# Exponential Families

MIT 18.655

Dr. Kempthorne

Spring 2016

# Outline

# Definition

Let $X$ be a random variable/vector with sample space $\mathcal{X} \subset R^q$ and probability model $P_\theta$. The class of probability models $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is a one-parameter exponential family if the density/pmf function $p(x \mid \theta)$ can be written:

$$p(x \mid \theta) = h(x) exp\{\eta(\theta)T(x) - B(\theta)\}$$

where

$$h : \mathcal{X} \to R$$
$$\eta : \Theta \to R$$
$$B : \Theta \to R.$$

**Note:**

- By the Factorization Theorem, $T(X)$ is sufficient for $\theta$

  $$p(x \mid \theta) = h(x) g(T(x), \theta)$$

  Set $g(T(x), \theta) = exp\{\eta(\theta)T(x) - B(\theta)\}$
- $T(X)$ is the *Natural Sufficient Statistic*.

# Examples

Poisson Distribution (1.6.1) : $X \sim Poisson(\theta)$, where $E[X] = \theta$.

$$\begin{aligned} p(x \mid \theta) &= \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, \ldots \\ &= \frac{1}{x!} exp\{(log(\theta)x - \theta\} \\ &= h(x) exp\{\eta(\theta) T(x) - B(\theta)\} \end{aligned}$$

where:

- $h(x) = \frac{1}{x!}$
- $\eta(\theta) = log(\theta)$
- $T(x) = x$

# Examples

Binomial Distribution (1.6.2) : $X \sim Binomial(\theta, n)$

$$
\begin{aligned}
p(x \mid \theta) &= \binom{n}{x} \theta^1 (1-\theta)^{n-x} \quad x = 0, 1, \ldots, n \\
&= \binom{n}{x} exp\{log(\tfrac{\theta}{1-\theta})x + nlog(1-\theta)\} \\
&= h(x) exp\{\eta(\theta) T(x) - B(\theta)\}
\end{aligned}
$$

where:

- $h(x) = \binom{n}{x}$
- $\eta(\theta) = log(\frac{\theta}{1-\theta})$
- $T(x) = x$
- $B(\theta) = -nlog(1-\theta)$

# Examples

Normal Distribution : $X \sim N(\mu, \sigma_0^2)$. (Known variance)

$$
\begin{aligned}
p(x \mid \theta) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2} \\
&= [\frac{1}{\sqrt{2\pi\sigma_0^2}} exp\{-\frac{x^2}{2\sigma_0^2}\}] \times exp\{\frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2}\} \\
&= h(x) exp\{\eta(\theta)T(x) - B(\theta)\}
\end{aligned}
$$

where:

- $h(x) = [-\frac{1}{\sqrt{2\pi\sigma_0^2}} exp\{-\frac{x^2}{2\sigma_0^2}\}]$
- $\eta(\theta) = \frac{\mu}{\sigma_0^2}$
- $T(x) = x$
- $B(\theta) = \frac{\mu^2}{2\sigma_0^2}$

# Examples

Normal Distribution : $X \sim N(\mu_0, \sigma^2)$. (Known mean)

$$\begin{aligned} p(x \mid \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_0)^2} \\ &= [\frac{1}{\sqrt{2\pi\sigma^2}}] \times exp\{-\frac{1}{2\sigma^2}(x-\mu_0)^2) - \frac{1}{2}log(\sigma^2)\} \\ &= h(x)exp\{\eta(\theta)T(x) - B(\theta)\end{aligned}$$

where:

- $h(x) = [\frac{1}{\sqrt{2\pi}}]$
- $\eta(\theta) = -\frac{1}{2\sigma^2}$
- $T(x) = (x-\mu_0)^2$
- $B(\theta) = \frac{1}{2}log(\sigma^2)$

# Samples from One-Parameter Exponential Family Distribution

Consider a sample: $X_1, \ldots, X_n$, where $X_i$ are iid $P$ where $P \in \mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ is a one-parameter exponential family distribution with density function

$$p(x \mid \theta) = h(x) exp\{\eta(\theta)T(x) - B(\theta)\}$$

The sample $\mathbf{X} = (X_1, \ldots, X_n)$ is a random vector with density/pmf:

$$\begin{aligned} p(\mathbf{x} \mid \theta) &= \prod_{i=1}^n \left(h(x_i) exp[\eta(\theta)T(x_i) - B(\theta)]\right) \\ &= [\prod_{i=1}^n h(x_i)] \times exp[\eta(\theta)\sum_{i=1}^n T(x_i) - nB(\theta)] \\ &= h^*(x) exp\{\eta^*(\theta)T^*(\mathbf{x}) - B^*(\theta)\} \end{aligned}$$

where:

- $h^*(x) = \prod_{i=1}^n h(x_i)$
- $\eta^*(\theta) = \eta(\theta)$
- $T^*(\mathbf{x}) = \sum_{i=1}^n T(x_i)$
- $B^*(\theta) = nB(\theta)$

Note: The Sufficient Statsitic $T^*$ is one-dimensional for all $n$.

