離散異分布の適合度検証法

慶應義塾大学理工学部
柴田　里程

独立だが異分布な標本の例

i 回目の海底浚渫で得られた個体数 $N_i, i = 1, 2, \ldots, n$

そのときの浚渫面積 $\alpha_i, i = 1, 2, \ldots, n$

1. Poisson

$$N_i \sim Po(\lambda \alpha_i)$$

2. Thomas ($Thomas\ Clustering\ Process$)

$$N_i = X_1 + X_2 + \cdots + X_Q \sim Tho(\lambda \alpha_i, \phi),$$

$$Q : Number\ of\ Patches \sim Po(\lambda \alpha_i),$$

$$X_j - 1 : Excess\ Count\ in\ Each\ Patch \sim Po(\phi)$$

ピアソンχ^2

$$X_1, X_2, \ldots, X_n : i.i.d., \quad p_i = P(X_k = i)$$

$$Y_{in} = \frac{\#\{X_k = i,\ k = 1, 2, \ldots, n\} - np_i}{\sqrt{np_i}},\ i = 1, 2, \ldots, m$$

$$\mathbf{Y}_n = (Y_{1n}, Y_{2n}, \ldots, Y_{mn})^T \simeq N\left(\mathbf{0},\ I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^T\right) : \mathbf{p} \text{ に依存}$$

$$\sqrt{\mathbf{p}}^T = (\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_m})$$

⟹ $\chi^2 = \|\mathbf{Y}_n\|^2 \simeq \chi^2(m-1) : \mathbf{p}$ に依存しない

- うまくできている → 広く用いられている適合度統計量としては唯一
- しかし不自由．たとえば部分的な2乗和の分布はpに依存してしまい，異分布の場合にはもともと通用しない

$\mathbf{Y}_n$を変換して分布$\mathbf{p}$に依存しない
極限分布を持つようにできないか？

$\Sigma_m^{-1/2}$は存在しない

$\mathbf{Z}_n = V\ \mathbf{Y}_n \simeq N(0, \Sigma_m)$: $\mathbf{p}$に依存しないΣ_m

$\left\langle \mathbf{Y}_n, \sqrt{\mathbf{p}} \right\rangle = 0$ ⇨ $\mathrm{rank}(\Sigma_m) \le m-1$

$\mathrm{rank}(\Sigma_m) = m-1$にできれば次元の縮約にともなう
情報の損失はない.

$\mathrm{Ker}(\Sigma_m) = Span\{\mathbf{r}\},\ \|\mathbf{r}\| = 1$とおけば

⇨ Σ_mの一つの標準形は直交射影 $\Sigma_m = I - \mathbf{r}\mathbf{r}^T$

$\Rightarrow \|\mathbf{Z}_n\|^2 \simeq \chi^2(m-1)$

E. Khmaladze (2011)

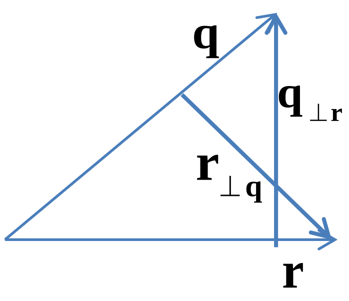
$$\mathbf{Z}_n = \mathbf{Y}_n - \left\langle \mathbf{Y}_n, \mathbf{r} \right\rangle \frac{\mathbf{r} + \sqrt{\mathbf{p}}}{1 + \left\langle \sqrt{\mathbf{p}}, \mathbf{r} \right\rangle} \simeq N(0, I - \mathbf{r}\mathbf{r}^T)$$

特に，$\mathbf{r} = \sqrt{\mathbf{p}}$ にとれば $\mathbf{Z}_n = \mathbf{Y}_n$

Lemma

$$U = \mathbf{r}\mathbf{q}^T \pm \frac{1}{\mu^2}\mathbf{q}_{\perp\mathbf{r}}\mathbf{r}_{\perp\mathbf{q}}^T \Longrightarrow U\text{はユニタリーで } U\mathbf{q} = \mathbf{r}$$

