Chapter 2: Parametric families of distributions

2.1 Exponential families (Severini 1.2)

(i) Defn: A parametric family $\{P_\theta; \theta \in \Theta\}$ with densities w.r.t. some σ-finite measure of the form
$f(x;\theta) = c(\theta)h(x)\exp(\Sigma_{j=1}^k \pi_j(\theta)t_j(x)) \quad -\infty < x < \infty$.

(ii) Examples: Binomial, Poisson, Gamma, Chi-squared, Normal, Beta, Negative Bionomial, Geometric ...

(iii) NOT: Cauchy, t-dsns, Uniform, any where support depends on θ.

(iv) Note $c(\theta)$ depends on θ through $\{\pi_j(\theta)\}$.
Defn: $(\pi_1, ..., \pi_k)$ is natural parametrization. – defined only up to linear combinations. We assume vector π is of full rank – no linear relationship between the π_j.

(v) Natural parameter space
$\Pi = \{\pi : 0 < \int h(x)\exp(\Sigma_{j=1}^k \pi_j(\theta)t_j(x))dx < \infty\}$
Lemma: Π is convex.

(vi) Thm: For any integrable ϕ, any π_0 in interior of Π, then $\int \phi(x)h(x)\exp(\Sigma_{j=1}^k \pi_j(\theta)t_j(x))dx$ is cts at π_0, has derivatives of all orders at π_0, and can get derivatives by differentiating under the integral sign.
Cor: With $\phi \equiv 1$, $c(\pi)$ is cts, diffble, etc.

(vii) Moments: $T = (t_1(X), ..., t_k(X))$
Note $\log f(x;\pi) \;=\; \log(h(x)) + \log c(\pi) + \Sigma_{j=1}^k \pi_j t_j(x)$.
Also $\frac{\partial f}{\partial \pi} = \frac{\partial \log f}{\partial \pi} f$.
Differentiating $\int f(x;\pi)dx = 1$ gives $\mathrm{E}(T) = -\frac{\partial}{\partial \pi}(\log c(\pi))$
Differentiating again gives $\mathrm{var}(T) = -\frac{\partial^2}{\partial \pi^2}(\log c(\pi))$

2.2 Transformation Group families (Severini 1.3)

(i) Groups $\mathcal{G}$ of transformations on $\Re$:
Contains identity, closed under inverses, and composition.

(ii) Location:
$\mathcal{G} = \{g_a : g_a(x) = x + a\}$ $x \in \Re, a \in \Re$.
Let $X \sim F$ **and** $X_a = X + a$,
$F_a(x) = \Pr(X_a \leq x) = \Pr(X \leq (x - a)) = F(x - a)$.
The set of dsns, for fixed F **and for all** $a \in \Re$ **is a location family. Examples: Normal, Cauchy, double exponential. Also Uniform, Exponential, ...**

(iii) Scale:
$\mathcal{G} = \{h_b : h_b(x) = bx\}$ $x \in \Re^+, b \in \Re^+$.
Let $X \sim F$ **and** $X_b = bX$,
$F_b(x) = \Pr(X_b \leq x) = \Pr(X \leq x/b) = F(x/b)$.
The set of dsns, for fixed F **and for all** $b \in \Re^+$ **is a scale family. Examples: exponential, gamma,**

(iv) We can combine location and scale: Normal, Cauchy, logistic, Uniform, ... see also Severini Pp 10-11.

(v) A rather large group family:
X_i **i.i.d. with cts df** F **and support the whole of** $\Re$.
$\mathcal{G} = \{g : g$ **cts strictly increasing,** $g(-\infty) = -\infty, g(\infty) = \infty\}$
$W_i = g(X_i)$, W_i **are also i.i.d. with cts df and support the whole real line. Family consists of all such dsns.**

(vi) A group family: $\mathcal{G} = \{g_{b,c} : g(x) = bx^{1/c}, b > 0, c > 0\}$.
If X **is standard exponential:** $F(x) = 1 - e^{-x}$, $g_{b,c}(X)$ **has Weibull dns with density** $cb^{-c}x^{c-1}\exp((-x/b)^c)$

2.3 Sufficiency, minimal sufficiency, and completeness (Severini 1.5) (THIS IS REVISION ONLY)

(i) Defn: Vector T is sufficient for θ w.r.t. $\{P_\theta; \theta \in \Theta\}$ if $P_\theta(X|T(X) = t)$ does not depend on θ

(ii) Factorization criterion.
$T(X)$ is sufficient for θ iff $f(x;\theta) \equiv h(x)g(T(X), \theta)$

(iii) Defn: T(X) is minimally sufficient if it is a function of every sufficient statistic. Idea: coarsest partition of the sample space that is sufficient.
Minimal sufficient statistics are essentially unique.