# Samples from One-Parameter Exponential Family Distribution

**Theorem 1.6.1** Let $\{P_\theta\}$ be a one-parameter exponential family of discrete distributions with pmf function:

$$p(x \mid \theta) = h(x) exp\{\eta(\theta)T(x) - B(\theta)\}$$

Then the family of distributions of the statistic $T(X)$ is a one-parameter exponential family of discrete distributions whose frequency functions are

$$P_\theta(T(x) = t) = p(t \mid \theta) = h^{**}(t) exp\{\eta(\theta)t - B(\theta)\}$$

where

$$h^{**}(t) = \sum_{\{x: T(x)=t\}} h(x)$$

**Proof:** Immediate

# Canonical Exponential Family

- Re-parametrize setting $\eta = \eta(\theta)$ – the *Natural Parameter*
- The density has the form
$$p(x, \eta) = h(x) exp\{\eta T(x) - A(\eta)\}$$
- The function $A(\eta)$ replaces $B(\theta)$ and is defined as the normalization constant:
$$log(A(\eta)) = \int \int \cdots \int h(x) exp\{\eta T(x)] dx$$
if $X$ continuous

or
$$log(A(\eta)) = \sum_{x \in \mathcal{X}} h(x) exp\{\eta T(x)]$$
if $X$ discrete
- The Natural Parameter Space $\{\eta : \eta = \eta(\theta), \theta \in \Theta\} = \mathcal{E}$ (Later, Theorem 1.6.3 gives properties of $\mathcal{E}$)
- $T(x)$ is the Natural Sufficient Statistic.

# Canonical Representation of Poisson Family

Poisson Distribution (1.6.1) : $X \sim Poisson(\theta)$, where $E[X] = \theta$.

$$\begin{aligned} p(x \mid \theta) &= \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, \ldots \\ &= \frac{1}{x!} exp\{(log(\theta)x - \theta\} \\ &= h(x) exp\{\eta(\theta) T(x) - B(\theta)\} \end{aligned}$$

where:

- $h(x) = \frac{1}{x!}$
- $\eta(\theta) = log(\theta)$
- $T(x) = x$

**Canonical Representation**

- $\eta = log(\theta)$.
- $A(\eta) = B(\theta) = \theta = e^{\eta}$.

# MGFs of Canonical Exponenetial Family Models

**Theorem 1.6.2** Suppose $X$ is distributued according to a canonical exponential family, i.e., the density/pmf function is given by

$$p(x \mid \eta) = h(x)exp[\eta T(x) - A(\eta)], \text{ for } x \in \mathcal{X} \subset R^q.$$

If $\eta$ is an interior point of $\mathcal{E}$, the natural parameter space, then

- The moment generating function of $T(X)$ exists and is given by
$$M_T(s) = E[e^{sT(X)} \mid \eta] = exp\{A(s+\eta) - A(\eta)\}$$
for $s$ in some neighborhood of 0.
- $E[T(X) \mid \eta] = A'(\eta)$.
- $Var[T(X) \mid \eta] = A''(\eta)$.

**Proof:**

$$\begin{aligned} M_T(s) &= E[e^{sT(X)}) \mid \eta] = \int \cdots \int h(x) e^{(s+\eta)T(x) - A(\eta)} dx \\ &= [e^{[A(s+\eta)-A(s)]}] \times \int \cdots \int h(x) e^{(s+\eta)T(x) - A(s+\eta)} dx \\ &= [e^{[A(s+\eta)-A(s)]}] \times 1 \end{aligned}$$

Remainder follows from properties of MFGs.

