Multinomial Distribution

→ Applies when the outcomes can take only a finite number L of values.

	City	Bypass
High	1 100	2 20
Toll	3 70	4 45

There are $n = 235$ observations

Let $\underline{x} = (x_1, x_2, x_3, x_4)$ be the vector of $L = 4$ possible outcomes.

$$x_\ell = \#\{\text{cases having outcome } \ell\}$$

$$\boxed{\underline{x} = (100, 20, 70, 45).}$$

→ It's convenient to code the outcomes in terms of coordinate vectors $\underline{e}_\ell$ of length L,

$$e_1 = (1, 0, 0, 0)$$
$$e_2 = (0, 1, 0, 0)$$
$$e_3 = (0, 0, 1, 0)$$
$$e_4 = (0, 0, 0, 1)$$

The multinomial distribution assumes observations are independent $\boxed{\text{given the probability of each outcome.}}$

$$\underbrace{\pi_\ell}_{\text{Prob. of outcome } e_\ell} = \Pr\{e_\ell\} \quad , \quad \ell = 1, \dots, L$$

and $\underline{\pi} = (\pi_1, \dots, \pi_L)^T$.

Probability Vector

Given all of this, the count vector $\underline{x}$ then follows the multinomial distribution

$$f_{\underline{\pi}}(\underline{x}) = \frac{n!}{x_1!\, x_2! \cdots x_L!} \cdot \prod_{\ell=1}^{L} \pi_\ell^{x_\ell}$$

Multinomial pmf.

probability mass function

Generalization of Binomial Distr.

If $L = 2$ and sample size $= n$

$\pi_2 = 1 - \pi_1$

$\underline{\pi} = (\pi_1, \pi_2) = (\pi_1, 1-\pi_1)$

$\underline{x} = (x_1, x_2) = (x_1, n-x_1)$

$$f_{\underline{\pi}}(\underline{x}) = \frac{n!}{x_1!\,(n-x_1)!} \pi_1^{x_1} (1-\pi_1)^{n-x_1}$$

$\Rightarrow$ This is the Binomial pmf!

We denote a Multinomial r.v. by

$$\underline{x} \sim Mult_L(n, \underline{\pi}),$$

$\rightarrow$ In this multiparameter distribution, the parameter space is not just an interval.

For example, the Binomial distribution has only parameter $p \in \Omega = [0, 1)$

Further, there is $\boxed{\text{dependency}}$ between the components of $\underline{\pi}$.

It is not Like having a $N(\mu, \sigma^2)$ with both μ and σ^2 unknown. In this case

$\mu \in (-\infty, \infty)$

$\sigma^2 \in [0, \infty)$

$$\Omega_L = \{\underline{\pi} : \underbrace{\pi_l \geq 0}_{\text{each component is nonneg.}} \text{ and } \underbrace{\sum_{l=1}^{L} \pi_l = 1}_{\text{sum of components} = 1.}\}$$

All vectors $\underline{\pi}$

such that

If $\underline{x} \sim \underline{Mult_L(\underline{\pi})}$

then $E(\underline{x}) = \mu = n\underline{\pi}$

Prob. of taking each

Mean vector.

Total # of cars

route.

Expected # of cars going through each route

# of cars taking each route combination

Question: Does the covariance look like this?

$$\Sigma = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix}$$

Nope → covariance between the outcomes!

$$\boxed{\Sigma = n\left[\mathrm{diag}(\underline{\pi}) - \underbrace{\underline{\pi}}_{L\times 1} \cdot \underbrace{\underline{\pi}^T}_{1\times L}\right]}$$

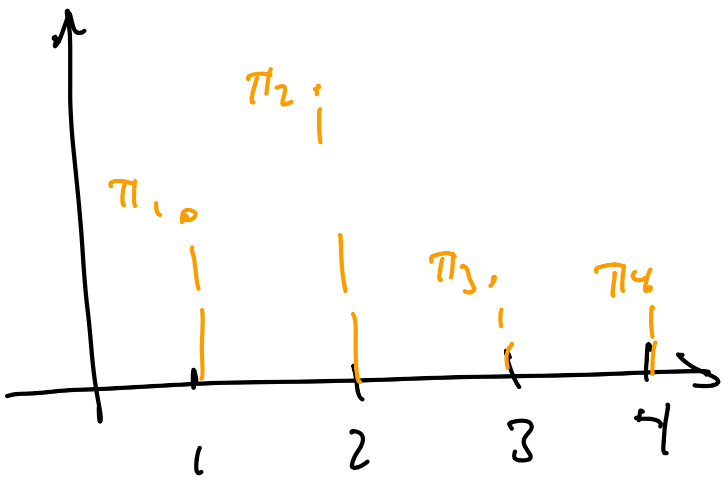
$\mathrm{diag}(\underline{\pi})$ is matrix w/ elements π_ℓ

