# 4 Hierarchical Models and Mixture Distributions

**Example 4.1** *(Binomial-Poisson hierarchy) Perhaps the most classic hierarchical model is the following. An insect lays a large number of eggs, each surviving with probability $p$. On the average, how many eggs will survive?*

*The large number of eggs laid is a random variable, often taken to be Poisson($\lambda$). Furthermore, if we assume that each egg's survival is independent, then we have Bernoulli trials. Therefore,, if we let $X$=number of survivors and $Y$=number of eggs laid, we have*

$$X|Y\, binomial(Y, p), \qquad Y \sim Poisson(\lambda),$$

*a hierarchical model.*

The advantage of the hierarchy is that complicated process may be modeled by a sequence of relatively simple models placed in a hierarchy.

**Example 4.2** *(Continuation of Example 4.1) The random variable $X$ has the distribution given by*

$$\begin{aligned}
P(X = x) &= \sum_{y=0}^{\infty} P(X = x, Y = y) = \sum_{y=0}^{\infty} P(X = x|Y = y)P(Y = y) \\
&= \sum_{y=x}^{\infty} [\binom{y}{x} p^x (1-p)^{y-x}][\frac{e^{-y}\lambda^y}{y!}] \quad (\textit{conditional probability is 0 if } y < x) \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda p},
\end{aligned}$$

*so $X \sim Poisson(\lambda)$. Thus, any marginal inference on $X$ is with respect to a Poisson($\lambda p$) distribution, with $Y$ playing no part at all. Introducing $Y$ in the hierarchy was mainly to aid our understanding of the model. On the average,*

$$EX = \lambda p$$

*eggs will survive.*

Sometimes, calculations can be greatly simplified be using the following theorem.

**Theorem 4.1** *If $X$ and $Y$ are any two random variables, then*

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

*provided that the expectations exist.*

PROOF: Let $f(x, y)$ denote the joint pdf of $X$ and $Y$. By definition, we have

$$EX = \int \inf x f(x, y) dx dy = \int [\int x f(x|y) dx] f_Y(y) dy$$
$$\int E(X|y) f_Y(y) dy = E(E(X|Y))$$

Replacing integrals by sums to prove the discrete case. □

Using Theorem 4.1, we have

$$EX = E(E(X|Y)) = E(pY) = p\lambda$$

for Example 4.2.

**Definition 4.1** *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.*

Thus, in Example 4.1 the Poisson($\lambda p$) distribution is a mixture distribution since it is the result of combining a binomial($Y, p$) with $Y \sim$ Poisson($\lambda$).

**Theorem 4.2** *(Conditional variance identity) For any two random variables $X$ and $Y$,*

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

*provided that the expectations exist.*

PROOF: By definition, we have

$$\begin{aligned}\mathrm{Var}X &= E([X - EX]^2) = E([X - E(X|Y) + E(X|Y) - EX]^2) \\ &= E([X - E(X|Y)]^2) + E([E(X|Y) - EX]^2) + 2E([X - E(X|Y)][E(X|Y) - EX]).\end{aligned}$$

The last term in this expression is equal to 0, however, which can easily be seen by iterating the expectation:

$$E([X - E(X|Y)][E(X|Y) - EX]) = E(E\{[X - E(X|Y)][E(X|Y) - EX]|Y\})$$

In the conditional distribution $X|Y$, $X$ is the random variable. Conditional on $Y$, E(X—Y) and $EX$ are constants. Thus,

$$E\{[X - E(X|Y)][E(X|Y) - EX]|Y\} = (E(X|Y) - E(X|Y))(E(X|Y) - EX) = 0$$

Since

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

and

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

Theorem 4.2 is proved. □

**Example 4.3** *(Beta-binomial hierarchy) One generalization of the binomial distribution is to allow the success probability to vary according to a distribution. A standard model for this situation is*

$$X|P \sim \textit{binomial}(P), \quad i = 1, \ldots, n,$$
$$P \sim \textit{beta}(\alpha, \beta).$$

*The mean of $X$ is then*

$$EX = E[E(X|p)] = E[nP] = \frac{n\alpha}{\alpha + \beta}.$$

*Since $P \sim beta(\alpha, \beta)$,*

$$Var(E(X|P)) = Var(np) = n^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

*Also, since $X|P$ is binomial$(n, P)$, $Var(X|P) = nP(1-P)$. We then have*

$$\begin{aligned} E[Var(X|P)] = nE[P(1-P)] &= n\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p(1-p)p^{\alpha-1}(1-p)^{\beta-1} dp \\ &= n\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} = \frac{n\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}. \end{aligned}$$

*Adding together the two pieces, we get*

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

## 5 Covariance and Correlation

In earlier sections, we have discussed the absence or presence of a relationship between two random variables, Independence or nonindependence. But if there is a relationship, the relationship may be strong or weak. In this section, we discuss two numerical measures of the strength of a relationship between two random variables, the covariance and correlation.

Throughout this section, we will use the notation $EX = \mu_X$, $EY = \mu_Y$, $\mathrm{Var}X = \sigma_X^2$, and $\mathrm{Var}Y = \sigma_Y^2$.

**Definition 5.1** *The covariance of $X$ and $Y$ is the number defined by*

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)).$$

**Definition 5.2** *The correlation of $X$ and $Y$ is the number defined by*

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

*The value $\rho_{XY}$ is also called the correlation coefficient.*

**Theorem 5.1** *For any random variables $X$ and $Y$,*

$$Cov(X,Y) = EXY - \mu_X \mu_Y.$$