n: # cases in whole area G

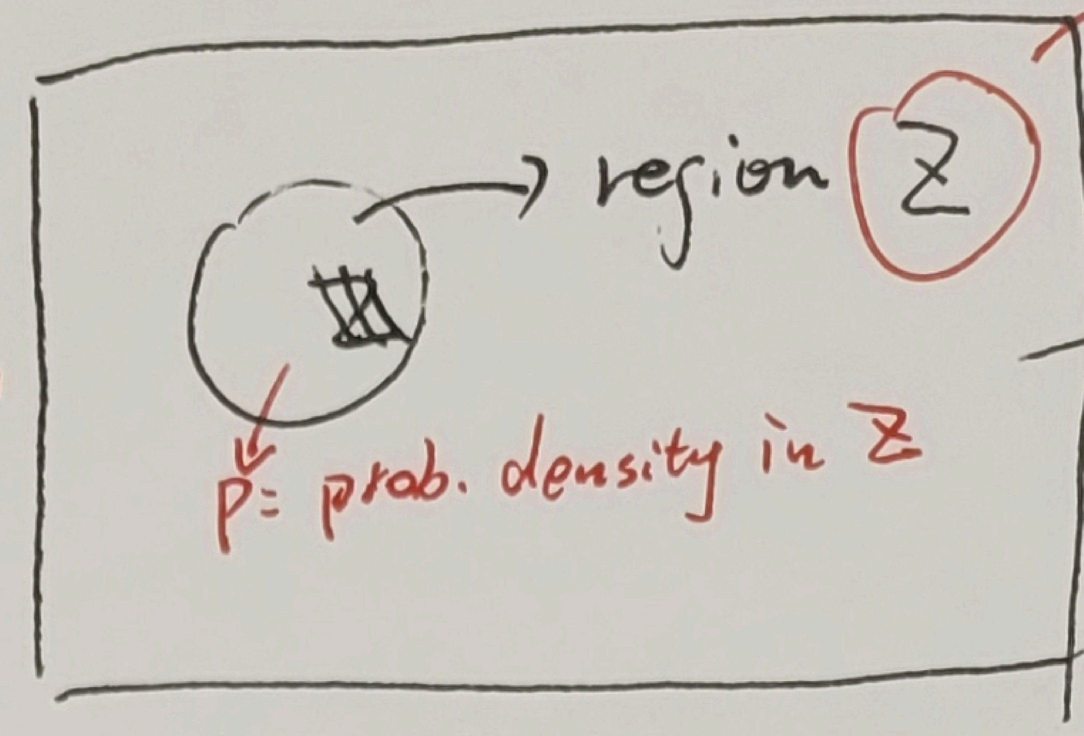
n_Z: # cases in region $Z \subset G$ ↳ means "a subset of"

$\boxed{N}$: total population

$\boxed{N_Z}$: total population in region $Z \subset G$

↳ unknow here: assume a homogeneous & continous distribution of underlying population

so $\frac{N_Z}{N} = \frac{A_Z}{A}$, where A_Z and A are areas of region Z and G, respectively



q: prob. density outside Z

whole study area G

Step 1: Get probability for observing n cases in the study area → non-spatial

Step 2: Further calculate probabilities ~~for~~ with respect to the locations of points → spatial (i.e., inside Z vs. outside Z)

Step 1: Two inputs needed for Poisson: observed #points, and, expected #points

Observed: n Expected: $p \cdot N_Z + q \cdot (N - N_Z)$

↳ interpret as prob. of being a ~~case~~ case point per unit population

$$\text{prob.} = \frac{e^{-p \cdot N_Z - q \cdot (N - N_Z)} \cdot [p \cdot N_Z + q \cdot (N - N_Z)]^n}{n!}$$

→ we are just plugging in the above quantities into Poisson's probability mass function (PMF)

Step 2:

Inside Z: prob. of each ~~point~~ location having an incident (case) is:

$$\frac{p \cdot N_x}{p \cdot N_Z + q(N - N_Z)}$$

N_x → population at location x

→ think about this as: there are an extremely large number of locations in the study area (i.e., a Bernoulli scenario where the distribution of control points is homogeneous or uniform, and has an extremly high density)

