Hierarchical models. Multivariate distributions

Anastasiia Kim

April 10, 2020
Hierarchical models

Complicated process may be modeled by a sequence of relatively simple models placed in a hierarchy

*Binomial-Poisson hierarchy.* An insect lays a large number of eggs, each surviving with probability $p$. On the average, how many eggs will survive? Assume that each egg’s survival is independent.

- Let $Y$ be the number of eggs and $X$ be the number of survivors ($X$ and $Y$ are random variables)
- First model the distribution of $Y$. Then model the distribution of $X$ given $Y$
  - the large number of eggs laid is often modeled with Poisson distribution, $Y \sim \text{Poisson}(\lambda)$, where $\lambda > 0$ is the average number of eggs laid.
  - given $Y$, the number of survivors can be modeled as $X|Y \sim \text{Binomial}(Y, p)$
Hierarchical models

*Binomial-Poisson hierarchy.* An insect lays a large number of eggs, each surviving with probability $p$. On the average, how many eggs will survive? Assume that each egg’s survival is independent.

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

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

is a hierarchical model. Find pdf of $X$ and $E(X)$.

Given that the conditional probability is 0 if $y < x$, the random variable $X$ has the distribution given by

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

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

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


Hierarchical models

If $X$ and $Y$ are any two r.v.s, then

$$E(X) = E[E(X|Y)]$$

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

In our example,

$$E(X) = E[E(X|Y)] = E[Yp] = pE(Y) = p\lambda$$

which is the expected value for $Poisson(\lambda p)$. 
Hierarchical models

- Hierarchical models can have more than two stages.
- The random variables in hierarchical models may be all discrete, all continuous, or some discrete and some continuous.

**Beta-Binomial hierarchy** One generalization of the binomial distribution is to allow the success probability to vary according to a distribution from trial to trial. A standard model for this situation is

\[
X_i | P_i \sim \text{Binomial}(n_i, P_i)
\]

\[
P_i \sim \text{Beta}(\alpha, \beta), \quad i = 1, \ldots, n
\]

- a certain machine produces defective and nondefective parts, but we do not know what proportion of defectives we would find among all parts that could be produced by the machine. \(X | P \sim \text{Binomial}(n, P)\). We might believe that \(P\) has a continuous distribution.
- when measuring the success of a drug on patients, it is better not to assume that the success probabilities are constant because the patients are different.
The Multinomial distribution is a generalization of the Binomial.

Whereas the Binomial distribution counts the successes in a fixed number of trials that can only be categorized as success or failure.

The Multinomial distribution keeps track of trials whose outcomes can fall into multiple categories: such as excellent, adequate, poor; or red, yellow, green, blue.
Multinomial distribution. Definition

- Each of n objects is independently placed into one of k categories
- An object is placed into category j with probability $p_j$, where $\sum_{j=1}^{k} p_j = 1$
- Let $X_1$ be the number of objects in category 1, $X_2$ the number of objects in category 2, etc., so that $X_1 + ... + X_k = n$
- Then r.v.s $X_1, X_2, ...$ have the Multinomial distribution with parameters $n$ and $p = (p_1, ..., p_k)$ and the joint probability mass function is

$$P(X_1 = n_1, X_2 = n_2, ..., X_k = n_k) = \frac{n!}{n_1!n_2!...n_k!}p_1^{n_1}p_2^{n_2}...p_k^{n_k}$$

for $n_1, n_2, ..., n_k$ satisfying $n_1 + n_2 + ... + n_k = n$. 
Example

Of 20 graduating students, how many ways are there for 12 to be employed in a job related to their field of study, 6 to be employed in a job unrelated to their field of study, and 2 unemployed?

\[
\binom{20}{12} \binom{8}{6} \binom{2}{2} = \frac{20!}{12!6!2!} = 3,527,160
\]

What if now probabilities are different

- probability of job related to field of study is 0.70
- probability of job unrelated to field of study is 0.20
- probability of no job is 0.10

Then this probability is (using multinomial joint pmf)

\[
\frac{20!}{12!6!2!} 0.70^{12} 0.20^6 0.10^2 = 0.03
\]

in R: `dmultinom(c(12, 6, 2), prob = c(.7, .2, .1))`
Example

- Given that a student finds a job, what is the probability that the job will be in the student's field of study?

\[
P(\text{Field} | \text{Job}) = \frac{P(\text{Field, Job})}{P(\text{Job})} = \frac{0.7}{0.7 + 0.2} = \frac{7}{9} = 0.78
\]

- Suppose we choose 30 students at random from those who found jobs. What is the probability that exactly \( s \) of them will be employed in their field of study, for \( s = 0, \ldots, 30 \)?

\[
P(s | \text{Job}) = \binom{30}{s} \left( \frac{7}{9} \right)^s \left( 1 - \frac{7}{9} \right)^{30-s}
\]
The Multivariate Normal (MVN) is a continuous multivariate distribution that generalizes the Normal distribution into higher dimensions.

- The r.v.s $X_1, X_2, \ldots$ have the MVN distribution if every linear combination of the $X_j$ has a Normal distribution. That is, we require

$$a_1 X_1 + \ldots + a_k X_k$$

to have a Normal distribution for any constants $a_1, \ldots, a_k$.

- An important special case is $k = 2$; this distribution is called the Bivariate Normal.

- The joint MVN depends on means and covariance matrix that gives the covariance between each pair of r.v.s.

- If $X_1, X_2, \ldots$ have the Multivariate Normal distribution, then the marginal distribution of each $X_j$ is Normal.

- The converse is false: it is possible to have Normally distributed r.v.s $X_1, \ldots, X_k$ such that $(X_1, \ldots, X_k)$ is not Multivariate Normal.
The multivariate normal distribution (MVN) is useful in analyzing the relationship between multiple normally distributed variables.

MVN has heavy application to biology and economics where the relationship between approximately-normal variables is of great interest.

MVN is used to learn the statistics of the local features (for example, in detecting faces in images).
Joint pdfs of two Bivariate Normal distributions

- If \((X, Y)\) is Bivariate Normal and \(\text{Corr}(X, Y) = 0\), then \(X\) and \(Y\) are independent.
- \(X\) and \(Y\) are marginally Normal
- In the figure, both \(X\) and \(Y\) are \(N(0, 1)\)
- On the left, \(X\) and \(Y\) are uncorrelated, so the level curves of the joint PDF are circles
- On the right, \(X\) and \(Y\) have a correlation of 0.75, so the level curves are ellipsoidal, reflecting the fact that \(Y\) tends to be large when \(X\) is large