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 mostly consider rv with finite expectation where 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 result is true for all rv though we consider only discrete or continuous.

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

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

- 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 (mgf) of rv X, $M_X : \Re \to \Re$, is defined by

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

- We say the mgf exists if $E[e^{tX}] < \infty$ for t in some interval around zero. When it exits we have

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

- The mgf may not exist for some random variables.
- The mgf uniquely determines the df

Recap: Characteristic Function

- The characteristic function of X is defined by

$$\phi_X(t) = E[e^{itX}] = \int e^{itx} \, dF_X(x) \quad (i = \sqrt{-1})$$

- Characteristic function always exists because $|e^{itx}| = 1, \forall t, x$
- We can get the moments of the random variable by differentiating the characteristic function also.
- Characteristic function also uniquely determines the df

Recap: Generating function

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

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

  This infinite sum converges (absolutely) for $|s| \leq 1$.

- We have

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

$$P'_X(1) = EX; \; P''_X(1) = E[X(X-1) \; \cdots$$

- 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.
- Suppose x is a quantile of order p. Then we have
  - $p \leq P[X \leq x] = F_X(x)$
  - $1 - p \leq 1 - P[X < x] = 1 - (P[X \leq x] - P[X = x])$
    $\Rightarrow \ \ 1 - p \leq 1 - F_X(x) + P[X = x]$
    $\Rightarrow F_X(x) \leq p + P[X = x]$
- Thus, x satisfies (if it is quantile of order p)

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

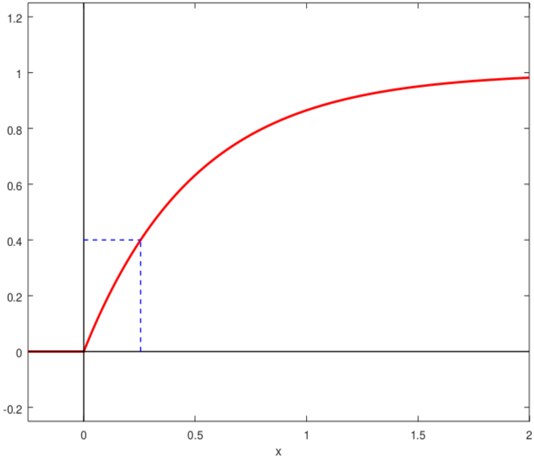
- Note that for a given p there can be multiple values for x to satisfy the above.

- If x is a quantile of order p then

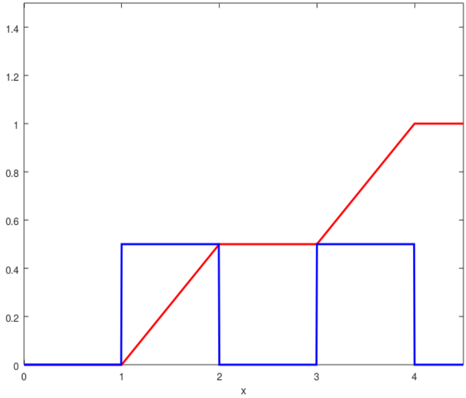
$$p \leq F_X(x) \leq p + P[X = x]$$

- If X is continuous rv, we need to satisfy $p = F_X(x)$.
- In general, for a given p, there may be multiple x that satisfy the above.
- Let us see some examples.

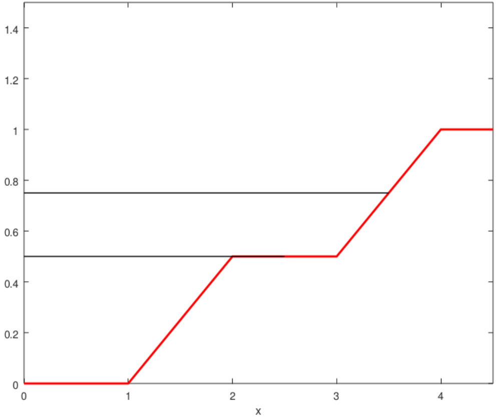
- Let X be continuous rv.
- If the df is strictly monotone then $F_X(x) = p$ would have a unique solution.



