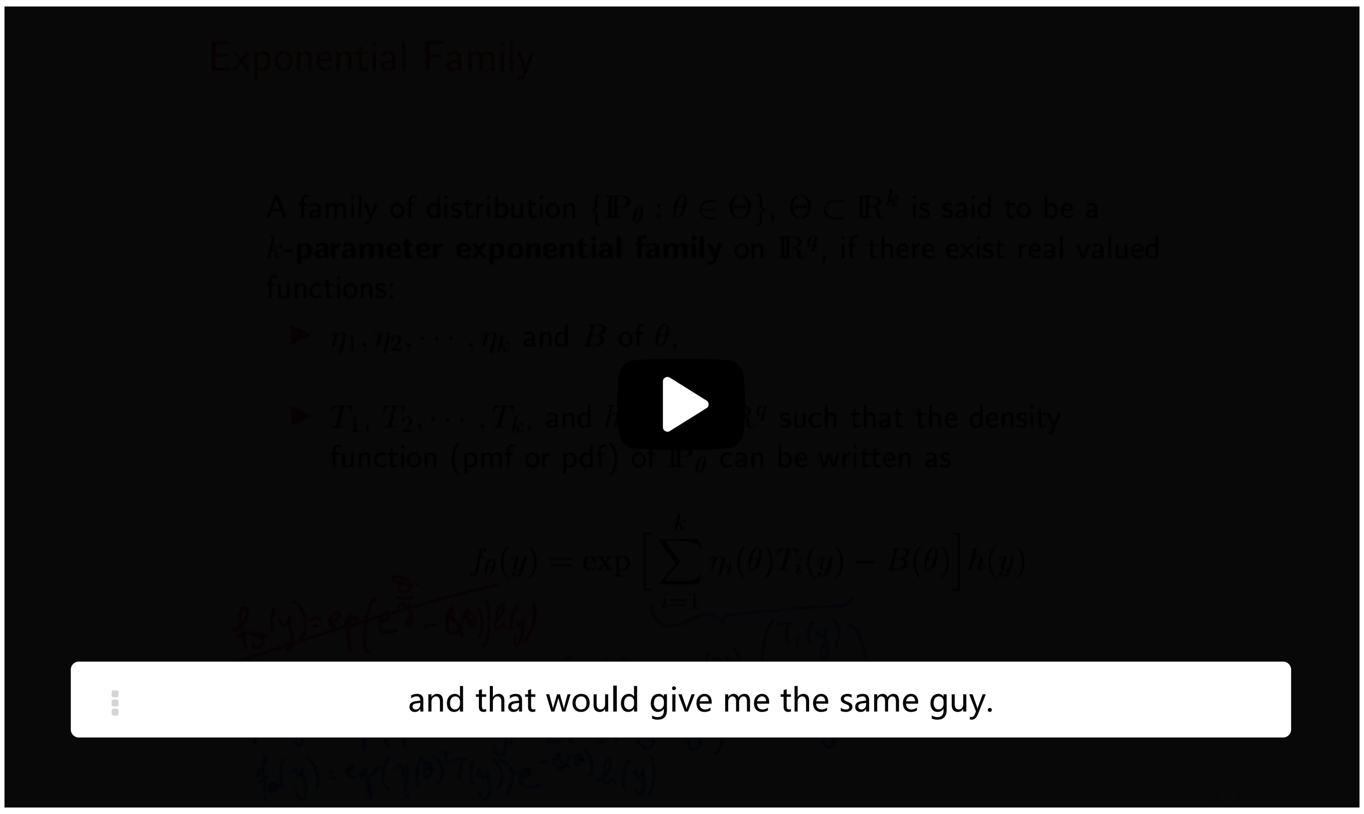
# 6. The Exponential Family

## Exponential Families: Definition




8:09 / 8:09 1.0x

to this very general family.

It's actually quite general, but it does not allow everything.

But the fact that I can take any function eta here

and any function t here just gives me a lot of flexibility.

It's not uniquely defined, right?

I could multiply my etas by 2 and divide my t by 2,

**and that would give me the same guy.**

End of transcript. Skip to the start.

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Recall from lecture that a family of distribution $\{\mathbf{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$, where the parameter space $\Theta \subset \mathbb{R}^k$ is $k$-dimensional, is called a ***k*-parameter exponential family** on $\mathbb{R}^1$ if the pmf or pdf $f_{\boldsymbol{\theta}} : \mathbb{R}^q \to \mathbb{R}$ of $\mathbf{P}_{\boldsymbol{\theta}}$ can be written in the form

$$f_{\boldsymbol{\theta}}(\mathbf{y}) = h(\mathbf{y}) \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(\mathbf{y}) - B(\boldsymbol{\theta})\right) \qquad \text{where} \begin{cases} \boldsymbol{\eta}(\boldsymbol{\theta}) = \begin{pmatrix} \eta_1(\boldsymbol{\theta}) \\ \vdots \\ \eta_k(\boldsymbol{\theta}) \end{pmatrix} & : \mathbb{R}^k \to \mathbb{R}^k \\ \mathbf{T}(\mathbf{y}) = \begin{pmatrix} T_1(\mathbf{y}) \\ \vdots \\ T_k(\mathbf{y}) \end{pmatrix} & : \mathbb{R}^q \to \mathbb{R}^k \\ B(\boldsymbol{\theta}) & : \mathbb{R}^k \to \mathbb{R} \\ h(\mathbf{y}) & : \mathbb{R}^q \to \mathbb{R}. \end{cases}$$

When $k = 1$, this reduces to

$$f_\theta(y) = h(y) \exp\left(\eta(\theta) T(y) - B(\theta)\right).$$

**Note:** The following exercises are similar to what will be presented in lecture, but we encourage you to first attempt these yourselves.

