Review

Exponential Family

$$f(x|\underline{\theta}) = h(x)c(\underline{\theta}) \exp(w_i(\underline{\theta})t_i(x))$$

Definition 3.4.5 The *indicator function* of a set A, most often denoted by $I_A(x)$, is the function

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

An alternative notation is $I(x \in A)$.

for exponential family sample space won't be a function of $\underline{\theta}$.

→ Sample space indicator part of $h(x)$.

→ exponential families we can interchange integration and integration, even if $-\infty < x < \infty$.

Theorem 2.4.1 (Leibnitz's Rule) *If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

$$\frac{d}{d\theta}\int_a^b f(x,\theta)\,dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

MS Theory I Lecture - 9
6

Review

More theorems in C&B to cover when $0 < x < \infty$ or $-\infty < x < \infty$

$a(\theta)$ "=" $-\infty$ or $b(\theta)$ "=" ∞

Can we assume:

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)\,dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)\,dx \;?$$

The property illustrated ~~for the exponential distribution~~ above holds for a large class of densities, which will be dealt with in Section 3.4.

↘ 'exponential families'

"obviously" the sample space / support won't be a ft'n of θ.

Similarly for 'exponential families' we can assume that we can interchange differentiation and summation.

$\frac{d}{d\theta}$ and $\sum_{i=1}^{\infty}$ or $\sum_{i=1}^{n}$.

Review

exponential family

Theorem 3.4.2 *If X is a random variable with pdf or pmf of the form (3.4.1), then*

$$\text{(3.4.4)}\quad \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol\theta)}{\partial\theta_j} t_i(X)\right) = -\frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta);$$

$$\text{(3.4.5)}\quad \mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol\theta)}{\partial\theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial\theta_j^2}\log c(\boldsymbol\theta) - \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol\theta)}{\partial\theta_j^2} t_i(X)\right).$$

Find Expectations by taking derivatives! ☺ & w/out mgfs!

Try Expected values for a normal (μ,σ^2) (Notes Page 2)

$\rightarrow$ Lecture-12: $N(\mu,\sigma^2)$ $k=2$; $\theta_1=\mu$; $\theta_2=\sigma^2$

Definition 3.4.7 A *curved exponential family* is a family of densities of the form (3.4.1) for which the dimension of the vector $\boldsymbol\theta$ is equal to $d<k$. If $d=k$, the family is a *full exponential family*. (See also Miscellanea 3.8.3.)

Curved exponential

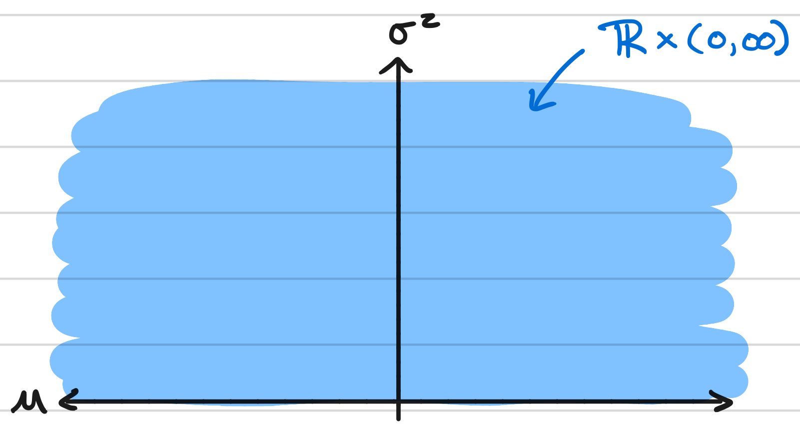
- $\dim(\theta) < k$ $\quad f(x|\underline{\theta}) = h(x)\,c(\underline{\theta})\exp\left(\sum_{i=1}^{k} w_i(\theta)\,t_i(x)\right)$

