X}(y|x) f_X(x)
$$

- We can similarly talk about the joint distribution of any finite number of rv's
- Let $X_1, X_2, \cdots, X_n$ be rv's on the same probability space.
- We denote it as a vector $\mathbf{X}$ or $\underline{X}$. We can think of it as a mapping, $\mathbf{X} : \Omega \to \Re^n$.
- We can write the joint distribution as

$$F_{\mathbf{X}}(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_i \leq x_i, \ i = 1, \cdots, n]$$

- We represent by $f_{\mathbf{X}}(\mathbf{x})$ the joint density or mass function. Sometimes we also write it as $f_{X_1 \cdots X_n}(x_1, \cdots, x_n)$
- We use similar notation for marginal and conditional distributions

Independence of multiple random variables

- Random variables $X_1, X_2, \cdots, X_n$ are said to be independent if the the events $[X_i \in B_i],\ i = 1, \cdots, n$ are independent.
  (Recall definition of independence of a set of events)
- Independence implies that the marginals would determine the joint distribution.
- If X, Y, Z are independent then
  $f_{XYZ}(x, y, z) = f_X(x) f_Y(y) f_Z(z)$
- For independent random variables, the joint mass function (or density function) is product of individual mass functions (or density functions)

Example

- Let a joint density be given by

$$f_{XYZ}(x, y, z) = K, \quad 0 < z < y < x < 1$$

First let us determine K.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)\; dz\; dy\; dx &= \int_0^1 \int_0^x \int_0^y K\; dz\; dy\; dx \\
&= K \int_0^1 \int_0^x y\; dy\; dx \\
&= K \int_0^1 \frac{x^2}{2}\; dx \\
&= K \frac{1}{6} \quad \Rightarrow K = 6
\end{aligned}$$

Example

- Let a joint density be given by

$$f_{XYZ}(x,y,z) = K, \quad 0 < z < y < x < 1$$

First let us determine K.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XYZ}(x,y,z)\, dz\, dy\, dx &= \int_{x=0}^{1}\int_{y=0}^{x}\int_{z=0}^{y} K\, dz\, dy\, dx \\
&= K \int_{x=0}^{1}\int_{y=0}^{x} y\, dy\, dx \\
&= K \int_{0}^{1} \frac{x^2}{2}\, dx \\
&= K \frac{1}{6} \quad \Rightarrow K = 6
\end{aligned}$$

$$f_{XYZ}(x,y,z) = 6, \quad 0 < z < y < x < 1$$

- Suppose we want to find the (marginal) joint distribution of X and Z.

$$\begin{aligned} f_{XZ}(x,z) &= \int_{-\infty}^{\infty} f_{XYZ}(x,y,z)\, dy \\ &= \int_{z}^{x} 6\, dy, \quad 0 < z < x < 1 \\ &= 6(x-z), \quad 0 < z < x < 1 \end{aligned}$$

- We got the joint density as

$$f_{XZ}(x,z) = 6(x-z), \quad 0 < z < x < 1$$

- We can verify this is a joint density

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XZ}(x,z)\, dz\, dx &= \int_0^1 \int_0^x 6(x-z)\, dz\, dx \\
&= \int_0^1 \left(6x\ z|_0^x - 6\ \frac{z^2}{2}\Big|_0^x \right)\ dx \\
&= \int_0^1 \left(6x^2 - 6\ \frac{x^2}{2} \right)\ dx \\
&= 3\ \frac{x^3}{3}\Big|_0^1 = 1
\end{aligned}$$

- The joint density of X, Y, Z is

$$f_{XYZ}(x, y, z) = 6, \quad 0 < z < y < x < 1$$

- The joint density of X, Z is

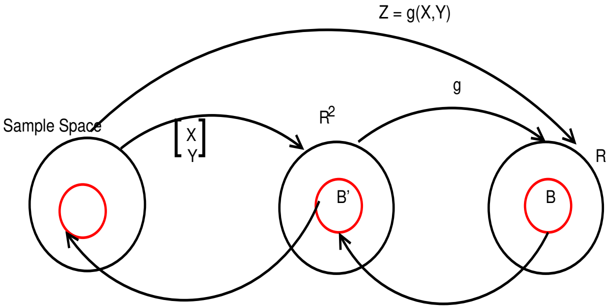
$$f_{XZ}(x, z) = 6(x - z), \quad 0 < z < x < 1$$

- Hence,

$$f_{Y|XZ}(y|x, z) = \frac{f_{XYZ}(x, y, z)}{f_{XZ}(x, z)} = \frac{1}{x - z}, \quad 0 < z < y < x < 1$$

Functions of multiple random variables

- Let X, Y be random variables on the same probability space.
- Let $g : \Re^2 \to \Re$.
- Let $Z = g(X, Y)$. Then Z is a rv
- This is analogous to functions of a single rv



- let $Z = g(X, Y)$
- We can determine distribution of Z from the joint distribution of X, Y

$$F_Z(z) = P[Z \leq z] = P[g(X, Y) \leq z]$$

- For example, if X, Y are discrete, then

$$f_Z(z) = P[Z = z] = P[g(X, Y) = z] = \sum_{\substack{x_i, y_j: \\ g(x_i, y_j) = z}} f_{XY}(x_i, y_j)$$

- Let X, Y be discrete rv's. Let $Z = \min(X, Y)$.

$$
\begin{aligned}
f_Z(z) &= P[\min(X,Y) = z] \\
&= P[X = z, Y > z] + P[Y = z, X > z] + P[X = Y = z] \\
&= \sum_{y>z} P[X = z, Y = y] + \sum_{x>z} P[X = x, Y = z] \\
&\qquad\qquad\qquad\qquad + P[X = z, Y = z] \\
&= \sum_{y>z} f_{XY}(z, y) + \sum_{x>z} f_{XY}(x, z) + f_{XY}(z, z)
\end{aligned}
$$

- Now suppose X, Y are independent and both of them have geometric distribution with the same parameter, p.
- Such random variables are called **independent and identically distributed** or **iid** random variables.

