Exponential Families II

MIT 18.655

Dr. Kempthorne

Spring 2016

Outline

Random Vectors: Expectation and Variance

$\mathbf{U}$ $(k \times 1)$ and $\mathbf{V}$ $(l \times 1)$ are random vectors

- If $\mathbf{A}$ $(m \times k)$, $\mathbf{B}$ $(m \times l)$ are nonrandom, and then
$$E(\mathbf{AU} + \mathbf{BV}) = \mathbf{A}E(\mathbf{U}) + \mathbf{B}E(\mathbf{V})$$
- If $\mathbf{U} = \mathbf{c}$ with probability 1 $E(\mathbf{U}) = \mathbf{c}$.
- For a random vector $\mathbf{U}$, if $E(|\mathbf{U}|^2) = \sum_{i=1}^{k} E(U_i^2) < \infty$ define the *variance* of $\mathbf{U}$ by
$$\begin{aligned} Var(\mathbf{U}) &= E\left[(\mathbf{U} - E(\mathbf{U}))(\mathbf{U} - E(\mathbf{U}))^T\right] \\ &= ||Cov(U_i, U_j)|| \quad (k \times k) \end{aligned}$$
- For $\mathbf{A}$ $(m \times k)$ as above:
$$Var(\mathbf{A}U) = \mathbf{A}Var(\mathbf{U})\mathbf{A}^T \quad (m \times m)$$
- For $\mathbf{c}$ $(k \times 1)$ a constant vector
$$Var(\mathbf{U} + \mathbf{c}) = Var(\mathbf{U})$$
- For $\mathbf{a}$ $(k \times 1)$ a constant vector,
$$\begin{aligned} Var(\mathbf{a}^T\mathbf{U}) &= Var(\textstyle\sum_{j=1}^{k} a_j U_j) \\ &= \mathbf{a}^T Var(\mathbf{U})\mathbf{a} = \textstyle\sum_{i,j} a_i a_j Cov(U_i, U_j) \end{aligned}$$

Random Vectors: Expectation and Variance

Proposition B.5.1 If $E[|\mathbf{U}|^2] < \infty$ then

$$Var(\mathbf{U}) \text{ is positive definite}$$

if and only if

$$P[\mathbf{a}^T\mathbf{U} + b = 0] < 1,$$

$$\text{for every } \mathbf{a} \neq \mathbf{0}, \text{ and } b \in R.$$

Proof. $Var(U)$ is not positive definite iff $\mathbf{A}^T Var(\mathbf{Y})\mathbf{a} = 0$ for some $\mathbf{a} \neq 0$ which is equivalent to $Var(\mathbf{a}^T\mathbf{U}) = 0$.

Random Vectors: Covariance

Definition: For random vectors **U** $(k \times 1)$ and **V** $(l \times 1)$ define the *Covariance* of U $(k \times 1)$ and **V** $(l \times 1)$ by

$$Cov(\mathbf{U}, \mathbf{V}) = E\left[(\mathbf{U} - E(\mathbf{U}))(\mathbf{V} - E(\mathbf{V}))^T\right] \quad (k \times l)$$

(must assume: $E|\mathbf{U}|^2 < \infty$ and $E|\mathbf{V}|^2 < \infty$)

- If **U** and **V** are independent
  $$Cov(\mathbf{U}, \mathbf{V}) = 0.$$
- For nonrandom A, a, B, b,
  $$Cov(A\mathbf{U} + a, B\mathbf{V} + b) = ACov(\mathbf{U}, \mathbf{V})B^T$$
- If **U** and **W** are random $(k \times 1)$ vectors, then
  $$Var(\mathbf{U} + \mathbf{W}) = Var(\mathbf{U}) + Cov(\mathbf{U}, \mathbf{W}) + Cov(\mathbf{W}, \mathbf{U}) + Var(\mathbf{W})$$
  and if **U** and **W** are independent
  $$Var(\mathbf{U} + \mathbf{W}) = Var(\mathbf{U}) + Var(\mathbf{W})$$

Random Vectors: Moment Generating Functions

Moment-Generating Function of a Random Vector

Let $\mathbf{T} = (T_1, T_2, \ldots, T_k)^T$ be a $(k \times 1)$ random vector.

