3

- Family - type of distribution whose members are decided by parameters
- Discrete Distribution if its sample set is countable

Discrete Uniform $(1,N)$ $\quad x = 1, 2, \dots N$

- This distribution puts equal weights on each outcome
- $P(X=x|N) = \frac{1}{N} \quad \forall\, x = (1, 2, \dots N)$

$$E(x) = \sum_{1}^{N} x\, P(X=x|N) = \frac{N+1}{2}$$
$$E(x^2) = \sum_{1}^{N} x^2 P(X=x|N) = \frac{(N+1)(2N+1)}{6}$$
$$Var(x) = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{(N+1)(N-1)}{12}$$

Hypergeometric Distribution

- This distribution is used in the 'urn model'
  - We have an urn w/ N balls (M red, N-M green)
    We then select K balls at random - what is the probability that x of the K balls are red
    # of samples of size K $\binom{N}{K}$
    # of ways x of M balls are red $\binom{M}{x}$ w/ K-x green
- $P(X=x|N,M,K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \quad x = 0, 1, \dots K$
  $M \geq x$ and $N-M \geq K-x$
- $E(x) = \sum_{0}^{K}\left[x\,\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\right] = \sum_{1}^{K}\left[x\,\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\right] = \frac{KM}{N}$ w/o loss
- $E(x^2) = \sum_{0}^{K}\left[x^2\,\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\right] = \sum_{1}^{K}\left[\frac{x^2\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}\right]$ w/o loss
- $Var(x) = \frac{KM}{N}\left(\frac{(N-M)(N-K)}{N(N-1)}\right)$

Binomial Distribution

Base on Bernoulli trials (Success/fail)

$$X = \begin{cases} 1 & \text{w/ prop } p \\ 0 & \text{w/ prob } 1-p \end{cases} \quad \begin{matrix} \text{//success} \\ \text{//failure} \end{matrix}$$

$$P(Y=y) = \frac{(n-y+1)}{y} \cdot \frac{p}{p-1} \cdot P(Y=y-1)$$

$E(x) = p$

$Var(x) = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p)$

* For n identical Bernoulli trials
  - $A_i = \{X = 1 \text{ on the } i^{th} \text{ trial}\}$ $i = 1, 2, \dots n$
  - A_i must be independent on i
  - Y = total # of successes in n trials
  - $P(Y=y)$ = prb of y success and $n-y$ fails $= p^y(1-p)^{1-y}$
  - All together $P(Y=y \mid n, p) = \binom{n}{y} p^y (1-p)^{n-y}$ $y = 0, 1, 2, 3, \dots n$

$M_x(t) = pe^t + (1-p)^n$

$E(x) = np$ $var(x) = np(1-p)$ — Y is a binomial (n, p) random variable

Theorem 3.2.2 Binomial Theorem: For any $x, y \in \mathbb{R}$ and $n \geq 0 \in \mathbb{Z}$

$$(x-y)^n = \sum_0^n \binom{n}{i} x^i y^{n-i}$$

Example: $P(1\ 6$ in 4 rolls)

$n = 4$, $p = 1/6$

X = total number of 6s in four rolls

$X \sim$ binomial $(4, 1/6)$

$E(x) = 4/6$

$var(x) = 4/6 - 4/36 = 20/36 = 10/18 = 5/6$

$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = .518$

Poisson Distribution

- used for wait time, spatial statistics
- $P(X=x \mid \lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, \dots$ λ = intensity

- $E(X) = \lambda$
- $Var(x) = \lambda$
- $M_x(t) = e^{\lambda(e^t - 1)}$

$$e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!}$$

<u>Example</u> Operator handles 5 calls every 3 minutes

P(No call in the next minute)? P(At least 2 calls?)