- For continuous rv, X, F_X need not be strictly monotone.
- Consider a pdf: $f_X(x) = 0.5, \; x \in [1, \; 2] \cup [3, \; 4]$
- The pdf and the corresponding df are:

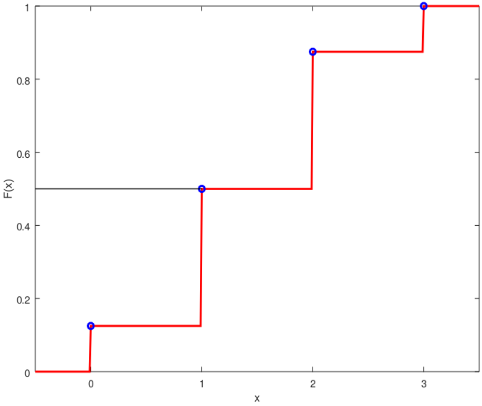


- For this df, for $p = 0.5$, the quantile of order p is not unique because there many x with $F_X(x) = 0.5$
  But for $p = 0.75$ it is unique.



- Let $X \in \{x_1, x_2, \cdots\}$
- Given a p we want to calculate quantile of order p
- Suppose there is a x_i such that $F_X(x_i) = p$.
- Then, for $x_i \leq x < x_{i+1}$, $F_X(x) = p$
- For $x_i \leq x \leq x_{i+1}$, we have $p \leq F_X(x) \leq p + P[X = x]$
- So, quantile of order p is not unique and all such x qualify.

- This situation is illustrated below

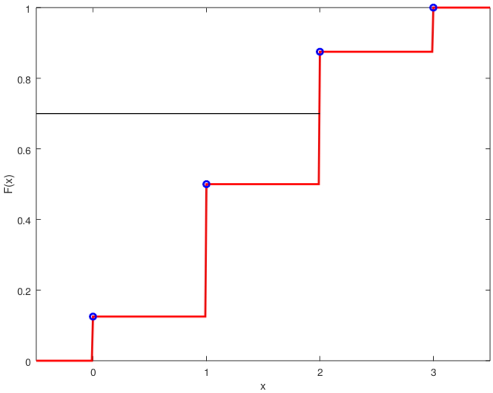


- Now suppose p is such that $F_X(x_{i-1}) < p < F_X(x_i)$.
- Let $F_X(x_{i-1}) = p - \delta_1$ and $F_X(x_i) = p + \delta_2$. (Note that $\delta_1, \delta_2 > 0$)
- Then $P[X = x_i] = F_X(x_i) - F_X(x_{i-1}) = \delta_2 + \delta_1$
- Hence we have

$$p < p + \delta_2 = F_X(x_i) < p + \delta_2 + \delta_1 = p + P[X = x_i]$$

- Hence, x_i is quantile of order p.
- For any $x < x_i$ we would have $F_X(x) \leq F_X(x_{i-1}) < p$.
- For any x, with $x_i < x < x_{i+1}$ we have $p + P[X = x] = p < F_X(x) = p + \delta_2$.
- Similarly, for $x \geq x_{i+1}$ we have $F_X(x) > p + P[X = x]$.
- Thus quantile of order p is unique here.

- This situation is illustrated below



Median of a distribution

- For $p = 0.5$ quantile of order p is called the median.
- For a continuous rv, median, x satisfies: $F_X(x) = 0.5$.
- For a discrete rv, it satisfies:
  $0.5 \leq F_X(x) \leq 0.5 + P[X = x]$.
- As we saw, median need not be unique.
- Recall that the (standard) Cauchy density is given by

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \; -\infty < x < \infty$$

- One can show that $\int_{-\infty}^{0} f_X(x) \, dx = 0.5$ and hence the median is at the origin.