- For $\mathbf{s} = (s_1, s_2, \ldots, s_k)^T \in R^k$, define
$$M(s) \equiv E[e^{\mathbf{s}^T \mathbf{T}}]$$
- $M(\mathbf{s})$ is the **moment-generating function (mgf) of** T
- The mgf may not exist for a given $\mathbf{T}$. If it does exist, it is defined for $\mathbf{s}$ in some ball centered at $\mathbf{s} = \mathbf{0}$.
- Define the **characteristic function (cf)** of $\mathbf{T}$:
$$\phi(\mathbf{s}) = E[e^{i\mathbf{s}^T \mathbf{T}}] = E[cos(\mathbf{s}^T \mathbf{T})] + iE[sin(\mathbf{s}^T \mathbf{T})]$$
- The cf always exists.

Random Vectors: Moment Generating Functions

Theorem B.5.1 Let $\mathcal{S} = \{\mathbf{s} : M(\mathbf{s}) < \infty\}$. Then

- $\mathcal{S}$ is convex.
- If $\mathcal{S}$ has a nonempty interior $\mathcal{S}^0$, (contains a sphere $\mathcal{S}(\mathbf{0}, \epsilon), \epsilon > 0$), then M is analytic on $\mathcal{S}^0$.
- If $\mathcal{S}^0 \neq \emptyset$, and $E[|\mathbf{T}|^p] < \infty$ for all p, then
  if $i_1 + i_2 + \cdots + i_k = p$,

$$\frac{\partial^p M(\mathbf{s})}{\partial {s_1}^{i_1} \cdots \partial {s_k}^{i_k}} |_{\mathbf{s}=\mathbf{0}} = E[T_1^{i_1} \cdots T_k^{i_i}]$$

$$||\frac{\partial M}{\partial s_j}(\mathbf{s} = \mathbf{0})|| \quad = \quad ||E(\mathbf{T}_j)\,|| = E[\mathbf{T}]$$

$$||\frac{\partial^2 M}{\partial s_i \partial s_j}(\mathbf{s} = \mathbf{0})|| \quad = \quad ||E(\mathbf{T}_i \mathbf{T}_j)\,|| = \; \mathrm{E}[\mathbf{T}\mathbf{T}^T]$$

- If $\mathcal{S}^0$ is nonempty, then $M(\mathbf{s})$ determines the distribution of $\mathbf{U}$ uniquely.

Random Vectors: Moment Generating Functions

Definition: The *Cumulant Generating Function* of the random vector $\mathbf{T}$ with mgf $M_{\mathbf{T}}(\mathbf{s})$ is

$$K(\mathbf{s}) = K_{\mathbf{T}}(\mathbf{s}) = \log M_{\mathbf{T}}(\mathbf{s}).$$

$$c_{i_1,\cdots,i_k} = c_{i_1,\cdots,i_k}(\mathbf{T}) = \frac{\partial^p}{\partial s_1^{i_1}\cdots\partial s_k^{i_i}} K(\mathbf{s})\Big|_{\mathbf{s}=\mathbf{0}}$$

- In the bivariate case ($k = 2$) where

$$\mu = E[\mathbf{T}], \text{ and } \tau_{i,j} = E[(T_1 - \mu_1)^i (T_2 - \mu_2)^j]$$

$$\begin{aligned}
c_{10} &= \mu_1 \\
c_{01} &= \mu_2 \\
c_{2,0} &= \tau_{2,0} = var(T_1) \\
c_{0,2} &= \tau_{0,2} = var(T_2) \\
c_{1,1} &= \tau_{1,1} = cov(T_1, T_2) \\
c_{3,0} &= \tau_{3,0} = E[(T_1 - \mu_1)^3] \\
c_{0,3} &= \tau_{0,3} = E[(T_2 - \mu_2)^3] \\
c_{4,0} &= \tau_{4,0} - 3\tau_{2,0}^2 \\
c_{0,4} &= \tau_{0,4} - 3\tau_{0,2}^2
\end{aligned}$$

