POBABILITY AND STATISTICS, Lessons 9-10.

- **Theory of point estimation**. *Likelihood function*: for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}$ and $\theta \in \Theta$ let $L_\theta(\mathbf{x}) = \mathbb{P}_\theta(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n p_\theta(x_i)$ in the discrete, and $L_\theta(\mathbf{x}) = \prod_{i=1}^n f_\theta(x_i)$ in the absolutely continuous case.

  Theorem (Neyman–Fisher Factorization) The statistic $T(\mathbf{X})$ is *sufficient* for θ if and only if $L_\theta(\mathbf{x}) = g_\theta(T(\mathbf{x})) \cdot h(\mathbf{x})$, $\forall \theta \in \Theta$, $\quad \mathbf{x} \in \mathcal{X}$ with some measurable, nonnegative real functions g and h. (A sufficient statistic contains all the important information for θ, and it is minimal if it is the function of any other sufficient statistic.)

  Let $(\Omega, \mathcal{A}, \mathcal{P})$ be parametric statistical space, $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. We want to estimate θ, or its measurable function $\psi(\theta)$ by means of the statistic $T(\mathbf{X})$ on the basis of the i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$. The point estimator is sometimes denoted by $\hat{\theta}$ or $\hat{\psi}$. Criteria for the „goodness" of an estimate:

  - $T(\mathbf{X})$ is an **unbiased** estimator of $\psi(\theta)$, if $\mathbb{E}_\theta(T(\mathbf{X})) = \psi(\theta), \quad \forall \theta \in \Theta$.
  - $T(\mathbf{X}_n)$ is an **asymptotically unbiased** estimator of $\psi(\theta)$, if $\lim_{n\to\infty} \mathbb{E}_\theta(T(\mathbf{X}_n)) = \psi(\theta), \quad \forall \theta \in \Theta$.
  - Let T_1 and T_2 be both unbiased estimators of $\psi(\theta)$. T_1 is **at least as efficient** than T_2, if $\mathbb{D}^2_\theta(T_1) \le \mathbb{D}^2_\theta(T_2)$, $\forall \theta \in \Theta$. An unbiased estimator is **efficient**, if it is at least as efficient than any other unbiased estimator. Efficient estimator does not always exist, but if yes, then it is unique (with probability 1).
  - $T(\mathbf{X}_n)$ is a weakly/strongly **consistent** estimator of $\psi(\theta)$, if $\forall \theta \in \Theta$:
    $T(\mathbf{X}_n) \to \psi(\theta)$ in probability/almost surely as $n \to \infty$.

  We want to give a lower bound for the variance of an unbiased estimator if $dim(\Theta) = 1$. The **Fisher informatin** contained in the i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$ is
  $I_n(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \ln L_\theta(\mathbf{X})\right)^2 \ge 0$. ($I_n(\theta) = nI_1(\theta)$ under the regularity conditions below.)

  Theorem (Cramér–Rao inequality) In the above setup let $T(\mathbf{X})$ be unbiased estimator of the differentiable parameter function $\psi(\theta)$, and suppose that $\mathbb{D}^2_\theta(T) < \infty$ ($\forall \theta \in \Theta$). Further, the following regularity conditions hold $\forall \theta \in \Theta$:

  $$\frac{\partial}{\partial\theta} \int L_\theta(\mathbf{x})\, d\mathbf{x} = \int \frac{\partial}{\partial\theta} L_\theta(\mathbf{x})\, d\mathbf{x} \quad \text{and} \quad \frac{\partial}{\partial\theta} \int T(\mathbf{x}) L_\theta(\mathbf{x})\, d\mathbf{x} = \int T(\mathbf{x}) \frac{\partial}{\partial\theta} L_\theta(\mathbf{x})\, d\mathbf{x}.$$

  Then $\mathbb{D}^2_\theta(T) \ge \frac{(\psi'(\theta))^2}{I_n(\theta)} = \frac{(\psi'(\theta))^2}{nI_1(\theta)}, \qquad \forall \theta \in \Theta.$

  Rao–Blackwell–Kolmogorov Theorem: In the above setup let $T(\mathbf{X})$ be a sufficient statistic, and $S(\mathbf{X})$ be an unbiased estimetor for $\psi(\theta)$. Then one can construct an unbiased estimator $U = g(T)$ that is at least as efficient az S. The construction of U („blackwellization"): $U := \mathbb{E}_\theta(S|T) = g(T(\mathbf{X}))$, $\forall \theta \in \Theta$. (The message of the theorem: find the efficient estimator among the functions of the minimal sufficient statistic.)

  Methods of point estimation

  - *Maximum likelihood (ML) principle*: maximize the likelihood or log-likelihood function in θ! (The ML-estimator is asymptotically unbiased, efficient, and strongly consistent.)
  - *Method of moments*: $dim(\Theta) := k$ and find the first k moments in the function of $\theta_1, \ldots, \theta_k$. The moment estimator $\hat{\theta}_j$ is the inverse function of the empirical moments.

- **Interval estimation**. The random interval $(T_1(\mathbf{X}), T_2(\mathbf{X}))$ is a *confidence interval* of level (at least) $1 - \varepsilon$ for $\psi(\theta)$, if $\mathbb{P}_\theta(T_1 < \psi(\theta) < T_2)(\ge) = 1 - \varepsilon$ ($\forall \theta \in \Theta$).