X = Number of calls $X \sim \text{poisson}(5/3)$

$P(\text{No calls in the next minute}) = P(X=0) = \frac{e^{-5/3}\left(\frac{5}{3}\right)^0}{0!} = .189$

$P(\text{At least two in the next min}) = 1 - P(X=0) - P(X=1) = .496$

$P(X=x) = \frac{\lambda}{x} P(X=x-1) \quad x = 1,2,3 \ldots$

<u>Poisson Approximation to the binomial</u>

$\frac{n-y+1}{y} \cdot \frac{p}{1-p} = \frac{np - p(y-1)}{y - py} \approx \frac{\lambda}{y}$

<u>Negative Binomial Distribution</u> opposite of binomial — here we count number of Bernoulli trials to see x successes

- Let X = trials at which the r^{th} success occurs
- $P(X=x \mid r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad x = r, r+1, \ldots$
  - x has a neg. binomial (r,p) distribution

<u>Equally</u>

- Let Y = failures before r^{th} success
- $P(Y=y) = \binom{r+y-1}{y} p^r (1-p)^y \quad y = 0,1,\ldots$
  - it can be shown $\binom{r+y-1}{y} = (-1)^y \binom{-r}{y}$

  which is close to the binomial

- $E(Y) = r\frac{(1-p)}{p}$
- $var(Y) = r\frac{(1-p)}{p^2}$

example $X \sim \text{negbinomial}(100, p)$

$P(X \geq N) = \sum_N^{\infty} \binom{x-1}{99} p^{100} (1-p)^{x-100} = 1 - \sum_N \binom{x-1}{99} p^{100} (1-p)^{x-100}$

Geometric distribution (special case of negative binomial r=1)

$\cdot P(X=x|p) = p(1-p)^{x-1} \quad x=1,2,3,\ldots$

$\sum_{1}^{\infty} a^{x-1} = \frac{1}{1-a}$

· Trial until first success

$x = Y+1$

$\cdot EX = EY+1 = \frac{1}{p} \quad Var(x) = \frac{1-p}{p^2}$ (use neg binomial formulas)

· memoryless: for $t, s \in \mathbb{Z}$

$P(X > s | X > t) = P(X > s-t)$

3.3

Uniform Distribution · spread mass uniformly over interval $[a,b]$

$\cdot f_X(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{ow} \end{cases}$

$E(X) = \frac{b+a}{2}$

$Var(X) = \frac{(b-a)^2}{12}$

Gamma Distribution · on $[0,\infty)$

· gamma function $= \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt \quad \alpha > 0 \in \mathbb{Z}^+$

$\cdot \Gamma(\alpha+1) = \alpha(\Gamma(\alpha)) \quad \alpha > 0$

$\cdot \Gamma(\frac{1}{2}) = \sqrt{\pi}$

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

· α shape parameter

· β scale parameter

$\cdot E(X) = \alpha\beta$

$\cdot Var(X) = \alpha\beta^2$

$\cdot M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha \quad t < \frac{1}{\beta}$

Example $P(X \geq x) = P(Y \geq \alpha)$ for $X \sim Gamma(\alpha,\beta)$; $Y \sim Poisson(x/\beta)$

Chi Squared w/p degrees of freedom

- Special case of gamma $\alpha = p/2$ $\beta = 2$
- $f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2 - 1} e^{-x/2}$ $0 < x < \infty$
- $E(x) = p$
- $Var(x) = 2p$
- $M_x(t) = \left(\frac{1}{1-2t}\right)^{p/2}$
- Used in statistical inference, sampling from normal dist.

Exponential

- Special case of gamma $\alpha = 1$ $\beta = \beta \leftarrow$ scale
- $f_x(x|\beta) = \frac{1}{\beta} e^{-x/\beta}$ $0 < x < \infty$
- $E(x) = \beta$
- $Var(x) = \beta^2$
- memoryless $\Rightarrow$ $s > t \geq 0$ $P(X > s | X > t) = P(X > s - t)$

Weibull

- $X \sim$ exponential(β)
- $Y = X^{1/\gamma} \sim$ weibull(γ, β)
- $f_Y(y|\gamma, \beta) = \frac{\gamma}{\beta} y^{\gamma - 1} e^{-y^\gamma/\beta}$ $0 < y < \infty$ $\gamma > 0$ $\beta > 0$
- modeling hazard functions

Normal $X \sim N(\mu, \sigma^2)$

$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$Z = \frac{(x-\mu)}{\sigma} \sim N(0,1)$ (Standard Normal)