- Now we can get pmf of Z as (note $Z \in \{1, 2, \cdots\}$)

$$
\begin{aligned}
f_Z(z) &= P[X = z, Y > z] + P[Y = z, X > z] + P[X = Y = z] \\
&= P[X = z]P[Y > z] + P[Y = z]P[X > z] + P[X = z]P[Y = z] \\
&= p(1-p)^{z-1}(1-p)^z * 2 + \left(p(1-p)^{z-1}\right)^2 \\
&= 2p(1-p)^{z-1}(1-p)^z + \left(p(1-p)^{z-1}\right)^2 \\
&= 2p(1-p)^{2z-1} + p^2(1-p)^{2z-2} \\
&= p(1-p)^{2z-2}(2(1-p) + p) \\
&= (2-p)p(1-p)^{2z-2}
\end{aligned}
$$

- We can show this is a pmf

$$\begin{aligned}
\sum_{z=1}^{\infty} f_Z(z) &= \sum_{z=1}^{\infty} (2-p)p(1-p)^{2z-2} \\
&= (2-p)p \sum_{z=1}^{\infty} (1-p)^{2z-2} \\
&= (2-p)p \, \frac{1}{1-(1-p)^2} \\
&= (2-p)p \, \frac{1}{2p-p^2} = 1
\end{aligned}$$

- Let us consider the $\max$ and $\min$ functions, in general.
- Let $Z = \max(X, Y)$. Then we have

$$\begin{aligned} F_Z(z) &= P[Z \leq z] = P[\max(X, Y) \leq z] \\ &= P[X \leq z, Y \leq z] \\ &= F_{XY}(z, z) \\ &= F_X(z) F_Y(z), \quad \text{if } X, Y \text{ are independent} \\ &= (F_X(z))^2, \quad \text{if they are iid} \end{aligned}$$

- This is true of all random variables.
- Suppose X, Y are iid continuous rv. Then density of Z is

$$f_Z(z) = 2F_X(z) f_X(z)$$

- Suppose X, Y are iid uniform over $(0,\ 1)$
- Then we get df and pdf of $Z = \max(X, Y)$ as

$$F_Z(z) = z^2, 0 < z < 1; \quad \text{and} \quad f_Z(z) = 2z, 0 < z < 1$$

$F_Z(z) = 0$ for $z \leq 0$ and $F_Z(z) = 1$ for $z \geq 1$ and $f_Z(z) = 0$ outside $(0, 1)$

- This is easily generalized to n radom variables.
- Let $Z = \max(X_1, \cdots, X_n)$

$$
\begin{aligned}
F_Z(z) &= P[Z \leq z] = P[\max(X_1, X_2, \cdots, X_n) \leq z] \\
&= P[X_1 \leq z, X_2 \leq z, \cdots, X_n \leq z] \\
&= F_{X_1 \cdots X_n}(z, \cdots, z) \\
&= F_{X_1}(z) \cdots F_{X_n}(z), \quad \text{if they are independent} \\
&= (F_X(z))^n, \quad \text{if they are iid} \\
&\qquad \text{where we take } F_X \text{ as the common df}
\end{aligned}
$$

- For example if all X_i are uniform over $(0, 1)$ and ind, then $F_Z(z) = z^n, \ 0 < z < 1$

- Consider $Z = \min(X, Y)$ and X, Y independent

$$F_Z(z) = P[Z \leq z] = P[\min(X, Y) \leq z]$$

- It is difficult to write this in terms of joint df of X, Y.
- So, we consider the following

$$\begin{aligned} P[Z > z] &= P[\min(X, Y) > z] \\ &= P[X > z, Y > z] \\ &= P[X > z]P[Y > z], \quad \text{using independence} \\ &= (1 - F_X(z))(1 - F_Y(z)) \\ &= (1 - F_X(z))^2, \quad \text{if they are iid} \end{aligned}$$

$$\text{Hence,} \quad F_Z(z) = 1 - (1 - F_X(z))(1 - F_Y(z))$$

- We can once again find density of Z if X, Y are continuous

- Suppose X, Y are iid uniform $(0,\ 1)$.
- $Z = \min(X, Y)$

$$F_Z(z) = 1 - (1 - F_X(z))^2 = 1 - (1 - z)^2, 0 < z < 1$$

- Notice that $P[X > z] = (1 - z)$.
- We get the density of Z as

$$f_Z(z) = 2(1 - z), \ \ 0 < z < 1$$

- $\min$ fn is also easily generalized to n random variables
- Let $Z = \min(X_1, X_2, \cdots, X_n)$

$$\begin{aligned} P[Z > z] &= P[\min(X_1, X_2, \cdots, X_n) > z] \\ &= P[X_1 > z, \cdots, X_n > z] \\ &= P[X_1 > z] \cdots P[X_n > z], \quad \text{using independence} \\ &= (1 - F_{X_1}(z)) \cdots (1 - F_{X_n}(z)) \\ &= (1 - F_X(z))^n, \quad \text{if they are iid} \end{aligned}$$

- Hence, when X_i are iid, the df of Z is

$$F_Z(z) = 1 - (1 - F_X(z))^n$$

where F_X is the common df