Statistics for Data Science

Lecture 8

Hierarchical Models and Mixture Distributions

Ken Cameron

Class Test

- 1.15pm Thursday 2nd Nov
- 1W 2.104
- 60 minutes
- Multiple choice (10 question)

Class Test

- Test will start promptly at 1.15pm.

- Desks will not be numbered.
  - You may sit where you wish.

- Collect question paper and script on arrival.
  - Don’t open question paper until instructed to do so.

- Calculators will be provided.
  - Collect on arrival.

Class Test

- You should hand in your answer script and calculator before leaving.
  - Don’t leave in the first 20 minutes.
  - Don’t leave in the last 10 minutes.
- No admittance after the first 20 minutes.

Content

- Hierarchical Models
- Mixture Distributions
- Multivariate Distributions

Hierarchical Models

- So far,
  - The random variables we’ve seen have a single distribution.
  - With parameters.
- In general,
  - A random variable has only one distribution.
  - However, it is often easier to model a situation as a hierarchy.

A Classic Example

- An insect lays a *large number* of eggs, each surviving with probability *p*.
  - On average how many eggs survive?
- The large number is a random variable.
  - Let's use a Poisson distribution (λ).
- And if the survival of eggs is independent,
  - The we have Bernoulli Trials.

A Classic Example

- Two random variables
  - X = number of survivors
  - Y = number of eggs laid
  - $X|Y \sim \text{binomial}(Y,p)$.
  - $Y \sim \text{Poisson}(\lambda)$

Complex model from simple models

- The hierarchy allows us to build a complex model
  - From a series of simple models.
- The hierarchy is built using
  - Conditional and marginal distributions.
  - And we know how to deal with those.

Example, continued.

- We're interested in X = number of survivors.

$$P(X = x) = \sum_{y=0}^{\infty} P(X = x, Y = y)$$

$$= \sum_{y=0}^{\infty} P(X = x \mid Y = y) P(Y = y)$$

$$= \sum_{y=x}^{\infty} \left[\binom{y}{x} p^{x} (1-p)^{y-x} \right] \left[\frac{e^{-\lambda} \lambda^{y}}{y!} \right]$$

- Given, x|Y = y is binomial(y, p) and Y is Poisson(λ).

Example, simplify.

$$P(X=x)=\sum_{y=x}^{\infty}\left[\binom{y}{x}p^{x}(1-p)^{y-x}\right]\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]$$

$$=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}\sum_{y=x}^{\infty}\frac{((1-p)\lambda)^{y-x}}{(y-x)!}$$

$$=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}\sum_{t=0}^{\infty}\frac{((1-p)\lambda)^{t}}{t!}$$

$$=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}e^{(1-p)\lambda}$$

$$=\frac{(\lambda p)^{x}}{x!}e^{-p\lambda}$$

Simplify, and multiply by λ^x / λ^x.

Note, t = y – x

So X ~ Poisson(λp)

Example

- Y no longer plays a part.
  - Introducing it was to make the model easier to follow.
- The parameter of X is now the product of two parameters.
  - Easier to understand both.
- Final answer:

$$EX = \lambda p$$

Theorem

- If X and Y are any two random variables, then

$$EX = E(E(X|Y))$$

- Provided expectations exist.

Proof

- Let $f(x,y)$ denote the joint pdf of X and Y.

- By definition,

$$EX = \iint xf(x,y)dxdy = \int \left[\int xf(x \mid y) \right] f_Y(y)dy$$

- Where $f(x|y)$ is the conditional pdf of X given Y =y
- And $f_Y(y)$ is the marginal pdf of Y

Proof

- Note, the inner integral is the conditional expectation E(X|y), so

$$EX = \int E(X \mid y)\, f_{Y(y)} dy = E(E(X \mid y))$$

- You can replace the integrals with sums to prove this for the discrete case as well.

Example, revisited

- So we can use this to find the answer for our previous example.

$$
\begin{aligned}
EX &= E(E(X \mid Y)) \\
&= E(pY) \\
&= p\lambda
\end{aligned}
$$

- Given X|Y ~ binomial(Y,p)
- And Y ~ Poisson(λ)

Theorem

- For any two random variables X and Y:

$$\text{Var } X = E(\text{Var}(X|Y) + \text{Var}(E(X|Y))$$

- Provided the expectations exist.

Proof

- By definition,
- $\text{Var } X = E([X - EX]^2) = E([X - E(X|Y) + E(X|Y) - EX]^2)$
- $\text{Var } X = E([X - E(X|Y)]^2 + E([E(X|Y) - EX]^2) + 2E([X - E(X|Y)][E(X|Y)\text{-}EX])$
  - Last term is zero.