$E(X) = \mu$

$Var(x) = \sigma^2$

* NORMAL APPROX OF BINOMIAL Binomial$(n,p) = n(np, np(1-p))$

for large n and small p

- heuristic: $\min(np, np(1-p)) \geq 5$

Example $X \sim$ Binomial$(25, .6) \approx Y \sim n((25)(.6), (25)(.6)(1-.6))$

$\sim n(15, 2.45)$

$P(X \leq 13) \approx P(Y \leq 13)$

continuity correction: $P(X \leq 13) \approx P(Y \leq 13.5)$

So: $P(X \leq x) \approx P(Y \leq x + 1/2)$

$P(X \geq x) \approx P(Y \geq x - 1/2)$

Beta Distribution = Beta(α, β)

- $f_X(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $0<x<1$ $\alpha>0$ $\beta>0$

$\ni B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$

- $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ ← related to gamma

- $E(x^n) = \frac{B(\alpha+n,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+n)(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$

- $E(X) = \frac{\alpha}{\alpha+\beta}$
- $var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- $(\alpha>1, \beta=1)$ strictly increasing
- $(\alpha=1, \beta>1)$ " decreasing
- $(\alpha<1, \beta<1)$ U-shaped
- $(\alpha>1, \beta>1)$ unimodal
- $(\alpha=\beta)$ symmetric at $\frac{1}{2}$ w/ var $= (4(2\alpha+1))^{-1}$

Cauchy Dist
- symmetric, bellshaped on $(-\infty, \infty)$
- $f_X(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$ $-\infty<x<\infty$ $-\infty<\theta<\infty$
- $E(X)$ DNE $\therefore$ no moments exist
- θ center of the distribution
  - $P(X\leq\theta) = .5$
  - $P(X\geq\theta) = .5$
- ratio of 2 standard normals is cauchy

Lognormal
- $X \sim$ lognormal if $\log(x) \sim n(\mu, \sigma^2)$

$f(x|\mu,\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)\left(\frac{1}{x}\right) e^{-(\log x - \mu)^2/(2\sigma^2)}$ $0<x<\infty$ $-\infty<\mu<\infty$ $\sigma>0$

$E(X) = e^{\mu + (\sigma^2/2)}$

$var(x) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$

- used for data skewed right

Double Exponential · Given by reflecting exponential dist around its mean

· $f_X(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}$ $\quad -\infty < x < \infty, \ -\infty < \mu < \infty \ \sigma > 0$,

· symmetric w/ fat tails (not bell shaped)

· $E(X) = \mu$

· $Var(X) = 2\sigma^2$

III 3.4

A family of pdfs or pmfs is called an exponential family if it can be expressed as:

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

$\Rightarrow$ $h(x) \geq 0$ and $t_1(x), \ldots t_k(x)$ are real valued functions of x they cannot depend on θ. $c(\theta) \geq 0$ and $w(\theta), \ldots w_k(\theta)$ are real valued functions of θ they cannot depend on x

Example: $n \in \mathbb{Z}^+$ We consider the binomial(n,p) family $0 \leq p \leq 1$

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{· binomial pdf}$$

$$= \binom{n}{x}(1-p)^n \left(\frac{p}{1-p}\right)^x \quad \text{· rearrange exponents}$$

$$= \binom{n}{x}(1-p)^n e^{\log\left(\left(\frac{p}{1-p}\right)^x\right)} \quad \cdot x = e^{\log(x)}$$

$$= \binom{n}{x}(1-p)^n e^{x\log\left(\frac{p}{1-p}\right)} \quad \cdot \log(x^n) = n\log(x)$$

$h(x) = \binom{n}{x}$ $\quad c(p) = (1-p)^n$ $\quad t_1(x) = x$ $\quad w_1(p) = \log\left(\frac{p}{1-p}\right)$

$\therefore$ $f(x|p) = h(x)c(p)e^{t_1(x)w_1(p)}$

$\therefore$ We can call the binomial an exponential family $k=1$

note for $x = 0, 1, \ldots, n$, $0 \leq p < 1$

<u>Theorem 3.4.2</u>: If X is a random variable w/ pdf or pmf
$$f(x|\theta) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta) t_i(x)\right) \text{ then}$$

$$E\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x)\right) = \frac{-\partial}{\partial \theta_j} \log(c(\theta))$$

and
$$\operatorname{var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x)\right) = \frac{-\partial^2}{\partial \theta_j^2} \log(c(\theta)) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(x)\right)$$

Is this right?

