

روش درستنما وزن در رودما را ناحه کوچک

ملحه عباس نژاد مشهد

گروه مار دانشکده علوم را دانشگاه فردوس مشهد

چکیده رودمای رای ناحه کوچک^۱ در سالهای اخیر مورد توجه زیادی قرار گرفته است. دلایل آن که تقابل رای رودگرهای معتبر نواحی کوچک هم از جای بخشهای دولتی و هم از جای بخشهای خصوصی افزایش یافته است از ررسهای نمونه ای می توان رودهای مستقیم معتبری رای پارامترهای کل جامعه دست ورد اما رودگرهای مستقیم رای هر ناحه کوچک معنی رودگرهایی که تنها رمینای داده های موجود در ناحه دست می بند دلایل کوچکی حجم نمونه در هر ناحه خاهاای استاندارد زرگی منجر می شوند. نارن روری نه ر می رسد روشهای رای آن منور جاگزین شوند که علاوه را ابلعات موجود در خود ناحه ا ابلعات سار نواحی ا ابلعات سرشماری را نیز مورد استفاده قرار دهند. در آن مقاله ابتدا نه غر مجاز ودن رودگر درستنمای ماکزیم مانگن نرمال در رودمای همزمان^۲ معنی \bar{X} اشاره ای کوتاه خواهم داشت روش درستنمای وزنی^۳ را در خانواده نمایی ورکلی آن کرده و سپس ن را در مساله رودمای رای ناحه کوچک مورد بحث قرار می دهیم.

واژه ها کلید ناحه کوچک درستنمای وزنی رودگر چمنز استان^۴

مقدمه

ناحه کوچک ا ناحه محلی نه ک ناحه جغرافیای کوچک مانند ک استان شهر ا ک بخش سرشماری الاق می گردد. همچن می توان ک ز جامعه کوچک همانند ک گروه از مردم درون ک ناحه جغرافیای زرگ را که از سن و جنس و نژاد معنی هستند نه عنوان ک ناحه کوچک در نر گرفت.

مساله رودمای رای ناحه کوچک مساله ای است که در سالهای اخیر توجه زیادی نه شده است. آن امر نه علت است که تقابل رای ماره های نواحی کوچک چه از جای بخشهای دولتی و چه از جای بخشهای خصوصی رشد زیادی داشته است نه عنوان مثال بخش دولتی رای رنامه رزی کلان در سطح کشور ناز نه ماره های از قبل نرخ بکاری متوس درمد خانوارها و... در استانها و شهرهای مختلف دارد نه عنوان مثال فرض کند که بی

1) Small Area 2) Simultaneous Estimation 3) Weighted Likelihood 4) James-stein Estimator

از نواحی جغرافیایی از رگروههایی از جامعه از قبل تعیین شده است و با قشر خاصی از افراد از رسته درمد پائینتر از یک مانگن قابل قبول و در حد آنتاراشند و واضح است که قبل از انجام هرکاری و هر نوع رنامه رزی لازم است که چنین نواحی از رگروههایی مشخص شوند داده های دست مده از رر سه های نمونه ای را می توان رای استخراج روردهای مستقیم معتبر رای نواحی زرگ اقل رروه های زرگ نواحی نمونه های زرگ استفاده نمود اما روردهای مستقیم یعنی روردهای که تنها ر مبنای داده های که از نمونه اخذ شده از ناحیه استخراج می شوند قرار دارند دارای خای سار زرگی خواهند بود که ناشی از کوچک بودن شان از اندازه حجم نمونه در نواحی است دلیل این امر این است که در ابتدا دقت مورد نظر رای روردها پارامتر کل جامعه در زر گرفته شده و حجم نمونه اصلی ر مبنای تعیین گردیده است یعنی در هنگام رنامه رزی و تعیین حجم نمونه فقط حجم نمونه کل مورد نیاز محاسبه شده است و بجا سهم هر ناحیه کوچک از کل حجم نمونه ناچیز است در نتیجه رای نواحی کوچک این روردها دقت کافی را نخواهند داشت

