Secs. 12.7 – 12.9

Families of Distributions

Individual Distributions versus Families

- Here is **one** distribution
  (**not** a family of distributions): $X \sim Exp(3)$

But if we let the parameter vary, then we have a parameterized family of distributions.

- $X \sim Exp(\lambda)$ where $\lambda > 0$
- $X \sim \text{Normal}(3, \sigma^2)$ where $\sigma^2 > 0$
- $X \sim \text{Normal}(\theta, 5)$ where $-\infty < \theta < \infty$
- Of course, we can have a
  family of dist'ns with multiple parameters
  $X \sim \text{Normal}(\theta, \sigma^2)$ where $-\infty < \mu < \infty$, $\sigma^2 > 0$

Caution about notation in parameterization

Here are two different ways to think about the parameterization of an exponential distribution:

pdf	name
$f(x \mid \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & x \geq 0,\ \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$	$X \sim Exp(\beta)$
$f(x \mid \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0,\ \lambda > 0 \\ 0 & \text{elsewhere} \end{cases}$	$X \sim Exp(\lambda)$

Do not count on the distinction between the letters β and λ to distinguish between these types.

Additional cautions

Other distributions where different versions of the parameterization are often used include

- Gamma
- Geometric
- Negative Binomial

Notation: A vector of parameters

When discussing results in theoretical statistics, sometimes we think of the several parameters as a vector of parameters.

For example, instead of $N(\mu, \sigma^2)$ we might use $N(\boldsymbol{\theta})$ where we think of this parameter as a vector: $\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$

Here the bold type for theta indicates that it denotes a vector.

In handwriting, we can't effectively indicate bold type: the standard handwriting notation for this vector is $\underset{\sim}{\theta}$ (a "tilde" beneath it.)

Notation: Indicator Functions

In theoretical work with statistics, sometimes we want to factor the pdf. When we do that, it is convenient to eliminate the notation of showing the function "in pieces." We do that with indicator functions. Here's the notation

$$I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Example 1: Binomial Distribution

$$P(X = x \mid p) = \begin{cases} p^x (1-p)^{1-x} & \text{for } x \in \{0, 1, \ldots, n\} \text{ and } p \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

is written as

$$P(X = x \mid p) = p^x (1-p)^{1-x} I_{\{0,1,\ldots,n\}}(x) I_{[0,1]}(p)$$

Example 2: Uniform Distribution 1

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

is written as

$$f(x \mid a,b) = \frac{1}{b-a}\; I_{[a,b]}(x)$$

or, if it is useful to have the parameters in separate factors,

$$f(x \mid a,b) = \frac{1}{b-a}\; I_{[a,\infty)}(x) I_{(-\infty,b]}(x)$$

Example: Uniform Distribution 2

$$f(x \mid a,b) = \frac{1}{b-a} \; I_{[a,\infty)}(x) I_{(-\infty,b]}(x)$$

If we are focusing on this as a function of the parameters (as we do when we think of a Likelihood function) then

$a \le x$ means $x \ge a$ and $x \le b$ means $b \ge x$, so we have

$$L(a,b \mid x) = \frac{1}{b-a} \; I_{[x,\infty)}(b) I_{(-\infty,x]}(a)$$

- New section

- Exponential Families

Types of Families

All probability distributions have certain properties in common, which you have studied.

When we talk about a family of dist'ns, those have additional properties in common.
We often think of the properties of a dist'n in the $N(\mu, \sigma^2)$ family.

This is a particular parameterized family.

Exponential Families

Now we will look at a different way of characterizing a family of dist'ns.

Part 1. Simplest form

Part II. More complex form

Exponential Families, Part I

A family of pdfs or pmfs is a full
one-parameter exponential family
if it can be expressed in the following form:

$$f(x \mid \theta) = h(x)c(\theta) \cdot \exp\left(\omega(\theta) t(x) \right)$$

where h and t are real-valued functions
of the observation x which don't depend on θ
and
c and ω are real-valued functions of θ
which don't depend on the observation x.
By convention, we are careful to define
both $h(x) > 0$ and $c(x) > 0$.

Exponents and Logarithms

Notice that this factorization requires that we have an exponential function as part of it.

Algebra Review: Recall the various logarithm rules, which we can use to find this result:

$$e^{n \cdot \log_e m} = e^{\log_e m^n} = m^n .$$

Thus, if we have a factor of the pdf which includes the data and the parameter with one the "base" and one in the "exponent" then we can use this to put them into a useful exponential function.