- If we want to find c to minimize $E\left[(X-c)^2\right]$ then the solution is $c = EX$.
- We saw this earlier.
- Suppose we want to find c to minimize $E\left[|(X-c)|\right]$
- Then we would get c to be the median.
  (Exercise: Show this for discrete and continuous rv)

- The value of x where $f_X(x)$ attains its maximum value is called the mode of a distribution.
- For a discrete random variable it is the value that the random variable takes with highest probability.
- Take X to be binomial (with parameters, n, p). Then the mode is that value of k for which $f_X(k)$ is maximum. This gives the 'most probable' number of heads when we toss this coin n times.
- For a continuous rv, we can say mode is the point with 'maximum likelihood'
- For the Gaussian density, the mode, the median and the mean are all same.

- We next consider some inequalities involving moments of a random variable.
- These help us bound the probabilities of some important events in terms of the moments.

Markov Inequality

- Let $g : \Re \rightarrow \Re$ be a non-negative function. Then
$$P[g(X) > c] \leq \frac{E[g(X)]}{c}, \quad (c > 0)$$
- **Proof**: We prove it for continuous rv. Proof is similar for discrete rv
$$\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \\
&= \int_{g(x) \leq c} g(x)\, f_X(x)\, dx \, + \int_{g(x) > c} g(x)\, f_X(x)\, dx \\
&\geq \int_{g(x) > c} g(x)\, f_X(x)\, dx \quad \text{because } g(x) \geq 0 \\
&\geq c \int_{g(x) > c} f_X(x)\, dx \, = \, c\, P[g(X) > c]
\end{aligned}$$
Thus, $P[g(X) > c] \leq \frac{E[g(X)]}{c}$

Markov Inequality

$$P[g(X) > c] \leq \frac{E[g(X)]}{c}, \quad (c > 0)$$

- In all such results an underlying assumption is that the expectation is finite.
- Let $g(x) = |x|^k$ where k is a positive integer. We have $g(x) \geq 0, \ \forall x$. Let $c > 0$.
- We know that $|x| > c \Rightarrow |x|^k > c^k$ and vice versa.
- Now we get,

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

  (For what k is this true?)
- Markov inequality is often used in this form.

Chebyshev Inequality

- Markov Inequality:

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

- Take $|X|$ as $|X - EX|$ and take $k = 2$

$$P[|X - EX| > c] \leq \frac{E\left[|X - EX|^2\right]}{c^2} = \frac{\mathsf{Var}(X)}{c^2}$$

- This is known as the Chebyshev inequality.
- An example of what are called concentration inequalities.

- The Chebyshev inequality is

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

- Let $EX = \mu$ and let $\mathsf{Var}(X) = \sigma^2$. Take $c = k\sigma$
- We call, σ, square root of variance, as standard deviation.
- Now, Chebyshev inequality gives us

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

- This is true for all random variables and the RHS above does not depend on the distribution of X.

- **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**

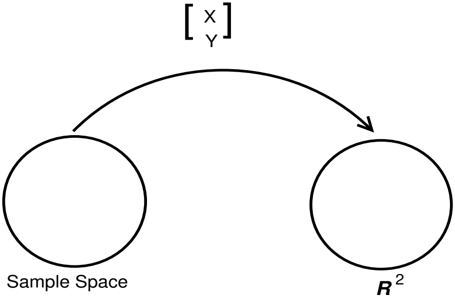
$$P[|X - EX| > c] \leq \frac{\mathsf{Var}(X)}{c^2}$$

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

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

A pair of random variables

- Let X, Y be random variables on the same probability space $(\Omega, \mathcal{F}, P)$
- Each of X, Y maps Ω to $\Re$.
- We can think of the pair of radom variables as a vector-valued function that maps Ω to $\Re^2$.