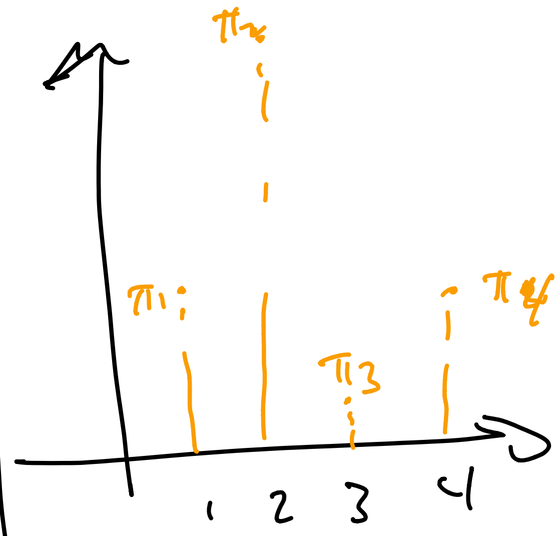
$$\mathrm{diag}(\underline{\pi}) = \begin{bmatrix} \pi_1 & & & 0 \\ & \pi_2 & & \\ & & \pi_3 & \\ 0 & & & \pi_4 \end{bmatrix}$$

μ and Σ generalize Binomial mean $n\pi$ and covariance $n\pi(1-\pi)$

$$n \cdot \left\{ \underbrace{\begin{bmatrix} \pi_1 & & & 0 \\ & \pi_2 & & \\ & & \pi_3 & \\ 0 & & & \pi_4 \end{bmatrix}}_{L\times L} - \underbrace{\begin{bmatrix} \pi_1\pi_1 & \pi_1\pi_2 & \pi_1\pi_3 & \pi_1\pi_4 \\ \pi_2\pi_1 & \ddots & & \\ \vdots & & \ddots & \\ \vdots & & & \ddots \end{bmatrix}}_{L\times L} \right\}$$

→ The multinomial family contains all discrete distributions for a given number of L categories.





→ In our example, consider

$\underline{\pi} = (.25, .25, .25, .25)$ Equally Likely outcomes.

Q1: How probable is it to observe our vector?

$\underline{x} = (100, 20, 70, 45)$

$$f_{\underline{\pi}}(\underline{x}) = \frac{235!}{100!\,20!\,70!\,45!}(.25)^{100}\cdot(.25)^{20}(.25)^{70}(.25)^{45}$$

$$= \frac{\Gamma(236)}{\Gamma(101)\Gamma(21)\Gamma(71)\Gamma(46)}(.25)^{100}(.25)^{20}(.25)^{70}(.25)^{45}$$

$$= 5.3713 \times 10^{-10}$$

→ Then, it would be almost impossible to observe these counts if probs were truly equal.

$$E(\underline{X}) = 235 \cdot \begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix} = \begin{bmatrix} 58.75 \\ 58.75 \\ 58.75 \\ 58.75 \end{bmatrix}$$

$$\Sigma = 235 \cdot \left\{ \begin{bmatrix} .25 & & & 0 \\ & .25 & & \\ & & .25 & \\ 0 & & & .25 \end{bmatrix} - \begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix} [.25, .25, .25, .25] \right\}$$

$$\hat{\theta}^{MLE} \approx \theta - \frac{\dot{\ell}_x(\theta)}{\ddot{\ell}_x(\theta)}$$

$\hat{\theta}^{MLE}$ is an estimator for θ, the true parameter value.

$$f_x(\theta)$$

$\downarrow$ multiplying

$$f_{\underline{x}}(\theta)$$

$\downarrow$ Log

$$\ell_{\underline{x}}(\theta)$$

$\downarrow$ Differentiating wrt θ

$$\cancel{\hat{\lambda}^{MLE}} = \dot{\ell}_{\underline{x}}(\theta)$$

$\downarrow$

$$\dot{\ell}_{x}(\theta) = 0$$

Equate to zero, to find max of $\ell_{\underline{x}}(\theta)$

$\downarrow$

Solve for $\theta \rightarrow \hat{\theta}^{MLE}$

Differentiating wrt θ

$$I_\theta = \int_{\mathcal{X}} f_x(\theta)\, \ddot{\ell}_x(\theta)\, dx \neq I(\underline{x})$$

$$\ddot{\ell}_{\underline{x}}(\theta)$$

$\downarrow$

$$\boxed{I(\underline{x}) = -\ddot{\ell}_{\underline{x}}(\hat{\theta}^{MLE})}$$

$$= V(\hat{\theta}^{MLE})$$

$\downarrow$

$$1/I(\underline{x})$$