- Probability spaces are triplets of $(\Omega, \mathcal{F}, \Pr(\cdot))$ consisting of a *sample space* Ω, set of *events* $\mathcal{F}$, and a *probability measure* $\Pr(\cdot)$
  - *Sample space*: any set Ω
  - *Events* $\mathcal{F}$: This is a set consisting of **subsets** of Ω satisfying:
    a. $\Omega \in \mathcal{F}$
    b. **Closed under complement**: $E \in \mathcal{F}$ implies $E^c \in \mathcal{F}$, and
    c. **Closed under countable union**: for any countably many subsets $E_1, \ldots, E_k, \ldots \in \mathcal{F}$, we have $\cup_{k=1}^{\infty} E_k \in \mathcal{F}$
  - *Probability measure* $\Pr(\cdot)$: is a function from $\mathcal{F}$ to $\mathbb{R}^+ := \{x \in \mathbb{R} \mid x \geq 0\}$ that satisfies:
    i. $\Pr(\Omega) = 1$, and
    ii. For a countably many subsets $\{E_k\}$ in $\mathcal{F}$ that is mutually disjoint (i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j$), we have

$$\Pr(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \Pr(E_k).$$

- Question: How to show that a function $X : \Omega \to \mathbb{R}$ is a random variable?
- There are several ways:
  a. By definition, showing that for all $B \in \mathcal{B}$, it suffices to show that $X^{-1}((-\infty, a]) \in \mathcal{F}$ for all $a \in \mathbb{R}$.
  b. Practical way: Let $g : \mathbb{R}^n \to \mathbb{R}$ be a continuous mapping and let $X_1, \ldots, X_n$ be r.v's. Then $X = g(X_1, X_2, \ldots, X_n)$ is a r.v. This allows us to construct new random variables from the old ones: for example if X, Y are random variables, $X+Y$, $X-Y$, $X \times Y$, X^Y, etc. are all random variables.

- For a r.v. X, we define the distribution function (or cumulative distribution function (CDF)) of X, to be the mapping $F_X : \mathbb{R} \to [0, 1]$ defined by $F(x) = \Pr(X^{-1}((-\infty, x]))$.
- Properties of Distribution Functions (see HW 3):
  a. F_X is non-decreasing.
  b. $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$.
  c. $F_X(\cdot)$ **is right-continuous**, i.e., for any $x \in \mathbb{R}$, $\lim_{y \to x^+} F_X(y) = F_X(x)$.
  d. Define $F_X(x^-) := \lim_{y \uparrow x} F_X(y)$, then

$$F_X(x^-) = \Pr(X < x) = \Pr(\{\omega \in \Omega \mid X(\omega) < x\}).$$

  e. For any $x \in \mathbb{R}$, we have $\Pr(X = x) = F_X(x) - F_X(x^-)$.

- Expected value of simple r.v.s: For $X = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}$, we define

$$\mathbb{E}[X] := \sum_{i=1}^{m} \alpha_i \Pr(A_i).$$

- For a positive random variable X (i.e., $X \geq 0$ almost surely), we define:

$$\mathbb{E}[X] := \sup\{\mathbb{E}[Y] \mid Y \leq X \text{ and } Y \text{ is a simple function}\}.$$

- Define positive and negative side of a random variable as: $X^+ = \mathbf{1}_{X \geq 0} X$ and $X^- = -\mathbf{1}_{X \leq 0} X$. Note that they are both non-negative r.v.s.
- We say that the expected value of X exists if either $\mathbb{E}[X^+] < \infty$ or $\mathbb{E}[X^-] < \infty$ and we let it be

$$\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

- If $X \geq 0$, then $\mathbb{E}[X] \geq 0$.
- *monotonicity*: if $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- $\mathbb{E}[|X|] = 0$ if and only if $X = 0$ almost surely.
- Markov Inequality: For a non-negative random variable X,

$$\Pr(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$$

  for any $\alpha > 0$.
- Jensen's Inequality: For a convex function $\Phi : \mathbb{R} \to \mathbb{R}$,

$$\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)].$$

  Since $-\Phi$ is a concave function, the reverse inequality holds for concave functions.
- Very^n important result: Monotone Convergence Theorem (MCT): Suppose that $X_1 \leq X_2 \leq \cdots \leq X = \lim_{k \to \infty} X_k$. Then,

$$\lim_{k \to \infty} \mathbb{E}[X_k] = \mathbb{E}[\lim_{k \to \infty} X_k] = \mathbb{E}[X].$$