Sums of Independent Random Vectors

If **U** and **V** are independent $(k \times 1)$ random vectors, then

$$\begin{aligned} M_{U+V}(\mathbf{s}) &= M_U(\mathbf{s}) \times M_V(\mathbf{s}) \\ K_{U+V}(\mathbf{s}) &= K_U(\mathbf{s}) + K_V(\mathbf{s}) \end{aligned}$$

Multivariate Normal Distributions

Definition B.6.1: A random vector $\mathbf{U}$ $(k \times 1)$ has a k-variate normal distribution iff $\mathbf{U}$ can be written as

$$\mathbf{U} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$$

where $\boldsymbol{\mu}$, $\mathbf{A}$ are constant and $\mathbf{Z} = (Z_1, \dots, Z_k)^T$; Z_i iid $N(0,1)$.

Definition B.6.2: A random vector $\mathbf{U}$ $(k \times 1)$ has a k-variate normal distribution iff for every $(k \times 1)$ nonrandom $\mathbf{a}$:

$$\mathbf{a}^T\mathbf{U} = \sum_{i=1}^{k} a_i U_i \text{ has a univariate normal distribution}$$

The moment generating function of U is

$$M_{\mathbf{U}}(\mathbf{s}) = exp\left\{\mathbf{s}^T\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{s}^T\Sigma\mathbf{s}\right\}$$

where $\boldsymbol{\mu} = E[\mathbf{U}]$, and $\Sigma = Cov(\mathbf{U}) = \mathbf{A}\mathbf{A}^T$.

Outline

Properties of Exponential Families

Properties of Exponential Families

Theorem 1.6.3 Let $\mathcal{P}$ be a canonical k-parameter exponential family generated by $(\mathbf{T}, h)$, with corresponding natural parameter space $\mathcal{E}$ and function $A(\boldsymbol{\eta})$. Then

- $\mathcal{E}$ is convex
- $A : \mathcal{E} \to R$ is convex
- If $\mathcal{E}$ has nonempty interior $\mathcal{E}^0 \subset R^k$, and $\eta_0 \in \mathcal{E}^0$, then $\mathbf{T}(X)$ has under $\boldsymbol{\eta}_0$ a mgf given by

  $$M(\mathbf{s}) = \exp\{A(\boldsymbol{\eta}_0 + s) - A(\boldsymbol{\eta}_0)\}$$

  valid for all $\mathbf{s}$ such that $\boldsymbol{\eta}_0 + \mathbf{s} \in \mathcal{E}$.
  (this set of $\mathbf{s}$ includes a ball about $\boldsymbol{\eta}_0$)

Corollary 1.6.1 Under the conditions of the theorem

$$\begin{aligned} E_{\eta_0}[\mathbf{T}(X)] &= \dot{A}(\boldsymbol{\eta}_0) \\ Var_{\eta_0}[\mathbf{T}(X)] &= \ddot{A}(\boldsymbol{\eta}_0) \end{aligned}$$

where $\dot{A}(\boldsymbol{\eta}_0) = |\frac{\partial A}{\partial \eta_j}(\boldsymbol{\eta}_0)|$ and $\ddot{A}(\boldsymbol{\eta}_0) = ||\frac{\partial^2 A}{\partial \eta_i \partial \eta_j}(\boldsymbol{\eta}_0)||$

Example: Multinomial Distribution

Multinomial Distribution

$X = (X_1, X_2, \ldots, X_q) \sim Multinomial(n, \theta = (\theta_1, \theta_2, \ldots, \theta_q))$

$p(x \mid \theta) = \frac{n}{x_1! \cdots x_q!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_q^{x_q}$ where

- q is a given positive integer,
- $\theta = (\theta_1, \ldots, \theta_q) : \sum_1^q \theta_j = 1$.
- n is a given positive integer
- $\sum_1^q X_i \ = n$.