# Moments of Canonical Exponential Family Distributions

**Poisson Distribution:** $A(\eta) = B(\theta) = \theta = e^{\eta}$.

$$E(X \mid \theta) = A'(\eta) = e^{\eta} = \theta.$$
$$Var(X \mid \theta) = A''(\eta) = e^{\eta} = \theta.$$

**Binomial Distribution:**

$$\begin{aligned} p(x \mid \theta) &= \binom{n}{x} \theta^{X} (1-\theta)^{n-x} \\ &= h(x) exp\{log(\tfrac{\theta}{(1-\theta)})x + nlog(1-\theta)\} \\ &= h(x) exp\{\eta x - nlog(e^{\eta}+1)\} \end{aligned}$$

So $A(\eta) = nlog(e^{\eta} + 1, )$ with $\eta = log(\frac{\theta}{1-\theta})$

$$A'(\eta) = n\frac{e^{\eta}}{e^{\eta}+1} = n\theta$$

$$\begin{aligned} A''(\eta) &= n\frac{1}{e^{\eta}+1} e^{\eta} + ne^{\eta} \times \frac{-1}{(e^{\eta}+1)^2} e^{\eta} \\ &= n[\frac{e^{\eta}}{e^{\eta}+1}] \times (1 - \frac{e^{\eta}}{(e^{\eta}+1)}) \\ &= n\theta(1-\theta) \end{aligned}$$

# Moments of the Gamma Distribution

$X_1, \ldots, X_n$ i.i.d $Gamma(p, \lambda)$ distribution with density

$$p(x \mid \lambda, p) = \frac{\lambda^p x^{p-1} e^{-\lambda x}}{\Gamma(p)}, \; 0 < x < \infty$$

where

$$\Gamma(p) = \int_0^\infty \lambda^p x^{p-1} e^{-\lambda x} dx$$

$$\begin{aligned} p(x \mid \lambda, p) &= [\tfrac{x^{p-1}}{\Gamma(p)}] exp\{-\lambda x + plog(\lambda)\} \\ &= h(x) exp\{\eta T(x) - A(\eta)\} \end{aligned}$$

where

- $\eta = -\lambda$
- $A(\eta) = -plog(\lambda) = -plog(-\eta)$

Thus

$$\begin{aligned} E(X) &= A'(\eta) = -p/\eta = p/\lambda \\ Var(X) &= A''(\eta) = (p/\eta^2) = p/\lambda^2 \end{aligned}$$

# Notes on Gamma Distribution

- $Gamma(p = n/2, \lambda = 1/2)$ corresponds to the Chi-Squared distribution with n degrees of freedom.
- $p = 2$ corresponds to the Exponential Distribution
- For $p = 1$, $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(p+1) = p\Gamma(p)$ for positive integer $p$.

**Rayleigh Distribution Sample** $X_1 \ldots, X_n$ iid with density function

$$\begin{aligned} p(x \mid \theta) &= \frac{x}{\theta^2} exp(-x^2/2\theta^2) \\ &= [x] \times exp\{-\frac{1}{2\theta^2}x^2 - log(\theta^2)\} \\ &= h(x) exp\{\eta T(x) - A(\eta)\} \end{aligned}$$

where

- $\eta = -\frac{1}{2\theta^2}$
- $T(X) = X^2$.
- $A(\eta) = log(\theta^2) = log(\frac{-1}{2\eta}) = -log(-2\eta)$

By the mgf

$$\begin{aligned} E(X^2) &= A'(\eta) = -\frac{2}{2\eta} = -\frac{1}{\eta} = 2\theta^2 \\ Var(X^2) &= A''(\eta) = +\frac{1}{\eta^2} = 4\theta^4 \end{aligned}$$

For the $n-$ sample: $\mathbf{X} = (X_1, \ldots, X_n)$

- $T(\mathbf{X}) = \sum_1^n X_i^2$
- $E[T(\mathbf{X})] = -n/\eta = 2n\theta^2$
- $Var[T(\mathbf{X})] = n\frac{1}{\eta^2} = 4n\theta^4$.

Note: $P(X \leq x) = 1 - exp\{-\frac{x^2}{2\theta^2}\}$ (Failure time model)

# Outline

1. Exponential Families
   - One Parameter Exponential Family
   - Multiparameter Exponential Family
   - Building Exponential Families

# Definition

$\{P_\theta, \theta \in \Theta\}$, $\Theta \subset R^k$, is a $k$-**parameter exponential family** if the the density/pmf function of $X \sim P_\theta$ is

$$p(x \mid \theta) = h(x) exp[\sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta)],$$

where $x \in \mathcal{X} \subset R^q$, and

- $\eta_1, \ldots, \eta_k$ and $B$ are real-valued functions mapping $\Theta \to R$.
- $T_1, \ldots, T_k$ and $h$ are real-valued functions mapping $R^q \to R$.

