Statistics for Data Science

Lecture 6

More Distribution Families

Ken Cameron

Admin

- Background maths tutorials
  - Differentiation cheat sheet now on moodle.
- Wednesdays 10.15am 3W 3.9
- Next topic: Integration

Content

- Challenge (and a correction)
- Exponential Families
- Location and Scale Families
- Chebychev’s Inequality

Challenge

- Assume you take delivery of a consignment of 25 disc drives.
  - As part of acceptance testing, you run the self test on 10 of them.
  - What is the probability of all 10 passing if 6 of the 25 are faulty?
- Start by picking the right distribution to model the problem.

Hypergeometric (corrected)

- Example: Bag containing N balls, M red, N-M green. Select K balls.
  - What is the probability that x are red?

$$P(X \mid N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, x = 0,1,\ldots,K.$$

M >= x and N-M >= K – x
M – (N – K) <= x <= M

Solution

- N=25, M=6, K=10

$$P(X = 0) = \frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}} = 0.028$$

- Not a very likely event if 6 (or more) are bad.

Exponential Families

- An exponential family is a family of pdfs or pmfs if it can be expressed as:

$$f(x \mid \theta) = h(x)c(\theta)\exp\left(\sum_{t=1}^{k} w_i(\theta)t_i(x)\right) \qquad [1]$$

- $h(x) >= 0$

- $c(\theta) > 0$

- $t_1(x),\ldots,t_k(x)$ are real valued functions of the observation x, they cannot depend on θ.

- $w_1(\theta),\ldots,w_k(\theta)$ are real valued functions of parameter θ, they cannot depend on x.

Exponential Families

- Many of the distributions from the previous lecture are exponential families.
  - Normal
  - Binomial
  - Poisson
  - Negative Binomial

Exponential Families

- To verify that a family of pmfs or pdfs is an exponential family we need to find the functions:

$$h(x),\ c(\theta),\ w_i(\theta) \text{ and } t_i(x)$$

- And that it takes form [1]

Example: Binomial

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}$$

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

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

Example: Binomial

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, \ldots, n \\ 0 & otherwise \end{cases}$$

$c(p) = (1 - p)^n$, $0 < p < 1$

$$w_1(p) = \log\left(\frac{p}{1-p}\right), 0 < p < 1$$

$t_1(x) = x$

Example: Binomial

- Therefore,

$$f(x|p) = h(x)c(p)exp[w_1(p)t_1(x)]$$

Which is of the form [1] with k = 1.

Only valid for $0 < p < 1$

Why bother?

- This form has a number of useful mathematical properties
  - In particular, statistical properties.
  - For example.... Another theorem.

Theorem

- If X is a random variable with pdf or pmf of the form [1] then:

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

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

- These may look complicated, but they work out well in many cases.
    - Allow us to replace integration/summation by differentiation.
    - And that can be more straight forward in many cases.

Example: Binomial

- From [1] we have,

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

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

Example: Binomial

- Calculate Expected Value

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

E(X) = np

- Variance works in a similar manner.
  - Left as an exercise.

Example: Normal

- Let’s look at a case where θ contains more than one element.

- Normal distribution: $n(\mu, \sigma^2)$, $\theta = (\mu, \sigma)$

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

$$= \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{\mu^2}{2\sigma^2} \right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} \right)$$

Example: Normal

- Find h(), c(), w_1(), w_2(), t_1() and t_1()

$$h(x) = 1 \; for \; all \; x. \qquad c(\theta) = c(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\mu^2}{2\sigma^2} \right), \quad -\infty < \mu < \infty, \quad \sigma > 0$$

$$w_1(\mu, \sigma) = \frac{1}{\sigma^2}, \sigma > 0 \qquad t_1(x) = -x^2 / 2$$

$$w_2(\mu, \sigma) = \frac{\mu}{\sigma^2}, \sigma > 0 \qquad t_2(x) = x$$

Example: Normal

$$f(x|\mu,\sigma^2) = h(x)c(\mu,\sigma)\exp[w_1(\mu,\sigma)t_1(x) + w_2(\mu,\sigma)t_2(x)]$$

- Which is of the form [1] with k=2

  Note that the parameter functions are defined only over the range of the parameter.

Location and Scale Families

- Three ways of construction families of distributions.
- The resulting families have ready physical interpretations which makes the useful in modelling.
- These are:
  - Location Families
  - Scale Families
  - Location and Scale Families

Theorem

- Let *f(x)* be any pdf and let σ > 0 and μ be any given constants.

- Then the function

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

is a pdf.

Location families

- Let *f(x)* be and pdf.

