Lecture 13

Revision

Multinomial distribution

Conditional independence

Examples

6: One-dimensional marginals

$$p_{X_1}(x) = \sum_{x_2} \cdots \sum_{x_n} p_{X_1,X_2,\ldots,X_n}(x, x_2, \ldots, x_n).$$

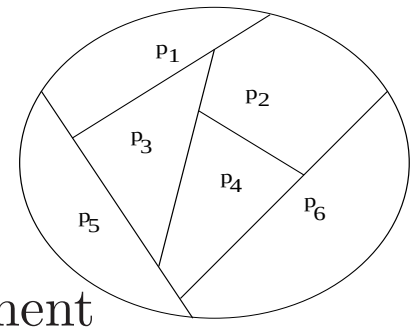
7: k-dimensional marginals

$$p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \sum_{x_{k+1}} \cdots \sum_{x_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

8: Conditional distributions

$$\begin{aligned} & p_{(X_1,\ldots,X_k)|(X_{k+1},\ldots,X_n)}(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n) \\ &= \frac{p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{p_{X_{k+1},\ldots,X_n}(x_{k+1}, \ldots, x_n)} \end{aligned}$$

Multinomial distribution



- $m \geq 2$ categories
- $n \geq 1$ items chosen at random, with replacement
- $p_k = P\{\text{Item of type } k \text{ chosen}\}, \quad k = 1, \ldots, m.$
- $N_k = \text{Number of type } k \text{ chosen}, \quad k = 1, \ldots, m.$

$$
\begin{aligned}
&P\{N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m\} \\
&\quad = \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\ldots-n_{m-1}}{n_m} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m} \\
&\quad = \frac{n!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}
\end{aligned}
$$

as long as $n_1 + n_2 + \ldots + n_m = n$.

- $m = 2$ gives the Binomal distribution

$$P\{N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m\} = \frac{n!}{n_1! n_2! \cdots n_m!} \, p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

- "Marginal" distributions are again multinomial.
- Example: One dimensional marginals.

$$\begin{aligned} P\{N_1 = n_1\} &= \sum_{n_2 + \ldots + n_m = n - n_1} P\{N_1 = n_1, \ldots, N_m = n_m\} \\ &= ???? \end{aligned}$$

Alternatively, pool all the "other" types together, so that the problem reduces to asking about n_1 "successes" out of n Bernoulli trials. Thus,

$$N_1 \sim \text{Bin}(n, p_1)$$

- The same pooling argument works for all other marginals.

More facts about the multinomial

- $\Rightarrow$ One dimensional marginals are binomial:
  i.e. $N_k \sim Bin(n, p_k), \quad k = 1, \ldots, m$
- All of this is obvious, and needs no computation!
- Also obvious (?) multidimensional <u>conditionals</u> are multinomial, but with what parameters?

$$P\{N_1 = n_1, \ldots, N_k = n_k \big| N_{k+1} = n_{k+1}, \ldots, N_m = n_m\}$$

$$= \frac{\frac{n!}{n_1! n_2! \cdots n_m!}\, p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}}{\frac{n!}{(n_1+\cdots+n_k)! n_{k+1}! \cdots n_m!}\, (p_1 + \cdots + p_k)^{n_1+\cdots+n_k} p_{k+1}^{n_{k+1}} \cdots p_m^{n_m}}$$

$$= \frac{(n_1 + \cdots + n_k)!}{n_1! n_2! \cdots n_k!} \left(\frac{p_1}{p_1 + \cdots + p_k}\right)^{n_1} \cdots \left(\frac{p_k}{p_1 + \cdots + p_k}\right)^{n_k}$$

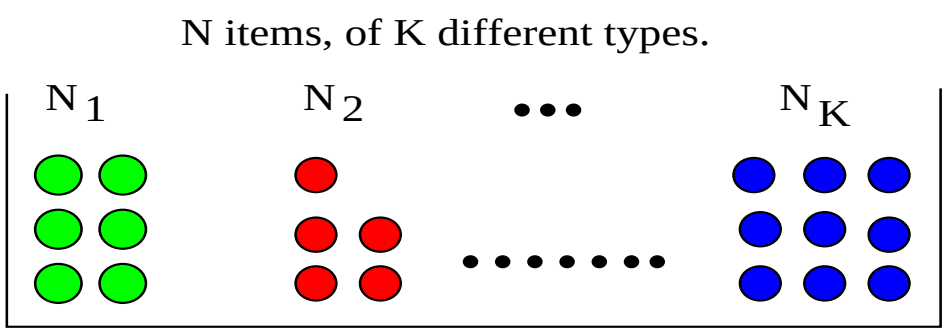
Still facts about the multinomial

- The N_k are dependent.

- The multinomial is a large population approximation to the hypergeometric probabilities, in which n items are chosen <u>without</u> replacement, and we saw (Lecture 9)

$$P\{n_j \text{ of type } j,\ j = 1, \ldots, m\ \} \ = \ \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_m}{n_m}}{\binom{N}{n}}$$

for $n = n_1 + n_2 + \ldots + n_m$, after setting $p_j = N_j/N$.



Conditional independence

Events Events A and B are called conditionally independent given C iff

$$P\{A \cap B \big| C\} \;=\; P\{A \big| C\} \cdot P\{B \big| C\}$$

Random variables Discrete random variables X and Y are called conditionally independent given Z iff

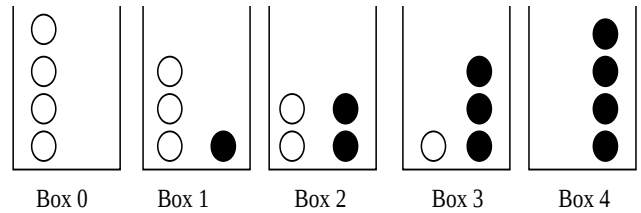
$$P\{X = x,\; Y = y \big| Z = z\} \;=\; P\{X = x \big| Z = z\} \cdot P\{Y = y \big| Z = z\}$$

for all x, y, z.