$\boxed{\text{Example}}$ From previous example

$$\frac{d}{dp} w_1(p) = \frac{d}{dp} \log\left(\frac{p}{1-p}\right) = \frac{1}{p(1-p)}$$

$$\frac{d}{dp} \log(c(p)) = \frac{d}{dp} n \log(1-p) = \frac{-n}{1-p}$$

$\therefore$ Theorem 3.4.2 gives us that

$$E\left(\frac{1}{p(1-p)} X\right) = \frac{n}{1-p} \Rightarrow E(x) = np$$

$$-\frac{d^2}{dp^2} n \log(1-p) = \frac{d}{dp} \frac{n}{1-p} = \frac{d}{dp} n(1-p)^{-1} = -\frac{n}{(1-p)^2}$$

$$\operatorname{var}\left(\frac{1}{p(1-p)} X\right) = \frac{-n}{(1-p)^2} - E\left(\frac{-n}{(1-p)^2} X\right)$$
$$= \frac{-n}{(1-p)^2} + \frac{n}{(1-p)^2} E(X)$$
$$= \frac{n}{(1-p)^2} (E(X) - 1)$$
$$= \frac{n^2 p - n}{(1-p)^2}$$

Example: $f(x|\mu,\sigma^2)$ be $N(\mu,\sigma^2)$ family of pdfs where $\theta = (\mu,\theta)$, $-\infty < \mu < \infty$, $\sigma > 0$ then

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}(\sigma)} e^{-[(x-\mu)^2/2\sigma^2]}$$
$$= \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-[x^2-2\mu x+\mu^2/2\sigma]}$$
$$= \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-(x^2\frac{1}{2\sigma})+(2x\frac{\mu}{2\sigma})-(\frac{\mu^2}{2\sigma})}$$

$h(x) = \frac{1}{\sqrt{2\pi}}$ $c(\mu,\sigma) = \frac{1}{\sigma}$ $t_1(x) = x^2$ $w_1(\mu,\sigma) = \frac{-1}{2\sigma}$

$t_2(x) = 2x$ $w_2(\mu,\sigma) = \frac{\mu}{2\sigma}$

$t_3(x) = 1$ $w_3(\mu,\sigma) = \frac{-\mu^2}{2\sigma}$

$k = 3$

* Is there a pref to minimize k?

Note: we can write as

$$= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma}} e^{-(x^2\frac{1}{2\sigma})+(2x\frac{\mu}{2\sigma})}$$

$h(x) = \frac{1}{\sqrt{2\pi}}$ $c(\mu,\sigma) = \frac{1}{\sigma} e^{-\frac{\mu^2}{2\sigma}}$ $t_1(x) = x^2$ $w_1(\mu,\sigma) = \frac{-1}{2\sigma}$

$t_2(x) = 2x$ $w_2(\mu,\sigma) = \frac{\mu}{2\sigma}$

$k = 2$

Definition 3.4.5 $I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$ Can be used to show support inline.

$f_x(x) = f_X(x)\, I(x \in \mathcal{X})$

* Example 3.4.6 ?

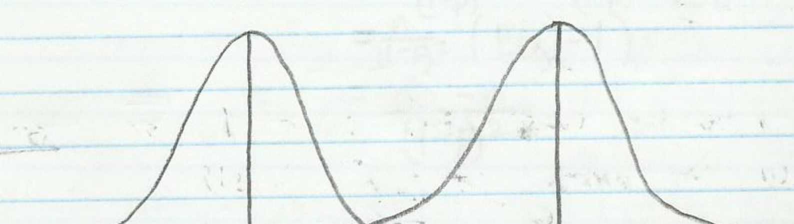
Def A *curved exponential family* is a family of densities of the form $f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$ for which the dimension of $\theta = d < k$, if $d = k$

* ↑ what about the multiple ways of defining it?
Normal $d=2$ I showed exp family $k=3$ and the book $k=2$
Is it assume we define with minimum k?