Definition 1. *Let $(\Omega, \mathcal{F}, \mathbf{Pr}(\cdot))$ be a probability space. The mapping $X : \Omega \to \mathbb{R}$ is called a random variable if the pre-image of any interval $(-\infty, a]$ belongs to $\mathcal{F}$, i.e.*

$$X^{-1}((-\infty, a]) \in \mathcal{F} \qquad \text{for all } a \in \mathbb{R}, \tag{1}$$

where

$$X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\}.$$

Important Result: If we have a random variable X, then (1) implies that

$$X^{-1}(B) \in \mathcal{F} \qquad \text{for all } B \in \mathcal{B}.$$

Notation: We use three notations interchangeably $\{\omega \in \Omega \mid X(\omega) \in B\} := \{X \in B\} := X^{-1}(B)$.

- We say that b is an upper bound for a sequence $\{\alpha_k\}$, if $\alpha_k \leq b$ for all k. Smallest such b is called is the supremum of $\{\alpha_k\}$ and is denoted by $\sup_{k \geq 1} \alpha_k$. We always assume that $+\infty$ is an upper bound for a sequence and hence, supremum always exists.
- Similarly, we say that b is a lower bound for a sequence $\{\alpha_k\}$, if $\alpha_k \geq b$ for all k. Largest such b is called is the infimum of $\{\alpha_k\}$ and is denoted by $\inf_{k \geq 1} \alpha_k$.
- We define:

$$\limsup_{k \to \infty} \alpha_k = \inf_{t \geq 1} \sup_{k \geq t} \alpha_k$$
$$\liminf_{k \to \infty} \alpha_k = \sup_{t \geq 1} \inf_{k \geq t} \alpha_k.$$

- Note that for a sequence of r.v.s $\{X_k\}$ and for an $\omega \in \Omega$, $\{X_i(\omega)\}$ is a sequence in $\mathbb{R}$. Then
  I. $X(\omega) = \sup_{k \geq 1} X_k(\omega)$ is a random variable,
  II. $X(\omega) = \inf_{k \geq 1} X_k(\omega)$ is a random variable,
  III. $\overline{X}(\omega) := \limsup_{k \to \infty} X_k(\omega)$ is a random variable,
  IV. $\underline{X}(\omega) := \liminf_{k \to \infty} X_k(\omega)$is a random variable,
  V. if $\overline{X}(\omega) = \underline{X}(\omega)$ for almost all $\omega \in \Omega$, X defined by $X = \lim_{k \to \infty} X_k(\omega)$ is a random

- For a r.v. X, we define the distribution function (or cumulative distribution function (CDF)) of X, to be the mapping $F_X : \mathbb{R} \to [0, 1]$ defined by $F(x) = \Pr(X^{-1}((-\infty, x]))$.
- Properties of Distribution Functions (see HW 3):
  a. F_X is non-decreasing.
  b. $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$.
  c. $F_X(\cdot)$ **is right-continuous**, i.e., for any $x \in \mathbb{R}$, $\lim_{y \to x^+} F_X(y) = F_X(x)$.
  d. Define $F_X(x^-) := \lim_{y \uparrow x} F_X(y)$, then

$$F_X(x^-) = \Pr(X < x) = \Pr(\{\omega \in \Omega \mid X(\omega) < x\}).$$

  e. For any $x \in \mathbb{R}$, we have $\Pr(X = x) = F_X(x) - F_X(x^-)$.

- If we are given a distribution F, the first question one may ask is *Does there exist a probability space $(\Omega, \mathcal{F}, \mathbf{Pr})$ and a function $X : \Omega \to \mathbb{R}$ such that X has the given distribution F?* And the following theorem is the answer to this question.