واضح است که از هر ناحیه نیز نمی توان نمونه ای زرگ اختیار کرد لذا ناچار مده دنبال راهی بود که دقت روردها را تا آنجا که ممکن است بالا برد رای این منظور می توان از داده های نواحی دیگر و اطلاعاتی که در نواحی راساس نمونه دست مده است در جهت بهبود بخشیدن روردهای پارامترهای نواحی کوچک استفاده نمود این ررق حجم نمونه موثر افزایش یافته و دقت روردها افزایش می مده در مقاله حاضر روش درستمای وزنی را مده این مورد بحث قرار می مدهم

روش درستمای وزنی

روش درستمای بدون شک یکی از قدرتمندترین ابزارها در مسال روردهای و زمون فرض می باشد اما زمانی که در روردهای همزمان مساله ترک اطلاعات روردهای مستقیم روش درستمای کلاسیک کار مده نخواهد بود اهمیت این مساله در مقاله استان مده شده در واقع مقاله او نشان داد که در روردهای همزمان مانگن های جوامه نرمال روردهای درستمای ماکزیمم یعنی مانگن نمونه غیر مجاز می باشد او نشان داد که رای روردهای جوامه نرمالی که از یک جهت شبه مده گردند می توان رای دست وردن روردهای که دقت بیشتری از مانگن های نمونه دارد از سار نمونه های مرتبه کمک گرفت

مساله ترک اطلاعات از مینا گوناگون خارج از روش درستمای قرار گرفته و در تا زمانی که تیش رانی و هاستی روش درستمای محلی⁵ را در محدوده رگرس ون ناپارامتری اراه نمودند هو روش درستمای محلی را خارج از حوزه رگرس ون ناپارامتری مورد ررسی قرار داد که این روش درستمای وزنی مرتبه⁶ REWL نام مده شده علاوه ران هو وزن کم

5) Local Likelihood 6) Relevance Weighted Likelihood

نشان دادند که چه می‌توان روردرگر معروف جمز استان را به عنوان روردرگر REWL استخراج نمود مشرو را ن که وزنهای از نمونه روردر شوند تا را ن می‌توان روش درستنمایی را رای ترک الاعات نمونه‌های جوام مختلف توسعه داد

روش درستنمایی وزنی را روش درستنمایی کلاسیک متفاوت است در روش کلاسیک فرض را ن است که مشاهدات از یک جامعه همسان خارج شده‌اند در عوض درستنمایی وزنی در استنباط پارامتری زمانی پیش می‌د که علاوه بر نمونه مورد ن را از جامعه مورد م العه نمونه‌های مرتبه اما مستقل از جوام دیگر ن در دسترس هستند که ان همان مساله م لمو در روردمای رای ناده کوچک می‌اشد روش درستنمایی وزنی را اختصاص وزنهای ان نمونه‌ها را اساس م زان ارتبایشان را الاعات انها را ترک می‌کند در ان مقاله م تولد روردرهای همزمان مانگن‌های چندن توز می‌پردازم که هر یک از ان توزیها متعلق م یک خانواده نمایی م یعنی م تا وارانس درجه دوم NEF-QVF⁷ هستند

روش درستنمایی وزنی در خانواده نما

فرض کنند m جامعه داریم بردارهای نمونه‌های مستقل $(y_{i1}, \dots, y_{in_i})'$ $i = 1, \dots, m$ نشان داده می‌شوند که $n_i \geq 1$ در عمل همیشه ان امکان وجود دارد که از یک ناده محلی هچ مشاهده‌ای گرفته نشود می‌دازم Y دارای توزی نمایی را پارامتر θ است اگر

$$f(y, \theta) = \exp[\theta y - \phi(\theta) + h(y)]$$

نارن اگر $\theta = (\theta_1, \dots, \theta_m)'$ و ω_{ij}^* وزن اختصاص یافته م نمونه زام در روردر پارامتر i ام اشد تا درستنمایی م صورت ز ر تعرف می‌شود

$$\begin{aligned} WL(\theta) &= WL(\theta_1, \dots, \theta_m) \\ &= \prod_{i=1}^m \prod_{j=1}^m \prod_{l=1}^{n_j} \exp[(\theta_i y_{jl}) - \phi(\theta_i) + h(y_{jl})]^{\omega_{ij}^* n_j^{-1}} \end{aligned}$$

