# Lecture 29: Nov. 7

## Last time

- Presentations
- Exponential families

## Today

- Exponential families
- Location and Scale families
- Chebychev's Inequality

**Exponential Families** A family of pdfs or pmfs with vector parameter $\boldsymbol{\theta}$ is called an *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right), \quad x \in S \subset \mathbb{R} \tag{1}$$

where $S$ is not defined in terms of $\boldsymbol{\theta}$, $h(x)$, $c(\boldsymbol{\theta}) \geqslant 0$ and the functions are just functions of the parameters specified; i.e. $h$ is free of $\boldsymbol{\theta}$, $c(\boldsymbol{\theta})$ is free of $x$, etc...

**Theorem** If $X$ is a random variable with pdf or pmf of the form 1, then

$$\begin{aligned}
E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) &= -\frac{\partial}{\partial \theta_j} \log c(\boldsymbol{\theta}) \\
Var\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) &= -\frac{\partial^2}{\partial \theta_j^2} \log c(\boldsymbol{\theta}) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)\right).
\end{aligned}$$

Although these equations may look formidable, when applied to specific cases they can work out quite nicely. Their advantage is that we can replace integration or summation by differentiation, which is often more straightforward.

**Example (Normal exponential family)** Let $f(x|\mu, \sigma^2)$ be the $n(\mu, \sigma^2)$ family of pdfs, where $-\infty < \mu < \infty, \sigma > 0$. Then

$$\begin{aligned}
f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right)
\end{aligned}$$

Define

$$\theta_1 = \tfrac{1}{\sigma^2} > 0, \quad \theta_2 = \tfrac{\mu}{\sigma^2} \in \mathbb{R}$$

Then

$$f_X(x) = \frac{\sqrt{\theta_1}}{\sqrt{2\pi}} \exp\left(-\frac{\theta_2^2}{2\theta_1}\right) \exp\left(-\theta_1 \frac{x^2}{2} + \theta_2 x\right)$$

and

$$\begin{aligned}
&h(x) = 1 \text{ for all } x; \\
&c(\boldsymbol{\theta}) = c(\theta_1, \theta_2) = \tfrac{\sqrt{\theta_1}}{\sqrt{2\pi}} \exp\left(-\tfrac{\theta_2^2}{2\theta_1}\right), && (\theta_1, \theta_2) \in (0, \infty) \times \mathbb{R} \\
&w_1(\boldsymbol{\theta}) = \theta_1 && t_1(x) = -x^2/2 \\
&w_2(\boldsymbol{\theta}) = \theta_2 && t_2(x) = x
\end{aligned}$$

Therefore, by the above theorem

$$\begin{aligned}
E(X) &= -\frac{\partial}{\partial \theta_2} \log c(\boldsymbol{\theta}) = \frac{\theta_2}{\theta_1} = \mu \\
Var(X) &= -\frac{\partial^2}{\partial \theta_2^2} \log c(\boldsymbol{\theta}) = -\frac{1}{\theta_1} = \sigma^2
\end{aligned} \tag{2}$$

## Location and Scale families

Let $Z$ be a continuous random variable with pdf $f(z)$. Define the class of rvs

$$X_{\mu,\sigma} = \sigma Z + \mu, \quad \mu \in \mathbb{R}, \sigma > 0$$

Then

1. $X_{\mu,\sigma}$ has pdf
$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$
2. $$E(X) = \sigma E(Z) + \mu, \quad Var(X) = \sigma^2 Var(Z)$$
3. The variable $Z = X_{0,1}$ is called the *generator* and is a member of the class.

### Location families and scale families

- The family of pdfs $f_{\mu,\sigma}(x)$ is called a *location-scale* family where $\mu$ is called the *location parameter*, and $\sigma$ is called the *scale parameter*.
- The family of pdfs
$$f_{\mu,1}(x) = f(x - \mu)$$
with $\sigma = 1$ is called a *location* family.
- The family of pdfs
$$f_{0,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$
with $\mu = 0$ is called a *scale* family.