3.5 Thm 3.5.1 Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant then the function
$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \text{ is a pdf}$$

proof: Let $y = \frac{x-\mu}{\sigma} \in \mathcal{X}$
$f(y) \geq 0$ as f is a pdf for $y \in \mathcal{X}$
$\int_{\mathcal{X}} f(y) = 1$ as f is a pdf for $y \in \mathcal{X}$

Def 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 w/ standard pdf $f(x)$* an μ is called the *location parameter* for the family



* Shape is unchanged
* Shifts graph over

vs. $f(x-\mu) = f(0)$ $x=\mu$
$f(x-\mu) = f(2)$ $x = \mu+2$

If X is a RV: $P(-1 \leq X \leq 2|0) = P(\mu-1 \leq X \leq \mu+2|\mu)$
where the pdf of X is $f(x-0) = f(x)$

Example: $f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \quad -\infty < x < \infty$

$X \sim N(0, \sigma^2)$

$\rightarrow f_x(x-\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x-\mu < \infty$

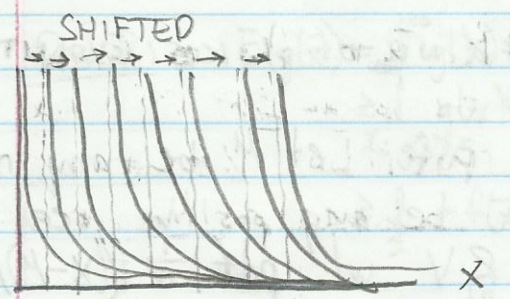
$X \sim N(\mu, \sigma^2)$ is the the location family

Can be thought as $f_x(z)$ where $z = x - \mu$

APPLICATION: Suppose we are measuring something like temp (x), but there is some measurement error involved in the observation. So the actual obs. is $z + \mu$ where z is the actual temp and μ is the error.

Mathy Example Let $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

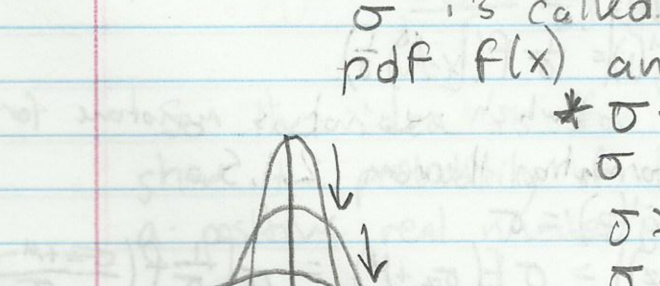
The location family is: $f(x-\mu) = \begin{cases} e^{-(x-\mu)} & x-\mu \geq 0 \\ 0 & x-\mu < 0 \end{cases}$



As μ denotes a bound on the range of X, μ is sometimes called a threshold parameter

Def 3.5.4 Let $f(x)$ be any pdf. Then for any $\sigma > 0$ the family of pdfs $\left(\frac{1}{\sigma}\right) f(x/\sigma)$ indexed by parameter σ is called the scale family w/ standard pdf $f(x)$ and σ is called the scale parameter

$* \sigma > 0 *$



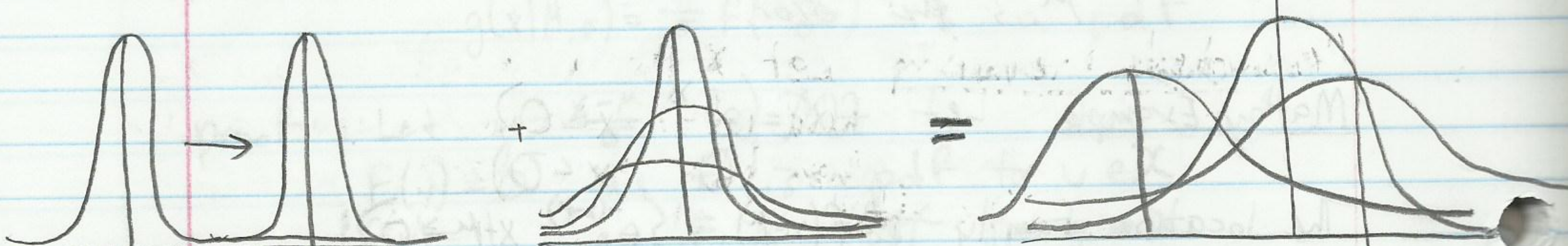
σ changes the shape/peakness of $f(x)$

$\sigma > 1$ stretches

$\sigma < 1$ compresses

Def 3.5.5: Let $f(x)$ be any pdf. Then for any μ $-\infty < \mu < \infty$ and any $\sigma > 0$ the family of pdfs $\left(\frac{1}{\sigma}\right) f((x-\mu)/\sigma)$ indexed by the parameter (μ, σ) is called the location-scale family w/ standard pdf $f(x)$; μ is called the location parameter and σ is called the scale parameter

And what a lovely couple location and scale families make. Their families get along well. At least a lot of the time.