ただし，$\|\mathbf{r}\| = \|\mathbf{q}\| = 1,\ \mu = \|\mathbf{q}_{\perp\mathbf{r}}\| = \|\mathbf{r}_{\perp\mathbf{q}}\|$



$$V = \mathrm{Proj}_{L^\perp} + U \qquad L = Span\{\mathbf{q}, \mathbf{r}\}$$

$$V\mathbf{q} = \mathbf{r} \qquad VV^T = \mathrm{Proj}_{L^\perp} + UU^T = I$$ （V：直交変換）

$\mathbf{y} \perp \mathbf{q}$ ならば

$$\mathrm{Proj}_{L^\perp}\mathbf{y} = \mathbf{y} - \frac{1}{\mu^2} < \mathbf{y}, \mathbf{r}_{\perp\mathbf{q}} > \mathbf{r}_{\perp\mathbf{q}}$$

$$U\mathbf{y} = (\mathbf{r}\mathbf{q}^T - \frac{1}{\mu^2}\mathbf{q}_{\perp\mathbf{r}}\mathbf{r}_{\perp\mathbf{q}}^T)\mathbf{y} = -\frac{1}{\mu^2}\mathbf{q}_{\perp\mathbf{r}} < \mathbf{r}_{\perp\mathbf{q}}, \mathbf{y} >$$

$$= -\frac{(\mathbf{r}+\mathbf{q})(1 - <\mathbf{q}, \mathbf{r}>) - \mathbf{r}_{\perp\mathbf{q}}}{\mu^2} < \mathbf{r}_{\perp\mathbf{q}}, \mathbf{y} >$$

$$= -\frac{\mathbf{r}+\mathbf{q}}{1 + <\mathbf{q}, \mathbf{r}>} < \mathbf{r}, \mathbf{y} > + \frac{\mathbf{r}_{\perp\mathbf{q}}}{\mu^2} < \mathbf{r}_{\perp\mathbf{q}}, \mathbf{y} >$$

$\mathbf{q} = \sqrt{\mathbf{p}}$ にとれば

$$V\,\mathbf{Y}_n = \mathbf{Y}_n - \frac{\mathbf{r} + \sqrt{\mathbf{p}}}{1 + <\sqrt{\mathbf{p}}, \mathbf{r}>} < \mathbf{r}, \mathbf{Y}_n >$$

異分布の場合

$$X_1, X_2, \ldots, X_n : i.d., \qquad p_{ki} = P(X_k = i)$$

$$\boldsymbol{\eta}_k = (\eta_{1k}, \eta_{2k}, \ldots, \eta_{mk})^T \qquad \eta_{ki} = \frac{I_{(X_k = i)} - p_{ki}}{\sqrt{p_{ki}}},\ i = 1, 2, \ldots, m,\ k = 1, 2, \ldots, n$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\eta}_k \simeq N\left(\mathbf{0},\ I - \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sqrt{\mathbf{p}_k} \sqrt{\mathbf{p}_k}^T\right), \qquad \sqrt{\mathbf{p}_k}^T = (\sqrt{p_{k1}}, \sqrt{p_{k2}}, \ldots, \sqrt{p_{km}})$$

$\frac{1}{n} \sum_{k=1}^{n} \sqrt{\mathbf{p}_k} \sqrt{\mathbf{p}_k}^T$ が収束すると限らないし，射影行列になるとも限らない，もちろんランクも定まらない

$$\mathbf{Z}_k = \boldsymbol{\eta}_k - \langle \boldsymbol{\eta}_k, \mathbf{r}_k \rangle \frac{\mathbf{r}_k + \sqrt{\mathbf{p}_k}}{1 + \langle \sqrt{\mathbf{p}_k}, \mathbf{r}_k \rangle},\ k = 1, 2, \ldots, n$$

$$\mathbf{W}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbf{Z}_k \simeq N\left(\mathbf{0},\ I - \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{r}_k \mathbf{r}_k^T\right)$$

特に $\mathbf{r}_k \equiv \mathbf{r},\ k = 1, 2, \ldots, n$ にとれば異分布でも

χ^2統計量が得られる

$$\mathbf{Z}_k = \boldsymbol{\eta}_k - \left\langle \boldsymbol{\eta}_k, \mathbf{r} \right\rangle \frac{\mathbf{r} + \sqrt{\mathbf{p}_k}}{1 + \left\langle \sqrt{\mathbf{p}_k}, \mathbf{r} \right\rangle},\ k = 1, 2, \ldots, n$$

$$\mathbf{W}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbf{Z}_k \simeq N\left(\mathbf{0},\ I - \mathbf{r}\mathbf{r}^T\right)$$