ا فرض

$$\bar{y}_j = \frac{1}{n_j} \sum_{l=1}^{n_j} y_{jl} \quad j = 1, \dots, m$$

7) Natural Exponential Family with Quadratic Variance Function

دارم

$$WL(\theta) = \exp\left[\sum_{i=1}^m [\theta_i \sum_{j=1}^m \omega_{ij}^* \bar{y}_j - \phi(\theta_i) \sum_{j=1}^m \omega_{ij}^* + \sum_{j=1}^m \omega_{ij}^* n_j^{-1} \sum_{l=1}^{n_j} h(y_{jl})]\right] \quad ()$$

نارایان اگر ω_{ij}^* ها معلوم باشند

$$\frac{\partial \ln WL(\theta)}{\partial \theta_i} = \sum_{j=1}^m \omega_{ij}^* \bar{y}_j - \phi'(\theta_i) \sum_{j=1}^m \omega_{ij}^* =$$

$$\frac{\partial^2 \ln WL(\theta)}{\partial \theta_i^2} = -\phi''(\theta_i) \sum_{j=1}^m \omega_{ij}^* \leq$$

روردگرهای درستنمایی ماکزیم وزنی MWLE's رای مانگ نهایی جامعه یعنی $\phi_i = \phi'(\theta_i)$ عبارتند از

$$\hat{\phi}_{iWL} = \frac{\sum_{j=1}^m \omega_{ij}^* \bar{y}_j}{\sum_{j=1}^m \omega_{ij}^*} = \sum_{j=1}^m \omega_{ij} \bar{y}_j \quad ()$$

که

$$\omega_{ij} = \frac{\omega_{ij}^*}{\sum_{j=1}^m \omega_{ij}^*}$$

اما در عمل چون وزنها معلوم نیستند باید از روی داده‌ها رورد شوند

۱۳ روش درستنمایی وزنی در مدل تعویض پذیر

در این حالت فرض کند

$$\omega_{ii} = \omega \quad i = 1, \dots, m \quad \omega_{ij} = \omega^* \quad j \neq i = 1, \dots, m$$

چنین فرضی این ایده را در بردارد که نمونه‌های سایر جوامع نسبت به اطلاعاتی که در مورد i پارامتر دارند تعویض پذیر^۸ هستند این فرض صریحاً به ورودگر جیمز استان در حالت نرمال زمانی که وزنها با استفاده از داده‌ها رورد می‌شوند منجر می‌شود تحت این فرض

$$\hat{\phi}_{iWL} = \omega \bar{y}_i + \omega^* \sum_{j \neq i} \bar{y}_j$$

8) Exchangable

..... هفتمین کنفرانس مارا ران

و چون مجموع وزنهای رارک می‌باشد پس

$$\omega + (m - 1)\omega = 1 \Rightarrow \omega^* = \frac{1}{m}$$

و

$$\hat{\phi}_{iWL} = \omega \bar{y}_i + \frac{1}{m-1} \sum_{j \neq i} \bar{y}_j \quad (1)$$

رای پیدا کردن ω بهتر است $E(\hat{\phi}_i - \phi_i)^2$ نسبت به ω می‌نیم شود

$$u_i = V(\bar{y}_i) = \frac{1}{n_i} \phi''(\theta_i) \quad \bar{u} = \frac{1}{m} \sum_{i=1}^m u_i$$

$$s_\phi^2 = \frac{1}{m-1} \sum_{i=1}^m (\phi_i - \bar{\phi})^2$$

داریم

$$g(\omega) = m[\omega^2 \bar{u} + (1-\omega)^2 (m-1)^{-1} \bar{u} + m(1-\omega)^2 (m-1)^{-1} s_\phi^2] \quad (2)$$

$$\Rightarrow \frac{\partial g(\omega)}{\partial \omega} = m[2\omega \bar{u} - 2(1-\omega)(m-1)^{-1} \bar{u} - 2m(1-\omega)(m-1)^{-1} s_\phi^2] =$$