Full exponential $\nwarrow$ statistical nirvana ☺

$\rightarrow$ A state of perfect happiness; an ideal or idyllic place

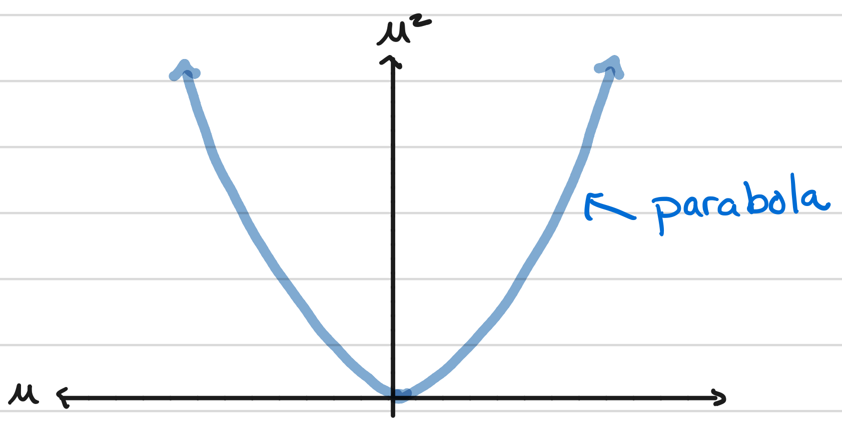
- $\dim(\theta) = k$

Parameter Space for $N(\mu,\sigma^2)$



Full exponential

Parameter Space for $N(\mu,\mu^2)$



Curved Exponential

Natural Parameterization

$$f(x|\eta) = h(x)c^*(\eta)\exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right).$$

Example $N(\mu,\sigma^2)$ assuming: $\sigma > 0$; $-\infty < \mu < \infty$

$h(x) = I_{(-\infty,\infty)}(x)$

$c(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-\mu^2}{2\sigma^2}\right)$ $-\infty < \mu < \infty$, $\sigma > 0$

$t_1(x) = -x^2/2$ $\qquad t_2(x) = x$

$w_1(\mu,\sigma) = \frac{1}{\sigma^2}\ \{\sigma > 0$ $\qquad w_2(\mu,\sigma) = \frac{\mu}{\sigma^2}\ \begin{cases} -\infty<\mu<\infty \\ \sigma > 0\end{cases}$

$\eta_1 = \frac{1}{\sigma^2}\ \{\eta_1 > 0$ $\qquad \eta_2 = \frac{\mu}{\sigma^2}\ \{-\infty<\eta_2<\infty$

- solve for old in terms of new $\qquad$ plug into $c(\theta) = c(\mu,\sigma^2) \Rightarrow c^*(\eta)$

$\sigma = \frac{1}{\sqrt{\eta_1}}$; $\eta_2 = \mu\cdot\eta_1$ $\quad \mu = \frac{\eta_2}{\eta_1}$

$c^*(\eta) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{\eta_2}{\eta_1}\right)^2\eta_1\right) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}}\exp\left(\frac{-\eta_2^2}{2\eta_1}\right)$

$$f(x|\eta_1,\eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}}\exp\left(\frac{-\eta_2^2}{2\eta_1}\right)\exp\left(\eta_1\left(\frac{-x^2}{2}\right) + \eta_2 x\right) * I_{(-\infty,\infty)}(x)$$

- $\eta_1 = \frac{1}{\sigma^2} \Rightarrow \eta_1 > 0$
- also η_1 must be >0 for $c^*(\eta)$ to be finite $\left[\exp\left(\frac{\eta_1}{2}(-x^2)\right)\right]$
  $\lim_{x\to\pm\infty}\exp\left(\frac{\eta_1}{2}(-x^2)\right) = 0$ if $\eta_1 > 0$
- if $\eta_1 > 0$ can have $-\infty < \eta_2 = \frac{\mu}{\sigma^2} < \infty$
- $\underbrace{\exp\left(-\eta_1 x^2/2 + \eta_2 x\right)}_{\text{order }(x^2)}$ finite $\eta_1 > 0$, $-\infty<\eta_2<\infty$
- "Although natural parameterizations provide a convenient mathematical formulation, they sometimes lack simple interpretations like mean and variance."

