

4.4 Mixture and hierarchical models

Idea :

Complex families of distributions may be composed in terms of simpler models of distributions, which are combined in a hierarchical way.

- Example :
- An insect lays a large number of eggs
 - Each egg survives with probability p .
 - Let X be the number of eggs that survive

How to model the (Random) behavior of X ?

Bring in the (latent) random Y , which represents the number of eggs laid:

Assume

$$Y \sim \text{Poisson}(\lambda)$$
$$X|Y \sim \text{binomial}(Y, p)$$

This defines a proper joint distribution for (X, Y) which induces a proper (marginal) dist. for X .

$x, y \in \mathbb{N}$

$$P(X=x, Y=y) = P(Y=y) P(X=x | Y=y)$$

So

Poisson (Conditionally binomial)

$$P(X=x) = \sum_{y \in \mathbb{N}} P(X=x, Y=y)$$

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

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

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

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

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

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

Hence

$X \sim \text{Poisson } (\lambda p).$

$$\Rightarrow \begin{cases} E[X] = \lambda_p \\ E[X|Y=y] = y_p \end{cases} \quad \forall y \in \mathbb{N}.$$

Note

Conditional expectation $E[X|Y]$ is a random variable,
and here, $E[X|Y] = Y_p$
thus $E(E[X|Y]) = E(Y_p) = p E[Y] = \lambda_p$.

Remark.

- Usually, marginal distribution for X may not share the same family as that of the latent Y .

- The latent variable Y helps to explain the role of parameters λ and p .

Theorem

then

if X and Y are any random variables

$$\mathbb{E}X = \mathbb{E}[\mathbb{E}[X|Y]]$$

provided that the expectations exist.

A.k.a. "law of iterated expectation".

Proof.

Suppose $(X, Y) \sim f_{XY}(x, y)$ (continuous setting)
Then

$$\begin{aligned}\mathbb{E}X &= \iint x \underbrace{f_{XY}(x, y)}_{f_{X|Y}(x|y) f_Y(y)} dx dy \\ &= \iint x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int \left(\int x f_{X|Y}(x|y) dx \right) f_Y(y) dy \\ &= \int \mathbb{E}[X|Y=y] f_Y(y) dy \\ &= \mathbb{E}[\mathbb{E}[X|Y]].\end{aligned}$$

□

Def.

A random variable X is said to have a mixture distribution if the distribution of X depends on a quantity that is also random.

Above:

$$X|Y \sim \text{Binomial}(Y)$$

$$Y \sim \text{Poisson}(\lambda)$$

This is called a Poisson-Binomial Mixture.

Another example:

Let

$$Y \sim \text{Bernoulli}(p)$$

$$X|Y=0 \sim \text{Normal}(\mu_1, \sigma_1^2)$$

$$X|Y=1 \sim \text{Normal}(\mu_2, \sigma_2^2)$$

Then

$$P(X \leq x) = P(X \leq x, Y=0) + P(X \leq x, Y=1)$$

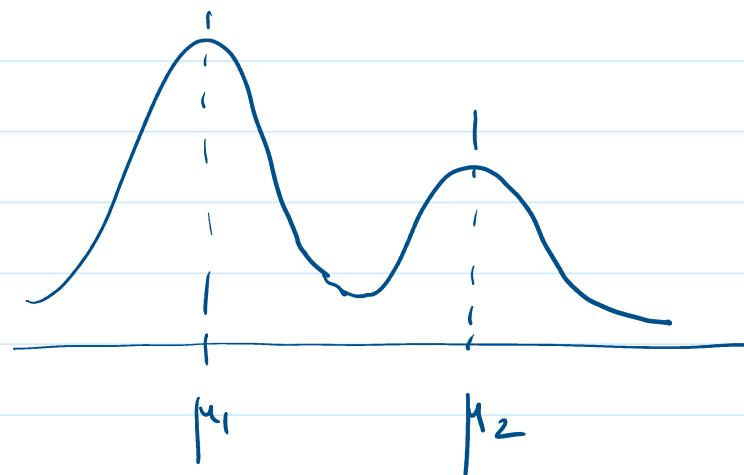
$$= P(Y=0) P(X \leq x | Y=0) + P(Y=1) P(X \leq x | Y=1)$$