ولذا

$$\omega_{opt} = (\bar{u} + m s_\phi^2) [m(\bar{u} + s_\phi^2)]^{-1} \quad (3)$$

اما جایگزینی در ω داریم

$$\hat{\phi}_{iWL}^{opt} = (1 - \beta_{opt}) \bar{y}_i + \beta_{opt} \bar{y} \quad (4)$$

که

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m \bar{y}_i \quad \beta_{opt} = \frac{\bar{u}}{\bar{u} + s_\phi^2}$$

از هم \bar{u} و s_ϕ^2 نامعلومند و باید ورود شوند
فرض کند

$$s_y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$$

نگاه $E(s_y^2) = \bar{u} + s_\phi^2$ اما چون جوامع دارای توزیع نرمال توابع واریانس درجه دوم هستند لذا

$$V(y_{jl}) = \nu_0 + \nu_1 \phi_j + \nu_2 \phi_j^2 = Q(\phi_j) \quad l = 1, \dots, n_j, \quad j = 1, \dots, m$$

$$E[Q(\bar{y}_j)] = \nu_0 + \nu_1 \phi_j + \nu_2 [V(\bar{y}_j) + \phi_j^2]$$

$$\Rightarrow E[Q(\bar{y}_j)] = Q(\phi_j) + \nu_2 n_j^{-1} Q(\phi_j) = (n_j + \nu_2) u_j \quad ()$$

پس

$$E\left[\frac{1}{m} \sum_{j=1}^m (n_j + \nu_2)^{-1} Q(\bar{y}_j)\right] = u$$

فرض کند

$$\hat{u} = \frac{1}{m} \sum_{j=1}^m (n_j + \nu_2)^{-1} Q(\bar{y}_j)$$

نارین β_{opt} را h و $s_{\hat{u}}$ را $\hat{\beta}$ قرار می‌دهیم که عبارتست از

$$\hat{\beta} = \min\left\{\frac{\hat{u} + 1/m}{s_{\hat{u}}^2 + 1/m}, \right\} \quad ()$$

اکنون ϕ_i h صورت زیر قرار می‌دهیم

$$\hat{\phi}_{iWL} = \left(1 - \hat{\beta}\right) \bar{y}_i + \hat{\beta} \bar{y} \quad i = 1, \dots, m \quad ()$$

تذکره حالت خاصی را در نظر بگیرید که نمونه‌های h حجم مساوی n و مستقل از توزیع‌های $N(\theta_i, \sigma^2)$ $i = 1, \dots, m$ استخراج شده‌اند و $\sigma^2 >$ معلوم است در این صورت $\nu_0 = \sigma^2, \nu_1 = \nu_2 = 1/n$ پس $\hat{u} = \frac{\sigma^2}{n}$ نارین h جایگذاری در h داریم

$$\hat{\phi}_{iWL} = \left(1 - \frac{\sigma^2/n}{s_y^2}\right) \bar{y}_i + \frac{\sigma^2/n}{s_y^2} \bar{y} \quad i = 1, \dots, m$$

۳ روش درست‌نمایی وزنی در مدل تعویض پذیر محلی

فرض کند $\omega_{ii} = \omega_i, \omega_{ij} = \omega_{i^*}, j \neq i = 1, \dots, m$ معنی فرض تعویض پذیر سراسری داشته شده اما فرض تعویض پذیر درون هر ناحیه محلی اقی می‌ماند نگاه

$$\hat{\phi}_{iWL} = \omega_i \bar{y}_i + \omega_{i^*} \sum_{j \neq i} \bar{y}_j$$