  Theorem 1. *Suppose that a function $F : \mathbb{R} \to [0, 1]$ satisfies the above properties (a), (b) and (c), then there exists a probability space $(\Omega, \mathcal{F}, \mathbf{Pr})$ and a r.v. X such that F is the distribution function of X.*

- We say that X is a continuous r.v., if $F_X(x)$ is continuous. If further, $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ for a non-negative function $f_X : \mathbb{R} \to \mathbb{R}^+$, we say that X has a probability density function (PDF) $f_X(x)$.
- **Very important**: For a (continuous) r.v. X with the pdf $f_X(x)$

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

- More generally (and important result), for any integrable function $g(\cdot)$, for the random variable $Z = g(X)$, we have

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

- We say that a random variable is a discrete random variable if $\Pr(X \in B) = 1$ for a (finite or) countable set $B = \{b_k \mid k \geq 1\}$.
- We define the *probability mass function* $p : \mathbb{R} \to [0, 1]$ of a discrete random variable X to be defined by:

$$p_X(x) = \begin{cases} \Pr(X = b_k) & \text{if } x = b_k \text{ for some } k \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- **Very important**: For a discrete r.v. X (see HW3)

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} b_k p_X(b_k).$$

- We say X, Y are two random variables, if $X^{-1}(B_1)$ and $Y^{-1}(B_2)$ are independent for any Borel sets $B_1, B_2 \in \mathcal{B}$, i.e.,

$$\Pr(X \in B \text{ and } Y \in B) = \Pr(X \in B)\Pr(Y \in B).$$

- **Important Fact (lemma)**: X, Y are independent if $X^{-1}((-\infty, \alpha])$ and $Y^{-1}((-\infty, \beta])$ are independent for all $\alpha, \beta \in \mathbb{R}$, i.e., it suffices to hold the above for sets of the form $(-\infty, \alpha]$. In other words, X, Y are independent if and only if

$$\Pr(X \leq \alpha, Y \leq \beta) = F_X(\alpha)F_Y(\beta).$$

- Similarly, we say that $X_1, \ldots, X_n$ are independent if for any collection of Borel-sets $B_1, \ldots, B_n$, the events $X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n)$ are independent.
- Again it follows from a result[1] that $X_1, \ldots, X_n$ are independent iff for any selection of real numbers $\alpha_1, \ldots, \alpha_n$:

$$\Pr(X_1 \leq \alpha_1, X_2 \leq \alpha_2, \ldots, X_n \leq \alpha_n) = F_{X_1}(\alpha_1) \cdots F_{X_n}(\alpha_n).$$

- Recall Markov's Inequality: For a non-negative rv X

$$\Pr(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha},$$

for any $\alpha > 0$.

- Define $Var(X) = \mathbb{E}[(X - \bar{X})^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. (if exists)
- **Chebyshev's inequality** (an extension of the Markov's inequality): For a random variable X we have

$$\Pr(|X - \bar{X}| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}.$$

- Now suppose that we have an independent process $\{X_k\}$. What can we say about?

$$P(\max_{1 \leq k \leq n} |S_k| \geq \alpha)$$

Theorem 2. *(Kolmogorov's Maximal Inequality) For a zero mean and independent process $\{X_k\}$ and any $n \geq 1$, we have*

$$P(\max_{1 \leq k \leq n} |S_k| \geq \alpha) \leq \frac{Var(S_n)}{\alpha^2}.$$

- Application: Estimating worst case scenario by day n of COVID

Theorem 4. *For an independent sequence of random variables $\{X_k\}$ with zero mean, if*

$$\sum_{k=1}^{\infty} Var(X_k) < \infty,$$

then $\lim_{n\to\infty} \sum_{k=1}^{n} X_n$ exists.