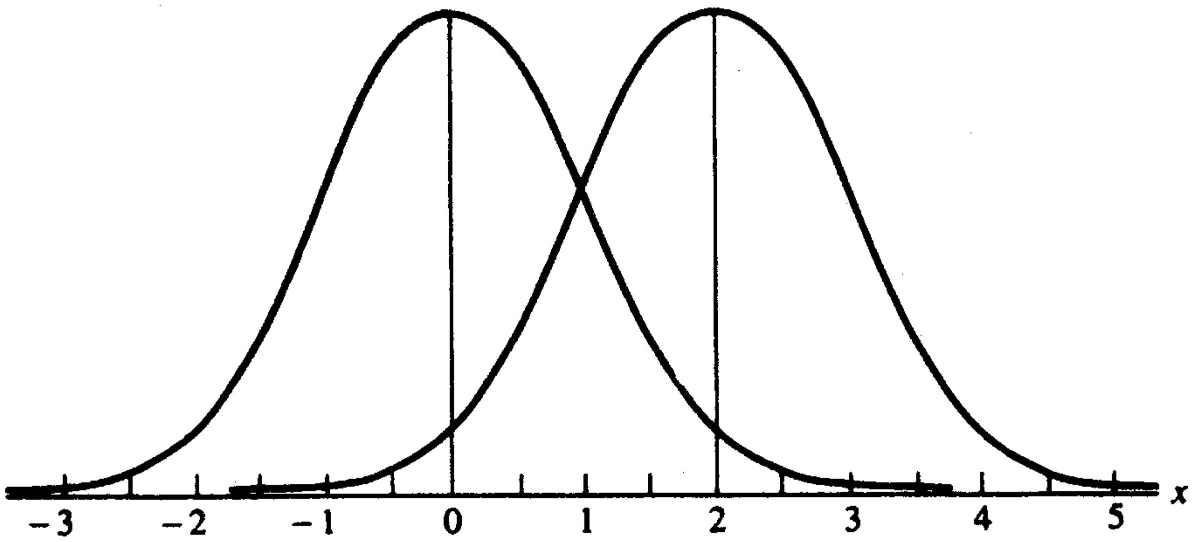
- The family of pdfs $f(x - \mu)$

  - index by the parameter μ

  - $-\infty < \mu < \infty$

- Is called the location family with standard pdf f(x) and location parameter μ.

Location families



Scale Families

- Let $f(x)$ be any pdf.

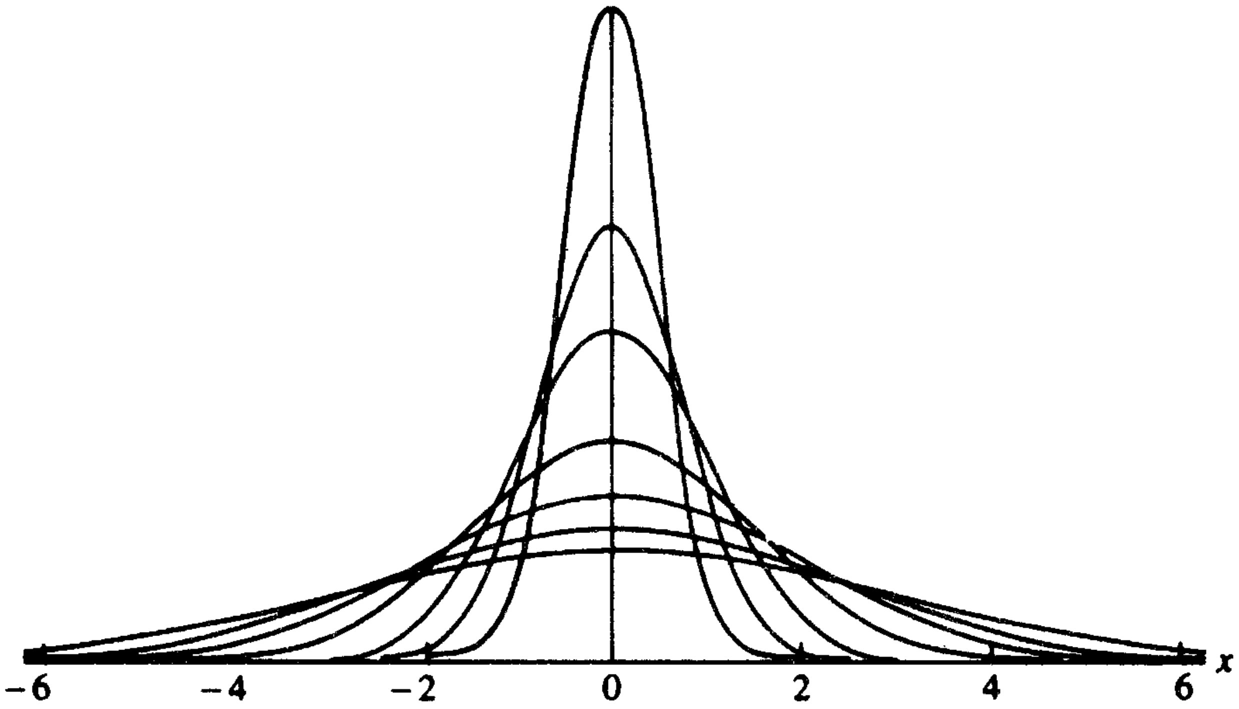
- The family of pdfs $(1/\sigma)f(x/\sigma)$