- Note: Any or all of X, Y, Z may be multidimensional
- Independence $\Rightarrow$ conditional independence.
  The reverse is not necessarily true.

Examples

- 5 boxes, with black and white balls as shown. Choose a box at random, and then balls, <u>with replacement</u>.



  - $C \in \{0, 1, 2, 3, 4\}$ is the number of the box.
  - $X_k = 0$ if k-th ball chosen is black. $X_k = 1$ if it is white
  - We have already seen (Lecture 4/Next slide) that

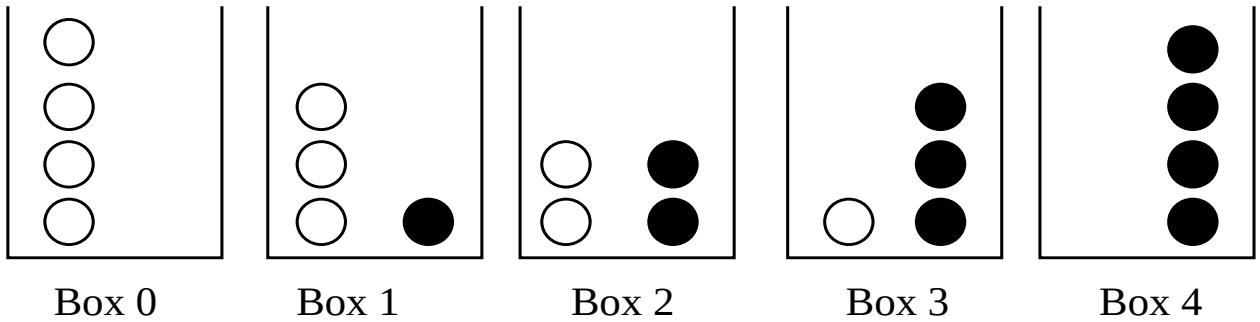
$$P\{X_1 = 1\} = P\{X_2 = 1\} = \frac{1}{2}, \qquad P\{X_1 = 1,\ X_2 = 1\} \ = \ \frac{3}{8},$$

    and so X_1 and X_2 are not independent.

  - But they are obviously independent given any value of C

Lecture 4 - Example 4

- 5 boxes, with black and white balls as shown. I choose a box at random, and then two balls, with replacement. What is the probability that <u>both</u> are black?



$$
\begin{aligned}
P\{\text{Both black}\} &= \sum_{k=0}^{4} P\{\text{Both black}|\text{Box } k\} \times P\{\text{Box } k\} \\
&= \sum_{k=2}^{4} \left(\frac{k}{4}\right)^2 \frac{1}{5} = \frac{1}{5}\left[\left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 + \left(\frac{3}{4}\right)^2 + 1\right] \\
&= \frac{3}{8}
\end{aligned}
$$

"Changing" dependencies

- C is more dependent on $(X_1, \ldots, X_{n+1})$ than it is on $(X_1, \ldots, X_n)$

$$
\begin{aligned}
& P\{C = k \big| X_1 = \ldots = X_n = 0\} \\
= \quad & \frac{P\{X_1 = \ldots = X_k = 0 \big| C = k\} \cdot P\{C = k\}}{\sum_{j=0}^{4} P\{X_1 = \ldots = X_n = 0 \big| C = j\} \cdot P\{C = j\}} \\
= \quad & \frac{\frac{1}{5}(k/4)^n}{\frac{1}{5}\sum_{0=1}^{4}(j/4)^n} \\
= \quad & \frac{(k/4)^n}{(1/4)^n + (2/4)^n + (3/4)^n + (4/4)^n} \\
\rightarrow \quad & \begin{cases} 1 & \text{if } k = 4 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- Correlation: Lecture 19

Order statistics (again)

- Choose 3 numbers at random, without replacement, from $\{1, 2, \ldots, N\}$
- $X = X_{(1)},\ Y = X_{(2)},\ Z = X_{(3)}$

$$
\begin{aligned}
P\{X = x, Y = y, Z = z\} &= \frac{1}{\binom{N}{3}} \\
P\{Y = y\} &= \frac{(y-1)\cdot(N-y)}{\binom{N}{3}} \\
P\{X = x, Z = z \big| Y = y\} &= \frac{P\{X = x, Y = y, Z = z\}}{P\{Y = y\}} \\
&= \frac{1}{(y-1)(N-y)} \qquad 1 \le x < y < z \le N.
\end{aligned}
$$

$$P\{X = x, Z = z \big| Y = y\} \quad = \quad \frac{1}{(y-1)(N-y)} \qquad 1 \le x < y < z \le N.$$

- Consequently

$$\begin{aligned} P\{X = x \big| Y = y\} &= \sum_{z=y+1}^{N} P\{X = x, Z = z \big| Y = y\} \\ &= \sum_{z=y+1}^{N} \frac{1}{(y-1)(N-y)} \\ &= 1/(y-1) \\ \text{and} \quad P\{Z = z \big| Y = y\} &= 1/(N-y) \end{aligned}$$

- Thus, given Y, X and Z are conditionally independent:

$$X \sim U[1, \ldots, Y-1], \qquad Z \sim U[Y+1, \ldots, N]$$