§ 3.5 Location and Scale Families

- **Location families:** $f(x-\mu)$ ← where ever x in pdf subtract μ.

example: $N(\mu, 1)$ $\quad f(x|\mu, \sigma=1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) * I_{(-\infty,\infty)}(x)$

- **Scale families:** $\frac{1}{\sigma} f(x/\sigma)$ ← wherever x in pdf divided by σ.
+ extra $\left(\frac{1}{\sigma}\right)$ out front.

example: $N(0, \sigma^2)$ $\quad f(x|\mu=0, \sigma^2) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) * I_{(-\infty,\infty)}(x)$

- **Location-Scale families:** $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ ← x in pdf, μ subtract, σ divide, $\frac{1}{\sigma}$ out front.

example: $N(\mu, \sigma^2)$ $\quad f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) * I_{(-\infty,\infty)}(x)$

- Location, Scale, Location-Scale generated from a standard pdf
↑
$\mu = 0, \sigma = 1$
location shift / scale shift

example: $f(x|\mu=0, \sigma^2=1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Theorem 3.5.1 *Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function*

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

Proof C+B: (page 116).
show $g(x|\mu, \sigma) \geq 0 \ \forall \ x, \mu, \sigma$
$\int_{-\infty}^{\infty} g(x|\mu, \sigma)\, dx = 1$
} assuming $f_X(x)$ is pdf.

Location Family: $f(x-\mu)$

- start with any pdf $f(x)$ and generate a family of pdfs by introducing a location parameter.

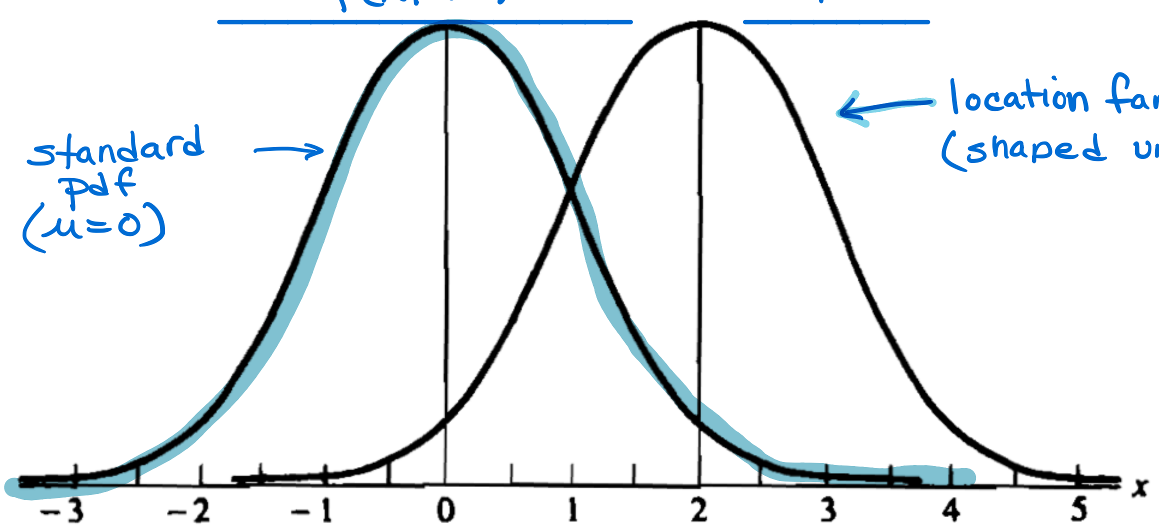
Definition 3.5.2 Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$ and μ is called the *location parameter* for the family.

$f_X(x)$ + μ $\left\{ \begin{array}{l} \text{location parameter (constant)} \\ -\infty < \mu < \infty \end{array} \right.$ $\rightarrow$ $f(x-\mu)$ $\left\{ \text{location family} \right.$

$f_X(x)$: standard pdf

$f(x|\mu=0)$ = $f(x-\mu|\mu)$

standard pdf ($\mu=0$)



location family ($\mu=2$) (shaped unchanged).