**Note:** By the Factorization Theorem (Theorem 1.5.1):

- $\mathbf{T}(X) = (T_1(X), \ldots, T_k(X))^T$ is sufficient.
- For a sample $X_1, \ldots, X_n$ iid $P_\theta$, the sample $\mathbf{X} = (X_1, \ldots, X_n)$ has a distribution in the $k$-parameter exponential family with *natural sufficient statistic*

$$T^{(n)} = (\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_1(X_n))$$

# Multiparameter Exponential Family Examples

**Example 1.6.5. Normal Family** $P_\theta = N(\mu, \sigma^2)$, with $\Theta = R \times R_+ = \{(\mu, \sigma^2)\}$ and density

$$p(x \mid \theta) = exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \; log(2\pi\sigma^2)\;\right)\right\}$$

a $k = 2$ multiparameter exponential family ($\mathcal{X} = R^1, q = 1$) and

- $\eta_1(\theta) = \frac{\mu}{\sigma^2}$ and $T_1(X) = X$
- $\eta_2(\theta) = \; -\frac{1}{2\sigma^2}$ and $T_2(X) = X^2$
- $B(\theta) = \frac{1}{2}(\frac{\mu^2}{\sigma^2} + \; log(2\pi\sigma^2))$
- $h(x) = 1$

**Note:**

- For an $n$-sample $\mathbf{X} = (X_1, \ldots, X_n)$ the *natural suffficient statistic* is $\mathbf{T}(\mathbf{X}) = (\sum_1^n X_i, \sum_1^n, X_i^2)$

# Canonical $k$-Parameter Exponential Family

Corresponding to

$$p(x \mid \theta) = h(x) exp[\sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta)],$$

consider

- *Natural Parameter:* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^T$
- *Natural Sufficient Statistic:* $\mathbf{T}(\mathbf{X}) = (T_1(X), \ldots, T_k(X))^T$
- Density function

  $$q(x \mid \boldsymbol{\eta}) = h(x) exp\{\mathbf{T}^T(x)\boldsymbol{\eta}) - A(\boldsymbol{\eta})\}$$

  where

  $$A(\boldsymbol{\eta}) = log \int \cdots \int h(x) exp\{\mathbf{T}^T(x)\boldsymbol{\eta}\} dx$$

  or

  $$A(\boldsymbol{\eta}) = log[\sum_{x \in \mathcal{X}} h(x) exp\{\mathbf{T}^T(x)\boldsymbol{\eta}\}]$$

- *Natural Parameter space:* $\boldsymbol{\mathcal{E}} = \{\boldsymbol{\eta} \in R^k : -\infty < A(\boldsymbol{\eta}) < \infty\}$.

# Canonical Exponential Family Examples ($k > 1$)

**Example 1.6.5. Normal Family (continued)** $P_\theta = N(\mu, \sigma^2)$, with $\Theta = R \times R_+ = \{(\mu, \sigma^2)\}$ and density

$$p(x \mid \theta) = exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + log(2\pi\sigma^2)\right)\right\}$$

- $\eta_1(\theta) = \frac{\mu}{\sigma^2}$ and $T_1(X) = X$
- $\eta_2(\theta) = -\frac{1}{2\sigma^2}$ and $T_2(X) = X^2$
- $B(\theta) = \frac{1}{2}(\frac{\mu^2}{\sigma^2} + log(2\pi\sigma^2))$ and $h(x) = 1$

**Canonical Exponential Density:**

$$q(x \mid \eta) = h(x)exp\{\mathbf{T}^T(x)\boldsymbol{\eta} - A(\boldsymbol{\eta})\}$$

- $\mathbf{T}^T(x) = (x, x^2) = (T_1(x), T_2(x))$
- $\boldsymbol{\eta} = (\eta_1, \eta_2)^T = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$
- $A(\boldsymbol{\eta}) = \frac{1}{2}[-\frac{\eta_1^2}{2\eta_2} + log(\pi \times \frac{-1}{\eta_2})]$
- $\mathcal{E} = R \times R_- = \{(\eta_1, \eta_2) : A(\boldsymbol{\eta}) \text{ exists}\}$

# Canonical Exponential Family Examples ($k > 1$)

**Multinomial Distribution**

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

$$p(x \mid \theta) = \frac{n}{x_1! \cdots x_q!} \text{ where}$$

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

Notes:

- What is $\Theta$?
- What is the dimensionality of $\Theta$
- What is the Multinomial distribution when $q = 2$?