Example 1. Pareto Dist'n

Consider the Pareto family

$$f(x \mid \alpha, \beta) = \frac{\beta \alpha^{\beta}}{x^{\beta+1}} \text{ for } \alpha < x < \infty, \;\; \alpha > 0, \; \beta > 0$$

a. If we fix the parameter α and so there is only one parameter β, is this an exponential family?

b. If we fix the parameter β, so there is only one parameter α, is this an exponential family?

Solution, beginning

Preparatory work:

Algebraic work	$\frac{\beta\alpha^{\beta}}{x^{\beta+1}} = \frac{\beta\alpha^{\beta}}{x^{\beta}} x^{-1} = \beta\alpha^{\beta} x^{-1} \cdot x^{-\beta}$ $= \beta\alpha^{\beta} x^{-1} \cdot e^{-\beta \log x}$
Complete with indicator functions	$= \beta\alpha^{\beta} x^{-1} \cdot e^{-\beta \log x} I_{(\alpha,\infty)}(x) I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta)$

Solution: continued

$$\beta\alpha^{\beta}x^{-1}\cdot e^{-\beta\log x}I_{(\alpha,\infty)}(x)I_{(0,\infty)}(\alpha)I_{(0,\infty)}(\beta)$$

Part a. Assume α is constant.

First we look at the part in the exponent

	$\omega(\beta)=-\beta$
	$t(x)=\log x$

Then we look at the part which is not in the exponent.

$h(x)=x^{-1}I_{(\alpha,\infty)}(x)$	$\omega(\beta)=-\beta$
$c(\beta)=\beta\alpha^{\beta}I_{(0,\infty)}(\alpha)\cdot I_{(0,\infty)}(\beta)$	$t(x)=\log x$

Part a. Conclusion: Yes, this is an exponential family, with the above functions.

Choices 1

$$\beta\alpha^{\beta}x^{-1}\cdot e^{-\beta\log x}I_{(\alpha,\infty)}(x)I_{(0,\infty)}(\alpha)I_{(0,\infty)}(\beta)$$

Note, this is not the unique way to assign the four functions. Here are some possible changes.

1. Because α is a constant here, not a parameter, the indicator function for it could be assigned to either function h or c.
2. Because the exponent has a negative sign, that could be assigned to either function in that exponent.

Choices 2

$$\beta\alpha^{\beta}x^{-1}\cdot e^{-\beta\log x}I_{(\alpha,\infty)}(x)I_{(0,\infty)}(\alpha)I_{(0,\infty)}(\beta)$$

3. The factor of x^{-1} was factored out and so it was left out of the exponent. That wasn't necessary. It would have been acceptable to not do that and then the ω function would be $\omega(\beta)=-(\beta+1)$
4. From an algebraic point of view, it would have been possible for various functions to have simply been the constant 1, if there wasn't a need for that part. But, to show it IS an exponential family, one would need to put some of the pdf into the exponent part.

Solution to part b.

$$\beta\alpha^{\beta}x^{-1}\cdot e^{-\beta\log x}I_{(\alpha,\infty)}(x)I_{(0,\infty)}(\alpha)I_{(0,\infty)}(\beta)$$

$h(x)=x^{-1}I_{(\alpha,\infty)}(x)$	$\omega(\beta)=-\beta$
$c(\beta)=\beta\alpha^{\beta}I_{(0,\infty)}(\alpha)\cdot I_{(0,\infty)}(\beta)$	$t(x)=\log x$

Part b. Assume β is a known constant.

Conclusion.
The algebraic part of the solution starts in the same way.
However, we immediately notice that it is not possible to separate the factors appropriately into the h and c functions, because one of the required indicator functions is a function of BOTH the data and the parameter.

When that is true, the family of dist'ns is NOT an exponential family.

New section   Exponential Families Part II

Exponential Families, Part II

Exponential families Part II

Now, suppose we have multiple parameters. Here, we denote the multiple parameters as a vector $\boldsymbol{\theta}$

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \cdot \exp \sum_{j=1}^{k} \left(\omega_j(\boldsymbol{\theta}) t_j(x) \right)$$

where h and t are real-valued functions of the observation x which don't depend on $\boldsymbol{\theta}$
and
c and ω are real-valued functions of $\boldsymbol{\theta}$ which don't depend on the observation x
and, by convention, we are careful to define both $h(x) > 0$ and $c(\boldsymbol{\theta}) > 0$.

Make it simple!

In the simplest case of this kind, the value k
giving the number of terms necessary in the summation
is no larger than the number of parameters
(the length of the parameter vector.)

We always want to minimize the k we use.
To use a larger k than necessary restricts the
usefulness of the result.

Example: Gamma Dist'n

Stop the video and do this on your own.

