Multinomial Distribution

We shall now study about 3 families of multivariate distributions. Among these, one will be a ~~discrete~~ family of discrete joint distn.s called the multinomial distribution, which is a generalization of binomial distribution.

Recall: Bernoulli trials and binomial distn.

Bernoulli trial has two possible outcomes - S (success) and F (failure). Let X be the number of successes out of n $(\in \mathbb{N})$ independent Bernoulli trials with constant success probability $p \in (0,1)$. Then we know that $X \sim Bin(n,p)$. In this situation $n-X \sim Bin(n, 1-p)$ and it counts the number of failures out of these n independent Bernoulli trials.

Define $X_1 = X$ and $X_2 = n - X_1 = n - X$.

Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Then $\underset{\sim}{X} = (X_1, X_2)$ is a discrete random vector with $\text{Range}(\underset{\sim}{X}) = \{(x_1, x_2) : x_1, x_2 \in \mathbb{N}_0, x_1 + x_2 = n\}$ $= \{(\ell, n-\ell) : \ell \in \{0, 1, 2, \ldots, n\}\} \doteq R$. Also, the joint pmf of $\underset{\sim}{X}$ is

$$p_{X_1,X_2}(x_1,x_2) = \begin{cases} \dfrac{n!}{x_1!\, x_2!} \; p^{x_1}(1-p)^{x_2} & \text{if } (x_1,x_2)\in R, \\ 0 & \text{if } (x_1,x_2)\notin R. \end{cases}$$

Multinomial distn is the k-dimensional ~~case~~ generalization of the above example.

Setup: Suppose a trial has $k\ (\geq 2)$ possible disjoint and exhaustive outcomes $1, 2, 3, \ldots, k$. (a prototypical example would be throwing a die, in which case $k=6$.) with probabilities $p_1, p_2, p_3, \ldots, p_k$, resp. Clearly, each $p_i \geq 0$ and $p_1+p_2+\cdots+p_k = 1$. Assume each $p_i > 0$. Suppose this trial is repeated n times $(n\in\mathbb{N})$ ~~$(n\in\mathbb{N})$~~ independently. Define

$X_1 :=$ no. of times the outcome 1 appears,

$X_2 :=$ no. of times the outcome 2 appears,

$\vdots$

$X_k :=$ no. of times the outcome k appears.

Defn: In the above setup, the k-dimensional discrete random vector $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$ is said to follow a multinomial distn. with parameters n and $p_1, p_2, \ldots, p_k$.

Notation: $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k) \sim \text{Mult}(n; p_1, p_2, \ldots, p_k)$.

Note that $X \sim \text{Bin}(n, p) \Rightarrow (X, n-X) \sim \text{Mult}(n; p, 1-p)$.

Clearly, $\text{Range}(\underset{\sim}{X}) = R$

$$:= \{(x_1, x_2, \ldots, x_k) \in \mathbb{N}_0^k : x_1 + x_2 + \cdots + x_k = n\}.$$

Here $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The joint pmf of $\underset{\sim}{X}$ is given by

$$p_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) = \begin{cases} \dfrac{n!}{x_1!\, x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} & \text{if } (x_1, x_2, \ldots, x_k) \in R, \\ 0 & \text{otherwise.} \end{cases}$$

Exc: Verify this.

Remark: Since the value of the above pmf is a term in the multinomial expansion of $(p_1 + p_2 + \cdots + p_k)^n$, this distn. is known as the multinomial distn.

Note that if the random vector $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k) \sim \text{Mult}(n; p_1, p_2, \ldots, p_k)$, then marginally, each $X_i \sim \text{Bin}(n, p_i)$ and hence $E(X_i) = np_i$ and $Var(X_i) = np_i(1-p_i)$ for each $i = 1, 2, \ldots, k$.

Exc: Using the "indicator method", show that $\forall$ i, j with $1 \leq i < j \leq k$,

$$Cov(X_i, X_j) = -np_i p_j .$$

In particular, $\forall$ i, j with $1 \leq i < j \leq k$,

$$\rho(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}} .$$

Interpret the sign of the above correlation.