# Example: Multinomial Distribution

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

where:

- $h(x) = \frac{n}{x_1! \cdots x_q!}$
- $\eta(\theta) = (\eta_1(\theta), \eta_2(\theta), \ldots, \eta_{q-1}(\theta))$
  $\eta_j(\theta) = log(\theta_j/(1 - \sum_1^{q-1} \theta_j)), \ j = 1, \ldots, q-1$
- $T(x) = (X_1, X_2, \ldots, X_{q-1}) = (T_1(x), T_2(x), \ldots, T_{q-1}(x))$.
- $B(\theta) = -nlog(1 - \sum_{j=1}^{q-1} \theta_j)$
- For the *canonical exponential* density:
  $A(\boldsymbol{\eta}) = +nlog(1 + \sum_{j=1}^{q-1} e^{\eta_j})$

# Outline

1. Exponential Families
   - One Parameter Exponential Family
   - Multiparameter Exponential Family
   - Building Exponential Families

# Building Exponential Families

**Definition: Submodels** Consider a $k$-parameter exponential family $\{q(x \mid \eta); \eta \in \mathcal{E} \subset R^k\}$. A **Submodel** is an exponential family defined by

$$p(x \mid \theta) = q(x \mid \eta(\theta))$$

where $\theta \in \Theta \subset R^{k^*}$, $k^* \leq k$, and $\eta : \Theta \to R^k$.

**Note:**

- The submodel is specified by $\Theta$.
- The natural parameters corresponding to $\Theta$ are a subset of the natural parameter space $\mathcal{E}^* = \{\eta \in \mathcal{E} : \eta = \eta(\theta), \theta \in \Theta\}$.

**Example**:$X$ is a discrete r.v.'s as $X$ with $\mathcal{X} = \{1, 2, \ldots, k\}$, and $X_1, X_2, \ldots, X_n$ are iid as $X$. Let $\mathcal{P}$ = set of distributions for $\mathbf{X} = (X_1, \ldots, X_n)$, where the distribution of the $X_i$ is a member of *any* fixed collection of discrete distributions on $\mathcal{X}$. Then

- $\mathcal{P}$ is exponential family (subset of Multinomial Distributions).

# Building Exponential Families

## Models from Affine Transformations: Case I

- Consider $\mathcal{P}$, the class of distributions for a r.v. $X$ which is a canonical family generated by the natural sufficient statistic $\mathbf{T}(X)$, a $(k \times 1)$ vector-statistic, and $h(\cdot) : \mathcal{X} \to R$. A distribution in $\mathcal{P}$ has density/pmf function:
$$p(X \mid \eta) = h(x) exp\{\mathbf{T}^T(x)\eta - A(\eta)\}$$
where
$$A(\eta) = log[\sum_{x \in \mathcal{X}} h(x) exp\{\mathbf{T}^T(x)\}]$$
or
$$A(\eta) = log[\int \cdots \int_{\mathcal{X}} h(x) exp\{\mathbf{T}^T(x)\} dx]$$
- $\mathbf{M}$: an affine tranformation form $R^k$ to $R^{k^*}$ defined by
$$\mathbf{M}(\mathbf{T}) = M\mathbf{T} + b,$$
where $M$ is $k^* \times k$ and $b$ is $k^* \times 1$, are known constants.

# Building Exponential Families

## Models from Affine Transformations (continued)

- Consider $\mathcal{P}^*$, the class of distributions for a r.v.$X$ generated by the natural sufficient statistic
$$M(\mathbf{T}(X)) = M\mathbf{T}(X) + b$$
- Since the distribution of $X$ has density/pmf: density/pmf function:
$$p(x \mid \eta) = h(x) exp\{\mathbf{T}^T(x)\eta - A(\eta)\}$$
we can write
$$\begin{aligned} p(x \mid \eta) &= h(x) exp\{[M(\mathbf{T}(x)]^T \eta^* - A^*(\eta^*)\} \\ &= h(x) exp\{[M\mathbf{T}(x) + b]^T \eta^* - A^*(\eta^*)\} \\ &= h(x) exp\{\mathbf{T}^T(x)[M^T \eta^*] + b^T \eta^* - A^*(\eta^*) \\ &= h(x) exp\{\mathbf{T}^T(x)[M^T \eta^*] - A^{**}(\eta^*) \end{aligned}$$
a subfamily pf $\mathcal{P}$ corresponding to
$$\Theta = \{\eta^* : \exists \eta \in \mathcal{E} : \eta = M^T \eta^*\}$$
- Density constant for level sets of $\mathbf{M}(\mathbf{T}(x))$