Outside Z: ....:

$$\frac{q \cdot N_x}{p \cdot N_Z + q \cdot (N - N_Z)}$$

Multiply probability for all case points:

$$\left(\prod_{x_i \in Z} \frac{p \cdot N_x}{p \cdot N_Z + q \cdot (N - N_Z)} \right) \times \left(\prod_{x_i \notin Z} \frac{q \cdot N_x}{p \cdot N_Z + q \cdot (N - N_Z)} \right)$$

→ inside Z → outside Z

$x_i \in Z$: for all ~~points~~ points in Z; total number is n_Z

$x_i \notin Z$: for all points NOT in Z: total number is $n - n_Z$

Step 1 & Step 2 together (multiply) → always multiply probabilities of independent events

Likelihood for observing our dataset is :
(probability)

$$L(Z, p, q) = \frac{e^{-p\cdot N_z - q\cdot(N-N_z)} \cdot [p\cdot N_z + q\cdot(N-N_z)]^n}{n!} \times \left(\prod_{x_i\in Z} \frac{p\cdot N_x}{p\cdot N_z + q(N-N_z)}\right)$$

L_1

cancelled out

$$\times \left(\prod_{x_i\notin Z} \frac{q\cdot N_x}{p\cdot N_z + q\cdot(N-N_z)}\right)$$

$$= \frac{e^{-p\cdot N_z - q\cdot(N-N_z)}}{n!} \cdot \frac{[p\cdot N_z + q\cdot(N-N_z)]^n}{[p\cdot N_z + q(N-N_z)]^{n_z} \cdot [p\cdot N_z + q(N-N_z)]^{(n-n_z)}}$$

see end of page 1

This has p and q as different terms → can be used for H_1

$$\cdot\, p^{n_z} \cdot q^{(n-n_z)} \cdot \prod_{x_i\in G} N_x$$

→ ~~G is the~~ this is the union of $x_i \in Z$ and $x_i \notin Z$

For H_0: assume $p=q$ ⇒ $L_0 = \frac{e^{-p\cdot N}}{n!} \cdot p^n \cdot \prod_{x_i\in G} N_x$

---

p and q are unknown

Find which p and q ~~values~~ values can be best supported by our data using Maximum Likelihood Estimation (MLE)

Taking derivative w.r.t. p and q using the original likelihoods L_1 and L_0 are difficult so use the log versions (does NOT change the solutions)

↳ if $A > B$, then $\log A > \log B$

$$\log L_1 = -p\cdot N_z - q\cdot(N-N_z) - \log n! + n_z\cdot\log p + (n-n_z)\cdot\log q + \log \prod_{x_i\in G} N_x$$

only these two terms contain p

$$\frac{d\log L_1}{dp} = \frac{d(-p\cdot N_z + n_z\cdot\log p)}{dp} = -N_z + n_z\cdot\frac{1}{p} \quad \left(\frac{d\log p}{dp} = \frac{1}{p}\right)$$

Set to 0 : $-N_z + \frac{n_z}{p} = 0 \Rightarrow p = \frac{n_z}{N_z}$ (similarly : $q = \frac{n-n_z}{N-N_z}$)

$\boxed{\text{DIY for } p \text{ in } H_0}$ → $p = \frac{n}{N}$ for L_0

Finally: Plug solutions (MLE estimated p, q for L_1 and p for L_0)
into likelihoods L_1 and L_0
and take the ratio:

$$LR = \frac{L_1}{L_0} = \left(\frac{n_z}{E_z}\right)^{n_z} \cdot \left(\frac{n-n_z}{n-E_z}\right)^{n-n_z}$$

$$\hookrightarrow E_z = n \cdot \frac{N_z}{N} = n \cdot \frac{A_z}{A}$$

A_z → area of Z

A → area of G

see definitions for N, N_z at the begining

Now with n_z, n, A_z, A

You can calulate the LR for each candidate