Figure 3.5.1. *Two members of the same location family: means at 0 and 2*
(same shape - shifted right by $\mu=2$)

$f(x-\mu) = f(x)$

Assume we know $f(x)$, want to get pdf of 'new' pdf with a location shift of μ.

new dist'n			standard
x	$f(x-\mu)$		$f(x)$
μ	$f(\mu-\mu)$	=	$f(0)$
$\mu+1$	$f((\mu+1)-\mu)$	=	$f(1)$
$\mu+2$	$f((\mu+2)-\mu)$	=	$f(2)$

Similarly areas: $P(\mu-1 \le X \le \mu+2 \,|\, \mu)$ ← location family (μ)

$= P(-1 \le X \le 2 \,|\, \mu=0)$ ← standard pdf

Location Family cont.

Cauchy dist'n:

$$f(x|\theta,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2} I_{(-\infty,\infty)}(x) \qquad -\infty<\theta<\infty, \ \sigma>0$$

θ is a location parameter

Double exponential:

$$f(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} I_{(-\infty,\infty)}(x) \qquad -\infty<\mu<\infty, \ \sigma>0$$

μ is location parameter

If $X \sim f(x-\mu)$ we can represent $X = Z+\mu$, $Z \sim f(z)$ standard pdf

Example (Exponential location family):

standard pdf: $f(x) = e^{-x} I_{[0,\infty)}(x)$

location family: $f(x-\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x)$

$$I_{[\mu,\infty)}(x) = \begin{cases} 1 & x-\mu \geq 0 \quad (x \geq \mu) \\ 0 & \text{else} \end{cases}$$

← positive part of graph starts at μ, rather than zero.

standard pdf ($\mu=0$)

Figure 3.5.2. *Exponential location densities*

Scale family $\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$

Definition 3.5.4 Let $f(x)$ be any pdf. Then for any $\sigma > 0$, the family of pdfs $(1/\sigma)f(x/\sigma)$, indexed by the parameter σ, is called the *scale family with standard pdf* $f(x)$ and σ is called the *scale parameter* of the family.

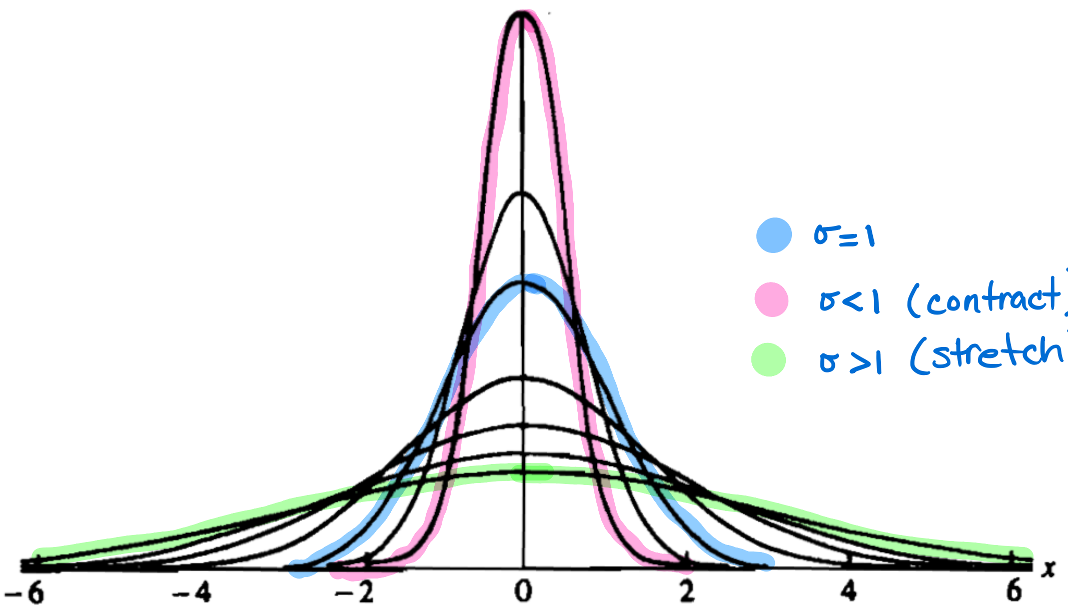


Figure 3.5.3. *Members of the same scale family*

Example: - exponential (β) $f(x|\beta) = \frac{1}{\beta} e^{-x/\beta} I_{[0,\infty)}(x)$ $\beta > 0$

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

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

β is a scale parameter.

