STAT 501 Fall 2022 Course Notes

Chapter 3: Common Families of Distributions (Sections 3.4-3.6)

3.4 Exponential Families

Families of pmfs or pdfs that belong to the exponential family have properties that make statistical inference in both the frequentist and Bayesian context simpler mathematically!

DEF 3.4.1 (Exponential Family) A family of pdfs or pmfs is called an exponential family if it can be expressed as:

$$f(x|\theta) = h(x)c(\boldsymbol{\theta})exp\left\{\sum_{i=1}^{k} t_i(x)w_i(\boldsymbol{\theta})\right\},$$

where $h(x) \geq 0$ and $t_1(x), t_2(x), ..., t_k(x)$ are real-valued functions of x alone (cannot depend on $\boldsymbol{\theta}$) and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ alone (cannot depend on x).

Notes:

- To verify that a family of pdfs or pmfs is an exponential family, we must identify the functions $h(x), c(\boldsymbol{\theta}), w_i(\boldsymbol{\theta})$, and $t_i(x)$ and show that the family has the form provided above in DEF 3.4.1.
- A **curved exponential family** is a family of densities of the form provided in DEF 3.4.1 for which the dimension of the vector $\boldsymbol{\theta}$ is $d < k$. If $d = k$ the family is a **full exponential family**.

Examples Determine whether the family of pdfs/pmfs is an exponential family...

- Let $X \sim Binomial(n, p)$. Assume $0 < p < 1$.

- Let $X \sim Normal(\mu, \sigma^2)$. Assume $\mu \in (-\infty, \infty)$ and $\sigma^2 > 0$.

- Let $X \sim Pareto(\alpha, \beta)$. Then $f_X(x|\alpha, \beta) = \dfrac{\beta\alpha^\beta}{x^{\beta+1}} I_{(\alpha,0)}(x)$; α, $\beta > 0$.

3.5 Location and Scale Families

We've discussed common families of continuous distributions, and in this section, we investigate different techniques for constructing families of distributions that are useful for modeling. These families include location families, scale families, and location-scale families, and the general process for constructing each of these three types of families is similar: specify a standard (or reference) pdf, and then transform that pdf in a specified way.

THM 3.5.1 (Theorem about PDFs) Let $f_X(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function $g(x|\mu, \sigma) = \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$ is a pdf.

Proof: On your own (see p. 116 in text)

THM 3.5.2 (Location Family) Let $f_X(x)$ be any pdf. Then the family of pdfs $f_X(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the **location family** with the standard pdf $f_X(x)$ and μ is called the location parameter for the family. That is, the standard (or reference) pdf is where $\mu = 0$.

Note: The location parameter μ shifts the graph of the pdf along the x-axis while keeping the same shape.

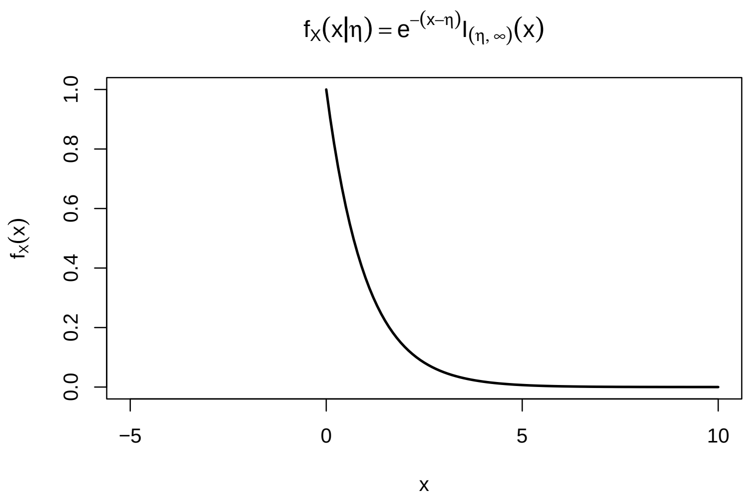
Examples Consider the two pdfs and the graphs of their standard pdf below. What does the parameter η represent for each distribution? What happens to the pdf if η is shifted?

