Recap: Joint Distribution Function

- Given X, Y rv on same probability space, joint distribution function: $F_{XY} : \Re^2 \to \Re$

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

- 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 $A1$ and $A2$ above is a joint pmf. (The 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

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

Example: Conditional pmf

- Consider the random experiment of tossing a coin n times.
- Let X denote the number of heads and let Y denote the toss number on which the first head comes.
- For $1 \leq k \leq n$

$$
\begin{aligned}
f_{Y|X}(k|1) &= P[Y=k|X=1] = \frac{P[Y=k, X=1]}{P[X=1]} \\
&= \frac{p(1-p)^{n-1}}{{}^nC_1 p(1-p)^{n-1}} \\
&= \frac{1}{n}
\end{aligned}
$$

- Given there is only one head, it is equally likely to occur on any toss.

- Let X, Y be continuous rv's with joint density, f_{XY}.
- We once again want to define conditional df

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

- But the conditioning event, $[Y = y]$ has zero probability.
- Hence we define conditional df as follows

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

- This is well defined if the limit exists.
- The limit exists for all y where $f_Y(y) > 0$ (and for all x)

- The conditional df is given by (assuming $f_Y(y) > 0$)

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

- We define conditional density of X given Y as

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

- 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

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

Example

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

- We saw that the marginal densities are

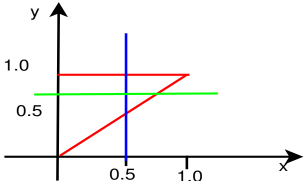
$$f_X(x) = 2(1-x),\ 0 < x < 1; \qquad f_Y(y) = 2y,\ 0 < y < 1$$

- Hence the conditional densities are given by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{y},\ 0 < x < y < 1$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{1-x},\ 0 < x < y < 1$$

- We can see this intuitively like this



- The identity $f_{XY}(x,y) = f_{X|Y}(x|y) f_Y(y)$ can be used to specify the joint density of two continuous rv's
- We can specify the marginal density of one and the conditional density of the other given the first.
- This may actually be the model of how the the rv's are generated.

Example

- Let X be uniform over $(0,\ 1)$ and let Y be uniform over 0 to X. Find the density of Y.
- What we are given is

$$f_X(x) = 1,\ 0 < x < 1; \quad f_{Y|X}(y|x) = \frac{1}{x}, 0 < y < x < 1$$

- Hence the joint density is: $f_{XY}(x, y) = \frac{1}{x},\ 0 < y < x < 1.$
- Hence the density of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ dx = \int_y^1 \frac{1}{x}\ dx = -\ln(y),\ 0 < y < 1$$

- We can verify it to be a density

$$-\int_0^1 \ln(y)\ dy = -y\ln(y)|_0^1 + \int_0^1 y\ \frac{1}{y}\ dy = 1$$

- We have the identity
$$f_{XY}(x,y) = f_{X|Y}(x|y)\ f_Y(y)$$
- By integrating both sides
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\ f_Y(y)\ dy$$
- This is a continuous analogue of total probability rule.
- But note that, since X is continuous rv, $f_X(x)$ is **NOT** $P[X=x]$
- In case of discrete rv, the mass function value $f_X(x)$ is equal to $P[X=x]$ and we had
$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y)$$
- It is as if one can simply replace pmf by pdf and summation by integration!!
- While often that gives the right result, one needs to be very careful

- We have the identity

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

- This gives rise to Bayes rule for continuous rv

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

- This is essentially identical to Bayes rule for discrete rv's. We have essentially put the pdf wherever there was pmf

- To recap, we started by defining conditional distribution function.
$$F_{X|Y}(x|y) = P[X \leq x | Y = y]$$
- When X, Y are discrete, we define this only for $y = y_j$. That is, we define it only for all values that Y can take.
- When X, Y have joint density, we defined it by
$$F_{X|Y}(x|y) = \lim_{\delta \to 0} P[X \leq x | Y \in [y, y + \delta]]$$
This limit exists and $F_{X|Y}$ is well defined if $f_Y(y) > 0$. That is, essentially again for all values that Y can take.
- In the discrete case, we define $f_{X|Y}$ as the pmf corresponding to $F_{X|Y}$. This conditional pmf can also be defined as a conditional probability
- In the continuous case $f_{X|Y}$ is the density corresponding to $F_{X|Y}$.
- In both cases we have: $f_{XY}(x, y) = f_{X|Y}(x|y) f_Y(y)$
- This gives total probability rule and Bayes rule for random variables

- 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 $f_{X|Y}(x|i) = f_i(x)$

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

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

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

- 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

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

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

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

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

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

- **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.