Lecture 18: Lagrange Multipliers and Neyman-Pearson

- I proved that if f and g are functions from a set $\mathcal{V}$ to $\mathbb{R}$ and there is a $\lambda > 0$ and a $v^* \in \mathcal{V}$ such that

$$v^* \text{ minimizes } H_\lambda(v) \equiv f(v) + \lambda g(v),$$

  and

$$g(v^*) = c$$

  then for all $v \in \mathcal{V}$ such that $g(v) \leq c$ we have

$$f(v) \geq f(v^*)$$

- I also proved that if f and g are functions from a set $\mathcal{V}$ to $\mathbb{R}$ and there is a $\lambda \in \mathbb{R}$ and a $v^* \in \mathcal{V}$ such that

$$v^* \text{ minimizes } H_\lambda(v) \equiv f(v) + \lambda g(v),$$

  and

$$g(v^*) = c$$

  then for all $v \in \mathcal{V}$ such that $g(v) = c$ we have

$$f(v) \geq f(v^*)$$

Neyman Pearson Lemma

- I applied this to prove the Neyman Pearson Lemma.
- A hypothesis (Θ_0 or Θ_1) is *simple* if it contains only one density.
- Otherwise it is composite.
- For data X with model $\{f_0, f_1\}$ with only 2 densities in it the probability of a Type II error is minimized, subject to a test having level no more than α by

$$\phi_{\lambda,\gamma}(x) = \begin{cases} 1 & f_1(x)/f_0(x) > \lambda \\ 0 & f_1(x)/f_0(x) < \lambda \\ \gamma & f_1(x)/f_0(x) = \lambda \end{cases}$$

provided that γ and λ are chosen so that the resulting test has level α.

The Proof

- V is the set of functions ϕ from $\mathcal{X}$ to [0,1].
- f is $\beta = \mathrm{E}_0(1-\phi(X)) = \int [1-\phi(x)] f_0(x) dx$.
- g is $\alpha = \mathrm{E}_0(\phi(X)) = \int \phi(x) f_0(x) dx$.
- So $H_\lambda = \beta + \lambda\alpha$.
- Parallel to the last lecture this is minimized by *any* ϕ of the form

$$\phi(x) = \begin{cases} 1 & f_x(x)/f_0(x) > \lambda \\ 0 & f_x(x)/f_0(x) < 1 \end{cases}$$

Finish the proof

- If $Y = f_1(X)/f_0(X)$ has a continuous distribution when X has density f_0 then the equation

$$P_0(Y > \lambda) = \alpha = P(Y \geq \alpha)$$

  has a solution.
- For discrete distributions γ might be needed.
- This finishes the proof.
- Then I did an example of Lagrange Multipliers.

Lagrange multipliers example

- Maximize $f(x) = x^T Qx$ subject to $x^T x = 1$.
- Lagrangian is $H_\lambda(x) = x^T(Q - \lambda I)x$.
- Take derivative wrt x and solve

$$(Q - \lambda I)x = 0$$

- If λ is not an eigenvalue of Q then the only critical point is $x = 0$ which does not satisfy the constraint.
- So assume λ is an eigenvalue and x is not 0.
- So x must be an eigenvector of Q and $H_\lambda(x) = 0$.

Example continued

- To satisfy the constraint we must simply divide x by its length so x will be a unit length eigenvector.
- The second derivative matrix of H_λ is $Q - \lambda I$.
- Let $\lambda_1 > \cdots > \lambda_p$ be the distinct eigenvalues of Q.
- The eigenvalues of H_{λ_1} are

$$0 > \lambda_2 - \lambda_1 > \cdots > \lambda_p - \lambda_1$$

- So H_{λ_1} has a non-positive definite, constant, Hessian.
- So any unit length eigenvector x for λ_1 maximizes H_{λ_1}.
- Any such x maximizes $x^T Qx$ subject to $x^T x = 1$.
- The maximized value is λ_1.

Coverage in the text

- Chapter 10.
- Course slides "Hypothesis Tests": 1-9, 13, 14, 18-24, 26, 27
- See "course notes" on web pages 130-131.