Statistics for Data Science - 2

Week 9 notes

- Let $X_1, \ldots, X_n \sim$ i.i.d.X, where X has the distribution described by parameters $\theta_1, \theta_2, \ldots$.
  - The parameters θ_i are unknown but a fixed constant.
  - Define the estimator for θ as the function of the samples: $\hat{\theta}(X_1, \ldots, X_n)$.

  Note:

  1. θ is an unknown parameter.
  2. $\hat{\theta}$ is a function of n random variables.

  Remark: Infinite number of estimators are possible for a parameter of a distribution.

- Estimation error: $\hat{\theta}(X_1, \ldots, X_n) - \theta$ is a random variable.
  - We expect the estimator random variable $\hat{\theta}(X_1, \ldots, X_n)$ to take values around the actual value of the parameter θ. So, the random variable 'Error' should take values close to 0.
  - Mathematically, it is expressed as $P(\mid \text{Error} \mid > \delta)$ should be small.
  - Chebyshev bound on error: $P(\mid \text{Error} - E[\text{Error}] \mid > \delta) \leq \dfrac{\text{Var}(\text{Error})}{\delta^2}$.
  - Good design: $P(\mid \text{Error} \mid > \delta)$ will fall with n.

- Good design principles:
  1. Error should be close to or equal to 0.
  2. $\text{Var}(\text{Error}) \to 0$ with n.

- Bias: The bias of the estimator $\hat{\theta}$

  for a parameter θ, denoted $\text{Bias}(\hat{\theta}, \theta)$ is defined as

$$\text{Bias}(\hat{\theta}, \theta) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$$

  1. Bias is the expected value of Error.
  2. An estimator with bias equal to 0 is said to be an unbiased estimator.

- Risk: The (squared-error) risk of the estimator $\hat{\theta}$ for a parameter θ, denoted $\text{Risk}(\hat{\theta}, \theta)$, is defined as

$$\text{Risk}(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2]$$

1. Risk is the expected value of "squared error" and is also called mean squared error (MSE) often.
2. Squared-error risk is the second moment of Error.

- Variance of estimator:

$$\text{Variance}(\hat{\theta}) = E[(\hat{\theta} - E[\theta])^2]$$

  $\text{Var}(\text{Error}) = \text{Var}(\hat{\theta})$

- Bias-Variance tradeoff: The risk of the estimator satisfies the following relationship:

$$\text{Risk}(\hat{\theta}, \theta) = \text{Bias}(\hat{\theta}, \theta)^2 + \text{Variance}(\hat{\theta})$$

- Estimator design approach:
  1. Method of moments
     (a) Sample moments: $M_k(X_1, \ldots, X_n) = \dfrac{1}{n} \sum_{i=1}^{n} X_i^k$
     (b) M_k is a random variable, and m_k is the value taken by it in one sampling instance. We expect that M_k will take values around $E[X^k]$
     (c) Procedure:
        - Equate sample moments to expression for moments in terms of unknown parameters.
        - Solve for the unknown parameters.
     (d) One parameter θ usually needs one moment
        - Sample moment: m_1
        - Distribution moment: $E[X] = f(\theta)$
        - Solve for θ from $f(\theta) = m_1$ in terms of m_1.
        - $\hat{\theta}$: replace m_1 by M_1 in above solution.
     (e) Two parameters θ_1, θ_2 usually needs two moments.
        - Sample moments: m_1, m_2
        - Distribution moment: $E[X] = f(\theta_1, \theta_2), E[X^2] = g(\theta_1, \theta_2)$
        - Solve for θ_1, θ_2 from $f(\theta_1, \theta_2) = m_1, g(\theta_1, \theta_2) = m_2$ in terms of m_1, m_2.
        - $\hat{\theta}$: replace m_1 by M_1 and m_2 by M_2in above solution.
  2. Maximum Likelihood estimators
     (a) Likelihood of i.i.d. samples: Likelihood of a sampling $x_1, x_2, \ldots, x_n$, denoted $L(x_1, x_2, \ldots, x_n)$

$$L(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i; \theta_1, \theta_2, \ldots)$$

        - Likelihood $L(x_1, x_2, \ldots, x_n)$ is a function of parameters.

– Maximum likelihood (ML) estimation

$$\theta_1^*, \theta_2^*, \ldots = arg \max_{\theta_1, \theta_2, \ldots} \prod_{i=1}^{n} f_X(x_i; \theta_1, \theta_2, \ldots)$$

We find parameters that maximize likelihood for a given set of samples.