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

$\boxed{\text{How to model the (Random) behavior of } X \text{ ?}}$

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.

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

$$P(X=x, Y=y) = \underbrace{P(Y=y)}_{\text{Poisson}} \; \underbrace{P(X=x \mid Y=y)}_{\text{(Conditionally binomial)}}$$

So

$$\begin{aligned}
P(X=x) &= \sum_{y \in \mathbb{N}} P(X=x, Y=y) \\
&= \sum_{y=0}^{\infty} P(Y=y) \underbrace{P(X=x \mid Y=y)} \\
&= \sum_{y=x}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \binom{y}{x} p^{x}(1-p)^{y-x} \\
&= \sum_{y=x}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{\cancel{y!}} \frac{\cancel{y!}}{(y-x)!\,x!} p^{x}(1-p)^{y-x} \\
&= \frac{\lambda^{x} p^{x}}{x!} e^{-\lambda} \sum_{y=x}^{\infty} \frac{\lambda^{y-x}}{(y-x)!} (1-p)^{y-x}
\end{aligned}$$

$t = y-x$

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

Hence $\boxed{X \sim \text{Poison}(\lambda p).}$

$$\Rightarrow \begin{cases} \mathbb{E}X = \lambda p \\ \mathbb{E}[X \mid Y=y] = yp \qquad \forall y \in \mathbb{N}. \end{cases}$$

Note

Conditional expectation $\mathbb{E}[X|Y]$ is a random variable, and here, $\mathbb{E}[X|Y] = Yp$

thus $\mathbb{E}\,\mathbb{E}[X|Y] = \mathbb{E}(Yp) = p\,\mathbb{E}Y = \lambda p.$

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 if X and Y are any Random variables then

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

provided than the expectations exist.

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

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

Then

$$\mathbb{E}X = \iint x \underbrace{f_{XY}(x,y)} \, dx \, dy$$

$$= \iint x \, f_{X|Y}(x|y) \, f_Y(y) \, dx \, dy$$

$$= \int \left(\int x f_{X|Y}(x|y) \, dx \right) f_Y(y) \, dy$$

$$= \int \mathbb{E}[X|Y=y] \, f_Y(y) \, dy$$

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

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

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

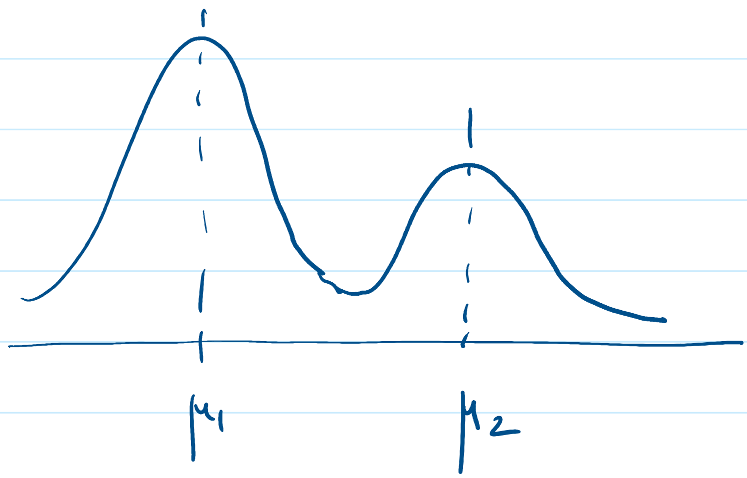
$$= P(Y=0)\, P(X \le x \mid Y=0) + P(Y=1)\, P(X \le x \mid Y=1)$$

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

Differentiating wrt x:

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

The RHS is the pdf of a mixture of normal distribution



Adding more "modes" to obtain arbitrarily complex distribution

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

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

and

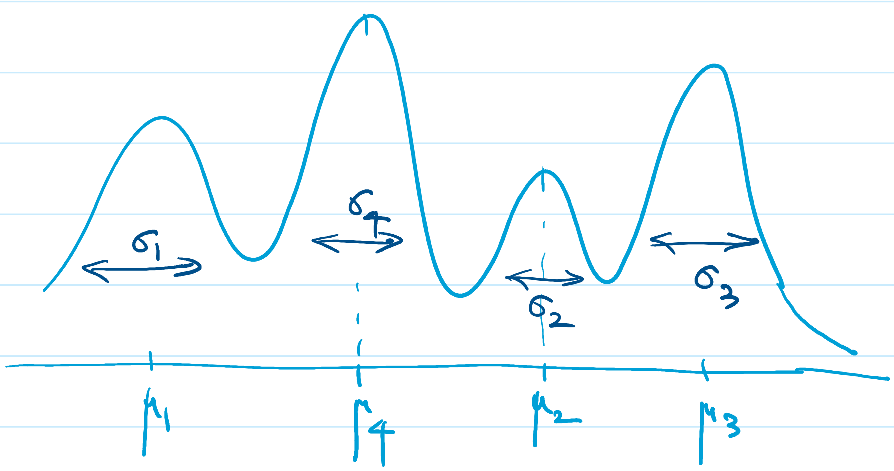
$$X \mid Y = i \sim \text{Normal}(\mu_i, \sigma_i^2).$$