COOL! IT'S KIND OF LIKE EVOLUTION!

Theorem 3.5.6 Let $f(\cdot)$ be any pdf. Let μ be any real number and σ be any positive real number then X is a RV w/ pdf $\frac{1}{\sigma} f((x-\mu)/\sigma)$ iff $\exists$ a RV Z with $X = \sigma Z + \mu$

Proof $\Rightarrow$ let $g(z) = \sigma z + \mu$ (which we note is monotone)

We then use transformation theorem 2.1.5

$g^{-1}(x) = \frac{x-\mu}{\sigma}$ $\quad \frac{d}{dx} g^{-1}(x) = \frac{1}{\sigma}$

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

$\Leftarrow$ let $g(x) = \frac{x-\mu}{\sigma}$ (which we not is monotone for fixed μ, σ)

We then use transformation theorem 2.1.5

$g^{-1}(z) = \sigma z + \mu$ $\quad \frac{d}{dz} g^{-1}(z) = \sigma$

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

$= f(z)$

Also $\sigma Z + \mu = \sigma g(X) + \mu = \sigma \left(\frac{X-\mu}{\sigma} \right) + \mu = X$

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

$$E(X) = \sigma E(Z) + \mu$$ via simple expectation rules

$$var(X) = \sigma^2 Var(Z)$$ via simple variance rules

3.6 We didn't cover this but Casella + Berger think it's important!

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

$$P(g(x) \geq r) \leq \frac{E(g(x))}{r}$$

Proof:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f_x(x)\, dx$$

$$\geq \int_{\mathcal{X}:\, g(x) \geq r} g(x) f_X(x)\, dx$$ • as restricted support will shrink integrals

$$\geq \int_{\mathcal{X}} r f(x)\, dx = r \int_{\mathcal{X}} f(x) dx$$ as $r \leq g(x)$

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

thus $E(g(x)) \geq r\, P(g(x) \geq r)$

divide by r

$$\frac{E(g(x))}{r} \geq P(g(x) \geq r)$$

So the idea is for non-negative functions we can show the probability of that function is greater than or equal to a positive real number r is $\frac{E(g(x))}{r}$. This is interesting.

<u>Example:</u> Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$ where $\mu = E(x)$ and $\sigma^2 = var(x)$

Let $r = t^2$

$$P\left(\frac{(x-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{E\left(\frac{(x-\mu)^2}{\sigma^2}\right)}{t^2} = \frac{1}{t^2}$$ ⇐ How?

$$P\left((x-\mu)^2 \geq \sigma^2 t^2\right) \leq$$

$$P\left(|x-\mu| \geq \sigma t\right) \leq \left(\frac{1}{\sigma}E(x) - \frac{\mu^2}{\sigma^2}\right)/t^2$$

$$\leq \left(\frac{\mu}{\sigma} - \frac{\mu^2}{\sigma^2}\right)/t^2$$

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

⋮ Some magic happens that's obvious to Casella & Berger we get:

$$P(|x-\mu| \geq t\sigma) \geq \frac{1}{t}$$

$$P(|x-\mu| \leq t\sigma) \leq 1 - \frac{1}{t}$$ complement!

Now we have a universal bound on the spread from the mean $P(|x-\mu| \geq t\sigma) < \frac{1}{t^2}$

$\boxed{\text{For } t=2}$

$$P(|x-\mu| \geq 2\sigma) = \frac{1}{4} = .25$$

- This is interpreted as ∃ 75% chance that the RV will be within 2σ of its mean regardless of the dist of X.

* Is this a confidence interval type of application

Example If z is standard normal $(N(0,1))$ then

$$P(|Z| \geq t) \leq \sqrt{\tfrac{2}{\pi}} \, \frac{e^{-t^2/2}}{t} \quad \forall\, t > 0$$

- Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$ $\mu = E(X)$ $\sigma^2 = var(x)$ (as in prev example)

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

$$\leq \frac{1}{\sqrt{2\pi}} \int_t^{\infty} \frac{x}{t} e^{-x^2/2} \, dx$$ since $\frac{x}{t} > 1$ for $x > t$ and our lower bound is t

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

- As z is standard normal

$P(|Z| \geq t) = 2P(Z \geq t)$ Yay symmetry!

