Theoretical Sequence Notes

Kenneth Tay

1 Overall Tricks

Tricks with multivariate normal:

1. Rotate to get identity covariance. If $X \sim \mathcal{N}(\mu, \Sigma)$, then $\Sigma^{-1/2}X \sim \mathcal{N}(\Sigma^{-1/2}\mu, I)$.
2. Change basis (using spherical symmetry). If U is orthogonal (i.e. $U^T U = I$), then if $X \sim \mathcal{N}(\mu, \Sigma)$, we have $U^T X \sim \mathcal{N}(U^T\mu, I)$.

   U^T is an invertible map, so hypotheses about μ correspond to hypotheses about $U^T\mu$.
3. Use the eigendecomposition of Σ: $\Sigma = VDV^T$, where D is a diagonal matrix with eigenvalues on the diagonal, and V is orthogonal with eigenvectors as columns.

   $\Sigma^{1/2} = VD^{1/2}V^T$, and if Σ is positive definite, $\Sigma^{-1/2} = VD^{-1/2}V^T$.

To find the MLE:

1. Differentiate the likelihood and set to 0. It is often easier to look at the log-likelihood.

To come up with an estimator:

1. Try the MLE.

2 Properties of Estimators

To compute the asymptotic distribution of an estimator (whether one- or multi-dimensional):

1. Make use of the multivariate CLT (VdV p16), especially if there is some averaging going on.
2. If it is a function of something whose asymptotic distribution you already know, try the delta method. (See VdV p26 for the Taylor expansion argument.)
3. If it is an MLE which is consistent, use the asymptotic normality result. (For exponential families, we may dispense with the consistency condition and use VdV Thm 4.6 directly.)

To show an estimator is consistent:

1. Bare hands approach: Law of large numbers.

2. From Chebyshev's inequality: $\mathbb{P}(|\hat{\theta}_n - \theta| \geq \varepsilon) \leq \dfrac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2}$. Thus, an estimator will be consistent if $\mathbb{E}[(\hat{\theta}_n - \theta)^2] \to 0$.

3. An estimator will be consistent if it is asymptotically unbiased (i.e. $\mathbb{E}_\theta[\hat{\theta}_n] - \theta \to 0$) and its variance goes to 0.

4. If θ can only take on finitely many values, the MLE is consistent (proved in Lec 3).

5. (TPE Thm 6.3.7 p447) Say we have a model family $\{P_\theta\}_{\theta \in \Theta}$ with density p_θ w.r.t. some μ. Suppose that:

   - The P_θ's have common support,
   - $X_1, \ldots, X_n \overset{iid}{\sim} P_\theta$,
   - Θ contains an open set ω such that true parameter $\theta_0 \in \text{int}\ \omega$,
   - For almost all x, $p_\theta(x)$ is differentiable w.r.t. θ in a neighborhood around θ_0.

   Then the solution to $\dfrac{\partial}{\partial \theta} \ell(\theta \mid x_1, \ldots, x_n) = \sum \dfrac{p'_\theta(x_i)}{p_\theta(x_i)} = 0$ (usually the MLE) is consistent.

6. If $\hat{\theta}_n$ is an M-estimator or Z-estimator, try to use the Argmax consistency theorem (Lec 15).

To show an estimator is efficient:

1. Bare hands approach: Use CLT to get the limiting distribution of $\hat{\theta}_n$, and show that the asymptotic variance is equal to the inverse Fisher information.

2. (TPE Thm 6.5.1 p463) Show that the **Cramér conditions** hold: Say we have a model family $\{P_\theta\}_{\theta \in \Theta}$ with density p_θ w.r.t. some μ. Suppose that:

   - The P_θ's have common support,
   - $X_1, \ldots, X_n \overset{iid}{\sim} P_\theta$,
   - There exists an open subset $\omega \subseteq \Theta$ which contains the true θ_0 such that for almost all x, $p_\theta(x)$ admits all third derivatives $\left(\dfrac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} p_\theta(x) \right)$ for all $\theta \in \omega$,
   - $\mathbb{E}_\theta \left[\nabla \log p_\theta(x) \right] = 0$, $I(\theta)_{jk} = \mathbb{E}_\theta \left[-\dfrac{\partial^2}{\partial \theta_j \partial \theta_k} \log p_\theta(x) \right]$,
   - $I(\theta)$ is finite and invertible for all $\theta \in \omega$, and
   - There are functions M_{jkl} such that $\left| \dfrac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} p_\theta(x) \right| \leq M_{jkl}(x)$ for all $\theta \in \omega$, and $\mathbb{E}_{\theta_0}[M_{jkl}(X)] < \infty$ for all j, k, l.