## Practice: Decomposing the exponent

4/4 points (graded)

For the two following pmfs with one parameter $\theta$ that are written in the form

$$f_\theta(y) = h(y)\, e^{w(\theta,y)},$$

first decompose $w(\theta, y)$ as

$$w(\theta, y) = \eta(\theta)\, T(y) - B(\theta),$$

then enter the product $\eta(\theta)\, T(y)$ below. Select the distribution that $f_\theta$ defines.

1. For $f_\theta(y) = e^{w(\theta,y)}$ where

$$w(\theta, y) = y \ln(\theta) + (1-y)\ln(1-\theta)$$

and $y = 0, 1,\ \theta \in (0,1)$ :

$\eta(\theta)\, T(y) =$ y*(ln(theta)-ln(1-theta)) ✔ **Answer:** y*(ln(theta)-ln(1-theta))

$y \cdot (\ln(\theta) - \ln(1-\theta))$

What distribution does the pmf $f_\theta(y)$ define?

- ○ $\mathcal{N}(\theta, 1)$
- ○ $\mathcal{N}(1, \theta)$
- ◉ $\text{Ber}(\theta)$ ✔
- ○ $\text{Poiss}(\theta)$
- ○ none of the above

2. For $f_\theta(y) = \frac{1}{y!} e^{w(\theta,y)}$ where $w(\theta, y) = -\theta + y\ln(\theta)$, and $y = 0, 1, 2, \ldots,\ \theta \in (0,1)$:

$\eta(\theta)\, T(y) =$ y*ln(theta) ✔ **Answer:** y*ln(theta)

$y \cdot \ln(\theta)$

What distribution does the pmf $f_\theta(y)$ define?

- ○ $\mathcal{N}(\theta, 1)$
- ○ $\mathcal{N}(1, \theta)$
- ○ $\text{Ber}(\theta)$
- ◉ $\text{Poiss}(\theta)$ ✔
- ○ none of the above

STANDARD NOTATION

**Solution:**

1. For $f_\theta(y) = e^{w(\theta,y)}$ where $w(\theta, y) = y\ln(\theta) + (1-y)\ln(1-\theta)$ and $y \in \{0, 1\},\ \theta \in (0,1)$:

$$w(\theta, y) = y\ln(\theta) + (1-y)\ln(1-\theta) = y(\ln(\theta) - \ln(1-\theta)) + \ln(1-\theta)$$

Hence, $\eta(\theta) T(y) = y(\ln(\theta) - \ln(1-\theta))$ and $B(\theta) = -\ln(1-\theta)$. Rewriting $f_\theta$:

$$f_\theta(y) = e^{y \ln(\theta) + (1-y)\ln(1-\theta)} = \theta^y (1-\theta)^{(1-y)},$$

we see that $f_\theta$ is the pmf of a Bernoulli distribution with parameter $\theta$.

2. For $f_\theta(y) = \frac{1}{y!} e^{w(\theta,y)}$ where $w(\theta, y) = -\theta + y\ln(\theta)$, and $y = 0, 1, 2, \ldots$, $\theta \in (0,1)$ Hence, $\eta(\theta) T(y) = y \ln(\theta)$ and $B(\theta) = \theta$. Rewriting $f_\theta$

$$f_\theta(y) = \frac{1}{y!} e^{-\theta + y\ln(\theta)} = e^{-\theta} \frac{\theta^y}{y!},$$

we recognize $f_\theta$ as the pmf of a Poisson distribution with parameter $\theta$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

## Practice: Normal distribution with known variance

1/1 point (graded)
The normal distribution $\mathcal{N}(\theta, 1)$ with with mean $\theta$ and known variance $\sigma^2 = 1$ has pdf

$$f_\theta(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2}}.$$

Rewrite $f_\theta$ in the form

$$f_\theta(y) = h(y) e^{\eta(\theta) T(y) - B(\theta)} \qquad \text{where } \eta(\theta), T(y) : \mathbb{R} \to \mathbb{R},$$

and enter the product $\eta(\theta) T(y)$ below.

$\eta(\theta) T(y) =$ theta*y ✔ **Answer:** y*theta

$\theta \cdot y$

STANDARD NOTATION

**Solution:**

$$\begin{aligned} f_\theta(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y^2 - 2y\theta + \theta^2)}{2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{+\frac{y\theta - \theta^2}{2}} \\ &= h(y) e^{\eta(\theta) T(y) - B(\theta)} \qquad \text{where} \begin{cases} \eta(\theta) T(y) & = (y)(\theta) \\ B(\theta) & = \frac{\theta^2}{2} \\ h(y) & = (e^{-\frac{y^2}{2}})/\sqrt{2\pi} \end{cases} \end{aligned}$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

## Discussion

Show Discussion