Example: Multinomial Distribution

$$\begin{aligned} p(x \mid \theta) &= \tfrac{n}{x_1!\cdots x_q!}\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_q^{x_q} \\ &= \tfrac{n}{x_1!\cdots x_q!} \times exp\{log(\theta_1)x_1 + \cdots + log(\theta_{q-1})x_{q-1} \\ &\qquad +log(1-\textstyle\sum_1^{q-1}\theta_j)[n-\sum_1^{q-1}x_j]\} \\ &= h(x)exp\{\textstyle\sum_{j=1}^{q-1}\eta_j(\theta)T_j(x) - B(\theta)\} \\ &= h(x)exp\{\textstyle\sum_{j=1}^{q-1}\eta_j T_j(x) - A(\eta)\} \end{aligned}$$

where:

- $h(x) = \frac{n}{x_1!\cdots x_q!}$
- $\eta(\theta) = (\eta_1(\theta), \eta_2(\theta), \ldots, \eta_{q-1}(\theta))$

  $$\eta_j(\theta) = log(\theta_j/(1-\textstyle\sum_1^{q-1}\theta_j)),\ j = 1, \ldots, q-1$$
- $T(x) = (X_1, X_2, \ldots, X_{q-1}) = (T_1(x), T_2(x), \ldots, T_{q-1}(x))$.
- $B(\theta) = -nlog(1-\sum_{j=1}^{q-1}\theta_j)$ and $A(\boldsymbol{\eta}) = +nlog(1+\sum_{j=1}^{q-1}e^{\eta_j})$

$$\dot{A}(\boldsymbol{\eta})_j = n\frac{e^{\eta_j}}{1+\sum_{j=1}^{q-1}e^{\eta_j}} = n\frac{\theta_j/(1-\sum_1^{q-1}\theta_k)}{1+\sum_1^{q-1}\theta_k/(1-\sum_1^{q-1}\theta_k)} = n\theta_j$$

$$\ddot{A}(\boldsymbol{\eta})_{i,j} = -n\theta_i\theta_j,\ (i \neq j) \text{ and } \ddot{A}(\boldsymbol{\eta})_{i,i} = n\theta_i(1-\theta_i),$$

Rank of Exponential Family

Defining the Rank of an Exponential Family

- Every k-parameter exponential family is also a k^*-parameter exponential family for any $k^* > k$.
- The *minimal* value of k defines the rank of the exponential family. Define *minimal* k as the rank when the generating statistic $\mathbf{T}(X)$ is k-dimentional, and the collection

  $$\{1, T_1(X), T_2(X), \ldots, T_k(X)\}$$

  are linearly independent with positive probability, i.e.,

  $$P[\sum_{j=1}^{k} a_j T_j(X) = a_{k+1} \mid \eta] < 1, \text{ unless all } a_j = 0.$$

  Note: the set of positive support on $\mathcal{X}$ does not depend on η.

Rank of Exponential Family

Theorem 1.6.4 Let $\mathcal{P} = \{q(x \mid \eta), \eta \in \mathcal{E}\}$ be a canonical exponential family generated by $(\mathbf{T}(X), h(X))$ with natural parameter space $\mathcal{E}$ such that $\mathcal{E}$ is open. Then the following statements are equivalent

- $\mathcal{P}$ is of rank k.
- $\boldsymbol{\eta}$ is an identifiable parameter.
- $Var(\mathbf{T} \mid \eta)$ is positive defnite
- $\boldsymbol{\eta} \to \dot{A}(\boldsymbol{\eta})$ is 1-to-1 on $\mathcal{E}$.
- $A(\boldsymbol{\eta})$ is strictly convex on $\mathcal{E}$.

Note: $\mathcal{E}$ open $\Longrightarrow$ $\dot{A}$ defined on all $\mathcal{E}$.

