Quiz 1 - Solutions

Maximum Likelihood and Sufficient Statistics

1. Let $\hat{\theta}_{\text{MLE}}$ be the maximum likelihood estimator for a parameter θ. If g is a one-to-one function, what is the maximum likelihood estimator for $\eta = g(\theta)$?

   Solution: $g(\hat{\theta}_{\text{MLE}})$.

   Derivation: Recall the definition of the MLE for the original parameter θ:

   $$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\theta; \mathcal{D})$$

   where $\mathcal{L}(\theta; \mathcal{D}) = p(\mathcal{D}|\theta)$ is the likelihood function, given the observations $\mathcal{D}$.

   Now consider the re-parameterization $\eta = g(\theta)$. Since g is one-to-one, we can invert it to write $\theta = g^{-1}(\eta)$. The likelihood function for the new parameter η, denoted $\mathcal{L}^*(\eta; \mathcal{D})$, is simply the original likelihood evaluated at the corresponding θ:

   $$\mathcal{L}^*(\eta; \mathcal{D}) = \mathcal{L}(g^{-1}(\eta); \mathcal{D})$$

   The MLE for η is defined as the value that maximizes this new likelihood:

   $$\hat{\eta}_{\text{MLE}} = \underset{\eta}{\operatorname{argmax}}\, \mathcal{L}^*(\eta; \mathcal{D}) = \underset{\eta}{\operatorname{argmax}}\, \mathcal{L}(g^{-1}(\eta); \mathcal{D})$$

   Since $\mathcal{L}(\theta; \mathcal{D})$ achieves its maximum at $\theta = \hat{\theta}_{\text{MLE}}$, the composite function $\mathcal{L}(g^{-1}(\eta); \mathcal{D})$ achieves its maximum when the input to $\mathcal{L}$, which is $g^{-1}(\eta)$, equals $\hat{\theta}_{\text{MLE}}$:

   $$g^{-1}(\hat{\eta}_{\text{MLE}}) = \hat{\theta}_{\text{MLE}} \implies \hat{\eta}_{\text{MLE}} = g(\hat{\theta}_{\text{MLE}})$$

2. Let $X_1, \ldots, X_n$ be i.i.d. uniform random variables on the interval $[0, \theta]$. Which of the following is the sufficient statistic for θ?

   Solution: The maximum of the sample $X_{(n)} = \max(X_1, \ldots, X_n)$.

   The likelihood function for a Uniform distribution on $[0, \theta]$ is:

   $$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{I}(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \mathbb{I}(0 \leq \max(x_i) \leq \theta)$$

   By the Fisher-Neyman Factorization Theorem, the likelihood depends on the data x only through the maximum value $X_{(n)}$. Therefore, $X_{(n)}$ is the sufficient statistic.

3. If $T(X)$ is a sufficient statistic for θ, which of the following statements regarding the Maximum Likelihood Estimator (MLE) $\hat{\theta}$ is always true?

   Solution: The MLE $\hat{\theta}$ is always a function of sufficient statistic $T(X)$.

   By the Factorization Theorem, the likelihood can be written as $\mathcal{L}(\theta; X) = h(X) g_\theta(T(X))$. When maximizing $\mathcal{L}(\theta; X)$ with respect to θ, the term $h(X)$ is constant and does not influence the location of the maximum. The maximization depends solely on $g_\theta(T(X))$. Therefore, the resulting estimator $\hat{\theta}$ will depend on the data only through $T(X)$.

4. Which of the following probability distributions does not belong to the one-parameter Exponential Family?

   Solution: Uniform distribution on the interval $[0, \theta]$ for $\theta > 0$.

   Either we recognize the other families are exponential families, or we use the following fact. For a distribution to belong to the exponential family, its support (the set of x with density satisfying $p(x) > 0$) must not depend on the parameter θ. Indeed, the parameter only intervenes in an exponential factor that cannot be equal to 0. $p(x)$ being equal to 0 is then equivalent to the other factor being equal to 0, and it does not depend on θ.

   - Poisson, Bernoulli, and Normal (with known variance) have supports that are independent of their parameters.
   - The Uniform distribution on $[0, \theta]$ has support $0 \leq x \leq \theta$, which depends directly on θ. Thus, it is not in the exponential family.

5. Consider a model for three binary random variables (x_1, x_2, x_3) where $x_i \in \{0, 1\}$. The joint probability mass function is given by:

   $$p(x_1, x_2, x_3) = \frac{1}{Z(\theta)} \exp(\theta(x_1 x_2 + x_2 x_3))$$

   Which of the following represents the sufficient statistic for θ?

   Solution: $T(x_1, x_2, x_3) = x_1 x_2 + x_2 x_3$.

   This distribution is already written in the canonical exponential family form:

   $$p(x|\theta) = h(x) \exp(\eta(\theta) \cdot T(x) - A(\theta))$$

   Here, the natural parameter is θ, and the term multiplying it in the exponent is the sufficient statistic $T(x) = x_1 x_2 + x_2 x_3$. We also have $h(x) = 1$ and $A(\theta) = \log Z(\theta)$.