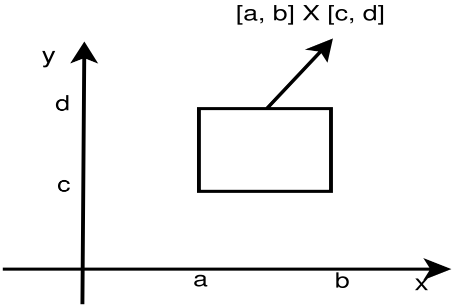


- Just as in the case of a single rv, we can think of the induced probability space for the case of a pair of rv's too.
- That is, by defining the pair of random variables, we essentially create a new probability space with sample space being $\Re^2$.
- The events now would be the Borel subsets of $\Re^2$.
- Recall that $\Re^2$ is cartesian product of $\Re$ with itself.
- So, we can create Borel subsets of $\Re^2$ by cartesian product of Borel subsets of $\Re$.

$$\mathcal{B}^2 = \sigma\left(\{B_1 \times B_2 \ : \ B_1, B_2 \in \mathcal{B}\}\right)$$

where $\mathcal{B}$ is the Borel σ-algebra we considered earlier, and $\mathcal{B}^2$ is the set of Borel sets of $\Re^2$.

- Recall that $\mathcal{B}$ is the smallest σ-algebra containing all intervals.
- Let $I_1, I_2 \subset \Re$ be intervals. Then $I_1 \times I_2 \subset \Re^2$ is known as a cylindrical set.



- $\mathcal{B}^2$ is the smallest σ-algebra containing all cylindrical sets.
- We saw that $\mathcal{B}$ is also the smallest σ-algebra containing all intervals of the form $(-\infty, x]$.
- Similarly $\mathcal{B}^2$ is the smallest σ-algebra containing cylindrical sets of the form $(-\infty, x] \times (-\infty, y]$.

- Let X, Y be random variables on the probability space $(\Omega, \mathcal{F}, P)$
- This gives rise to a new probability space $(\Re^2, \mathcal{B}^2, P_{XY})$ with P_{XY} given by

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

(Here, $B \subset \Re^2$)

- Recall that for a single rv, the resulting probability space is $(\Re, \mathcal{B}, P_X)$ with

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

(Here, $B \subset \Re$)

- In the case of a single rv, we define a distribution function, F_X which essentailly assigns probability to all intervals of the form $(-\infty,\ x]$.
- This F_X uniquely determines $P_X(B)$ for all Borel sets, B.
- In a similar manner we define a joint distribution function F_{XY} for a pair of random varibles.
- $F_{XY}(x,y)$ would be $P_{XY}((-\infty,x] \times (-\infty,y])$.
- F_{XY} fixes the probability of all cylindrical sets of the form $(-\infty,x] \times (-\infty,y]$ and hence uniquely determines the probability of all Borel sets of $\Re^2$.

Joint distribution of a pair of random variables

- 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 $[Y \leq y]$.

Properties of Joint Distribution Function

- Joint distribution function:

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

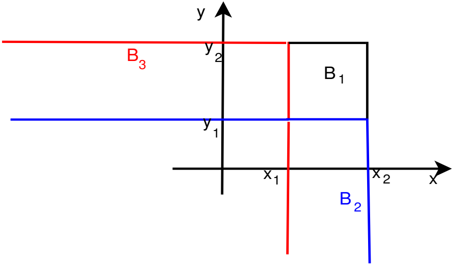
  - $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \forall x, y;$
    $F_{XY}(\infty, \infty) = 1$
    (These are actually limits: $\lim_{x \to -\infty} F_{XY}(x, y) = 0, \forall y$)
  - F_{XY} is non-decresing in each of its arguments
  - F_{XY} is right continuous and has left-hand limits in each of its arguments
- These are straight-forward extensions of single rv case
- But there is another crucial property satisfied by F_{XY}.