Location-scale family $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$

Definition 3.5.5 Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$, and any $\sigma > 0$, the family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the *location–scale family with standard pdf* $f(x)$; μ is called the *location parameter* and σ is called the *scale parameter*.

Location - Scale cont.

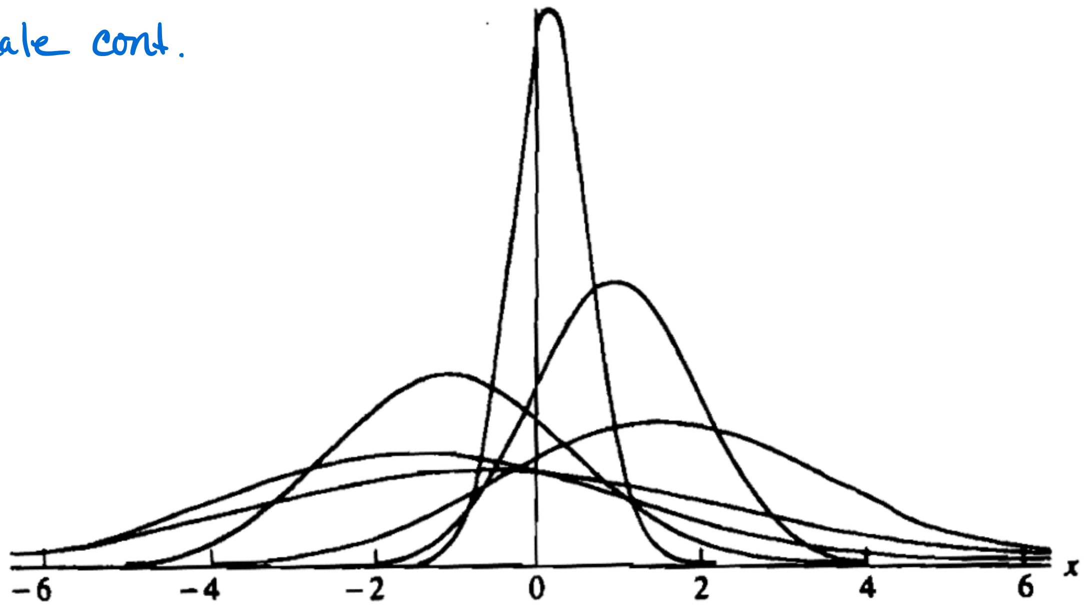


Figure 3.5.4. *Members of the same location–scale family*

- example - Cauchy(θ, σ)

$$f(x|\theta,\sigma) = \frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2} * I_{(-\infty,\infty)}(x)$$

θ is location parameter
σ is scale "

- Double exponential (μ, σ)

$$f(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} * I_{(-\infty,\infty)}(x)$$

Theorem 3.5.6 *Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.*

$$f_X(x) = \frac{1}{\sigma} f\left(\left(\frac{x-\mu}{\sigma}\right)\right) \overset{\text{iff}}{\Rightarrow} f(z) \text{ and } X = \sigma Z + \mu$$
$$\Leftarrow$$

Proof use thm 2.1.5: $f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left|\frac{d}{dy} g^{-1}(y)\right| & y \in \mathcal{Y} \\ 0 & \text{else} \end{cases}$

'if': $\Leftarrow$ Let $X = g(Z) = \sigma Z + \mu \quad g^{-1}(x) = \frac{(x-\mu)}{\sigma} \quad \left|\frac{d}{dx} g^{-1}(x)\right| = \frac{1}{\sigma}$

$$\Rightarrow f_X(x) = f_Z(g^{-1}(x)) \left|\frac{d}{dx} g^{-1}(y)\right| = f\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$$

'only if': $\Rightarrow$ $g(x) = \left(\frac{x-\mu}{\sigma}\right) \quad Z = g(X); \quad g^{-1}(z) = \sigma z + \mu; \quad \left|\frac{d}{dz} g^{-1}(z)\right| = \sigma$

$$f_Z(z) = f_X(g^{-1}(y)) \left|\frac{d}{dz} g^{-1}(z)\right| = \frac{1}{\cancel{\sigma}} f\left(\frac{(\sigma z+\mu)-\mu}{\cancel{\sigma}}\right)\cancel{\sigma} = f(z)$$