(iv) Likelihood ratio criterion: define

$$x \sim x' \ \text{ iff } \ \frac{f(x;\theta_1)}{f(x;\theta_2)} = \frac{f(x';\theta_1)}{f(x';\theta_2)} \ \forall \ \theta_1, \theta_2 \in \Theta$$

T is minimal sufficient iff $T(x) = T(x') \Leftrightarrow x \sim x'$

(v) Defn: Sufficient statistic T is (boundedly) complete if for any measurable real-valued (bounded) function g
$\mathrm{E}_\theta(g(T)) = 0 \ \forall \theta \in \Theta \quad \Rightarrow \quad P_\theta(g(T) = 0) = 1 \ \forall \theta \in \Theta$
(Completeness provides uniqueness of unbiased estimators of $\xi(\theta)$.)

(vi) Lehmann-Scheffé Thm: for sufficient T
T complete $\Rightarrow$ T min suff.

(vii) Basu's Thm: for sufficient T
T complete, V dsn not depending on θ
$\Rightarrow \quad \forall P_\theta, T, V$ independent.

2.4 Vector exponential families

(i) Density on some subset of $\Re^n$ w.r.t. some σ-finite measure: $f(x;\theta) = c(\theta)h(x)\exp(\Sigma_{j=1}^k \pi_j(\theta)t_j(x)) \quad \forall\ x \ \in\ \Re^n$
For example: X_i **i.i.d. from scalar exponential family** $\Rightarrow$ **vector** $X^{(n)}$ **from (vector) exponential family.**

(ii) For X_i **i.i.d. from scalar exponential family,** $\{T_j \equiv \Sigma_{i=1}^n t_j(X_i); j = 1, ..., k\}$ **are sufficient. (Use 2.3(ii)).**

(iii) $\{T_j; j = 1, ..., k\}$ **is natural sufficient statistic. If there are no affine relationships among the** $\{t_j(x)\}$**, then** π_j **are identifiable, and family is of full rank. Note the dimension** k **of suff. statistic does not depend on** n**.**

(iv) For X_i **i.i.d. from scalar exponential family,** $\{T_j; j = 1, ..., k\}$ **is also from a (vector) exponential family.**

(v) For X_i **i.i.d. from scalar exponential family,** $((T_1, ..., T_l)|(T_{l+1}, ..., T_k))$ **is also from a (vector) exponential family.**

(vi) $m_T(s) = \mathrm{E}(\exp(s'T)) = c(\pi)/c(s + \pi)$ **where** $c()$ **is the c-fn for** $T = (T_1, ..., T_k)$**.**

(vii) Provided Π **contains an open rectangle in** $\Re^k$
(a) Natural sufficient $(T_1, ..., T_k)$ **is minimal sufficient**
(b) Natural sufficient $(T_1, ..., T_k)$ **is complete.**

(viii) Rank vs dimension: Rank refers to affine relationships among $\{\pi_j\}$ **or** $\{T_j\}$**. Full rank needed for minimal sufficiency (see Sev.P.18). Dimension refers to** Π **containing open rectangle in** $\Re^k$**, and is needed for uniqueness of Laplace transforms, and hence for completeness.**

2.5 Multivariate Normal distribution (JAW 1.13, Sev 1.7)

Defn: $Y = (Y_1, ..., Y_n)$ **is jointly Normal with mean 0 if** $\exists$ $Z_1, ..., Z_k$ **i.i.d.** $N(0,1)$ **s.t.** $Y = AZ$ **for some** $n \times k$ **matrix** A.

(i) $\mathrm{var}(Y) = \mathrm{E}(YY') = \mathrm{E}(AZZ'A') = AA' \equiv \Sigma$

(ii) Σ **symmetric and non-negative definite** $\Rightarrow$ $\exists$ $n \times n$ **matrix** A **with** $\Sigma = AA'$.