- Recall that, for the case of a single rv, the probability of X being in any interval is given by the difference of F_X values at the end points of the interval.
- Let $x_1 < x_2$. Then

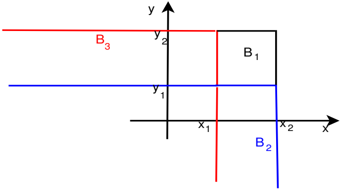
$$P[x_1 < X \leq x_2] = F_X(x_2) - F_X(x_1)$$

- The LHS above is a probability.
  Hence the RHS should be non-negative
  The RHS is non-negative because F_X is non-decreasing.
- We will now derive a similar expression in the case of two random variables.
- Here, the probability we want is that of the pair of rv's being in a cylindrical set.

- Let $x_1 < x_2$ and $y_1 < y_2$. We want $P[x_1 < X \leq x_2, y_1 < Y \leq y_2]$.
- Consider the Borel set $B = (-\infty, x_2] \times (-\infty, y_2]$.



$$
\begin{aligned}
B &\triangleq (-\infty, x_2] \times (-\infty, y_2] = B_1 + (B_2 \cup B_3) \\
B_1 &= (x_1,\ x_2] \times (y_1,\ y_2] \\
B_2 &= (-\infty, x_2] \times (-\infty, y_1] \\
B_3 &= (-\infty, x_1] \times (-\infty, y_2] \\
B_2 \cap B_3 &= (-\infty, x_1] \times (-\infty, y_1]
\end{aligned}
$$



$$
\begin{aligned}
F_{XY}(x_2, y_2) &= P[X \leq x_2, Y \leq y_2] = P[(X,Y) \in B] \\
&= P[(X,Y) \in B_1 + (B_2 \cup B_3)] \\
&= P[(X,Y) \in B_1] + P[(X,Y) \in (B_2 \cup B_3)]
\end{aligned}
$$

$$
\begin{aligned}
P[(X,Y) \in B_2] &= P[X \leq x_2, Y \leq y_1] = F_{XY}(x_2, y_1) \\
P[(X,Y) \in B_3] &= P[X \leq x_1, Y \leq y_2] = F_{XY}(x_1, y_2) \\
P[(X,Y) \in B_2 \cap B_3] &= P[X \leq x_1, Y \leq y_1] = F_{XY}(x_1, y_1)
\end{aligned}
$$

$$
\begin{aligned}
P[(X,Y) \in B_1] &= F_{XY}(x_2, y_2) - P[(X,Y) \in (B_2 \cup B_3)] \\
&= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1)
\end{aligned}
$$

- What we showed is the following.
- For $x_1 < x_2$ and $y_1 < y_2$

$$\begin{aligned} P[x_1 < X \leq x_2, y_1 < Y \leq y_2] &= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) \\ &\quad - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1) \end{aligned}$$

- This means F_{XY} should satisfy

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

  for all $x_1 < x_2$ and $y_1 < y_2$
- This is an additional condition that a function has to satisfy to be the joint distribution function of a pair of random variables

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.

- Let X, Y be two discrete random variables (defined on the same probability space).
- Let $X \in \{x_1, \cdots x_n\}$ and $Y \in \{y_1, \cdots, y_m\}$.
- 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 ω is a real number between 0 and 1
- Let $X(\omega)$ be the digit in the first decimal place in ω 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 $X, Y \in \{0, 1, \cdots, 9\}$.
- 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

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

Example

- Consider the random experiment of rolling two dice. $\Omega = \{(\omega_1, \omega_2) \ : \ \omega_1, \omega_2 \in \{1, 2, \cdots, 6\}\}$
- Let X be the maximum of the two numbers and let Y be the sum of the two numbers.
- Easy to see $X \in \{1, 2, \cdots, 6\}$ and $Y \in \{2, 3, \cdots, 12\}$
- 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 $X \in \{x_1, x_2, \cdots\}$ 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

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

- 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

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

  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 $m = 1, \cdots, 6$ and $n = 2, \cdots, 12$. 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 $P[Y = X + 2]$.

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