$\Rightarrow$ If $Z = \frac{(X-\mu)}{\sigma}$; $\quad X = \sigma Z + \mu$

$$f_Z(z) = \frac{1}{\sigma} f_X\left(\frac{(\sigma z+\mu)-\mu}{\sigma}\right) * \sigma = \frac{1}{1} f\left(\frac{z-0}{1}\right) = f(z)$$

Theorem 3.5.7 *Let Z be a random variable with pdf $f(z)$. Suppose $\mathrm{E}Z$ and $\mathrm{Var}\, Z$ exist. If X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$, then*

$$\mathrm{E}X = \sigma \mathrm{E}Z + \mu \quad \text{and} \quad \mathrm{Var}\, X = \sigma^2 \mathrm{Var}\, Z.$$

In particular, if $\mathrm{E}Z = 0$ and $\mathrm{Var}\, Z = 1$, then $\mathrm{E}X = \mu$ and $\mathrm{Var}\, X = \sigma^2$.

Proof (By Thm 3.5.6): By Thm 3.5.6, there is RV Z^* w/ pdf $f(z)$

$$X = \sigma Z^* + \mu$$

$$E[X] = E[\sigma Z^* + \mu] = \sigma E[Z] + \mu$$

$$Var[X] = \sigma^2 Var(Z^*) = \sigma^2 Var(z)$$

Location scale families w/ finite mean & variance

- Find a standard pdf s.t. $E[Z]=0$ $Var(z)=1$

- example: $N(\mu, \sigma^2)$ $\quad \frac{(x-\mu)}{\sigma} \sim N(0,1)$

here $E[X] = \mu \rightarrow$ mean

$Var[X] = \sigma^2 \rightarrow$ variance

But Double exponential (HW-5)

$\mu = E[X] \rightarrow$ mean

$\underline{\underline{2}}\sigma^2 = Var[X] \rightarrow$ variance

$$\left(\frac{X-\mu}{\sigma}\right) \sim DE(0,2) \Rightarrow \left(\frac{X-\mu}{\sqrt{2}\sigma}\right) \sim DE(0,1)$$

§ 3.6 Inequalities and Identities

Theorem 3.6.1 (Chebychev's Inequality) *Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,*

$$P(g(X) \geq r) \leq \frac{\mathrm{E}g(X)}{r}.$$

Proof: $$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \overset{g(x)\geq 0}{\geq} \int_{\{x:\, g(x)\geq r\}} g(x) f_X(x)\,dx \overset{r>0}{\geq} r \int_{\{x:\, g(x)\geq r\}} f_X(x)\,dx$$

$$= r\, P(g(x) \geq r)$$

Example: Cheychev's (mean & variance)

let $g(X) = \frac{(X-\mu)^2}{\sigma^2}$ (≥ 0) where $\mu = E[X]$, $\sigma^2 = Var[X]$ let $r = t^2$ (>0)

$$P\left(\frac{(X-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2} E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{t^2}$$

$$\rightarrow P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

$$\rightarrow P(|X-\mu| < t\sigma) \geq 1 - \frac{1}{t^2}$$

For dist'n with finite mean $\rightarrow$ bound on $|X-\mu|$ in terms of σ.

$t = 2$

$$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = .25$$

$\rightarrow$ Probability random variable with 2σ of mean $= .75$

- No matter dist'n of X.

Inequality based on Standard Normal $(\mu = 0, \sigma^2 = 1)$

Example (Normal dist'n):

integrating over $x > t$, so $\frac{x}{t} > 1$

$$P(Z \geq t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2} dx \leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{x}{t} e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}$$

$$P(|Z| \geq t) \leq 2\left(\frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}\right) = \frac{\sqrt{2}}{\sqrt{\pi}} \frac{e^{-t^2/2}}{t}$$

$$P(|Z| \geq 2) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-2^2/2}}{2} = \frac{1}{\sqrt{2\pi}} e^{-2} \cong .054$$

Prob R.V. with 2σ of mean $\cong (1 - .054) = .946$