Statistical Theorems

CS6464 - Concepts in Statistical Learning Theory

Prof. Sukhendu Das

Cramér's theorem

- Cramér's theorem is the result that if X and Y are independent real-valued random variables whose sum X + Y is a normal random variable, then both X and Y must be normal as well.
- By induction, if any finite sum of independent real-valued random variables is normal, then the summands must all be normal.
- Thus, while the normal distribution is infinitely divisible, it can only be decomposed into normal distributions (if the summands are independent).

Karhunen–Loève theorem

- Let X_t be a zero-mean square-integrable stochastic process defined over a probability space (*Ω, F, P*), [*Ω* – sample space, *F* – set of events, *P* – assignment of probabilities of events] and indexed over a closed and bounded interval [*a, b*], with continuous covariance function $K_X(s, t)$.
- Then $K_X(s,t)$ is a Mercer kernel and letting e_k be an orthonormal basis on $L^2([a,b])$ formed by the eigenfunctions of T_{KX} with respective eigenvalues λ_k , X_t admits the following representation:

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

- where the convergence is in L^2, uniform in t and:

$$Z_k = \int_a^b X_k e_k(t)\, dt$$

Law of total variance

- In probability theory, the law of total variance states that if X and Y are random variables on the same probability space, and the variance of Y is finite, then:

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

Law of total covariance

- In probability theory, the law of total covariance states that if X, Y, and Z are random variables on the same probability space, and the covariance of X and Y is finite, then

$$cov(X,Y) = E\big(cov(X,Y|Z)\big) + cov(E(X|Z), E(Y|Z))$$

Neyman–Pearson Lemma

- The Neyman–Pearson lemma states that when performing a hypothesis test between two simple hypotheses $H_0\colon \theta = \theta_0$ and $H_1\colon \theta = \theta_1$, the likelihood-ratio test which rejects H_0 in favour of H_1 when:

$$\Lambda(x) = \frac{L(x|\theta_0)}{L(x|\theta_1)} \leq \eta$$

- where L denotes likelihood and,

$$P(\Lambda(x) \leq \eta | H_0) = \alpha$$

- is the most powerful test at significance level α for a threshold η.

Shannon–Hartley theorem

- The Shannon–Hartley theorem states the channel capacity C, meaning the theoretical tightest upper bound on the information rate of data that can be communicated at an arbitrarily low error rate using an average received signal power S through an analog communication channel subject to additive white Gaussian noise of power N:

$$C = B \log_2\left(1 + \frac{S}{N}\right)$$

- where,
  - B is the bandwidth of the channel in hertz (passband bandwidth in case of a bandpass signal);
  - S/N is the signal-to-noise ratio (SNR) or the carrier-to-noise ratio (CNR) of the communication signal to the noise and interference at the receiver (expressed as a linear power ratio, not as logarithmic decibels).