- Exponential distribution with $\beta = 1$ and location parameter η:

$$f_X(x|\eta) = e^{-(x-\eta)} I_{(\eta,\infty)}(x)$$

  Standard pdf:

```
curve(dexp(x, rate = 1), from = 0, to = 10, lwd = 2,
  xlim = c(-5,10), ylab = expression(f[X](x)),
  main = latex2exp::TeX(
    "$f_X(x|\\eta) = e^{-(x-\\eta)}I_{(\\eta, \\infty)}(x)$"))
```



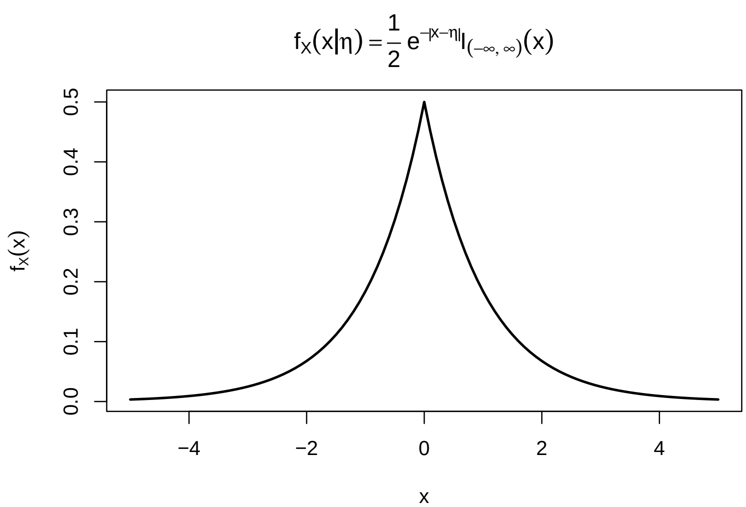
```r
# Add pdfs with differing values of eta
curve(dexp(x-2, rate = 1), from = 2, to = 10,
  lty = 2, lwd = 2, add = TRUE)
curve(dexp(x+3, rate = 1), from = -3, to = 10,
  lty = 3, lwd = 2, add = TRUE)
legend("topright",
   c(expression(paste(eta," = -3")),
     expression(paste(eta," = 0")),
     expression(paste(eta," = 2"))),
   lty = c(3, 1, 2), lwd = c(1, 1, 1))
```

- Double exponential distribution with $\mu = \eta$ and $\sigma = 1$:

$$f_X(x|\eta) = \frac{1}{2}e^{-|x-\eta|}I_{(-\infty,\infty)}(x)$$

Standard pdf:

```r
curve(extraDistr::dlaplace(x), from = -5, to = 5, lwd = 2,
  ylab = expression(f[X](x)),
  main = latex2exp::TeX(
  "$f_X(x|\\eta) = \\frac{1}{2}e^{-abs(x-\\eta)}I_{(-\\infty, \\infty)}(x)$"))
```