つまり $\|\mathbf{W}_n\|^2 \simeq \chi^2(m-1)$

パラメータ推定を含む場合

$p_i = p_i(\theta), \quad \theta \in \mathbb{R}$, 関数 $p_i(\cdot)$ は既知

$$Y_{in}(\theta) = \frac{\#\{X_k = i,\ k = 1, 2, \ldots, n\} - np_i(\theta)}{\sqrt{np_i(\theta)}},\ i = 1, 2, \ldots, m$$

$$\hat{\theta}:\ \sum_{i=1}^{m} Y_{in}(\theta)^2 \to \min$$

$$\mathbf{Y}_n(\hat{\theta}) \approx \mathbf{Y}_n(\theta) - \frac{1}{g^2}\left\langle \mathbf{Y}_n(\theta),\ \frac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}} \right\rangle \frac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}}, \quad g^2 = \left\| \frac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}} \right\|^2$$

右辺 $\perp \sqrt{\mathbf{p}(\theta)},\ \dfrac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}}, \quad \sqrt{\mathbf{p}(\theta)} \perp \dfrac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}}$

$$\hat{\mathbf{Y}}_n = \mathbf{Y}_n(\hat{\theta}) \simeq N\left(\mathbf{0}, I - \mathbf{q}_1\mathbf{q}_1^T - \mathbf{q}_2\mathbf{q}_2^T\right),\ \mathbf{q}_1 = \sqrt{\mathbf{p}(\theta)},\ \ \mathbf{q}_2 = \frac{1}{g}\frac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}}$$

$$\mathbf{Z}_n = V\,\hat{\mathbf{Y}}_n \simeq N(0, \Sigma_m):\ \mathbf{q}_1, \mathbf{q}_2 \text{に依存しない} \Sigma_m$$

$$Ker(\Sigma_m) = Span\{\mathbf{r}_1, \mathbf{r}_2\},\ \mathbf{r}_1 \perp \mathbf{r}_2, \|\mathbf{r}_1\| = \|\mathbf{r}_2\| = 1$$

$$\mathbf{Z}_n = \hat{\mathbf{Y}}_n - \left\langle \hat{\mathbf{Y}}_n, \boldsymbol{\alpha} \right\rangle(\boldsymbol{\alpha} - \boldsymbol{\beta}) + \left\langle \hat{\mathbf{Y}}_n, \boldsymbol{\gamma} \right\rangle(\boldsymbol{\gamma} - \boldsymbol{\delta}) \simeq N(\mathbf{0},\ I - \mathbf{r}_1\mathbf{r}_1^T - \mathbf{r}_2\mathbf{r}_2^T)$$

$$\boldsymbol{\alpha} = \frac{(\mathbf{r}_1)_{\perp \mathbf{q}_1\mathbf{q}_2}}{\left\|(\mathbf{r}_1)_{\perp \mathbf{q}_1\mathbf{q}_2}\right\|},\ \ \boldsymbol{\beta} = \frac{(\mathbf{r}_2)_{\perp \mathbf{r}_1\mathbf{q}_1\mathbf{q}_2}}{\left\|(\mathbf{r}_2)_{\perp \mathbf{r}_1\mathbf{q}_1\mathbf{q}_2}\right\|},\ \ \boldsymbol{\gamma} = \frac{(\mathbf{q}_1)_{\perp \mathbf{r}_1\mathbf{r}_2}}{\left\|(\mathbf{q}_1)_{\perp \mathbf{r}_1\mathbf{r}_2}\right\|},\ \ \boldsymbol{\delta} = \frac{(\mathbf{q}_2)_{\perp \mathbf{q}_1\mathbf{r}_1\mathbf{r}_2}}{\left\|(\mathbf{q}_2)_{\perp \mathbf{q}_1\mathbf{r}_1\mathbf{r}_2}\right\|}$$