و چون $\omega_i = \omega_i + (-m)\omega_{i^*}$ دارم

$$\hat{\phi}_{iWL} = \omega_i \bar{y}_i + (-\omega_i)(-m)^{-1} \sum_{j \neq i} \bar{y}_j$$

رای دست وردن ω_i های نه اکنون $g(\omega_i) = E(\hat{\phi}_{iWL} - \phi_i)^2$ را نسبت به ω_i ازای هر $i = 1, \dots, m$ می‌نماید.

$$g(\omega_i) = \omega_i^2 u_i + (-\omega_i)^2 (m-1)^{-1} (m\bar{u} - u_i) + m^2 (m-1)^{-2} (-\omega_i)^2 (\phi_i - \bar{\phi})^2$$

$$g'(\omega_i) = 2\omega_i u_i - 2(-\omega_i)(m-1)^{-1} (m\bar{u} - u_i) - 2m^2 (m-1)^{-2} (-\omega_i)(\phi_i - \bar{\phi})^2$$

$$g''(\omega_i) = 2u_i + \frac{2}{m-1} (m\bar{u} - u_i) + \frac{2m^2}{(m-1)^2} (\phi_i - \bar{\phi})^2 >$$

پس

$$\omega_{i_{opt}} = \frac{m\bar{u} - u_i + m^2 (\phi_i - \bar{\phi})^2}{(m-1)^2 u_i + m\bar{u} - u_i + m^2 (\phi_i - \bar{\phi})^2}$$

نارن

$$\hat{\phi}_{iWL}^{opt} = (-\beta_{i_{opt}}) \bar{y}_i + \beta_{i_{opt}} \bar{y}$$

()

که

$$\beta_{i_{opt}} = \frac{(m-1)u_i}{(m-1)u_i + \bar{u} + m(\phi_i - \bar{\phi})^2}$$

اما در عمل β_i ها معلوم نیستند و اندرورد شوند توجه کنید که

$$\begin{aligned} E(\bar{y}_i - \bar{y})^2 &= V(\bar{y}_i - \bar{y}) + (\phi_i - \bar{\phi})^2 \\ &= \frac{1}{m} [(m-1)u_i + \bar{u} + m(\phi_i - \bar{\phi})^2] \end{aligned}$$

همچنین

$$E[Q(\bar{y}_i)] = \nu_0 + \nu_1 \phi_i + \nu_2 (n_i + \nu_2) = (n_i + \nu_2) u_i$$

نارن $E[Q(\bar{y}_i)] / (n_i + \nu_2) = u_i$ از این رو $\beta_{i_{opt}}$ و $\hat{\beta}_i$ را می‌شود که

$$\hat{\beta}_i = \max \left[\frac{m-1}{m} \frac{Q(\bar{y}_i)(n_i + \nu_2)^{-1} + m^{-1}}{(\bar{y}_i - \bar{y})^2 + m^{-1}}, 1 \right]$$

ولذا

$$\hat{\phi}_{iWL} = (-\hat{\beta}_i) \bar{y}_i + \hat{\beta}_i \bar{y}$$

۳۳ روش درست‌نمایی وزنی بدون فرض تعوض پذیر

حالتی را در نظر بگیرد که هیچگونه فرض تعوض پذیر وجود ندارد اگر $\hat{\phi}_i = \sum_{j=1}^m \omega_{ij} \bar{y}_j$ پس ω_{ij} های بهینه می‌تواند با هم‌زمان کردن $E(\sum_{j=1}^m \omega_{ij} \bar{y}_j - \phi_i)^2$ نسبت به ω_{ij} دست می‌نند این شرط که $\sum_{j=1}^m \omega_{ij} = 1$ فرض کند $\Omega_i = (\omega_{i1}, \dots, \omega_{im})'$ و فرض کند

$$g(\Omega_i) = E\left(\sum_{j=1}^m \omega_{ij} \bar{y}_j - \phi_i\right)^2 - \lambda_i \left(\sum_{j=1}^m \omega_{ij} - 1\right)$$