# Building Exponential Families

**Models from Affine Transformations: Case II**

- Consider $\mathcal{P}$, the class of distributions for a r.v. $X$ which is a canonical family generated by the natural sufficient statistic $\mathbf{T}(X)$, a $(k \times 1)$ vector-statistic, and $h(\cdot) : \mathcal{X} \to R$. A distribution in $\mathcal{P}$ has density/pmf function:
$$p(X \mid \eta) = h(x) exp\{\mathbf{T}^T(x)\eta - A(\eta)\}$$
- For $\Theta \subset R^{k^*}$, define
$$\eta(\theta) = B\theta \in \mathcal{E} \subset R^k,$$
where $B$ is a constant $k \times k^*$ matrix.
- The submodel of $\mathcal{P}$ is a submodel of the exponential family generated by $B^T\mathbf{T}(X)$ and $h(\cdot)$.

# Models from Affine Transformations

**Logistic Regression**. $Y_1, \ldots, Y_n$ are independent

$Binomial(n_i, \lambda_i)$, $i = 1, 2, \ldots, n$

**Case 1:** Unrestricted $\lambda_i$: $0 < \lambda_i < 1$, $i = 1, \ldots, n$

- $n$-parameter canonical exponential family
- $\mathcal{Y}_i = \{0, 1, \ldots, n_i\}$
- Natural sufficient statistic: $\mathbf{T}(Y_1, \ldots, Y_n) = \mathbf{Y}$.
- $h(\mathbf{y}) = \prod_{i=1}^{n} \binom{n_i}{y_i} \mathbf{1}(\{0 \leq y_i \leq n_i\})$
- $\eta_i = log(\frac{\lambda_i}{1-\lambda_i})$
- $A(\boldsymbol{\eta}) = \sum_{i=1}^{n} n_i log(1 + e^{\eta_i})$
- $p(\mathbf{y} \mid \boldsymbol{\eta}) = h(\mathbf{y}) exp\{\mathbf{Y}^T \boldsymbol{\eta} - A(\boldsymbol{\eta})\}$

# Logistic Regression (continued)

**Case 2:** For specified levels $x_1 < x_2 < \cdots < x_n$ assume
$\eta_i(\theta) = \theta_1 + \theta_2 x_i,\ i = 1, \ldots, n$ and
$\theta = (\theta_1, \theta_2)^T \in R^2$.

- $\boldsymbol{\eta}(\theta) = B\theta$, where $B$ is the $n \times 2$ matrix

$$B = [\mathbf{1}, \mathbf{x}] = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

- Set $M = B^T$, this is the 2-parameter canonical exonential family generated by

$$M\mathbf{Y} = (\sum_{i=1}^n Y_i, \sum_{i=1}^n x_i Y_i)^T \text{ and } h(\mathbf{y}) \text{ with}$$

$$A(\theta_1, \theta_2) = \sum_{i=1}^n n_i log(1 + exp(\theta_1 + \theta_2 x_i)).$$

# Logistic Regression (continued)

## Medical Experiment

- $x_i$ measures toxicity of drug
- $n_i$ number of animals subjected to toxicity level $x_i$
- $Y_i$ = number of animals dying out of the $n_i$ when exposed to drug at level $x_i$.
- Assumptions:
  - Each animal has a random toxicity threshold $X$ and death results iff drug level at or above $x$ is applied.
  - Independence of animals' response to drug effects.
  - Distribution of $X$ is *logistic*

$$P(X \leq x) = [1 + exp(-(\theta_1 + \theta_2 x))]^{-1}$$

$$log\left(\frac{P[X \leq x]}{1 - P(X \leq x)}\right) = \theta_1 + \theta_2 x$$

# Building Exponential Models

**Additional Topics**

- Curved Exponential Families, e.g.,
  - Gaussian with Fixed Signal-to-Noise Ratio
  - Location-Scale Regression
- Super models
  - Exponential structure preserved under random (iid) sampling