ただし， $\mathbf{q}_1 = \sqrt{\mathbf{p}(\theta)},\ \ \mathbf{q}_2 = \frac{1}{g}\frac{\mathbf{p}'(\theta)}{\sqrt{\mathbf{p}(\theta)}}$ の θ は $\hat{\theta}$ で置き換える

異分布の場合

$p_{ki} = p_{ki}(\theta),\ \theta \in \mathbb{R}$, 関数 $p_{ki}(\cdot)$ は既知

$$\eta_{ki}(\theta) = \frac{I_{(X_k = i)} - p_{ki}(\theta)}{\sqrt{p_{ki}(\theta)}},\ i = 1, 2, \ldots, m,\ k = 1, 2, \ldots, n$$

$$\hat{\theta}:\ \sum_{i=1}^{m} \left(\sum_{k=1}^{n} \eta_{ki}(\theta)^2 \right) \to \min$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\eta}_k(\hat{\theta}) \approx \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\eta}_k(\theta) - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{g_k{}^2} \left\langle \boldsymbol{\eta}_k(\theta),\ \frac{\mathbf{p}'_k(\theta)}{\sqrt{\mathbf{p}_k(\theta)}} \right\rangle \frac{\mathbf{p}'_k(\theta)}{\sqrt{\mathbf{p}_k(\theta)}}, \quad g_k{}^2 = \left\| \frac{\mathbf{p}'_k(\theta)}{\sqrt{\mathbf{p}_k(\theta)}} \right\|^2$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\eta}_k(\hat{\theta}) \simeq N\left(\mathbf{0},\ I - \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (\mathbf{q}_{1k}\mathbf{q}_{1k}^T + \mathbf{q}_{2k}\mathbf{q}_{2k}^T) \right),$$

$$\mathbf{q}_{1k} = \sqrt{\mathbf{p}_k(\theta)},\quad \mathbf{q}_{2k} = \frac{1}{g_k} \frac{\mathbf{p}'_k(\theta)}{\sqrt{\mathbf{p}_k(\theta)}}$$

$$\mathbf{Z}_k = \boldsymbol{\eta}_k(\hat{\theta}) - \left\langle \boldsymbol{\eta}_k(\hat{\theta}), \boldsymbol{\alpha}_k \right\rangle (\boldsymbol{\alpha}_k - \boldsymbol{\beta}_k) + \left\langle \boldsymbol{\eta}_k(\hat{\theta}), \boldsymbol{\gamma}_k \right\rangle (\boldsymbol{\gamma}_k - \boldsymbol{\delta}_k)$$

$$\boldsymbol{\alpha}_k = \frac{(\mathbf{r}_1)_{\perp \mathbf{q}_{1k}\mathbf{q}_{2k}}}{\left\|(\mathbf{r}_1)_{\perp \mathbf{q}_{1k}\mathbf{q}_{2k}}\right\|}, \boldsymbol{\beta}_k = \frac{(\mathbf{r}_2)_{\perp \mathbf{r}_1\mathbf{q}_{1k}\mathbf{q}_{2k}}}{\left\|(\mathbf{r}_2)_{\perp \mathbf{r}_1\mathbf{q}_{1k}\mathbf{q}_{2k}}\right\|}, \quad \boldsymbol{\gamma}_k = \frac{(\mathbf{q}_{1k})_{\perp \mathbf{r}_1\mathbf{r}_2}}{\left\|(\mathbf{q}_{1k})_{\perp \mathbf{r}_1\mathbf{r}_2}\right\|}, \quad \boldsymbol{\delta}_k = \frac{(\mathbf{q}_{2k})_{\perp \mathbf{q}_{1k}\mathbf{r}_1\mathbf{r}_2}}{\left\|(\mathbf{q}_{2k})_{\perp \mathbf{q}_{1k}\mathbf{r}_1\mathbf{r}_2}\right\|}$$

ただし，$\mathbf{q}_{1k} = \sqrt{\mathbf{p}_k(\theta)}$, $\mathbf{q}_{2k} = \frac{1}{g}\frac{\mathbf{p}'_k(\theta)}{\sqrt{\mathbf{p}_k(\theta)}}$ の θ は $\hat{\theta}$ で置き換える

$$\mathbf{W}_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} \mathbf{Z}_k \simeq N\left(\mathbf{0},\ I - \mathbf{r}_1\mathbf{r}_1^T - \mathbf{r}_2\mathbf{r}_2^T\right)$$

