POBABILITY AND STATISTICS, Lesson 7. (not a topic of Final)

- **STOCHASTIC PROCESSES (TIME SERIES)**. Random variables changing in time: $\{X_t \,|\, t \in \mathbb{R}\}$, where t is the time and X_t is r.v. *Trajectory*: a realization of the stochastic process, as the function of t. E.g., economic, financial processes, empirical distribution function. *Phase state*: range of X_t's.

  1. Stochastic process with continuous time: $t \in \mathbb{R}$ or $t \in T$, where T is an interval.
  2. Stochastic process with discrete time: $t = t_1, t_2, \dots$. In case of equidistant observations: $X_0, X_1, X_2, \dots$, where X_0 is the starting value (in general, X_t's are not i.i.d. r.v.'s).

- $\{X_t : t \in \mathbb{R}\}$ has *independent increments*, if $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent for any $t_1 < t_2 < \cdots < t_n$. It has *stationary increments*, if the distribution of the increment $X_{t+h} - X_t$ depends merely on h for any t, h.

- E.g., the standard **Wiener-process** (Brownian motion): $\{W_t : t \geq 0\}$ is such that

  1. $W_0 = 0$ and the trajectories of W_t are continuous with probability 1.
  2. For any $0 \leq t_1 < t_2 < \cdots < t_n$ the joint distribution of $(W_{t_1}, W_{t_2}, \dots, W_{t_n})$ is n-dimensional Gauss.
  3. It has independent, stacionary increments: $W_{t+h} - W_t \sim \mathcal{N}(0, h)$, $\forall t, h > 0$.

- $\{X_t : t \in \mathbb{R}\}$ is a *martingale*, if $\mathbb{E}(|X_t|) < \infty$, and for any $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$

$$\mathbb{E}(X_{t_{n+1}} \,|\, X_{t_1} = a_1, X_{t_2} = a_2, \dots, X_{t_n} = a_n) = a_n$$

  for all $a_1, a_2, \dots a_n \in \mathbb{R}$. E.g., fair plays.

- $\{X_t : t \in \mathbb{R}\}$ is a *Markov-process*, if for any $t_1 < t_2 < \cdots < t_n < t$

$$\mathbb{P}(a < X_t \leq b \,|\, X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n) = \mathbb{P}(a < X_t \leq b \,|\, X_{t_n} = x_n)$$

  for all $a < b$ and $x_1, x_2, \dots x_n \in \mathbb{R}$. If the trajectories are continuous (with probability 1), then it is called *diffusion process*. E.g., the Wiener-process.

  In case of Markov-processes with discrete time (*Markov-chains*) and finite phase state, the process is uniquely defined by the transition probability matrix (see last Chapter of S. Ross' book).

- $\{X_t : t \in \mathbb{R}\}$ is a **Poisson-process**, if X_t measures the number of happenings in $[0, t)$ s.t.

  1. The number of happenings is independent on disjoint time intervals.
  2. X_t has independent, stationary increments.
  3. Sparsity: $\mathbb{P}$(there is exactly 1 happening on an interval of length h) $= \lambda h + o(h)$ and $\mathbb{P}$(more than 1 happenings on an interval of length h) $= o(h)$.

  E.g., earthquakes, accidents in $[0, t)$. Let $Y_1, Y_2, \dots \sim \mathcal{E}xp(\lambda)$ i.i.d. be consecutive exponential lifetimes, and let X_t denote the number of lifetimes expired in $[0, t)$. Then $X_t \in \mathcal{P}(\lambda t)$ is a Poisson-process, and vice versa: if X_t is a Poisson-process with λ happenings on average within unit time (there are λt happenings in t time), and Y denotes the time elapsed between two consecutive happenings, then $Y \sim \mathcal{E}xp(\lambda)$:

$$\mathbb{P}(Y < t) = \mathbb{P}(\text{at least 1 happening on length } t) = 1 - \mathbb{P}(X_t = 0) = 1 - e^{-\lambda t}.$$

- **INFORMATION THEORY, ENTROPY**.

Definition: The entropy of the distribution embodied by the r.v. X is

$$1. \quad H(X) = -\sum_x p(x)\log p(x) \qquad 2. \quad H(X) = -\int f(x)\log f(x)\,dx,$$

where $p(x)$ or $f(x)$ is the p.m.f. or p.d.f. of X, and $\log = \log_2$; further, the convention $0\log 0 = 0$ is used. As $H(X) = \mathbb{E}(-\log p(X))$ or $H(X) = \mathbb{E}(-\log f(X))$, entropy of X is the average amount of surprise one receives upon learning the value of X, or it is the average amount of information received when the value of X is observed. Among discrete distributions with n possible values, the entropy of the uniform distribution is the maximum. Among absolutely continuous distributions with given variance, the entropy of the normal distribution is the maximum.

Let X take on finitely many values with probabilities $p_1, \dots, p_n$. In view of the weak law of large numbers, if we make N independent experiments (in each of which we observe the value of X), then there will be about $Np_1, \dots, Np_n$ occurrences of the individual values, with high probability, provided N is large. We call these outcomes typical. The probability of a typical outcome is about $p_1^{Np_1}\cdots p_n^{Np_n}$. Therefore, almost all outcomes are typical and the expected number of typical outcomes is

$$\frac{1}{p_1^{Np_1}\cdots p_n^{Np_n}} = 2^{-N(p_1\log p_1+\cdots+p_n\log p_n)} = 2^{N\cdot H(X)}.$$

Noisiless coding theorem: Let X take on the values $x_1, \dots, x_n$ with respective probabilities $p_1, \dots, p_n$. Then, for any binary coding of X that assigns n_i bits to x_i: $\sum_{i=1}^n n_i p_i \ge H(X)$. (Here n_i is the length of the binary sequence assigned to x_i in such a way that the 0/1 codes assigned to different x_i's are not extensions of each other.)

- **RANDOM WALKS**.

1-dimensional symmetric random walk: Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. Then $S_0 = 0$, $S_n = X_1 + \cdots + X_n$ is the position at epoch n of a particle starting from 0 and performing a random walk ($n = 1, 2, \dots$). We remark that S_n is a Markov-chain. Let Y_n denote the number of X_i's taking on the value 1 up to n (the number of „right" steps among the first n steps). As $Y_n \sim \mathcal{B}in_n(1/2)$ and $S_n = Y_n - (n - Y_n) = 2Y_n - n$,

$$\mathbb{P}(S_n = 2k - n) = \binom{n}{k}(\frac{1}{2})^n, \quad k = 0, \dots, n.$$

Therefore, $\mathbb{P}(S_n = 0) = 0$ if n is odd and for even n:

$$\mathbb{P}(S_n = 0) = \mathbb{P}(Y_n = \frac{n}{2}) = \frac{n!}{(\frac{n}{2})!(\frac{n}{2})!}(\frac{1}{2})^n \sim \frac{\sqrt{2}}{\sqrt{\pi}\sqrt{n}}$$

by the Stirling formula. Using tis fact, the following theorem can be proved for the recurrence of this random walk. The theorem will be formulated more generally, for multidimensional random walks.

d-dimensional symmetric random walk: A particle is moving on the d-dimensional grid: in each step it moves toward one of the 2^d possible directions with equal probabilities.

Theorem (György Pólya): The probability that a particle performing a random walk on the d-dimensional grid returns infinitely many times to its starting position is 1 if $d = 1$ or $d = 2$, and 0 if $d \ge 3$.