$$= p F_{N(\mu_1, \sigma_1^2)}(x) + (1-p) F_{N(\mu_2, \sigma_2^2)}(x)$$

Differentiating wrt x :

$$f_x(x) = p f_{N(\mu_1, \sigma_1^2)}(x) + (1-p) f_{N(\mu_2, \sigma_2^2)}(x).$$

The RHS is the pdf of a mixture of normal distribution



Adding more "modes" to obtain arbitrarily complex distribution

Suppose $Y \in \{1, \dots, k\}$

$$P(Y = i) = p_i \quad i = 1, \dots, k$$

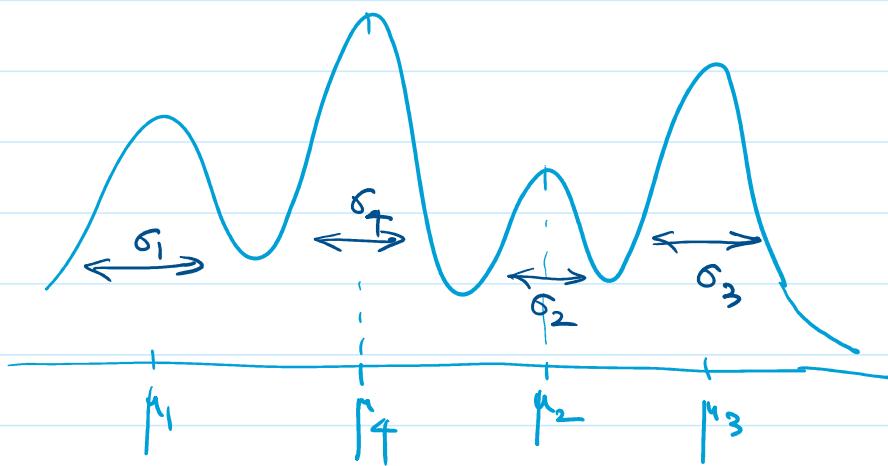
and

$$X|Y=i \sim \text{Normal}(\mu_i, \sigma_i^2).$$

Then the induced marginal distribution for X has the pdf

$$\begin{aligned} f_X(x) &= \sum_{i=1}^k p_i N(x | \mu_i, \sigma_i^2) \\ &= \sum_{i=1}^k p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x - \mu_i)^2\right) \end{aligned}$$

This is a mixture of k normal components.



if a probability distribution for X may be obtained by multiple stage of conditional distributions, then we obtain a **hierarchical model**.

Example : Let

$$\begin{aligned} X | Y &\sim \text{binomial}(Y, p) & (\# \text{ of surviving eggs}) \\ Y | \lambda &\sim \text{Poisson}(\lambda) & (\# \text{ of eggs laid}) \\ \lambda &\sim \text{Exponential}(\beta) \end{aligned}$$

The Randomness of λ captures the variation across the (insect) mothers.

$$\forall x \in \mathbb{N}$$

$$\begin{aligned} P(X=x) &= \sum_y P(X=x, Y=y) \\ &= \sum_{y=0}^{\infty} P(X=x | Y=y) P(Y=y) \end{aligned}$$

where, for $y \in \mathbb{N}$:

$$\begin{aligned} P(Y=y) &= P(Y=y, 0 < \lambda < \infty) \\ &= \int_0^{\infty} P(Y=y | \lambda = \lambda) f_{\lambda}(\lambda) d\lambda \\ &= \int_0^{\infty} f_Y(y | \lambda) f_X(\lambda) d\lambda \end{aligned}$$

$$\begin{aligned}
 &= \int_0^\infty e^{-\lambda} \frac{\lambda^y}{y!} \frac{1}{\beta} e^{-\lambda/\beta} d\lambda \\
 &= \frac{1}{\beta^y y!} \int_0^\infty \lambda^y e^{-\lambda(1+\frac{1}{\beta})} d\lambda \\
 &= \frac{1}{\beta^y y!} \cancel{\Gamma(y+1)} \left(\frac{\beta}{1+\beta}\right)^{y+1}. \\
 &= \frac{1}{1+\beta} \left(\frac{\beta}{1+\beta}\right)^y = p(1-p)^y.
 \end{aligned}$$