$\mathbf{r}_1, \mathbf{r}_2$ の例：

$$\mathbf{r}_1 = \left(\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}, \ldots, \frac{1}{\sqrt{m}}\right)^T, \ \mathbf{r}_2 = \left(\frac{1}{\sqrt{m}}, \frac{-1}{\sqrt{m}}, \frac{1}{\sqrt{m}}, \frac{-1}{\sqrt{m}}, \ldots, \frac{1}{\sqrt{m}}\right)^T$$

残された問題

可算無限個の値をとるとき

- ある範囲以上は併合
- 無限次元のまま扱う

乱数実験

実際問題への適用

$\mathbf{r}_1, \mathbf{r}_2$ の選び方による検出力の変化の評価

$\boldsymbol{\eta}_k$ の特性関数

$$\phi_k(\mathbf{t}) = E(e^{i\mathbf{t}^T \boldsymbol{\eta}_k}) = \sum_{j=1}^{m} p_{kj} e^{it_j/\sqrt{p_{kj}}} \cdot e^{-i\mathbf{t}^T \sqrt{\mathbf{p}_k}}$$

$W_k = \dfrac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\eta}_k$ の特性関数の対数

$$\sum_{k=1}^{n} \log\left(\phi_k(\mathbf{t}/\sqrt{n})\right) = \sum_{k=1}^{n} \left(\log\left(\sum_{j=1}^{m} p_{kj} e^{it_j/\sqrt{np_{kj}}} \right) - i\mathbf{t}^T \sqrt{\mathbf{p}_k} / \sqrt{n} \right)$$

$$\log\left(\sum_{j=1}^{m} p_{kj} e^{it_j/\sqrt{np_{kj}}} \right) = \sum_{j=1}^{m} it_j \frac{\sqrt{p_{kj}}}{\sqrt{n}} - \frac{1}{2n} \mathbf{t}^T (I - \sqrt{\mathbf{p}_k} \sqrt{\mathbf{p}_k}^T) \mathbf{t} + O\left(\frac{1}{n\sqrt{n}} \right)$$

$$\sum_{k=1}^{n} \log\left(\phi_k(\mathbf{t}/\sqrt{n})\right) = -\frac{1}{2} \mathbf{t}^T (I - \frac{1}{n} \sum_{k=1}^{n} \sqrt{\mathbf{p}_k} \sqrt{\mathbf{p}_k}^T) \mathbf{t} + O\left(\frac{1}{\sqrt{n}} \right)$$

Goodness of fit test of non-homogenious Thomas Distribution

Non-homogenious count data

1. Poisson

$$N_i \sim Po(\lambda \alpha_i)$$

2. Thomas (*Thomas Clustering Process*)

$$N_i = X_1 + X_2 + \cdots + X_Q \sim Tho(\lambda \alpha_i, \phi),$$

$$Q : Number\ of\ Patches \sim Po(\lambda \alpha_i),$$

$$X_j - 1 : Excess\ Count\ in\ Each\ Patch \sim Po(\phi)$$

Bettter explain zero counts than other distribution, like negative binomial does

$$Po(\lambda \alpha_i) \subset Tho(\lambda \alpha_i, \phi)$$

Goodness of fit test vs Model Selection

Model Section: Relative comparison of models
NO assurance of goodness of the model selected

Goodness of fit test of a model

⇓ passed

Selection of a model from the submodels

Jittered PP plot for disrete dist.

$$Y \sim U(F(X-1), F(X))$$

$$Scatter\ Plot\ of\ \left(\frac{k}{n}, Y_{(k)}\right), k = 1, 2, \ldots, n$$

Goodness of fit test of non-homogenious Poisson

$$N_i \sim Po(\lambda\alpha_i),\ i = 1,2,\ldots,k,\ \ \alpha_1,\alpha_2,\ldots,\alpha_k : known$$

$$N_1, N_2,\ldots,N_k \mid N \sim M_N\left(p_1, p_2,\ldots,p_k; N\right) \qquad p_i = \frac{\alpha_i}{\sum_i \alpha_i}, i = 1,2,\ldots,k$$