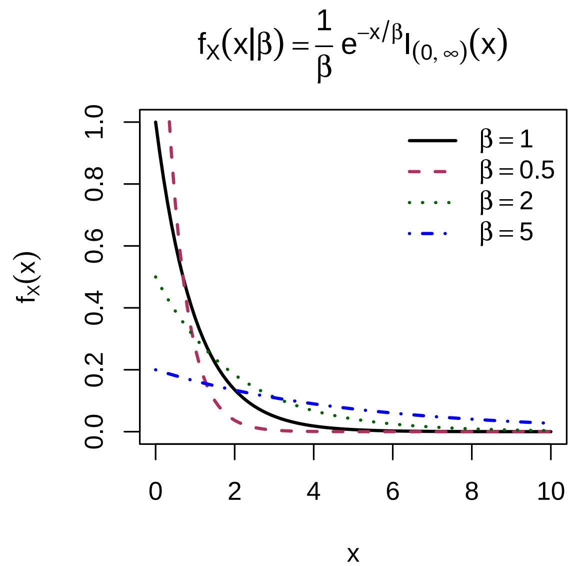


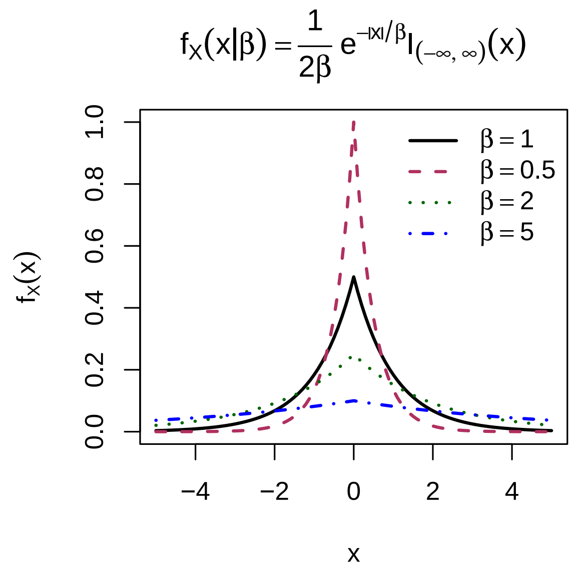

```r
# Add pdfs with differing values of eta
curve(extraDistr::dlaplace(x-2),
  lty = 2, lwd = 2, add = TRUE)
curve(extraDistr::dlaplace(x+3),
  lty = 3, lwd = 2, add = TRUE)
legend("topright",
   c(expression(paste(eta," = -3")),
     expression(paste(eta," = 0")),
     expression(paste(eta," = 2"))),
   lty = c(3, 1, 2), lwd = c(1, 1, 1))
```

What are some commonly used location families?

THM 3.5.4 (Scale Family) Let $f_X(x)$ be any pdf. Then the family of pdfs $\frac{1}{\sigma} f_X\left(\frac{x}{\sigma}\right)$, indexed by the parameter σ, $\sigma > 0$, is called the **scale family** with the standard pdf $f_X(x)$ and σ is called the scale parameter for the family. That is, the standard (or reference) pdf is where $\sigma = 1$.

Note: The scale parameter σ either stretches ($\sigma > 1$) or contracts ($\sigma < 1$) the graph of $f_X(x)$, while keeping the same basic shape and the center the same.





What are some examples of commonly used scale families?

THM 3.5.6 (Generating location, scale, and location-scale families) Let μ be any real number and let σ be any positive real number Then, X is a random variable with pdf $f_X(x) = \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$ if and only if there exists a random variable Z with pdf $f_Z(z)$ and $X = \sigma Z + \mu$.

Proof: On your own (see p. 120 in text)

Take home message: We can start with *any* pdf $f_X(x)$ and generate a family of pdfs by introducing a location and/or scale parameter in the manner described above! Essentially, we can generate location, scale, and location-scale families by subjecting a random variable with standard pdf to a family of transformations.

3.6 Inequalities and Identities

One of the most famous inequalities is introduced at the end of this chapter, Chebychev's Inequality. We will cover several other useful inequalities involving more than one random variable at the end of Chapter 4, including the Cauchy-Schwarz Inequality and Jensen's Inequality.

THM 3.6.1 (Chebychev's Inequality) Let X be a random variable and let $g(x)$ be a non-negative function. Then for any $r > 0$,

$$P(g(X) \geq r) \leq \frac{E[g(X)]}{r}$$

Proof:

Example A special case of Chebychev's Inequality involves the following non-negative function: $g(x) = \dfrac{(x-\mu)^2}{\sigma^2}$, where $\mu = E(X)$ and $\sigma^2 = Var(X)$.

Note: Chebychev's Theorem (particularly in its classic expression shown above) enables us to find bounds for probabilities that may be difficult or tedious to obtain. Often, we can also use it to obtain means and variances of random variables without specifying the distribution of the variable.

Example Experience has shown that the length of time X (in minutes) required to conduct a periodic maintenence check on a machine follows a gamma distribution with $\alpha = 3.1$ and $\beta = 2$. A new maintenance worker takes 22.5 minutes to check the machine. Does this length of time to perform the check seem unusually long?

Stein's Lemma (Lemma 3.6.5) Let $X \sim N(\mu, \sigma^2)$ and let g be a differentiable function satisfying $E[|g'(X)|] < \infty$. Then,

$$E[g(X)(X - \mu)] = \sigma^2 E[g'(X)].$$

- Note: This lemma is most useful for calculating higher-order moments.
- Proof: On your own (see p. 124 in text)

Example: Let $X \sim N(\mu, \sigma^2)$. Find $E(X^3)$.