Recap: Random Variables

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

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

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

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

Recap: Distribution function of a random variable

- Let X be a random variable. It distribution function, $F_X : \Re \rightarrow \Re$, 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 $X \in \{x_1, x_2, \cdots\}$
- 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 $X \in \{x_1, x_2, \cdots\}$.
- 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 $f_X : \Re \to \Re$ 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

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

- 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
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \;\; \forall x$$
- It satisfies
  1. $f_X(x) \geq 0, \;\; \forall x$
  2. $\int_{-\infty}^{\infty} f_X(t)\, dt = 1$
- We can, in principle, compute probability of any event as
$$P[X \in B] = \int_B \, f_X(t)\, dt, \;\; \forall B \in \mathcal{B}$$
- 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 $g : \Re \to \Re$ 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 $X \in \{x_1, x_2, \cdots\}$.
- Suppose $Y = g(X)$.
- Then Y is also discrete and $Y \in \{g(x_1), g(x_2), \cdots\}$.
- 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

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

  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 $X \in \{x_1, x_2, \cdots\}$. 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.
- $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$
- $E\left[(X - c)^2\right] \geq E\left[(X - EX)^2\right],\ \forall c$

Recap: Variance of random variable

- $$\mathsf{Var}(X) = E\left[(X - EX)^2\right] = E[X^2] - (EX)^2$$
- Properties of Variance:
  - $\mathsf{Var}(X) \geq 0$
  - $\mathsf{Var}(X + c) = \mathsf{Var}(X)$
  - $\mathsf{Var}(cX) = c^2 \mathsf{Var}(X)$

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 – $M_X : \Re \rightarrow \Re$

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

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

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

- We get the pmf from it as

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

- 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
$$P[|X| > c] \leq \frac{E\left[|X|^k\right]}{c^k}$$
- **Chebyshev inequality**
$$P[|X - EX| > c] \leq \frac{\mathsf{Var}(X)}{c^2}$$
- With $EX = \mu$ and $\mathsf{Var}(X) = \sigma^2$, 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 Ω to $\Re^2$.
- This gives rise to a new probability space $(\Re^2, \mathcal{B}^2, P_{XY})$ with P_{XY} given by

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

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 $F_{XY} : \Re^2 \to \Re$, defined by

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

- The joint distribution function is the probability of the intersection of the events $[X \leq x]$ and $[Y \leq y]$.
- Given the joint distribution function, probability of any event involving the pair of random variables can be (in principle) calculated.

Recap: Properties of Joint Distribution Function

- 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

- Let $X \in \{x_1, x_2, \cdots\}$ and $Y \in \{y_1, y_2, \cdots\}$ 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

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

Recap: Joint pmf

- Given sets $\{x_1, x_2, \cdots\}$ and $\{y_1, y_2, \cdots\}$.
- 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

Recap: 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,

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

Joint density function

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

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

  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.

properties of joint density

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

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

- 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

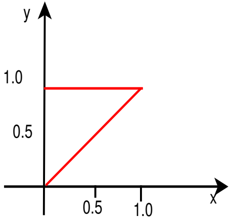
- Consider the function

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

- Let us show this is a density

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

- We can say this density is uniform over the region



The figure is not a plot of the density function!!

properties of joint density

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

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

- 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

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

- 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 $f_{XY} : \Re^2 \to \Re$ 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

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

- Let us consider the example

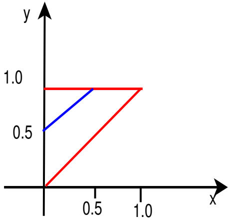
$$f(x, y) = 2, \quad 0 < x < y < 1$$

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

$$\begin{aligned}
P[Y > X + 0.5] &= P\left[(X, Y) \in \{(x, y) : y > x + 0.5\}\right] \\
&= \int_{\{(x,y)\,:\,y > x + 0.5\}} f_{XY}(x, y)\ dx\ dy \\
&= \int_{0.5}^{1} \int_{0}^{y - 0.5} 2\ dx\ dy \\
&= \int_{0.5}^{1} 2(y - 0.5) dy
\end{aligned}$$

$$= 2\ \frac{y^2}{2}\bigg|_{0.5}^{1} - y|_{0.5}^{1} = 1 - 0.25 - 1 + 0.5 = 0.25$$

- 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 $X \in \{x_1, x_2, \cdots\}$ and $Y \in \{y_1, y_2, \cdots\}$
- 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 $[Y = y_j],\ j = 1, \cdots$, 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, $f_X(m) = \sum_n f_{XY}(m,n), \; m = 1, \cdots, 6.$
- Given m, for what values of n, $f_{XY}(m,n) > 0$?
  We can only have $n = m+1, \cdots, 2m$.
- 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

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

- 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

$$\int_{-\infty}^{\infty} f_X(x)\ dx = \int_0^1 2(1 - x)\ dx = \left.(2x - x^2)\right|_0^1 = 1$$

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

$$\int_0^1 2y\; dy = 2\; \frac{y^2}{2}\Big|_0^1 = 1$$

- 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

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

- 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: $X \in \{x_1, x_2, \cdots\}$ 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.
- Since X is a discrete rv, we can write the above as

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

Conditional mass function

- We got

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

- We define the conditional mass function of X given Y as

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

- Note that

$$\sum_i f_{X|Y}(x_i|y_j) = 1, \ \forall y_j; \quad \text{and} \quad F_{X|Y}(x|y_j) = \sum_{i:x_i \leq x} f_{X|Y}(x_i|y_j)$$

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.

- 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 proability rule for discrete 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 $B_1, \cdots$ form a partition)

Bayes Rule for discrete Random Variable

- 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

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

- 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:

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

- 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

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

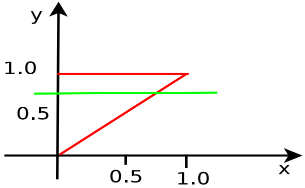
- We saw that the marginal densities are

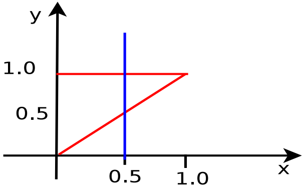
$$f_X(x) = 2(1-x), \ 0 < x < 1; \qquad f_Y(y) = 2y, \ 0 < y < 1$$

- Hence the conditional densities are given by

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

- 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