⇓

- $Test\ of\ N = \sum_i N_i \sim Po(\lambda \sum_i \alpha_i)$
- $Conditional\ Test\ of\ N_1, N_2,\ldots,N_k \mid N \sim M_N\ p_1, p_2,\ldots,p_k; N$

  $independ\ of\ \lambda$, not necessarily $\sum_i \alpha_i \to \infty\ as\ k \to \infty$

  $therefore,\ \chi^2\ test\ would\ not\ appropriate$

  $Usually\ \mathrm{E}(N) \to \infty\ \ as\ \ k \to \infty.$

$$F_n(x)\ \ vs\ \ F(x) = x,\ 0 \le x \le 1 \qquad F_n\left(\sum_{i=1}^{k} p_i\right) = \frac{1}{N}\sum_{i=1}^{k} N_i$$

Extension to Thomas distribution

$$N_i = X_{\sum_{j=1}^{i-1} Q_j + 1} + X_{\sum_{j=1}^{i-1} Q_j + 2} + \cdots + X_{\sum_{j=1}^{i} Q_j} \sim Tho(\lambda \alpha_i, \phi), \;\; i = 1, 2, \ldots, k$$

$$Q_i \sim Po(\lambda \alpha_i), \; i = 1, 2, \ldots, k$$

$$X_j - 1 \sim Po(\phi)$$

$$N = \sum_i N_i \sim Tho(\lambda \sum_i \alpha_i, \phi)$$

$$P(N_1 = n_1, N_2 = n_2, \ldots, N_k = n_k \,\Big|\, N = n)$$

$$= E^Q \left((n - Q)! e^{N-Q} E\left(\prod_{i=1}^{k} \frac{Q_i^{n_i - Q_i}}{(n_i - Q_i)!} \middle| Q \right) \right), \; Q = \sum_i Q_i$$

Extension to Stopped Poisson distribution

$$N_i = X_{\sum_{j=1}^{i-1} Q_j + 1} + X_{\sum_{j=1}^{i-1} Q_j + 2} + \cdots + X_{\sum_{j=1}^{i} Q_j}, \; i = 1, 2, \ldots, k \qquad Q_i \sim Po(\lambda \alpha_i), \; i = 1, 2, \ldots, k$$

$$X_j \sim Po(\phi)$$

$$P(N_1 = n_1, N_2 = n_2, \ldots, N_k = n_k \mid N = n, \; Q_1 = q_1, Q_2 = q_2, \ldots, Q_k = q_k)$$

$$= \frac{n!}{n_1! n_2! \cdots n_k!} \left(\frac{q_1}{q} \right)^{n_1} \cdots \left(\frac{q_k}{q} \right)^{n_k}, \; q = \sum_{i=1}^{k} q_i$$

- $N = \sum_{i=1}^{Q} X_i \sim Compound \;\; Poisson, \; Q \sim Po(\lambda \sum_{i=1}^{k} \alpha_i), \;\; X_j \sim Po(\phi)$
- $F_n \left(\sum_{i=1}^{k} \left(\frac{q_i}{q} \right) \right) = \frac{1}{N} \sum_{i=1}^{k} N_i \;\; vs \;\; F(x) = x \;\; \Big| \;\; N = n, \; Q_1 = q_1, \ldots, Q_k = q_k$

- $N = \sum_{i=1}^{Q} X_i \sim Compound\ \ Poisson,\ Q \sim Po(\lambda \sum_{i=1}^{k} \alpha_i),\ \ X_j \sim Po(\phi)$

- $F_n\left(\sum_{i=1}^{k}\left(\frac{q_i}{q}\right)\right) = \frac{1}{N}\sum_{i=1}^{k} N_i \ \ vs\ \ F(x) = x\ \Bigg|\ N = n,\ Q_1 = q_1, \ldots,\ Q_k = q_k$

$$E\left(\frac{1}{N}\sum_{i=1}^{k} N_i\ \Big|\ N = n,\ Q_1 = q_1, \ldots,\ Q_k = q_k\right) = \sum_{i=1}^{k}\left(\frac{q_i}{q}\right)$$