# Inference 2, 2023, lecture 5

Rolf Larsson

November 15, 2023

# Today

Chap. 3.3. Inference Principles (continued):

- Sufficiency and the Exponential family

# Sufficiency

Recall:

**Definition (3.7)**

A statistic $T$ is said to be **sufficient** for the statistical model $\{\mathrm{P}_\theta : \theta \in \Theta\}$ of $\mathbf{X}$ if the conditional distribution of $\mathbf{X}$ given $T = t$ is independent of $\theta$ for all $t$.

**Theorem (3.7)**

**The Factorization criterion**:
*Let $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ be a statistical model with probability function $p(\cdot; \theta)$.*
*A statistic $T$ is sufficient for $\mathcal{P}$ if and only if there exist nonnegative functions $g(\cdot; \theta)$ and $h$ such that*

$$p(\mathbf{x}; \theta) = g\{T(\mathbf{x}); \theta\} h(\mathbf{x}).$$

# Minimal sufficiency

Recall:

**Definition (3.8)**

A sufficient statistic $T$ is **minimal sufficient** if.f. it is a function of any other sufficient statistic.

Theorem 3.8:

If $\frac{L(\theta;\mathbf{x})}{L(\theta;\mathbf{y})}$ is no function of $\theta$ if.f. $T(\mathbf{x}) = T(\mathbf{y})$, then $T$ is minimal sufficient.

# The Exponential Family

Recall: $X$ belongs to the exponential family if

$$p(x;\theta) = A(\theta)\exp\left\{\sum_{j=1}^{k}\zeta_j(\theta)\,T_j(x)\right\}h(x).$$

For a sample $X_1, \ldots, X_n$ of independent random variables distributed as $X$, with observations $\mathbf{x} = (x_1, \ldots, x_n)$,

$$p(\mathbf{x};\theta) = \prod_{i=1}^{n} A(\theta)\exp\left\{\sum_{j=1}^{k}\zeta_j(\theta)\,T_j(x_i)\right\}h(x_i)$$

$$= A(\theta)^n \exp\left\{\sum_{j=1}^{k}\zeta_j(\theta)\sum_{i=1}^{n} T_j(x_i)\right\}\prod_{i=1}^{n} h(x_i).$$

By the factorization theorem, a sufficient statistic is given by $T_{(n)}(\mathbf{x}) = \{\sum_{i=1}^{n} T_1(x_i), \ldots, \sum_{i=1}^{n} T_k(x_i)\}$.

# The Exponential Family

## Theorem (3.9)

*For a sample of i.i.d. random variables from a strictly k-dimensional exponential family it holds that*

1. *The statistic* $T_{(n)}(\mathbf{x}) = \{\sum_{i=1}^{n} T_1(x_i), ..., \sum_{i=1}^{n} T_k(x_i)\}$ *is* **minimal** *sufficient.*
2. *The distribution of* $T_{(n)}(\mathbf{x})$ *belongs to a k-parameter exponential family.*

# The Exponential Family

Relax the independence condition:

## Corollary (3.4)

*For a sample* $\mathbf{X} = (X_1, ..., X_n)$ *from a strictly k-dimensional exponential family with*

$$p(\mathbf{x};\theta) = A(\theta)\exp\left\{\sum_{j=1}^{k}\zeta_j(\theta)\,T_j(\mathbf{x})\right\}h(\mathbf{x}),$$

*it holds that*

1. *The statistic* $T(\mathbf{x}) = \{T_1(\mathbf{x}), ..., T_k(\mathbf{x})\}$ *is minimal sufficient.*
2. *The distribution of* $T$ *belongs to a k-parameter exponential family.*

# The Exponential Family

Suppose we have an *i.i.d.* sample from any of the following distributions. Using theorem 3.9, derive minimal sufficient statistics for the parameters.

1. Exponential with intensity $\beta$
2. $N(\mu, \sigma^2)$

# News of today

For the exponential family:

- "Read off" minimal sufficiency directly from the likelihood.
- The minimal sufficient statistic belongs to an exponential family.
- Solve problems!