Show that $\text{Gamma}(\alpha, \beta)$ is an exponential family and can be written in the with $k = 2$.

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \text{ for } 0 \le x < \infty,\ \alpha > 0,\ \beta > 0$$

Solution: Gamma Dist'n

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \text{ for } 0 \le x < \infty,\ \alpha > 0,\ \beta > 0$$

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{(\alpha-1)\log x} e^{-x/\beta} I_{[0,\infty)}(x) I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta)$$

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{(\alpha-1)\log x - x/\beta} I_{[0,\infty)}(x) I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta)$$

Solution: Gamma Dist'n 2

$$f(x|\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}e^{(\alpha-1)\log x-x/\beta}I_{[0,\infty)}(x)I_{(0,\infty)}(\alpha)I_{(0,\infty)}(\beta)$$

First we look at the part in the exponent

	$\omega_1(\alpha,\beta)=\alpha-1$	$\omega_2(\alpha,\beta)=1/\beta$
	$t_1(x)=\log x$	$t_2(x)=-x$

Then we look at the part which is not in the exponent.

$h(x)=I_{[0,\infty)}(x)$	
$c(\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}I_{(0,\infty)}(\alpha)\cdot I_{(0,\infty)}(\beta)$	

Example: Binomial Dist'n

Turn off your video and do this problem.

Show that $\mathrm{Bin}(n, p)$ with n a given constant,
is an exponential family and can be written with $k = 1$.

(Hint: To do that, you'll need to combine two factors into one factor in order to have an expression where x occurs in only one exponent instead of two exponents.)

$$f(x \mid n, p) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad \text{for} \quad x \in \{0, 1, \ldots, n\}, 0 \leq p \leq 1$$

Binomial Dist'n solution page 1

$$f(x|n,p)=\binom{n}{x}p^x(1-p)^{(n-x)}\ \text{for}\ x\in\{0,1,\ldots,n\},0\le p\le 1$$

$$f(x|n,p)=\binom{n}{x}p^x(1-p)^n(1-p)^{-x}I_{\{0,1,\ldots,n\}}(x)\cdot I_{[0,1]}(p)$$

$$f(x|n,p)=\binom{n}{x}p^x\frac{1}{(1-p)^x}(1-p)^nI_{\{0,1,\ldots,n\}}(x)\cdot I_{[0,1]}(p)$$

$$f(x|n,p)=\binom{n}{x}\frac{p^x}{(1-p)^x}(1-p)^nI_{\{0,1,\ldots,n\}}(x)\cdot I_{[0,1]}(p)$$

$$f(x|n,p)=\binom{n}{x}\left(\frac{p}{1-p}\right)^x(1-p)^nI_{\{0,1,\ldots,n\}}(x)\cdot I_{[0,1]}(p)$$

$$f(x|n,p)=\binom{n}{x}\exp\left(x\log\left(\frac{p}{1-p}\right)\right)(1-p)^nI_{\{0,1,\ldots,n\}}(x)\cdot I_{[0,1]}(p)$$

Binomial Dist'n Solution page 2

$$f(x \mid n, p) = \binom{n}{x} \exp\left(x \log\left(\frac{p}{1-p} \right) \right) (1-p)^n I_{\{0,1,\ldots,n\}}(x) \cdot I_{[0,1]}(p)$$

First we look at the part in the exponent

	$\omega(p) = \log\left(\frac{p}{1-p}\right)$
	$t(x) = x$

Then we look at the part which is not in the exponent.

$h(x) = \binom{n}{x} I_{\{0,1,\ldots,n\}}(x)$	$\omega(p) = \log\left(\frac{p}{1-p}\right)$
$c(p) = (1-p)^n I_{[0,1]}(p)$	$t(x) = x$

How are exponential families useful?

- Some shortcuts for computing moments. (Replacing integration with differentiation)
- Finding conjugate priors in Bayesian statistics
- Some results in the theory of estimation about finding sufficient statistics and finding unique best estimators of parameters

Additional ideas

- In some applications, the "natural parameter" is relevant. The natural parameter is the function $\omega_i(\theta)$ of the parameter (when you have made k as small as possible.)
- The "sufficient statistic" for the parameter(s) is based on the function(s) $t_i(\theta)$.

Different “levels”

Some useful results depend on the exponential family having the value *k* being the same dimension as the parameter vector.

Other results can be proved even when *k* is larger than the dimension of the parameter vector.

These are described in two different ways. Exponential families where the value of *k* is the same as the dimension of the parameter vector are called “full exponential families” or “families where the parameter space contains an open set in $\Re^k$.”