Remark: Note that if $(X_1, X_2) \sim \text{Mult}(12; p_1, p_2)$, then $\rho(X_1, X_2) = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}} = -1$. This complete negative linear association, ~~arises~~ is observed because, in this case, $X_1 + X_2 \equiv 12$.

Dirichlet Distribution

This is the first family of jointly continuous distⁿs that we shall study. In order to define this family, we need the multivariate change of joint density formula stated below.

Thm: (Change of Multivariate Joint Density Formula)

Suppose $I, J \subseteq \mathbb{R}^k$ are two open path-connected sets and $g : I \to J$ is a bijective and "smooth" (as described below) map with $g^{-1} : J \to I$ of the form

$$g^{-1}(\underset{\sim}{y}) = (h_1(\underset{\sim}{y}), h_2(\underset{\sim}{y}), \ldots, h_k(\underset{\sim}{y})), \quad \underset{\sim}{y} \in J.$$

In other words, for each $i \in \{1, 2, \ldots, k\}$, $h_i : J \to \mathbb{R}$ is the ith component of the map g^{-1}. We assume that all the partial derivatives $\frac{\partial h_i}{\partial y_j}$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, k$ exist and are cont on J, and the determinant

$$\frac{d g^{-1}(\underset{\sim}{y})}{d \underset{\sim}{y}} := \det\left(J_{g^{-1}}(\underset{\sim}{y})\right) \neq 0 \quad \forall\, \underset{\sim}{y} \in J,$$

where $J_{g^{-1}}(\underset{\sim}{y}) \coloneqq \left(\left(\frac{\partial h_i}{\partial y_j}\right)\right)_{\substack{1 \le i \le k \\ 1 \le j \le k}}$ is the Jacobian matrix of the map g^{-1}. ~~If $\underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$~~

If $\underset{\sim}{X} = (X_1, X_2, \dots, X_k)$ is a cont random vector with a joint pdf $f_{\underset{\sim}{X}}$ that vanishes on I^c (this means $\text{Range}(\underset{\sim}{X}) \subseteq I$), then

$$\underset{\sim}{Y} = (Y_1, Y_2, \dots, Y_k) := g(\underset{\sim}{X}) = g(X_1, X_2, \dots, X_k)$$

is also a cont random vector with a joint pdf

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \begin{cases} f_{\underset{\sim}{X}}(g^{-1}(\underset{\sim}{y})) \left| \dfrac{dg^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}} \right| & \text{if } \underset{\sim}{y} \in J, \\ 0 & \text{if } \underset{\sim}{y} \notin J. \end{cases}$$

~~Here~~ Here $\dfrac{dg^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}}$ is the determinant defined in Pg 231.

Remarks: 0 This is simply the k-dimensional generalization of change of bivariate joint density formula given in Pg 133.

1 Please revisit Remarks 1 - 4 of Pg 134 - they all have k-dimensional analogues in light of the theorem stated in Pg 231 - 232.

Example: Suppose $k \geq 2$ and $X_1, X_2, \dots, X_k$ are ind r.v.s such that for each $i \in \{1, 2, \dots, k\}$, $X_i \sim \text{Gamma}(\alpha_i, \lambda)$.

Define $S := X_1 + X_2 + \dots + X_k$,

$$Y_1 := \frac{X_1}{S} = \frac{X_1}{X_1 + X_2 + \dots + X_k},$$

$$Y_2 := \frac{X_2}{S} = \frac{X_2}{X_1 + X_2 + \dots + X_k},$$

$$\vdots$$

$$Y_{k-1} := \frac{X_{k-1}}{S} = \frac{X_{k-1}}{X_1 + X_2 + \dots + X_k}.$$

Find the joint distn of $(Y_1, Y_2, \dots, Y_{k-1}, S)$.