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

$\leq .054$

$\therefore \exists$ .946 chance that the RV will be within 2σ of its mean

$\therefore$ We have more power knowing the distribution!!

Identities - Think of as recurrance relations

$X \sim$ Poisson

$P(X = x+1) = \frac{\lambda}{x+1} P(X = x)$ where $P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}$

ie $P(X=0) = \frac{e^{-\lambda}(1)}{(1)} = e^{-\lambda}$

$P(X=1) = \frac{e^{-\lambda}\lambda}{1} = e^{-\lambda}\lambda$ $P(X=1) = P(X=0+1) = \frac{\lambda}{0+1} e^{-\lambda} = e^{-\lambda}\lambda$

↑ This is **awesome**!

Theorem 3.6.4 Let $X_{\alpha,\beta}$ denote a gamma(α,β) RV w/ pdf $f(x|\alpha,\beta)$ where $\alpha > 1$ then for any constants a and b

$$P(a < X_{\alpha,\beta} < b) = \beta\left(f(a|\alpha,\beta) - f(b|\alpha,\beta)\right) + P(a < X_{\alpha-1,\beta} < b)$$

Proof: $P(a < X_{\alpha\beta} < b) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_a^b x^{\alpha-1} e^{-x/\beta} dx$ = area under Gamma

- By int by parts $\rightarrow = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(-x^{\alpha-1}\beta e^{-x/\beta}\Big|_a^b + \int_a^b (\alpha-1)x^{\alpha-2}\beta e^{-x/\beta} dx\right)$

  $u = x^{\alpha-1}$ $dv = e^{-x/\beta} dx$

- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \rightarrow \beta\left[\left(\frac{-b^{\alpha-1}e^{-b/\beta}}{\Gamma(\alpha)\beta^\alpha}\right) - \left(\frac{-a^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\right)\right] + \int_a^b (\alpha-1)x^{\alpha-2}\beta e^{-x/\beta} \frac{1}{\Gamma(\alpha)\beta^\alpha}$

$$= \beta\left(f(a|\alpha,\beta) - f(b|\alpha,\beta)\right) + \int_a^b \frac{(\alpha-1)\beta}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-2} e^{-x/\beta}$$

$$= \beta\left(f(a|\alpha,\beta) - f(b|\alpha,\beta)\right) + \int_a^b \frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}} x^{\alpha-2} e^{-x/\beta}$$

$$= \beta\left(f(a|\alpha,\beta) - f(b|\alpha,\beta)\right) + P(a < X_{\alpha-1,\beta} < b)$$

Lemma 3.6.5 Stein's lemma

- Let $X \sim N(\theta, \sigma^2)$ and let g be a differentiable function satisfying $E|g'(x)| < \infty$ then

$$E[g(x)(x-\theta)] = \sigma^2 E(g'(x))$$

- This is proved along similar lines as 3.6.4 p124

Example 3.6.6: $X \sim n(\theta, \sigma^2)$ we can use stein's Lemma to find higher order moments

$$E(g(x)(x-\theta)) = \sigma^2 E(g'(x))$$

· This is extended to artful rewrite

$$E(x^3) = E(x^2(x-\theta+\theta)) \qquad g(x)=x^2 \quad g'(x)=2x$$
$$= E(x^2(x-\theta)) + \theta E(x^2)$$
$$= \sigma^2 E(2x) + \theta E(x^2) \qquad \text{By Stein's Lemma}$$
$$= 2\sigma^2 E(2x) + \theta E(x^2)$$
$$= 2\sigma^2\theta + \theta(\sigma^2+\theta^2)$$
$$= 3\sigma^2\theta + \theta^3$$

Theorem 3.6.7 Let χ_p^2 be a chi-squared RV w/ p degrees of freedom for any function $h(x)$

$$E(h(\chi_p^2)) = pE\left(\frac{h(\chi^2_{(p+2)})}{\chi^2_{(p+2)}}\right) \text{ provided that all the expectations exist}$$