So $Y \sim \text{negbinomial}(p = \frac{1}{1+\beta}, r=1)$.
 (geometric)

Hence, the three-stage hierarchical model is equivalent mixture via (X, Y) (with λ integrated out)

$$\begin{aligned}
 Y &\sim \text{negbinomial}(p = \frac{1}{1+\beta}, r=1) \\
 X|Y &\sim \text{binomial}(Y, p).
 \end{aligned}$$

Theorem. "Iterated variance formula."

For any two Random variables X, Y

$$\text{var } X = \mathbb{E} \text{ var}(X|Y) + \text{var } \mathbb{E}(X|Y).$$

Proof.

Note

$$\text{var } X = \mathbb{E} X^2 - (\mathbb{E} X)^2 \quad (0)$$

$$\Rightarrow \text{var}(X|Y) = \mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2$$

$$\Rightarrow \mathbb{E} \text{var}(X|Y) = \mathbb{E} X^2 - \mathbb{E}(\mathbb{E}[X|Y])^2 \quad (1)$$

$$\text{var } \mathbb{E}[X|Y] = \mathbb{E}(\mathbb{E}[X|Y])^2 - \underbrace{(\mathbb{E} \mathbb{E}[X|Y])^2}_{(\mathbb{E} X)^2} \quad (2)$$

Thus

$$\begin{aligned} \text{RHS} &= (1) + (2) \\ &= (0) \end{aligned}$$

□.

Example

Beta-Binomial hierarchy

$$\begin{aligned} \text{Let } X | \theta &\sim \text{Binomial}(n, \theta) \quad \theta \in (0, 1) \\ \theta &\sim \text{beta } (\alpha, \beta) \end{aligned}$$

Then

$$\begin{aligned} P(X=x) &= \int P(X=x|\theta) f_{\Theta}(\theta) d\theta \\ &= \int \binom{n}{x} \theta^x (1-\theta)^{n-x} f_{\Theta}(\theta) d\theta \\ &= \int \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta \\ &= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int \theta^{x+\alpha-1} (1-\theta)^{n+\beta-x-1} d\theta \\ &= \frac{n!}{x!(n-x)!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)} \\ &= \frac{n!}{x!(n-x)!} \frac{(x+\alpha-1)\dots\alpha \times (n-x+\beta-1)\dots\beta}{(n+\alpha+\beta-1)\dots(\alpha+\beta)} \end{aligned}$$

Direct calculation of $\text{E}X$ and $\text{Var } X$ possible
but quite complicated.

Now

$$\begin{aligned} \text{E}X &= \mathbb{E}[\underbrace{\mathbb{E}[x|\theta]}_{n\theta}] \\ &= \mathbb{E}[n\theta] \\ &= n \mathbb{E}\theta \\ &= n\alpha / (\alpha+\beta) \end{aligned}$$

$$\text{var } \mathbb{E}[X|_0] = \text{var } [n\theta] \\ = n^2 \text{var } \theta = \frac{n^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \quad (*)$$

$$\mathbb{E} \text{var}[x|_0] = \mathbb{E} [n\theta(1-\theta)]$$

$$= n \mathbb{E}\theta - n \mathbb{E}\theta^2$$

$$= n \frac{\alpha}{\alpha + \beta} - n \frac{\alpha(\alpha+1)}{(\alpha + \beta)(\alpha + \beta + 1)} \quad (\text{see Sec 3.2 c})$$

$$= \frac{n \alpha \beta}{(\alpha + \beta)(\alpha + \beta + 1)} \quad (*)_f$$

So

$$\text{var } X = (*) + (*)_f$$

$$= \frac{n^2 \alpha \beta + n \alpha \beta (\alpha + \beta)}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

$$= \frac{n \alpha \beta (\alpha + \beta + n)}{(\alpha + \beta)^2 (\alpha + \beta + 1)}.$$

□