Remark: For $k=2$, this example boils down to the last exc given in Pg 139:

$$\text{ind} \begin{cases} X_1 \sim \text{Gamma}(\alpha_1, \lambda) \\ X_2 \sim \text{Gamma}(\alpha_2, \lambda) \end{cases} \Longrightarrow \text{ind} \begin{cases} S := X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda) \\ Y_1 := \dfrac{X_1}{X_1 + X_2} = \dfrac{X_1}{S} \sim \text{Beta}(\alpha_1, \alpha_2) \end{cases}$$

(the independence of S and Y_1 in this exc can be shown to follow from Basu's theorem, a deep and important result in theoretical statistics).

Solution: Note that because of independence of $X_1, X_2, \cdots, X_k$,

$$\text{Range}(X_1, X_2, \cdots, X_k) = \prod_{i=1}^{k} \text{Range}(X_i) = (0,\infty)^k.$$

We shall, therefore, take $I := (0,\infty)^k \subseteq \mathbb{R}^k$ and it is clearly open and path-connected.

Define $g : I \longrightarrow \mathbb{R}^k$ by

$$g(x_1, x_2, \cdots, x_k) = \left(\frac{x_1}{s}, \frac{x_2}{s}, \cdots, \frac{x_{k-1}}{s}, s\right),$$

where $s = x_1 + x_2 + \cdots + x_k$. In other words,

$$g(x_1, x_2, \cdots, x_k) = \left(\frac{x_1}{\sum_{i=1}^{k} x_i}, \frac{x_2}{\sum_{i=1}^{k} x_i}, \cdots, \frac{x_{k-1}}{\sum_{i=1}^{k} x_i}, \sum_{i=1}^{k} x_i\right),$$

$$(x_1, x_2, \cdots, x_k) \in I = (0,\infty)^k.$$

Question: What is $g(I) \doteqdot J$?

Clearly, $g(I)$

$$\subseteq \{(y_1, y_2, \cdots, y_{k-1}, s) \in (0,\infty)^k : y_1 + y_2 + \cdots + y_{k-1} < 1\}$$

$$= \{(y_1, y_2, \cdots, y_{k-1}) \in (0,\infty)^{k-1} : y_1 + y_2 + \cdots + y_{k-1} < 1\} \times (0,\infty).$$

Note that for all $(y_1, y_2, \dots, y_{k-1}) \in (0,\infty)^{k-1}$ such that $y_1+y_2+\cdots+y_{k-1} < 1$ and for all $s \in (0,\infty)$, solving the equation (for $\underset{\sim}{x} = (x_1, x_2, \dots, x_k)$)

$$g(x_1, x_2, \dots, x_k) = (y_1, y_2, \dots, y_{k-1}, s),$$

we get

$$g\left(y_1 s, y_2 s, \dots, y_{k-1} s, \left(1-\sum_{i=1}^{k-1} y_i\right) s\right) = (y_1, y_2, \dots, y_{k-1}, s),$$

and moreover

$$\left(y_1 s, y_2 s, \dots, y_{k-1} s, \left(1-\sum_{i=1}^{k-1} y_i\right) s\right) \in I = (0,\infty)^k.$$

This, together with the last observation in Pg 234, shows that $g(I) = J$ (open and path-connected)

$$:= \left\{(y_1, y_2, \dots, y_{k-1}) \in (0,\infty)^{k-1} : \sum_{i=1}^{k-1} y_i < 1\right\} \times (0,\infty)$$

$$= \left\{(y_1, y_2, \dots, y_{k-1}) \in (0,1)^{k-1} : \sum_{i=1}^{k-1} y_i < 1\right\} \times (0,\infty).$$

In the process (of computation of J), we have also found out the inverse map g^{-1}, namely,

$$g^{-1}(y_1, y_2, \dots, y_{k-1}, s) = \left(y_1 s, y_2 s, \dots, y_{k-1} s, \left(1-\sum_{i=1}^{k-1} y_i\right) s\right),$$

$$(y_1, y_2, \dots, y_{k-1}, s) \in J.$$