(iii) The mgf of Y **is** $m_Y(s) = \mathrm{E}(\exp(s'Y)) = \mathrm{E}(\exp(s'AZ))$
$= \mathrm{E}(\exp((A's)'Z)) = \mathrm{E}(\Pi_j \exp((A's)_j Z_j)) = \Pi_j \mathrm{E}(\exp((A's)_j Z_j))$
$= \Pi_j m_{Z_j}((A's)_j) = \Pi_j \exp(\frac{1}{2}((A's)_j)^2) = \exp(\frac{1}{2}\Sigma_j (A's)_j)^2)$
$= \exp(\frac{1}{2}(s'A)(A's)) = \exp(\frac{1}{2}s'\Sigma s)$

(iv) If Σ **is non-singular, then** A **is** $n \times n$ **non-singular, let** $Y = \mu + AZ$**, then the pdf of** $Y \sim N_n(\mu, \Sigma)$ **is**
$f_Y(y) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp((y-\mu)'\Sigma^{-1}(y-\mu))$

(v) If $Y \sim N_n(0, \Sigma)$ **and** Σ **is partitioned into dimensions** k **and** $n-k$ **as** Σ_{ij}, $i, j, = 1, 2$**, then, using the mgf,**
(a) $(Y_1, ..., Y_k) \sim N_k(0, \Sigma_{11})$.
(b) If $\Sigma_{12} = 0$, $Y^{(1)} = (Y_1, ..., Y_k)$ **is independent of** $Y^{(2)} = (Y_{k+1}, ..., Y_n)$.
(c) If $(X_1, X_2)'$ **are jointly Normal vectors they are indep, iff** $\mathrm{Cov}(X_1, X_2) = 0$.
(d) Linear combinations of Normals are Normal.

(vi) If $Y \sim N_n(\mu, \Sigma)$**, and** $\mu' = (\mu^{(1)}, \mu^{(2)})$**,** Σ **partitioned as in (v), and** Σ_{22} **non-singular then** $(Y^{(1)}|Y^{(2)}) \sim N_k(\mu^{(1.2)}, \Sigma_{11.2})$**, where** $\mu^{(1.2)} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu^{(2)})$**,** $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

(vii) $Y^{(1)} - \mathrm{E}(Y^{(1)}|Y^{(2)})$ **is independent of** $Y^{(2)}$
Proof: Check the covariance.

2.6 Chi-squared and non-central chi-sqared dsns (JAW 1.15-1.16)

(i) Defn: If X_i are i.i.d $N(0,1)$, $\Sigma_1^n X_i^2$ is χ_n^2.

(ii) If $X \sim N(0,1)$, $Y = X^2$, $m_Y(s) = \mathrm{E}(\exp(sX^2)) = (1-2s)^{-1/2}$
If $V \sim \mathcal{E}(1)$, $m_V(s) = (1-s)^{-1}$
If $W \sim G(\alpha, \beta)$, $m_W(s) = (1-\beta s)^{-\alpha}$

(iii) Hence χ_n^2 is Gamma, $G(n/2, 2)$.

(iv) $Y \sim N_n(0, \Sigma) \quad \Rightarrow \quad Y'\Sigma^{-1}Y \;\sim\; \chi_n^2$

(v) If $X \sim N(\mu, 1)$, $Y = X^2$,
$m_Y(s) = \mathrm{E}(\exp(sX^2) = (1-2s)^{-1/2}\exp(\mu^2 s/(1-2s))$.
If $(W|K = k) \sim \chi^2_{2k+1}$, $\quad K \sim \mathcal{P}(\mu^2/2)$, **then** $m_W(s) = \mathrm{E}(\mathrm{E}(e^{sW})|K) = \mathrm{E}((1-2s)^{-(2K+1)/2}) = (1-2s)^{-1/2} m_K(\log(1/(1-2s))$
which is same if we do the sums right!

(vi) Now let $X_1 \sim N(\mu, 1)$ and $X_{2,\ldots,n}$ i.i.d $N(0,1)$ indep of X_1, $Y = \Sigma_1^n X_i^2$. Mgf of Y is multiplied by $(1-2s)^{-(n-1)/2}$, so now Y has dsn of W, where $W|K \sim G((2k+n)/2, 2)$, K as above.
We define this dsn to be non-central χ^2 with n deg freedom and non-centrality $\delta = \mu^2$.
(Note: if $\mu = 0$, $P(K = 0) = 1$, and we regain $\chi_n^2 \equiv G(n/2, 2)$.)

(vii) Now $X \sim N_n(\mu, I)$, then $Y = X'X \sim \chi_n^2(\delta)$ with $\delta = \mu'\mu$.

(viii) $X \sim N_n(\mu, \Sigma)$, $Y = X'\Sigma^{-1}X$, **then** $Y \sim \chi_n^2(\delta)$, **where** $\delta = \mu'\Sigma^{-1}\mu$.