Example (Exponential location family) Let $f(x) = e^{-x}, x \geqslant 0$, and $f(x) = 0, x < 0$. To form a location family we replace $x$ with $x - \mu$ to obtain

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x - \mu \geqslant 0 \\ 0 & x - \mu < 0 \end{cases}$$
$$= \begin{cases} e^{-(x-\mu)} & x \geqslant \mu \\ 0 & x < \mu \end{cases}$$

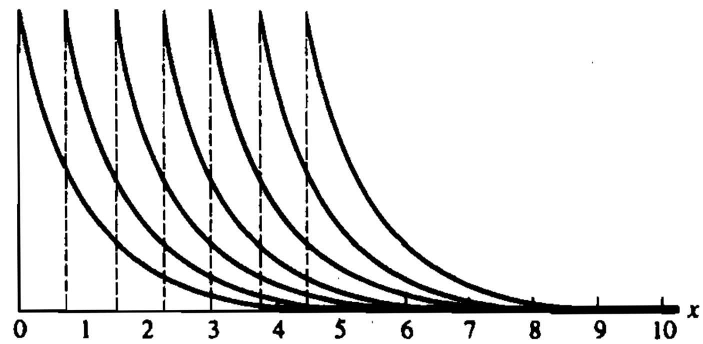


Figure 3.5.2. *Exponential location densities*

Figure 29.1: Figure 3.5.2. Exponential location densities.

As shown in the above graph, the densities are shifted. Now the positive part of the density starts at $\mu$ rather than at 0. If $X$ measures time, then $\mu$ might be restricted to be nonnegative so that $X$ will be positive with probability 1 for every value of $\mu$. In this type of model, where $\mu$ denotes a bound on the range of $X$, $\mu$ is sometimes called a *threshold parameter*.

The effect of introducing the scale parameter $\sigma$ is either to stretch ($\sigma > 1$) or to contract ($\sigma < 1$) the graph of $f(x)$ while still maintaining the same basic shape of the graph. This is illustrated in the Figure below.

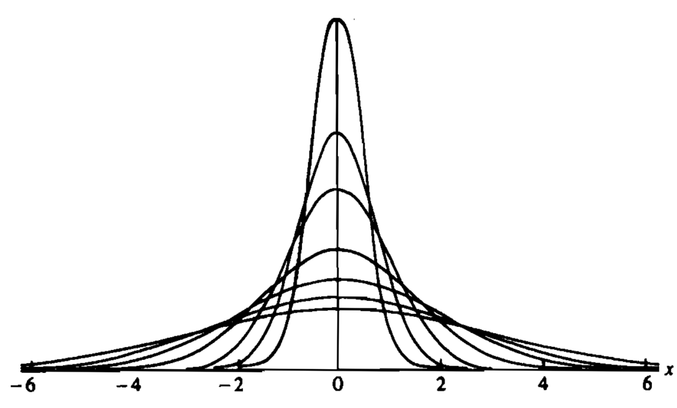


Figure 29.2: Figure 3.5.3. Members of the same scale family

## Probability Inequalities

The most famous, and perhaps most useful, probability inequality is Chebychev's Inequality.

**Theorem (Chebychev's Inequality)** Let $X$ be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,

$$\Pr\left(g(X) \geqslant r\right) \leqslant \frac{Eg(X)}{r}.$$

*Proof:*

$$\begin{aligned} Eg(X) &= \int\limits_{-\infty}^{\infty} g(x) f_X(x) dx \\ &\geqslant \int\limits_{\{x: g(x) \geqslant r\}} g(x) f_X(x) dx \qquad (g \text{ is nonnegative}) \\ &\geqslant r \int\limits_{\{x: g(x) \geqslant r\}} f_X(x) dx \\ &= r \Pr(g(X) \geqslant r) \end{aligned}$$

**Example** The most widespread use of Chebychev's Inequality involves means and variances. Let $g(x) = (x-\mu)^2/\sigma^2$, where $\mu = EX$ and $\sigma^2 = Var(X)$. For convenience write $r = t^2$.Then

$$\Pr\left(\frac{(X-\mu)^2}{\sigma^2} \geqslant t^2\right) \leqslant \frac{1}{t^2} E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{t^2}.$$

This means

$$\Pr(|X - \mu| \geqslant t\sigma) \leqslant \frac{1}{t^2}$$

and its companion

$$\Pr(|X - \mu| < t\sigma) \geqslant 1 - \frac{1}{t^2},$$

which give a universal bound on the deviation $|X - \mu|$ in terms of $\sigma$. For example, taking $t = 2$, we get

$$\Pr(|X - \mu| \geqslant 2\sigma) \leqslant \frac{1}{2^2} = 0.25,$$

so there is at least a 75% chance that a random variable will be within $2\sigma$ of its mean. Have you heard of Six Sigma?