Proof

- Thus

$$E([X - E(X|Y)]^2) = E(E\{[X - E(X|Y)]2|Y\}) = E(Var(X|Y))$$

- And,

$$E([E(X|Y) - EX]^2) = Var(E(X|Y))$$

- Establishing,

$$Var\ X = E(Var(X|Y) + Var(E(X|Y))$$

Mixture distribution

- A random variable X is said to have a

mixture distribution

- if the distribution of X depends on a quantity that also has a distribution.
- Hierarchies lead to mixture distributions.
- (This is the popular definition. There is no standardised one.)

Multi-stage Hierarchies

- Hierarchies are not limited to two distributions.
  - We can treat multi-stage hierarchies as a series of two-stage.
- Or,
  - In some cases we may keep them as explicit stages.
  - Sometimes that’s easier.

Multivariate Distributions

- We've been focused on bivariate distributions.

- But, as mentioned at the start of the previous lecture, all the ideas we've covered apply when there's more than two random variables.

- We can generalise this as a vector of random variables:

$$\mathbf{X} = (X_1, X_2, ..., X_n)$$

Sample Space

- The random vector **X** has a sample space that is a subset of R^n.

- If it's a discrete random vector, the sample space is countable.

- If **X** = $(X_1, X_2, ..., X_n)$ is a discrete random vector
  - Then the joint pmf of $(X_1, X_2, ..., X_n)$ is
  - The function defined by $f(x) = f(x_1, x_2, \ldots, x_n) = P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n)$
  - For each $(x_1, x_2, \ldots, x_n) \in R^n$

- Discrete

$$P(\boldsymbol{X} \in A) = \sum_{x \in A} f(x)$$

- Continuous

$$P(\boldsymbol{X} \in A) = \int \dots \int_A f(x)dx = \int \dots \int_A f(x_1, \dots, x_n)dx_1 \dots dx_n$$

- (For any A $\subset$ R^n)

Definition

- Let n and m be positive integers
- And let $p_1, \ldots, p_n$ be numbers satisfying $0 <= p_i <= 0$, $i = 0, \ldots, n$
- And $\sum_{i=1}^{n} p_i = 1$.

Definition

- Then the random vector (X_1, ..., X_n) has a multinomial distribution with m trials and cell probabilities p_1, …, p_n.

- If the joint pmf of (X_1, ..., X_n) is

$$f(x_1, \dots, x_n) = \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n} = m! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}$$

Multinomial Theorem

- Let m and n be positive integers.

- Let A be the set of vectors x = (x_1, x_2, …, x_n)

  - Such that each x_i is a nonnegative integer

  - And $\prod_{i=1}^{n} x_i$ = m

- Then for any real numbers p_1, …, p_n:

$$(p_1 + \ \ldots + p_n)^m = \sum_{x \in A} \frac{m!}{x_1! \ldots x_n!} p_1^{x_1} \ldots p_n^{x_n}$$

Definition

- Let $\mathbf{X_1}, \dots, \mathbf{X_n}$ be random vectors with joint pdf or pmf $f(x_1, \dots, x_n)$.
- Let $f_{Xi}(x_i)$ denote the marginal pdf or pmf of $\mathbf{X_i}$.
- Then $\mathbf{X_1}, \dots, \mathbf{X_n}$ are called mutually independent random vectors
  - If for every $f(x_1, \dots, x_n)$

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

- If the X_is are all one dimensional then $X_1, \dots., X_n$ are called mutually independent random variables.

Theorem

- Let $X_1, \ldots, X_n$ be mutually independent random variables.
- Let $g_1, \ldots, g_n$ be real valued functions
  - Such that $g_i(x_i)$ is a function only if x_i, $i=1,\ldots,n$.
- Then,

$$E(g_1(X_1)\ldots g_n(X_n)) = E(g_1(X_1)\ldots g_n(X_n))$$

Theorem

- Let $X_1, \ldots, X_n$ be mutually independent random variables
  - With mgfs $M_{X1}(t), \ldots, M_{Xn}(t)$.
- Let $Z = X_1 + \ldots + X_n$.
- Then the mgf of Z is

$$M_z(t) = M_{X1}(t) \ldots M_{Xn}(t).$$

- In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_x(t)$
- Then

$$M_z(t) = (M_x(t))^n$$

Theorem

- Let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be random vectors.

- Then $\mathbf{X_1}, \ldots, \mathbf{X_n}$ are mutually independent random vectors

  - If and only if there exists functions $g_i(x_i)$, $i=1,\ldots,n$

  - Such that the joint pdf or pmf of $(\mathbf{X_1}, \ldots, \mathbf{X_n})$ can be written as:

$$f(x_1, \ldots, x_n) = g_1(x_1)\ldots g_n(x_n)$$

Theorem

- Let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be independently random vectors.
- Let $g_i(x_i)$ be a function only of x_i, i=1,…,n.
- Then the random variables $U_i = g_i(\mathbf{X_i})$, i=1,…n,

are mutually independent.