Main idea of the proof: show that $\sum_{k=1}^{\infty} X_k$ is a Cauchy sequence almost surely by utilizing the Maximal inequality:

$$\Pr(\sup_{M \geq m} |S_M - S_m| \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{k=m}^{\infty} \text{Var}(X_k).$$

Theorem 5. *Sum of an independent random process $\{X_k\}$ converges almost surely if and only if for any $\alpha > 0$, if we let $Y_k = X_k \mathbf{1}_{|X_k| \leq \alpha}$, the following three (deterministic) series converges:*

- $\sum_{k=1}^{\infty} P(|X_k| \geq \alpha) < \infty$,
- $\sum_{k=1}^{\infty} \mathbb{E}[Y_k]$ *converges, and*
- $\sum_{k=1}^{\infty} Var(Y_k)$ *converges.*

- We say that a DT or a CT random process $\{X_t\}$ is
  1. An independent process: if any finite collection $X_{t_1}, \ldots, X_{t_n}$ are independent for any $n \geq 2$ and $t_1 < t_2 < \ldots < t_n$.
  2. An independent increment process: if for any $n \geq 2$, and $a_1 < b_1 \leq a_2 < b_2 \leq \ldots \leq a_n < b_n$, the increments $X_{b_1} - X_{a_1}$, $X_{b_2} - X_{a_2}$, ..., $X_{b_n} - X_{a_n}$ are independent.

Theorem 1. *(Kolmogorov's 0-1 Law) A tail event of an independent process is a trivial event, i.e., $\Pr(E) = 0$ or $\Pr(E) = 1$!*

- Note that the above result holds irrespective of the distributions of any of them!
- Implications:
  a. Probability of giant component on percolation process in $\mathbb{Z}^2$ is either 0 or 1.
  b. Probability of giant component on percolation process in $\mathbb{Z}^2$ that is edge dependent is either 0 or 1 irrespective of probability of each edge.
  c. $\Pr(\lim_{k\to\infty} S_k$ exists) is 0 or 1 for partial sums of independent processes.

- We say that a function $h : \mathbb{R}^n \to \mathbb{R}$ is measurable if $h^{-1}(B) \in \mathcal{B}$ for all Borel sets $B \in \mathcal{B}^n$.
- Important classes of measurable functions:
  - continuous functions,
  - piece-wise continuous functions, and
  - convex (concave) functions.
- Lemma 1: If X and Y are independent random variables, then $g_1(X)$ and $g_2(Y)$ are independent random variables for any measurable functions $g_1(\cdot), g_2(\cdot)$.
- More generally: for independent $X_1, \ldots, X_k, X_{k+1}, \ldots, X_n$, $g_1(X_1, \ldots, X_k)$ and $g_2(X_{k+1}, \ldots,$ would be independent for any measurable functions g_1, g_2 of appropriate dimensions.
- Lemma 2: If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Theorem 3. *For an independent sequence of random variables $\{X_k\}$ with zero mean, we have:*

$$\Pr(\max_{1 \leq k \leq n} |S_k| \geq \alpha) \leq \frac{Var(S_n)}{\alpha^2}.$$

Proof.
- Define $A_k = \{\omega \mid |S_k| \geq \alpha, |S_1|, \ldots, |S_{k-1}| < \alpha\}$.
- $A_1, \ldots, A_n$ are mutually exclusive and we have:

$$\begin{aligned}
\text{Var}(S_n) = \mathbb{E}[|S_n|^2] &\geq \mathbb{E}[(\mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \ldots + \mathbf{1}_{A_n})|S_n|^2] \\
&= \sum_{k=1}^{n} \mathbb{E}[\mathbf{1}_{A_k}|S_n|^2] \\
&= \sum_{k=1}^{n} \mathbb{E}[\mathbf{1}_{A_k}(S_n - S_k + S_k)^2] \\
&= \sum_{k=1}^{n} \mathbb{E}[\mathbf{1}_{A_k}((S_n - S_k)^2 + 2(S_n - S_k)S_k + S_k^2)] \\
&\geq \sum_{k=1}^{n} 2\mathbb{E}[\mathbf{1}_{A_k}S_k(S_n - S_k)] + \mathbb{E}[\mathbf{1}_{A_k}(S_k^2)] \\
&\geq \alpha^2 \sum_{k=1}^{n} \Pr(\mathbf{1}_{A_k}) \\
&= \alpha^2 \Pr(\max_{1 \leq k \leq n} |S_k| \geq \alpha).
\end{aligned}$$