  - Index by the parameter σ

  - Where any $\sigma > 0$

- Is called the scale family with standard pdf $f(x)$ and scale parameter σ.

Scale Families



Location Scale Families

- Let *f(x)* be any pdf.

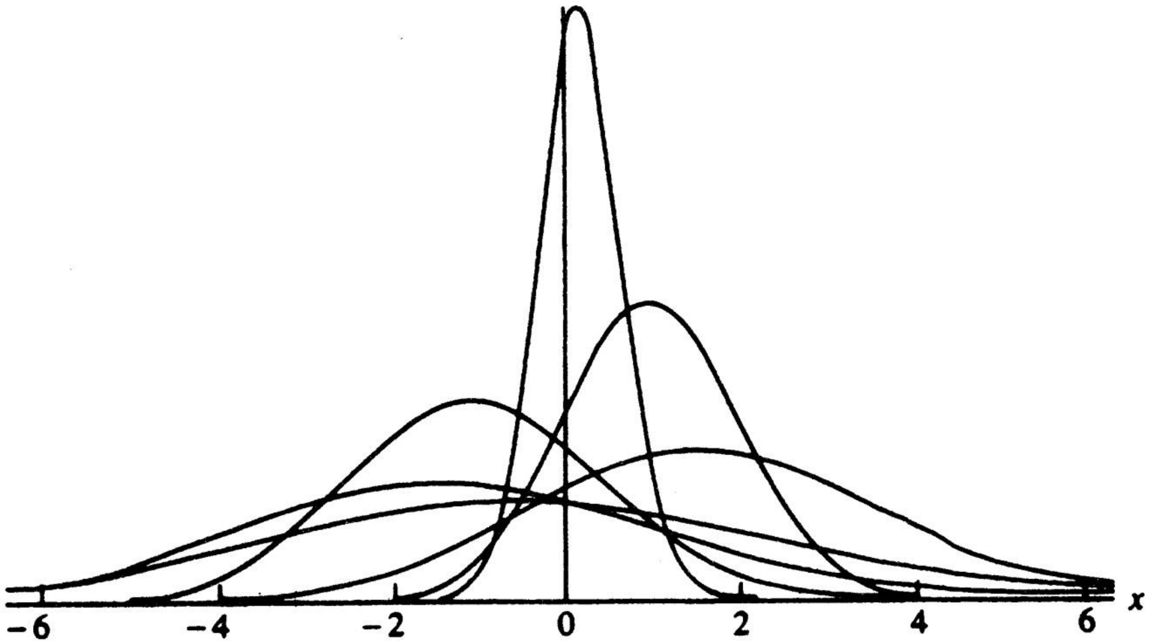
- The family of pdfs $(1/\sigma)f((x - \mu)/\sigma)$

    - Index by the parameter (σ, μ)

    - Where any $\sigma > 0$
    - And $-\infty < \mu < \infty$

- Is called the location-scale family with standard pdf f(x) with location parameter μ and scale parameter σ.

Location and Scale Families



Chebychev’s Inequality

- Let X be a random variable and let *g(x)* be a non-negative function.

- Then for any r > 0:

$$P(g(X) \geq r) \leq \frac{Eg(X)}{r}$$

Proof

$$
\begin{aligned}
Eg(X) &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\
&\geq \int_{\{x:g(x)\geq r\}} g(x) f_X(x)dx \\
&\geq r \int_{\{x:g(x)\geq r\}} f_X(x)dx \\
&= rP(g(X) \geq r)
\end{aligned}
$$

Example

- Let $g(x) = (x - \mu)^2/\sigma^2$
  - Where $\mu = EX$
  - $\sigma^2 = \mathrm{Var}\, X$
- Then writing r as t^2,

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

Example

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

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

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

Example

- This gives a universal bound on the deviation of $|X - \mu|$ in terms of σ

- For example, let $t = 2$,

$$P(|X - \mu| >= 2\sigma) <= 1/2^2 = 0.25$$

- Or, there's a 75% chance that a random variable will be within 2σ of its mean.

  - No matter what the distribution of X.