STATISTICAL THEORY – EXAMPLES 2/2

Part III, Michaelmas 2013, RN

Ex. 1 Consider $X_1, \dots, X_n$ i.i.d. copies of a random variable X, and suppose you model X as $X|\theta \sim N(\theta, 1)$ in a Bayesian way, where the prior distribution is $\theta \sim N(0,1)$. Show that the posterior distribution is

$$\theta|X_1, \dots, X_n \sim N\left(\frac{\sum_{i=1}^n X_i}{n+1}, \frac{1}{n+1}\right).$$

Ex. 2 In the setting of the Bernstein-von Mises theorem from lectures, let C_n be an Euclidean ball in $\mathbb{R}^p$ centred at the MLE $\hat{\theta}_n$ such that $\Pi(C_n|X_1, \dots, X_n) = 1 - \alpha$ for all n (C_n is a *credible set* for the posterior distribution). Show that $P_{\theta_0}(\theta_0 \in C_n) \to 1 - \alpha$ as $n \to \infty$ (that is, C_n is a frequentist confidence set).

Ex. 3 Derive the formula $\hat{\theta} = (X^T X)^{-1} X^T Y$ for the least squares estimator in the standard Gaussian linear model

$$Y = X\theta + \varepsilon,$$

when $p \leq n$, X is a $n \times p$ matrix of full column rank p, and $\varepsilon \sim N(0, \sigma^2 I_n), \sigma > 0$. Show that $X\hat{\theta} = PY$ where P is the projection matrix that projects onto the span of the column vectors of X and deduce $E\|X\hat{\theta} - X\theta\|^2 = \sigma^2 p$. Now let X be partitioned as (X^M, X^{M^c}) where X^M is a $n \times k$ matrix, $k < p$, and consider the least squares predictor $P_M Y = X\hat{\theta}^M$ from sub-model M, where P_M projects onto the linear span of the column vectors of X_M. For

$$\hat{\sigma}^2 = \frac{1}{n-p}\|Y - PY\|^2$$

show that $E\hat{\sigma}^2 = \sigma^2$ and that Mallow's C_p criterion

$$crit_{C_p}(M) = \|Y - P_M Y\|^2 + 2\hat{\sigma}^2 k - n\hat{\sigma}^2,$$

is an unbiased estimator of the prediction risk

$$E\|X\hat{\theta}^M - X\theta\|^2$$

of the least squares predictor from the restricted model M.

Ex. 4 Prove that every solution $\tilde{\theta}_{LASSO}$ of the LASSO criterion function generates the same fitted value $X\tilde{\theta}_{LASSO}$ and the same ℓ_1-norm $\|\tilde{\theta}_{LASSO}\|_1$.

Ex. 5 For $Z \sim N(0,1)$ prove that for all $x > 0$ we have $\Pr(|Z| > x) \leq e^{-x^2/2}$.

Ex. 6 For a $p \times p$ symmetric matrix Φ, show that the maximal absolute eigenvalue $\phi_{\max} = \max_i |\phi_i|$ is equal to $\sup_{\|v\|_2 \leq 1} |v^T \Phi v|$. Show further that the minimal absolute eigenvalue corresponds to $\inf_{\|v\|_2 \leq 1} |v^T \Phi v|$.

Ex. 7 Let B be the unit ball in a k-dimensional Euclidean space. Let $N(\delta), \delta > 0$, be the minimal number of closed balls of radius δ with centers in B that are required to cover B. Show that for some constant $A > 0$ and every $0 < \delta < A$ we have

$$N(\delta) \leq (A/\delta)^k.$$

Ex. 8 In the high-dimensional linear model generated from θ^0 under random design X, the 'signal to noise ratio' is defined as $SNR = \|X\theta^0\|_2/\sqrt{n}\sigma$. If $\hat{\sigma}^2 = Y^TY/n$ (and assuming $EY = 0$ for simplicity), show that for all $t > 0$ and with probability at least $1 - \exp\{-t^2/2\}$ we have

$$\frac{\hat{\sigma}^2}{\sigma^2} \in \left[1 + SNR \cdot (SNR \pm 2t/\sqrt{n}) \pm b_n\right], \quad b_n \equiv \left|\frac{\varepsilon^T\varepsilon}{n\sigma^2} - 1\right|.$$

Ex. 9 * In the setting of Corollary 2 from lecture notes, prove that with probability at least $1 - e^{-t^2/2}$ one has

$$\|\tilde{\theta} - \theta^0\|_2^2 \lesssim \frac{k}{n}\log p,$$

assuming in addition that for all θ considered in that corollary,

$$\|\theta_{\mathcal{N}} - \theta^0\|_2^2 \leq r_1(\theta - \theta^0)^T\hat{\Sigma}(\theta - \theta^0)$$

for some $r_1 > 0$, where $\theta_{\mathcal{N}}$ is the vector consisting of zeros except for those θ_j's for which $j \in S_0$ joined by those θ_j's with indices corresponding to the k largest $|\theta_j|'s$ for $j \notin S_0$. [Hint: Show that for all $\theta \in \mathbb{R}^p$, $\|\theta_{\mathcal{N}^c}\|_2 \leq k^{-1/2}\|\theta_{S_0^c}\|_1$.]