   Then, with probability tending to 1 as $n \to \infty$, there exist solutions $\hat{\theta}_n$ to the likelihood equations such that

   - $\hat{\theta}_{jn}$ is consistent for estimating θ_j,
   - $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, I(\theta)^{-1})$,
   - $\hat{\theta}_{jn}$ is asymptotically efficient, i.e. $\sqrt{n}(\hat{\theta}_{jn} - \theta_j) \xrightarrow{d} \mathcal{N}(0, [I(\theta)]^{-1}_{jj})$.

3. For QMD families, can refer to VdV Theorem 5.39 p65.

To show an estimator is not efficient:

1. Bare hands approach: Use CLT to get the limiting distribution of $\hat{\theta}_n$, and show that the asymptotic variance is not equal to the inverse Fisher information.

3 Confidence Intervals and Tests

To compute a confidence interval/standard error for θ:

1. Use the plug-in estimate. For example, if $\hat{s} \sim \text{Bin}(n, s)$, then $\text{Var } \hat{s} = ns(1-s) \approx n\hat{s}(1-\hat{s})$.
2. Find an estimator $\hat{\theta}_n$ for θ and determine its asymptotic distribution, i.e. $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, then construct the CI from there: $\left(\hat{\theta}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \hat{\theta}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$ for an asymptotic level $1 - 2\alpha$ CI.
3. Invert the exact distribution. For example, see Applied Qual 2015 Qn 3.
4. Invert a test. For example, with a generalized likelihood test, the confidence region at level α would look something like $\left\{\theta : \ell(\hat{\theta}) - \ell(\theta) \leq \frac{1}{2}\chi_1^2(1-\alpha)\right\}$.
5. For MLEs, can use the Fisher information estimate. If $\hat{\mu}$ is the MLE, the standard error can be approximated given by $\sqrt{-1/\ddot{\ell}(\hat{\mu})}$.
6. Use a parametric bootstrap. Get repeated bootstrap estimates $\hat{\theta}^{*1}, \ldots, \hat{\theta}^{*B}$, then construct the CI using the quantiles of the $\hat{\theta}^*$'s.
7. For linear regression, use the fact that $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^TX)^{-1})$.

To come up with a test:

1. If MLE is difficult to compute, consider using the score test.
2. If we have already computed the restricted MLE, consider using the score test.
3. The score test and Le Cam's Third Lemma (finding limiting distribution under alternative regime) often come together.

4 Concentration Inequalities

1. The bounded differences inequality is very handy here. Usually there is an envelope or Lipschitz condition from which you can derive the conditions for the inequality.

5 LAN Stuff

To show contiguity of measures:

1. Use the definition of continguity directly.

2. Use Le Cam's First Lemma.

To show a model family is QMD:

1. Exponential families are QMD (Lec 19).
2. (TSH Thm 12.2.1 p486) Suppose Θ is an open subset of $\mathbb{R}$, fix $\theta_0 \in \Theta$. If
   - $\sqrt{p_\theta(x)}$ is an absolutely continuous function of θ in some neighborhood of θ_0 for μ-almost all x,
   - $\dfrac{\partial p_\theta(x)}{\partial \theta}$ exists at $\theta = \theta_0$ for μ-almost all x,
   - Fisher information $I(\theta)$ is finite and continuous in θ at θ_0,

   then $\{P_\theta\}$ is QMD at θ_0. (See slightly different set of conditions in TSH Cor 12.2.1 p487, basically relaxing condition 2. VdV Lem 7.6 also has a slightly different set of conditions.)
3. (TSH Thm 12.2.2 p488) Suppose Suppose Θ is an open subset of $\mathbb{R}^k$, and that P_θ has density p_θ w.r.t. μ. Assume $p_\theta(x)$ is continuously differentiable in θ for μ-almost all x, and that $I(\theta)$ exists and is continuous in θ. Then the family is QMD.

To find the limiting distribution of a statistic in the local asymptotic regime:

1. Show that the family of distributions is LAN, then use Le Cam's Third Lemma. (To show the joint normality condition in Le Cam's Third Lemma, use the multivariate CLT.)
2. To show LAN, hope that your family is QMD.

6 Quantiles (VdV Ch 21)

- The **quantile function** of a CDF F is the generalized inverse: $F^{-1}(p) = \inf\{x : F(x) \geq p\}$. It is left-continuous with range equal to the support of F.
- $F^{-1}(U)$ has CDF F if $U \sim \text{Unif}(0,1)$.
- (VdV Lem 21.2) For any sequence of CDFs, $F_n^{-1} \Rightarrow F^{-1}$ if and only if $F_n \Rightarrow F$.
- (VdV Cor 21.5) Fix $0 < p < 1$. If F is differentiable at $F^{-1}(p)$ with positive derivative $f(F^{-1}(p))$, then we have asymptotic normality of the empirical quantiles: $\sqrt{n}\left(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\right) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{p(1-p)}{f^2(F^{-1}(p))}\right)$.