See Proof P 125

Theorem 3.6.8 Let $g(x)$ be a function w/ $-\infty < Eg(x) < \infty$ and $-\infty < g(-1) < \infty$ Then

(a) If $X \sim \text{Poisson}(\lambda)$
$$E(\lambda g(x)) = E(x\, g(x-1))$$

(b) If $X \sim \text{negbin}(r,p)$
$$E((1-p)g(x)) = E\left(\frac{x}{r+x-1}\, g(x-1)\right)$$

proved p 126

Example $X \sim \text{poisson}(\lambda)$, $g(x) = x^2$
use stein's Lemma $E(\lambda x^2) = E(x(x-1)^2) = E(x^3 - 2x^2 + x)$

$$\therefore E(x^3) = \lambda E(x^2) + 2E(x) - E(x)$$
$$= \lambda(\lambda+\lambda^2) + 2(\lambda+\lambda^2) - \lambda$$
$$= \lambda^3 + 3\lambda^2 + \lambda$$

I don't get this result?

3.8 Poisson Postulates

Theorem 3.8.1: For each $t \geq 0$, let N_t be an integer-valued RV w/ the following propertres

- N_t can denote the number of arrivals in the time period from time 0 to time t

i) $N_0 = 0$ *(Start w/ no arrivals)*

ii) $s < t \to N_s$ and $N_t - N_s$ are independent (arrivals in disjoint time periods are independent)

iii) N_s and $N_{t+s} - N_t$ are identically distributed (# of arrivals depends only on period length)

iv) $\lim_{t \to 0} \frac{P(N_t = 1)}{t} = \lambda$ (arrival probability proportional to period length, if length is small)

v) $\lim_{t \to 0} \frac{P(N > 1)}{t} = 0$ (No simultaneous arrivals)

If i-v hold $\forall n \in \mathbb{Z}$ then $P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$
that is $N_t \sim$ Poisson(λt)

↑ a really long, detailed explanation of a Poisson RV

Theorem If $0 < \sigma < \infty$ then

a) If $n=1$ chebyshev is attainable for $k \geq 1$ and not for $0 < k < 1$

b) If $n=2$ the chebyshev is attainable iff $k=1$

c) If $n \geq 3$ the chebyshev isn't attainable

↑ this is proved using the special case of Chebyshev:
$P(|\bar{X}_n - \mu| \geq k\sigma) \leq \frac{1}{nk^2}$

Lemma 3.8.3 Markov's Inequality: If $P(Y \geq 0) = 1$ and $P(Y = 0) < 1$ then $\forall\ r > 0$

$$P(Y \geq r) \leq \frac{E(Y)}{r} \quad \text{and } P(Y=r) = p = 1 - P(Y=0) \quad 0<p\leq 1$$

Chebyshev is weak b/c there are no assumptions on the underlying distribution

Theorem 3.8.4 Gauss Inequality Let $X \sim f$ where f is unimodal w/ mode v and define $\tau^2 = E(X-v)^2$ then

* Note this has heavy assumptions so it isn't very useful

$$P(|X-v| > \epsilon) \leq \begin{cases} \frac{4\tau^2}{9\epsilon^2} & \forall\ \epsilon \geq \sqrt{\frac{4}{3}}(\tau) \\ 1 - \frac{\epsilon}{\tau\sqrt{3}} & \forall\ \epsilon > \sqrt{\frac{4}{3}}(\tau) \end{cases}$$

Theorem 3.85 Some Russian Guy's Inequality (Vysochanskii-Petunin)

$X \sim f$ where f is unimodal and $\xi^2 = E(X-\alpha)^2$ for an arbitrary point α, Then

$$P(|X-\alpha| > \epsilon) \leq \begin{cases} \frac{4\xi^2}{9\epsilon^2} & \forall\ \epsilon \geq \sqrt{8/3}(\xi) \\ \frac{4\xi^2}{9\epsilon^2} - \frac{1}{3} & \forall\ \epsilon \leq \sqrt{8/3}(\xi) \end{cases}$$

NOTE: For $\alpha = \mu = E(X)$ and $\epsilon = 3\sigma$ where $\sigma^2 = var(x)$ yields

$$P(|X-\mu| > 3\sigma) \leq \frac{4}{81} < .05$$

* this is the "three sigma rule - the probability $< .05$ that x is more than 3σ from the μ