Corollary 1.6.2 If $\mathcal{P}$ is of rank k under Theorem 1.6.4, then

- $\mathcal{P}$ may be uniquely parametrized by
$$\boldsymbol{\mu}(\boldsymbol{\eta}) \equiv E[\mathbf{T}(X) \mid \boldsymbol{\eta}].$$
- $\log[q(x, \boldsymbol{\eta})]$ is strictly concave in $\boldsymbol{\eta}$ on $\mathcal{E}$.

p-Variate Gaussian Family

Let $\mathbf{Y}$ be a $(p \times 1)$ random vector with a p-variate Gaussian distribution

$$\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \Sigma)$$

where $\boldsymbol{\mu} = E[\mathbf{Y}]$ and $\Sigma = Var(\mathbf{Y})$ is positive definite, rank p.

The density of $\mathbf{Y}$ is

$$p(\mathbf{y}, \mu, \Sigma) = |det(\Sigma)|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}} \exp\{-\tfrac{1}{2}(\mathbf{Y} - \mu)^T \Sigma^{-1}(\mathbf{Y} - \mu)\}$$

Taking logs:

$$\begin{aligned}\log[p(\mathbf{y}, \mu, \Sigma)] &= -\tfrac{1}{2}\mathbf{Y}^T\Sigma^{-1}\mathbf{Y} + [\Sigma^{-1}\mu]^T\mathbf{Y} \\ &\quad -\tfrac{1}{2}\mu^T\Sigma^{-1}\mu - \tfrac{1}{2}log[det(\Sigma)] - \tfrac{p}{2}log[2\pi]]\end{aligned}$$

Defining $\Sigma^{-1} = ||\sigma^{i,j}||$, we can write the first 2 terms as

$$-[\textstyle\sum_{i<j} \sigma^{i,j} Y_i Y_j + \frac{1}{2}\sum_i \sigma^{i,i} Y_i^2] + \sum_{i=1}^p [\sum_{j=1}^p \sigma^{i,j}\mu_j] Y_i$$

- The parameter space dimension is
$$k = p + p(p+1)/2 = p(p+3)/2$$
- The sufficient statistics are
$$[(Y_1, \ldots, Y_p), \{Y_i Y_j, 1 \le i \le j \le p\}],$$

- $h(Y) \equiv 1$
- $\theta = (\mu, \Sigma)$
- $B(\theta) = \frac{1}{2}\left(log[|det(\Sigma)|] + \mu^T \Sigma^{-1} \mu\right)$

For a sample $Y_1, \ldots, Y_n$ of iid $N_p(\mu\Sigma)$ r.vectors, the data

$$\mathbf{X} = (Y_1, Y_2, \ldots, Y_n)$$

follows the $k = p(p+3)/2$ parameter exponential family with

$$\mathbf{T} = (\sum_i \mathbf{Y}_i, LowerTriangle(\sum_i \mathbf{Y}_i \mathbf{Y}_i^T))$$

($LowerTriangle(\cdot)$ refers to matrix elements along and below the diagonal)

Conjugate Families of Prior Distributions

Let $X_1, \ldots, X_n$ be a sample from the k-parameter exponential family
$p(x \mid \theta) = [\prod_{i=1}^{n} h(x_i)] exp\{\sum_{j=1}^{k} \eta_j(\theta) \sum_{i=1}^{n} T_j(x_i) - nB(\theta)\}$
where θ is k-dimensional.

- Treat θ as the variable of interest in $p(x \mid \theta)$
- Treat n and T_j as parameters in $p(x \mid \theta)$
- Find a normalizing function:
  $\omega(\mathbf{t}) = \int \cdots \int \exp\left\{\sum_{j=1}^{k} t_j \eta_j(\theta) - t_{k+1} B(\theta)\right\} d\theta_1 \cdots d\theta_k$
  and set
  $\Omega = \{(t_1, \ldots, t_{k+1}) : 0 < \omega(t_1, \ldots, t_{k+1}) < \infty\}$

Proposition 1.6.1 The $(k+1)$-parameter exponential family given by

$$\pi_t(\theta) = \exp\left\{\sum_{j=1}^{k} t_j \eta_j(\theta) - t_{k+1} B(\theta) - \log[\omega(t)]\right\}$$

where $t = (t_1, \ldots, t_{k+1}) \in \Omega$, is a conjugate prior to $p(x \mid \theta)$.

18.655 Mathematical Statistics
Spring 2016