که λ_i ها را لاگرانژ می‌باشند

$$g(\Omega_i) = \sum_{j=1}^m \omega_{ij}^2 u_j + \left(\sum_{j=1}^m \omega_{ij} \phi_j - \phi_i\right)^2 - \lambda_i \left(\sum_{j=1}^m \omega_{ij} - 1\right)$$

ناربان

$$\frac{\partial g}{\partial \omega_{ij}} = \omega_{ij} u_j + \left(\sum_{j=1}^m \omega_{ij} \phi_j - \phi_i\right) \phi_j - \lambda_i$$

$$\frac{\partial^2 g}{\partial \omega_{ij}^2} = u_j + \phi_j^2 > \quad \frac{\partial^2 g}{\partial \omega_{ij} \partial \omega_{kl}} = 0 \quad j \neq l$$

از این رو ماتریس هسسیان برای g معین مثبت است و ω_{ij} های بهینه با حل معادلات $\frac{\partial g}{\partial \omega_{ij}} = 0$ ، $i, j = 1, \dots, m$ دست می‌نند با استفاده از نماد ماتریسی این معادلات به صورت زیر نوشته می‌شوند

$$M\Omega_i = \lambda_i \mathbf{1}_m + \phi_i \Phi$$

که

$$M = D + \Phi \Phi', \quad D = \text{Diag}(u_1, \dots, u_m), \quad \Phi = (\phi_1, \dots, \phi_m)'$$

ناربان Ω_i بهینه به صورت زیر دست می‌نند

$$\Omega_i^{opt} = M^{-1}(\lambda_i \mathbf{1}_m + \phi_i \Phi) \quad ()$$

چون $\sum_{j=1}^m \omega_{ij} = 1$ ناربان

$$= \mathbf{1}_m' \Omega_i^{opt} = \lambda_i \mathbf{1}_m' M^{-1} \mathbf{1}_m + \phi_i \mathbf{1}_m' M^{-1} \Phi$$

$$\lambda_i = \frac{-\phi_i' M^{-1} \Phi}{\phi_i' M^{-1} \Phi} \quad ()$$

اکنون با استفاده از و دارم

$$\Omega_i^{opt} = \frac{-\phi_i' M^{-1} \Phi}{\phi_i' M^{-1} \Phi} (M^{-1} \Phi) + \phi_i M^{-1} \Phi$$

و

$$M^{-1} = D^{-1} - \frac{D^{-1} \Phi \Phi' D^{-1}}{\Phi' D^{-1} \Phi}$$

حال چون $E(\bar{y}_i) = \phi_j$ ، $E[(n_j + \nu_r)^{-1} Q(\bar{y}_j)] = u_j$ که

$$Q(\bar{y}_j) = \nu_0 + \nu_1 \bar{y}_j + \nu_2 \bar{y}_j^2$$

دارم

$$\hat{u}_j = (n_j + \nu_r)^{-1} Q(\bar{y}_j)$$

$$\hat{\Omega}_i^{opt} = \frac{-\bar{y}_i' \hat{M}^{-1} Z}{\bar{y}_i' \hat{M}^{-1} Z} (\hat{M}^{-1} Z) + \bar{y}_i \hat{M}^{-1} Z$$

که

$$Z = (\bar{y}_1, \dots, \bar{y}_m)' , \quad \hat{M} = (\hat{D} + ZZ')^{-1} , \quad \hat{D} = \text{Diag}(\hat{u}_1, \dots, \hat{u}_m)$$

مراج

- [1] Ghosh, M., and Rao, J.N.K. (1994), Small area estimation: an appraisal (with discussion), *Statistical Science*, 9, 65-93.
- [2] Hu, F. and Zidek, J.V. (2000), The relevance weighted likelihood with applications. In *Empirical Bayes and Likelihood Inference. Lecture Notes in Statistics*. Eds. S.E. Ahmed and N. Reid. Springer-Verlag, New York, pp 211-235.
- [3] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, V1. University of California Press, pp 197-206.
- [4] Tibshirani. R. and Hastie, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.*, 82, 559-567.