STA 6351, Report.2.7

Carson Slater *Baylor University*

2.7

Dynkin's theorem offers a likelihood-based characterization of minimal sufficiency. Unlike the factorization theorem, it does not rely on explicit algebraic decomposition of the joint PDF; rather, it identifies the minimal sufficient statistic directly from the shape of the likelihood ratio.

(a) *What does Dynkin's theorem add conceptually beyond the factorization theorem?* Why might it be preferable in theoretical development, even though both lead to the same sufficient statistics?

Dynkin's theorem gives a likelihood-based characterization of minimal sufficiency by defining an equivalence relation on sample points: two observations $\mathbf{x}$ and $\mathbf{y}$ are equivalent exactly when the likelihood ratio $L(\boldsymbol{\theta} \mid \mathbf{x})/L(\boldsymbol{\theta} \mid \mathbf{y})$ is constant in $\boldsymbol{\theta}$. This equivalence classes map is (up to a bijection) the minimal sufficient statistic. The result replaces the need to find an explicit algebraic factorization of the joint density with a direct criterion based on indistinguishability of samples for all parameter values. It is often preferable in abstract theory because it is model-agnostic and measure-theoretic, directly yields minimality and uniqueness, and connects sufficiency to the likelihood principle, simplifying arguments when explicit factorization is awkward or unavailable.

(b) *How does Dynkin's theorem embody the likelihood principle?* Explain why $L(\boldsymbol{\theta} \mid \mathbf{x})/L(\boldsymbol{\theta}_0 \mid \mathbf{x})$ captures all information about θ needed for inference, and why proportional likelihoods imply equivalent inference.

Dynkin's formulation implements the likelihood principle by treating the likelihood function up to a multiplicative constant as the complete datum for inference about $\boldsymbol{\theta}$. Fix a reference parameter value $\boldsymbol{\theta}_0$. For any observation $\mathbf{x}$ we may write

$$L(\boldsymbol{\theta} \mid \mathbf{x}) = L(\boldsymbol{\theta}_0 \mid \mathbf{x}) \, \frac{L(\boldsymbol{\theta} \mid \mathbf{x})}{L(\boldsymbol{\theta}_0 \mid \mathbf{x})},$$

so all $\boldsymbol{\theta}$-dependent variation in the likelihood is contained in the ratio $\Lambda_{\mathbf{x}}(\boldsymbol{\theta}) := L(\boldsymbol{\theta} \mid \mathbf{x})/L(\boldsymbol{\theta}_0 \mid \mathbf{x})$. Thus $\Lambda_{\mathbf{x}}(\boldsymbol{\theta})$ is a complete summary of the evidence in $\mathbf{x}$ about $\boldsymbol{\theta}$ up to an $\boldsymbol{\theta}$-independent scaling.

If two samples $\mathbf{x}$ and $\mathbf{y}$ satisfy

$$L(\boldsymbol{\theta} \mid \mathbf{x}) = c(\mathbf{x}, \mathbf{y}) \, L(\boldsymbol{\theta} \mid \mathbf{y}) \qquad \text{for all } \boldsymbol{\theta},$$

then $\Lambda_{\mathbf{x}}(\boldsymbol{\theta}) = \Lambda_{\mathbf{y}}(\boldsymbol{\theta})$ and every likelihood-based quantity is identical for $\mathbf{x}$ and $\mathbf{y}$. Concretely:

- Bayesian posterior: $\pi(\boldsymbol{\theta} \mid \mathbf{x}) \propto L(\boldsymbol{\theta} \mid \mathbf{x})\pi(\boldsymbol{\theta})$. Proportional likelihoods give $\pi(\boldsymbol{\theta} \mid \mathbf{x}) \propto c(\mathbf{x}, \mathbf{y}) \, L(\boldsymbol{\theta} \mid \mathbf{y})\pi(\boldsymbol{\theta})$, which normalizes to the same posterior as for $\mathbf{y}$.
- Point estimation by maximization: $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}(\mathbf{x}) = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \mathbf{x})$ is invariant to multiplicative constants, so proportional likelihoods yield the same MLE.

- Likelihood-based intervals and tests (likelihood-ratio statistics, evidence ordering) depend only on ratios of likelihood values, hence are identical for $\mathbf{x}$ and $\mathbf{y}$.

Therefore Dynkin's equivalence relation (samples with proportional likelihoods are equivalent) formalizes the likelihood principle: data that induce the same likelihood function up to a constant carry the same information about $\boldsymbol{\theta}$, so any valid inference based solely on the likelihood must coincide.

(c) In exponential-family models, canonical statistics are often minimal sufficient. Why does Dynkin's result make this intuitive? (Hint: How does the likelihood ratio depend on the canonical statistic?)

Write the exponential-family form

$$p(\mathbf{x} \mid \boldsymbol{\theta}) = h(\mathbf{x}) \exp\Big\{\eta(\boldsymbol{\theta})^{\top} T(\mathbf{x}) - A(\boldsymbol{\theta})\Big\}.$$

Hence for two samples $(\mathbf{x}, \mathbf{y})$

$$\frac{L(\boldsymbol{\theta} \mid \mathbf{x})}{L(\boldsymbol{\theta} \mid \mathbf{y})} = \frac{h(\mathbf{x})}{h(\mathbf{y})} \exp\Big\{\eta(\boldsymbol{\theta})^{\top} \big(T(\mathbf{x}) - T(\mathbf{y})\big)\Big\}.$$

All $\boldsymbol{\theta}$-dependence enters through $\eta(\boldsymbol{\theta})^{\top}\big(T(\mathbf{x}) - T(\mathbf{y})\big)$; therefore the likelihood ratio is constant in $\boldsymbol{\theta}$ exactly when $T(\mathbf{x}) = T(\mathbf{y})$ (assuming $\eta(\boldsymbol{\theta})$ is not degenerate). By Dynkin's criterion the equivalence classes determined by $T(\cdot)$ are the minimal sufficient statistic, so the canonical statistic $T(\mathbf{x})$ is minimal sufficient (up to a one-to-one transform).

(d) Why is minimal sufficiency important in practice? Discuss how using the minimal sufficient statistic can simplify Bayesian updating or likelihood-based inference without loss of information.

Minimal sufficiency is important because it provides the smallest possible data reduction that preserves all information about the parameter. If $T(X)$ is (minimal) sufficient then, by factorization,

$$L(\boldsymbol{\theta} \mid \mathbf{x}) = h(\mathbf{x})\, g\big(T(\mathbf{x}), \boldsymbol{\theta}\big),$$

so Bayesian updating and all likelihood-based inferences depend on $\mathbf{x}$ only through $T(\mathbf{x})$:

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) \propto g\big(T(\mathbf{x}), \boldsymbol{\theta}\big)\pi(\boldsymbol{\theta}).$$

Consequences: lower-dimensional summaries for storage and computation, simpler analytic or conjugate updates, identical MLEs and likelihood-ratio tests using T alone, and straightforward elimination or conditioning on nuisance parameters—achieving maximal data reduction without information loss.