Then the induced marginal distribution for X has the pdf

$$f_X(x) = \sum_{i=1}^{k} p_i \, N(x \mid \mu_i, \sigma_i^2)$$

$$= \sum_{i=1}^{k} p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(-\frac{1}{2\sigma_i^2}(x - \mu_i)^2\right)$$

This is a $\boxed{\text{mixture of } k \text{ normal components.}}$



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

Example. Let

$$X \mid Y \sim \text{binomial}(Y, p) \qquad (\#\text{ of surviving eggs})$$
$$Y \mid \Lambda \sim \text{Poisson}(\Lambda) \qquad (\#\text{ of eggs laid})$$
$$\Lambda \sim \text{Exponential}(\beta)$$

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

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

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

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

$$P(Y=y) = P(Y=y,\ 0<\Lambda<\infty)$$
$$= \int_0^{\infty} P(Y=y \mid \Lambda=\lambda)\, f_\Lambda(\lambda)\, d\lambda$$
$$= \int_0^{\infty} f_Y(y \mid \lambda)\, f_\lambda(\lambda)\, d\lambda$$

$$= \int_0^\infty e^{-\lambda} \frac{\lambda^y}{y!} \frac{1}{\beta} e^{-\lambda/\beta} \, d\lambda$$

$$= \frac{1}{\beta y!} \int_0^\infty \lambda^y e^{-\lambda(1+\frac{1}{\beta})} \, d\lambda$$

$$= \frac{1}{\beta \cancel{y!}} \cancel{\Gamma(y+1)} \left(\frac{\beta}{1+\beta}\right)^{y+1}.$$

$$= \frac{1}{1+\beta} \left(\frac{\beta}{1+\beta}\right)^y = p(1-p)^y.$$

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

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

$$Y \sim \text{negbinomial}\left(p = \frac{1}{1+\beta}, r = 1\right)$$
$$X|Y \sim \text{binomial}(Y, p).$$

Theorem. "Iterated variance formula."

For any two random variables X, Y

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

Proof.

Note

$$\operatorname{var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 \qquad (0)$$

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

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

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

Thus

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

$\square$.

Example Beta-Binomial hierachy

Let $X | \theta \sim \text{Binomial}(n, \theta) \qquad \theta \in (0,1)$

$\theta \sim \text{beta}(\alpha, \beta)$

Then

$$P(X=x) = \int P(X=x \mid \theta)\, f_{\Theta}(\theta)\, d\theta$$

$$= \int \binom{n}{x} \theta^{x}(1-\theta)^{n-x}\, f_{\Theta}(\theta)\, d\theta$$

$$= \int \binom{n}{x} \theta^{x}(1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int \theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}\, dx$$

$$= \frac{n!}{x!(n-x)!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$= \frac{n!}{x!(n-x)!} \frac{\overbrace{(x+\alpha-1)\cdots\alpha}^{x-1} \times \overbrace{(n-x+\beta-1)\cdots\beta}^{n-x \text{ terms}}}{\underbrace{(n+\alpha+\beta-1)\cdots(\alpha+\beta)}_{n-1 \text{ terms}}} .$$

Direct calculation of $\mathbb{E}X$ and $\operatorname{var} X$ ... possible but quite complicated.

Now

$$\mathbb{E}X = \mathbb{E}[\underbrace{\mathbb{E}[X \mid \theta]}]$$

$$= \mathbb{E}[n\theta]$$

$$= n\,\mathbb{E}\theta$$

$$= n\alpha/(\alpha+\beta) .$$

$$\operatorname{var} \mathbb{E}[X|\theta] = \operatorname{var}[n\theta]$$

$$= n^2 \operatorname{var}\theta = \frac{n^2 \alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (*)$$

$$\mathbb{E}\operatorname{var}[X|\theta] = \mathbb{E}[N\theta(1-\theta)]$$

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

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

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

So

$$\operatorname{var}X = (*) + (**)$$

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

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